CLOUD ACCOUNTABILITY PROJECT

D:C-7.2: Privacy Design Guidelines for Accountability Tools

Deliverable Number: D:37.2 (D:C-7.2)
Work Package: WP:C-7
Version: Final
Deliverable Lead Organisation: EURECOM & KAU
Dissemination Level: PU
Contractual Date of Delivery (release): 31/05/2014
Date of Delivery: 31/05/2014

Editor
Melek Önen (EURECOM), Tobias Pulls (KAU)

Contributors
Simone Fischer-Hübner (KAU), David Núñez (UMA), Melek Önen (EURECOM), Tobias Pulls (KAU)

Reviewers
Ronal E. Leenes (Tilburg University), Mario Südholt (EMN)

Executive Summary

In order to increase trust in the cloud ecosystem, A4Cloud aims to deliver tools for a range of different stakeholders (data subjects, auditors, regulators, etc.) to support accountability requirements such as transparency, assurance and responsibility. The design of such tools may however raise new challenges and issues with respect to privacy. This deliverable aims at assisting an accountability tool developer to implement a privacy friendly accountability tool. To this end, each of the planned A4Cloud tools is analysed in terms of privacy. Based on this initial analysis some privacy-by-design guidelines are further proposed in order to help any accountability tool developer (including A4Cloud developers) to discover and address privacy issues the expected tool may raise.

The evaluation of A4Cloud tools started with an initial survey conducted within work package C-7. Each tool owner was first contacted by a C-7 partner and was required to provide details on the tool and to raise some initial privacy concerns. Based on this survey and the analysis of the three A4Cloud business use cases described in work package B-3, the study raised many different privacy issues ranging from basic data confidentiality to user anonymity within data processing. Therefore, each tool was also analysed within the context of all business use cases. In particular, since Business Use Case 3 (BUC3), “Rights and relevant obligations in a multi-tenant cloud scenario”, is the most complex use case and involves two different data subjects (one of them being also an employee of a business), the deliverable provides an extensive privacy analysis of each tool within BUC3.

Based on this privacy analysis, the deliverable proposes privacy by design guidelines for accountability tools. The proposed guidelines not only include some inescapable steps such as the *functional requirements analysis*, the *modelling of attackers, threats and privacy risks* but also elaborate how a *multilateral security requirements analysis* could be conducted for addressing the inherent conflict between privacy and accountability.

Furthermore, the proposed approach also defines an additional step for the *identification and selection of PETs for data protection*. The deliverable suggests suitable practical and research PETs, which are classified into two categories: *application PETs* for protecting privacy at the application layer, and *communication PETs* for communication between different actors. We distinguish between practical PETs, that have reasonably mature open source implementations available, and research PETs, whose design is published but no mature implementations are available for to the best of our knowledge.

Finally, we apply the proposed guidelines to the functional description of the Audit Agent System, a proposed A4Cloud tool, and describe our findings on some further issues raised thanks to the use of our approach.

Contents

List of Abbreviations

BUC	Business Use Case
CA	Cloud Auditor
CB	Cloud Broker
CC	Cloud Customer
CP	Cloud Provider
DC	Data Controller
DP	Data Processor
DPA	Data Protection Authority
DS	Data Subject
PETs	Privacy-Enhancing Technologies

1. Introduction

In this section, we first discuss how privacy is defined in A4Cloud and briefly remind the set of accountability tools within the A4Cloud environment.

1.1. Privacy and Accountability Tools

The A4Cloud glossary [A4C14a] is the most appropriate reference to understand the relationship between privacy and accountability tools. This glossary defines *accountability* as follows (emphasis added):

> Accountability for **an organization** consists of **accepting responsibility** for the **stewardship of personal and/or confidential data** with which it is entrusted in a **cloud environment**, for processing, storing, sharing, deleting and otherwise using the data **according to contractual and legal requirements** from the time it is collected until when the data are destroyed (including onward **transfer to and from third parties**). It involves **committing to legal and ethical obligations**, policies, procedures and mechanisms, explaining and demonstrating ethical implementation to internal and external stakeholders and remedying any failure to act properly.

Hence, an organisation is accountable for is the use of *personal and/or confidential data* in a *cloud environment*. The A4Cloud glossary further defines privacy as follows (emphasis added):

> The claim of individuals, groups, or institutions **to determine for themselves** when, how, and to what extent **information about them** is communicated to others [Wes70]. The ability to control the collection and sharing of information about oneself.

On the other hand, *personal data* is defined[1] as (emphasis added):

> 'Personal data' shall mean any information relating to an **identified or identifiable** natural person ('**data subject**'); an identifiable person is one who can be identified, **directly or indirectly**, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

In other words, we can state that privacy in A4Cloud is about data subjects *controlling* when, how and to what extent data which is or could be related to them are shared with others. In a sense, accountability is all about organisations accepting the right to privacy of cloud customers (individual data subjects) by enabling them to control information about them in the cloud.

Furthermore, accountability mechanisms are defined in the A4Cloud glossary as follows (emphasis added):

> Diverse processes, non-technical mechanisms and **tools** that support accountability practices.

[1] The glossary uses the definition from Article 2 of the EU Data Protection Directive 95/46/EC.

Finally, accountability practices are defined as:

> Emergent behaviour characterising accountable organisations.

Therefore, the goal of accountability tools is to support organisations to be accountable, and we know from the earlier definition of accountability that this is centred around the appropriate treatment of personal and confidential data. When creating these tools to achieve accountability for how personal and confidential data are treated, it is paramount to *minimise* the creation of any *new* personal or confidential data created as a *result* of the tools being used. Conceptually, an accountable organisation should be a responsible steward for *all* personal and confidential data for which it is entrusted, including any new data created as a consequence trying to be accountable. To prevent the need for *accountability tools for accountability tools*, privacy by design and Privacy-Enhancing Technologies (PETs) for accountability tools play a vital role; proper privacy protections *stop the accountability tool loop*. In particular, the principle of *data minimisation* is of utmost importance. An ideal accountability tool, from a privacy perspective, should not generate any new personal or confidential data for which an organisation has to be accountable. Furthermore, it should not leak any personal or confidential data to any other third party, which may or may not be interested in being held accountable. As we know from the definition of accountability, an accountable organisation is also responsible for data that is 'transferred to and from third parties'.

In our work on analysing privacy issues with respect to accountability tools, a central goal is to identify personal data created and processed as a consequence of the use of the tools themselves. This includes any personal data that may be stored for "legitimate use", such as metadata (that is also personal data) for auditors in an auditing tool, or new personal data that is observable from the behaviour of different stakeholders in another tool. While the proliferation of personal data in accountability tools *may* be in the best interest of data subjects, there is an inherent risk in the processing of personal data; PETs play a vital role in minimising this risk. The goal is to create accountability tools that enable data subjects to exercise control over their personal data without, ideally, creating any new personal data in the process; if personal data is created then it should also be under the control of the corresponding data subject.

1.2. A4Cloud Actor and Roles

Figure 1 contains a simple flowchart for determining the roles of a cloud actor from WP C-2. The first step is to identify if the actor is an individual or an organisation. Next, the specific cloud computing roles are set, followed by the data protection roles.

For our purposes in this deliverable, one core activity is defining the data protection roles with respect to the personal data processed or created by accountability tools. Cloud computing roles are *secondary*. Assigning data protection roles to different actors in cloud computing roles for the *business use cases* is not a task of WPC-7. When conducting the initial privacy analysis of A4Cloud tools, we have tried to follow the terminology as defined by WPC-2, but since both the terminology and the tools have evolved throughout the course of the project there may be some inconsistencies. At the time of writing, the impact on our work due to evolving definitions should be minor. The biggest factor impacting the quality of our work throughout the duration of WPC-7 has been the vague and changing descriptions of tools.

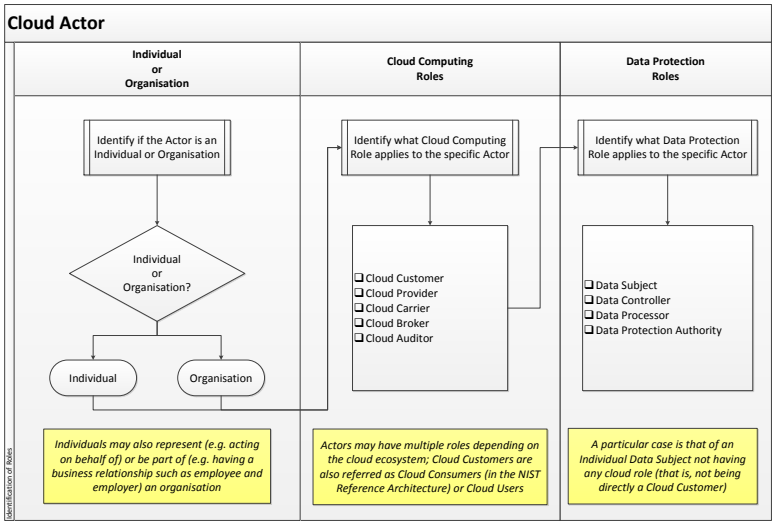

Figure 1: How to determine cloud actors (from C-2)

1.3. A4Cloud Tools

Table 1 summarises all the A4Cloud tools identified for analysis after an initial C-7 survey. The table has been updated at M18 thanks to another survey conducted by WP D-2. The table lists the name of the tool, the corresponding work package, the main stakeholders/A4Cloud actors using the tool and finally the purpose of the tool. Stakeholders are defined in terms of cloud customers (CCs), cloud providers (CPs), and cloud auditors (CAs). While some tools were on their initial design phase, some others are more mature (like the WP D-5 tools). Therefore, the details on the privacy analysis are depending on the details of the tool design and thus some tools such as the Incident Response tool (IRT) are very briefly analysed. On the other hand, some of the tools like the Data Protection Impact Assessment Tool (DPIAT) do not process or create any personal data and hence do not raise any privacy issues, based on the descriptions of the tool available to us at the time of writing. Therefore such tools have not been further analysed. We stress that this may change as tools evolve.

1.4. Contributions

Our contributions can be summarized as follows:

- We define privacy in the context of accountability tools, answering questions such as “what is an ideal accountability tool from a privacy perspective?”
- We present the initial privacy analysis of A4Cloud tool concepts in the early stages of the project, and how the tools’ use may impact privacy in the A4Cloud Business Use Cases. We also perform an up-to-date privacy analysis of current A4Cloud tools (as planned in spring 2014) in relation to the third business use case.

Tool	WPs	Stakeholders	Purpose
Accountability Laboratory (AccLab)	D-3	Organisations as CCs, CPs and CAs	Help writing abstract accountability obligations, to check for consistency and compliance and then generate A-PPL policies
Accountable-PPL Engine (A-PPLE)	D-3	Organisations as CCs and CPs	Enforcement of obligations expressed in policies
Audit Agent System (AAS)	C-8	Organisations as CCs, CPs and CAs	Enable stakeholders to perform audits of their cloud infrastructures
Assertion Tool (AT)	D-6	Organisations as CCs, CPs, CBs	A meta-tool that enables the validation of the correct working of other A4Cloud tools
Cloud Offerings Advisory Tool (COAT)	D-4	Organisations and individuals as CCs	Guidance to potential CCs wrt. security, privacy, and data protection considerations
Data Track (DT)	D-5	Individuals as data subjects (maybe CCs), and organisations as CPs	Enabling data subjects to exercise their rights online to access, correct and request removal of their data at data controllers
Data Protection Impact Assessment Tool (DPIAT)	C-6	Organisations as CCs	Help identify the risks of carrying out a certain business transaction such as buying a new cloud service
Data Subject Access Request tool (DSAR)	D-4	Individuals as data subjects, potentially CCs	Help data subjects to file an offline request for a copy of their data at data controllers
Data Transfer Monitoring Tool (DTMT)	D-3	Organisations as CCs, CPS, or CAs, potentially as data controller	To support accountable data localization and transfer across cloud software, platform and infrastructure services
Incident Response Tool (IRT)	D-4	Individuals as CCs and data subjects, and organisations as CCs	Allow users to respond to privacy and security incidents in cloud scenarios
Plug-in for Assessment of Policy Violation	C-5	Individuals as CCs and data subjects	Provide an assessment on the relevance of the violation of policies, so that stakeholders receive information about the most relevant policy violations
Remediation & Redress Tool (R&RT)	D-4	Individuals or organisations as CCs	Respond to (perceived) incidents
Transparency Log (TL)	D-5	Individuals as data subjects (maybe CCs), and organisations as CPs	Provide secure and privacy-friendly asynchronous one-way communication to recipients such as (passive) data subjects

Table 1: All tools planned for development in A4Cloud as of M18.

- We present privacy by design guidelines tailored to accountability tools and consequently suggest suitable privacy-enhancing mechanisms. We provide examples of our initial suggestions to two A4Cloud accountability tools. We also evaluate the impact of the use of privacy-enhancing mechanisms on the attributes of accountability.
- We finally perform a design iteration on an example A4Cloud accountability tool.

1.5. Organisation

The deliverable is organised as follows. In Section 2, we present our privacy analysis of A4Cloud tools, as planned in the early stages of the project, and provide an up-to-date analysis of currently planned A4Cloud tools in regard to the third A4Cloud business use-case. Section 3 presents our privacy by design guidelines for accountability tools. Section 4 presents a selection of privacy-enhancing mechanisms suitable for accountability tools. Section 5 discusses how to balance privacy and security protection goals. Section 6 presents one iteration of our guidelines for an A4Cloud accountability tool. Section 7 concludes this deliverable.

The document also provides some appendices which describe the template for our initial privacy analysis of A4Cloud tools (Appendix A), our initial privacy analysis of tools (Appendix B), our initial privacy analysis in relation to two A4Cloud business use-cases (Appendix C), and our initial privacy-enhancing mechanisms suggestions to tool owners (Appendix D).

2. Privacy Analysis

Within the first project year, we conducted an initial privacy analysis of all A4Cloud tools based on tool specifications that were available at this time (mainly obtained by interviewing the tool developers). Besides, a privacy analysis of those tools was also conducted in regard to the first two business use cases (BUCs) on “Health Care” (BUC1) and “Enterprise Resource Planning” (BUC2). These analyses were part of a first WP C-7 milestone report at M15 that were communicated to the tool owners together with an initial suggestion of privacy-enhancing technologies (PETs) that may be used.

As the list of A4Cloud tools and their specifications have evolved since our initial analysis to a large extent, and thus the analysis became to some part outdated, we decided not to include the outcome of most of our work in the main part of this deliverable but rather to move it to Appendix B and Appendix C for reference. Instead, we present a detailed privacy analysis of two A4Cloud tools whose specifications have not had any major changes since the end of the first project year and which seem interesting with regard to their privacy implications. The two tools we present are the Audit Agent System and the Transparency Log. In addition, we present a privacy analysis based on the latest tool specifications in regard to BUC3 on “Rights and relevant obligations in a multi-tenant cloud scenario” in section 2.4. This privacy analysis provides a summary of updated privacy analyses of all A4Cloud tools.

2.1. Structure of our initial Privacy Analysis of Tools

Our initial privacy analysis of the A4Cloud tools conducted in the first project year was structured into the following parts (Appendix A provides further details):

1. A summary of the tool specification, which also specifies the type of personal data created and processed by the tool, cloud actor roles, as well as purposes of data processing and data retention periods.

2. The information flow during the data’s life cycle, answering the following questions: By whom are the data collected? How are personal data transferred / flowing between different parties? Where/when are the data destroyed or anonymised?

3. Initial privacy concerns: What privacy issues arise by the tool that need to be addressed?

4. Privacy and data protection principles: How far are basic legal privacy requirements an issue or how far are they already considered/complied with? What means of protection are already taken by the tool design?

The first two steps are part of a functional requirements analysis and the last two steps are part of a risk analysis that should be conducted when following a privacy by design approach, as we will discuss in Section 3.

We set up the structure for our initial privacy analysis to cover the essential steps to be conducted during the process of a privacy impact assessment (PIA) that were meaningful and doable, especially given the provisional and rather high-level tool specification documentations

that were available within the first 15 months of the project. Hence, our goal was to perform a "light PIA" with our initial privacy analysis.

For determining essential steps of a PIA that our initial privacy analysis should cover, we referred to the final deliverable of the PIAF project[2] that identifies the following key elements of a PIA [HKW12]:

1. Determining whether a PIA is necessary
2. Identifying the PIA team and setting terms of reference
3. Description of the proposed project
4. Analysis of the information flows and other privacy impacts
5. Consultation with stakeholders
6. Risk management
7. Legal compliance check
8. Formulation of recommendations
9. Preparation and publication of the report
10. Implementation of recommendations
11. External review and/or audit
12. Revisiting PIA if the project in question changes.

The first part on "Summary of tools specifications" of our privacy analysis that we derived by interviews, i.e. in consultation with the respective tool developers, included descriptions of the proposed tool development projects (PIA key element 3). The parts on "The information flows during the data's lifecycle" and the analysis of "Initial privacy concerns" match with the PIA key element 4. The part "Privacy and data protection" comprises a legal compliance check (PIA key element 7). After this initial analysis, we presented recommendations (PIA key element 8) on some privacy-enhancing mechanisms, which were part of the WP C-7 Milestone report at M15, and are discussed in Section 4 and presented. In the next section we describe the outcome of the proposed analysis with respect to two A4Cloud tools, namely, the Audit Agent System (AAS) and the Transparency Log (TL). The analysis of the other tools is described in appendix B.

2.2. Example: The Audit Agent System

2.2.1. Summary of the Tool Specification

In a survey conducted by work package D-2 in the project month M11, the Audit Agent System was defined as:

[2] http://www.piafproject.eu

> "An automatic audit system based on software agents. These agents can be deployed across different architectural layers in cloud environments with the purpose to collect and process evidence, to generate audit reports and to aggregate new evidence.
>
> The cloud audit agent system, enables customers of cloud services to perform audits of their cloud infrastructures (i.e., virtual machines in IaaS) using custom defined policies to check against. By rolling out software agents to the virtual machines these tools are executed and relevant output and results is collected. By checking against thresholds and constraints defined in audit policies, audit results are generated and presented to the issuer of the audit process."

Table 2 provides a summary of the Audit Agent Tool and of its privacy-related characteristics.

Audit Agent System	
WP	C-8
Contact	Christoph Reich and Thomas Rüebsamen
Type of Personal Data Processed	Data stored in logs, traffic flow, system environment data etc. The data is primarily meta data about, and produced by, the service offered by the primary service provider. This meta data is also personal data. The type of personal data processed by the tool depends on the type of personal data processed by the primary service provider.
Data Subjects	All data subjects of the personal data processed by the primary service provider, that are monitored by or interacting with the tool.
Purpose of Data Processing	Auditing
Data Controller	Primarily the primary service provider, but also depending on how the tool is used, the end users of the tool, which may or may not also be data subjects.
Data Processors	None specific for the tool to function (like a dedicated storage provider), but any data processor used by the primary service provider would also become data processors of the tools data.
Stakeholders (tool users)	CAs (external and internal), potentially CPs, CCs (data controllers, potentially data subjects)
Retention Period	Unknown, presumably this depends on how the tool is used. Data stored by audit agents need to be retained at least until end users accesses it, possibly longer due to legal requirements. The description of work for C-8 states that gathered evidence should be legally binding, which suggests that the retention period may be significant. The retention period is also highly dependent on the nature of the audit task.

Table 2: Summary for the Audit Agent System.

2.2.2. Information Flow During the Data's Life-Cycle

Data is collected by audit agents, which either (i) run continuously at the primary service provider on all of its provisioned cloud services, or (ii) is triggered by end users to gather data at the primary service provider by executing code on some or all of the providers provisioned cloud services that end users can access. The collected data is minimized by the agent (according to the requirements of the audit task; e.g., data is trimmed to match a specific time

frame or audit goal defined in the audit task) is stored by audit agents in a per-tenant evidence store from which another component (evidence processor and presenter) generates audit reports containing the audit result as well as supporting evidence (e.g., log snippets depicting a violation). The reports are later on presented to the different stakeholders in various degrees of detail. Anonymisation of data is not (yet) considered.

2.2.3. Initial Privacy Concerns

The conservative approach is to treat the data gathered by the tool as at least as sensitive as the type of personal data processed by the primary service provider. The fact that traffic data is within scope and that retention period is unknown is of utmost concern. How is (traffic) data gathered (e.g., NetFlow) and who can access what information? Ensuring that no information is leaked as a consequence of sharing (traffic) data with end users is paramount.

2.2.4. Privacy and Data Protection Principles

The design of the tool is in its early stages in A4Cloud. Before A4Cloud, work as been done built on JADE (the Java Agent DEvelopment Framework). Fundamental privacy and data protection principles are relevant, but so far not part of the tool's design (cf. Table 3).

Audit Agent System	
Informed Consent	Currently no, but presumably end users are informed as part of the privacy policy of the primary service provider.
User Control	Data subjects will get access to part of aggregated data in the form of reports, to what extent is still unclear. No direct access to the stored data is provided.
Purpose Binding	The stored data is only processable by the Evidence Processor and Presenter component, which will limit usage to audit purposes only.
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	Data minimization components for the audit agents are planned. The main driving principle for the Audit Agents is to only collect data necessary to process an audit task. The audit task also dictates degree of anonymisation and/or pseudonymisation.
Transparency & Data Subject Rights	Presumably end users of the tool are informed in the privacy policy.
Confidentiality, Integrity & Availability	Different protection for different components, to be further investigated. Ideally, all components should be made tamper-proof.

Table 3: Privacy and data protection principles for the Audit Agent System.

2.3. Example: The Transparency Log

2.3.1. Summary of the Tool Specification

In a survey conducted by work package D-2 in the project month M11, the Transparency Log is defined as follows:

> "The transparency log (TL, more fancy name pending) is a tool from D-5 that enables the transfer of log data from data processors to data subjects while preserving the privacy of data subjects. This is needed by the Data Track (D-5), to:
>
> - enable data subjects to detect policy violation by matching log data,
> - to view detailed information about how the data has been shared and used by the data controller,
> - and to get an overview of prior data disclosures and associated metadata like privacy policy, time, etc."

Table 4 provides a summary of the Audit Agent Tool and of its privacy-related characteristics.

Transparency Log	
WP	D-5
Contact	Tobias Pulls
Type of Personal Data Processed	Any data the CP or CC running part of the tool wishes to inform a recipient about
Data Subjects	Individual end-users, i.e., the per definition data subjects
Purpose of Data Processing	Transparency purpose towards data subjects, audit data towards auditors (optional) for audit purposes
Data Controller	The CC in a data controller role
Data Processors	CPs and optionally non-individual CCs
Stakeholders (tool users)	CCs and CAs
Retention Period	Configurable global value per instance, set by other tools using this tool

Table 4: Summary for the Transparency Log.

2.3.2. Information Flow During the Data's Life-Cycle

Data is collected by a data processor (CP or CC) and sent to a specific individual end-user (ideally, the data subject of the data being sent) and one or more optional CAs by first encrypting the data. The encrypted data is then stored at an intermediate server, which may or may not be run by the data processor itself. The data is stored and transferred in such a way to minimise any creation of new personal or confidential data. The encrypted data is destroyed after a configurable time and can (in the case of no CAs being used) be destroyed by a data subject at any point in time by deleting the corresponding decryption key, with "forward secrecy"[3]. Any "data anonymisation" would have to be done by the processor before sending the data, and is out of scope of this tool. How long data is kept by the recipient CCs and CAs is also out of scope of this tool.

[3] In the sense that data received in the past remains secret from compromise of key material in the future.

2.3.3. Initial Privacy Concerns

Within the scope of the tool, the primary focus is leaking personal or confidential data to a third party. The Transparency Log can be seen as an unidirectional communication channel, and thus questions like confidentiality, integrity, and recipient anonymity arise. Looking from the perspective of a *log*, of course a detailed record of personal or confidential data pose a privacy threat to data subjects.

Outside the scope of the tool, questions like *what* data is sent to *whom*, and does the "*whom*" also include "trusted" CAs as recipients are paramount. In other words, how the tool is used (by other tools in A4Cloud) is a key concern.

2.3.4. Privacy and Data Protection Principles

The tool is a PET, so the goal of it is to provide a privacy-friendly solution. Table 5 summarises how privacy principles are addressed by this tool. The latest publication is [PPW13], further research and development work is ongoing as part of C-7.

Transparency Log	
Informed Consent	For sending data to data subjects, they are explicitly required to participate in the setup process. For sending data to CAs, presumably this is part of a privacy policy.
User Control	The primary purpose of the tool is to give data subjects more data, so yes, they have access to their data and only them unless CAs are involved
Purpose Binding	Yes, strong encryption together with algorithms/protocols that destroys or prevents the creation of new confidential or personal data.
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	In the tool, users are identified by *transaction* pseudonyms, which they generate themselves (in essence, their "public key"). Any two data items are *unlinkable*, which implies in the setting that for example data items and user identifiers are also unlinkable. All data items are encrypted and they are received by data subjects over an anonymous channel, i.e., reconstruction is also unlinkable to their natural identity, and not even linkable to the transaction pseudonym to which a data item belongs. If any data item for a particular user exists or not is *undetectable*, i.e., the tool provides *recipient unobservability*. Several of these properties are for data items created *before compromise* of the sender (processor), i.e., it makes sense to talk about *forward* recipient unobservability or forward unlinkability of data items. Servers are untrusted.
Transparency & Data Subject Rights	The tool is a *solution* for providing data subjects with information concerning data processing *after* data disclosure, i.e., it is an ex-post Transparency-Enhancing Tool
Confidentiality, Integrity & Availability	Data items are confidential, their integrity is protected (and verifiable), and availability is ensured by preventing servers storing data item from selectively blocking access for a particular user

Table 5: Privacy and data protection principles for the Transparency Log.

2.4. An Up-to-Date Analysis of A4Cloud Tools in regard to BUC3

In this section, we provide an up-to-date analysis of all A4Cloud tools with respect to BUC3. BUC3 in A4Cloud is called “Rights and relevant obligations in a multi-tenant cloud scenario” as described in the milestone report MSB-3.1. It is characterised by:

Personal and confidential data interaction Cloud services are used by individuals both to store personal data (for which they are the data subject) and confidential business data.

Bring Your Own Device (BYOD) Individuals bring their own devices as employees to the enterprise environment. These devices are used to access both personal and business related cloud services.

Multi-tenancy Devices, as well as other forms of resources like services, may be used by multiple users and belong to multiple organisations.

Data governance When personal data is moved to cloud services there is an increase in complexity of data governance which also increases the risk of loss of governance.

Figure 2 gives an overview of the actors and data flows in BUC3. There are no less than six different companies (Company A–F) involved in offering cloud services with different interdependencies. These cloud services are used by a business enterprise, its employees, and the employees privately for their personal data. From our C-7 perspective, we are interested in personal data and data subjects. BUC3 has two data subjects: the customer of the business enterprise and Employee A. The customer shares his or her personal data with the business enterprise, which in turn uses a number of cloud services provided by four of the companies. Employee A, from the same device, uses the business enterprise software as part of his or her job, and cloud services for healthcare and social network services for his or her private affairs. Since data protection is only relevant for personal data, we will focus our analysis on the impact on the customer’s and the Employee A’s privacy as a consequence of the deployment of A4Cloud tools throughout the cloud and at the enterprise. For sake of readability, we will name Employee A *Alice* and the customer *Bob*.

2.4.1. Accountability Laboratory (AccLab)

AccLab allows its user to write abstract accountability obligations and checks their consistency before generating the corresponding machine readable policies. In addition to a privacy officer, AccLab helps the Data Subject to enter user preferences. As an employee, Alice may have access to some other subjects’ preferences while she is not authorized to do it and such information can be considered as personal data since it includes information about health-care.

Furthermore, as this is the case for any tool, since Alice is considered as both an employee and a Data Subject, there may be a risk on the leakage of Alice’s personal data due to misconfiguration.

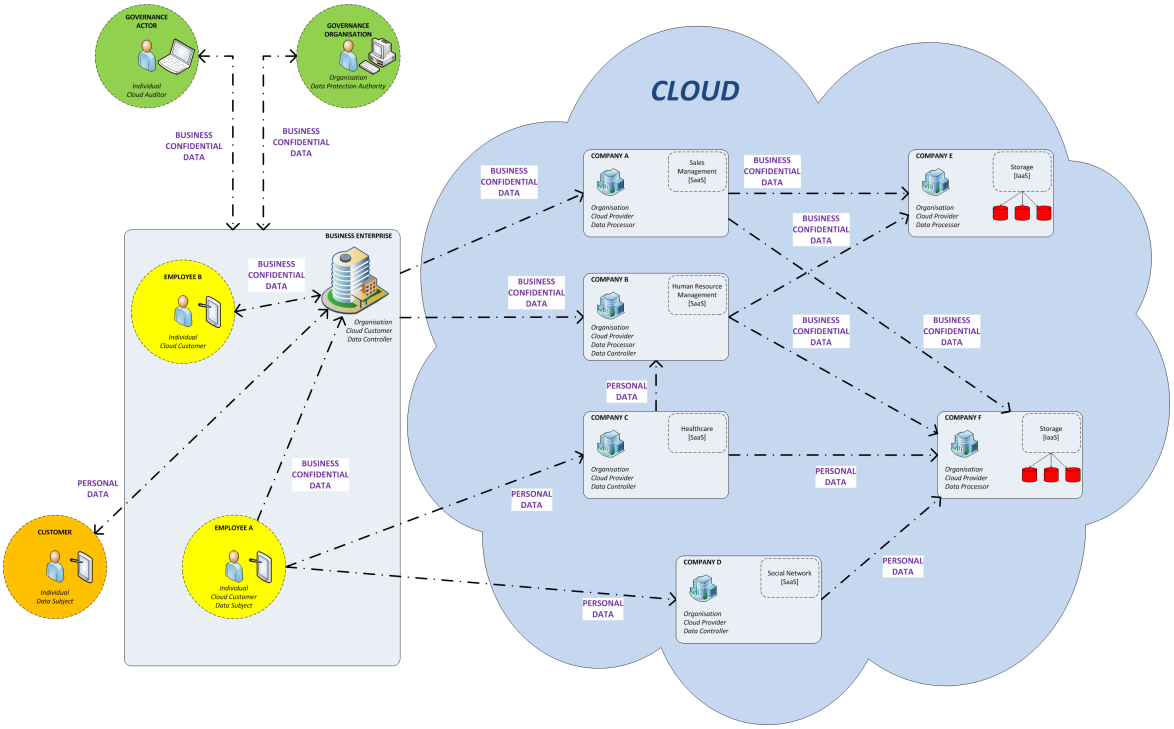

Figure 2: Overview of BUC3 actors and data flow (from MSB-3.1).

2.4.2. Accountable-PPL Engine (A-PPLE)

The main goal of the A-PPLE engine is to enforce A-PPL policies on the correct access and usage of personal data and to notify or report some actions or events. First of all, similarly to AccLab, if A-PPL policies stored in the A-PPLE engine include some user preferences, by accessing the policy as an employee, Alice can have access to other data subjects' preferences which can be considered as a privacy sensitive information. Also, whenever Alice uses/communicates with A-PPLE as an Employee, this may produce some logs which can be considered as privacy sensitive since she is also considered as a Data Subject. A-PPLE will provide and enforce the access to personal and business confidential data. Therefore the control of the access to and the usage of this specific tool should definitely be the very major privacy requirement. Furthermore, the access to the data by either Bob or Alice may result on the creation of additional privacy sensitive data such as access logs. Hence, A-PPLE also generates some internal logs which may also include some privacy sensitive information such as personally identifiable information (PII). Even if the original PII is protected through encryption mechanisms, an adversary may discover that some data was requested several times. Although there is no information on the content of such data, the fact that the same data was accessed several times may imply some leakage. Therefore, this tool requires the integration of a proper access control mechanism with some encryption or anonymization technique to protect the logs. Finally, another feature of A-PPLE is of course notification and the content of such "message" of course can leak some privacy sensitive information such as unauthorized

access logs.

2.4.3. Audit Agent System (AAS)

We assume that AAS will have the capability to be run at each company and the business enterprise. Furthermore, one key consideration is if AAS agents may be deployed to employee devices. If this is the case, then the interaction between AAS and data subject tools like the Data Track and Transparency Log significantly increases the risk of privacy violations for data subjects like Alice who use their own devices in their role as employees.

BUC3 has one important actor not depicted on Figure 2: Michael, an external auditor of all cloud providers. Auditors like Michael, together with the cloud providers themselves and organisational customers of cloud providers, are the primary users of AAS. AAS is deployed per-tenant, with an evidence store and aggregated data in audit reports. Beyond general considerations around the risk of further processing personal data and so on (as previously discussed in Section 2.2), one key consideration in BUC3 is data flow (of AAS data and audits) between the different actors. An audit using AAS of Company B, which is used both by the business enterprise and Company C (the healthcare SaaS), may reveal that Alice also uses the Healthcare service for her personal data, and even worse, may reveal (sensitive) personal data about Alice. Similar risks are in place in the interaction between most companies in BUC3. Another consideration is how AAS handles an audit of Company A when Company A uses Company E and F for storage. AAS agents require (to the best of our knowledge) arbitrary code execution rights to function and access to all data to perform a complete audit, as part of its runtime. It may very well be the case then that an audit of Company A risks revealing personal data of Alice due to her use of services at Company C and D, who in turn also rely on Company E and F. Presumably the one instance of AAS at Company E and F can access all data, since the instance should also be possible to use by Company F for internal audits, as well as for all other companies (A, B, C, and D) that use the service.

2.4.4. Assertion Tool (AT)

AT is a meta-tool and processes no personal data, so it has no impact on privacy in BUC3.

2.4.5. Cloud Offerings Advisory Tool (COAT)

COAT, when used by a data subject, will receive a data subject's preferences with respect to their location, context, needs, and requirements on the cloud provider. This is all personal data. COAT will log the advice offered by the tool and the user's decision for "accountability purposes". Assuming this log is only stored locally on the data subject's device, in general, the impact on the data subject's privacy may appear minimal. However, for Alice in BUC3, this may become an issue if her device is "compromised" due to her role as an employee of the business enterprise. Presumably, the COAT log would then contain sensitive personal data, if Alice used the tool for example to get advice with respect to healthcare services (company C).

2.4.6. Data Track (DT)

DT is primarily a data subject tool and its main component would therefore be running at Alice's and Bob's devices. The other major component of DT is the DT API, which is a component running at service providers to enable some core functionality of DT. The DT API integrates with A-PPLE (see Section 2.4.2) and TL (see Section 2.4.13). A-PPLE facilitates sending notifications (generated by A-PPLE and other A4Cloud tools) through TL to DT. In BUC3, we assume that A-PPLE, and therefore also the DT API, will be running at all companies and the business enterprise. We therefore focus our analysis on the implications of the deployment of the DT API at all service providers and the use of DT by Alice and Bob.

The main role of the DT API is to enable users of DT to exercise their rights with respect to their personal data at different service providers. This involves exposing API calls to: (i) download all the data subject's data at the service provider, (ii) request data to be corrected, and (iii) request that all data should be deleted (or at least revoke consent to processing). While the integration with TL provides strong authentication of the DT API, and support for Tor and TLS provides some resistance to traffic analysis as well as encryption of communication, the deployment of the DT API still introduces some new risks:

- Calls to the API can be profiled by the service provider. At the very least, the service provider can tell when specific calls are made by a particular data subject. This data is also personal data.
- The deployment of the API opens up (potentially) previously closed systems to the Internet. This increases the risk that personal data may leak, for example due to misconfiguration or simply the normalisation of the flow of personal data through the API.

DT, and the DT API, increases the impact of compromise on the privacy of the data subject using the DT. For starters, DT provides an overview of past data disclosures made by the data subject. This will (per definition) include personal data. Potentially even worse, the compromise of DT implies the compromise of the credentials to access the DT API. This means that the DT with its credentials (based on TL) and DT API together lays the necessary foundation for attackers to automatically harvest personal data of a compromised data subject at all service providers that support the DT API.

For Alice, as an employee of the business enterprise using her own device, her personal data is additionally exposed. Beyond the increased attack surface introduced by the business enterprise's software, obligations as part of her role as an employee may increase the risk of her device becoming compromised or seized by IT administrators. This would expose Alice's personal data stored both in DT, and available through the DT API at the healthcare and social network services she is using.

2.4.7. Data Protection Impact Assessment Tool (DPIAT)

DPIAT is used by, for example, the business enterprise before any processing of personal data. Therefore it has no impact on privacy in BUC3.

2.4.8. Data Subject Access Request tool (DSAR)

One objective of the Data Track tool is to allow data subjects, such as Alice and Bob in our example, to exercise their rights to access, and to request deletion or correction of their data at Cloud service provider sides online. As described above, for allowing data subjects to exercise those rights electronically, the respective service providers need to run the DT API. For the cases that the DT API for electronic online access by data subjects is not supported by a service provider, the DSAR tool can support the data subjects in drafting letters to be sent to this service providers by email or traditional (non-electronic) mail. Alice and/or Bob have to prove in this case (i.e. they have to be authenticated) to the service provider that they are the data subjects of the data that they want to access. For such authentication proofs, there is a risk that the data subject has to provide/prove more identifying personal information (e.g. the unique Personal Number in Scandinavian countries) than the service provider already knows about the data subject. If the DSAR is used as a plugin of the Data Track, the Data Track could provide pseudonymous authentication proofs to the DSAR tools for inclusion into the mail to be sent to the respective service provider.

Furthermore, mails for exercising data subject rights and replies to them by the service providers will typically contain personal information that need to be adequately protected from access by unauthorised parties, e.g. by encryption. Nonetheless, if mails drafted by the DSAR tool are sent electronically, traffic analysis may reveal further information about the data subject.

2.4.9. Data Transfer Monitoring Tool (DTMT)

The main purpose of the Data Transfer Monitoring Tool is monitoring personal data transfers. Information on location and flow of data can considered as privacy sensitive and their storage and processing should therefore be protected against malicious entities. Such information is even more exposed in the case of BUC3 whereby six companies are involved in the cloud environment and data may move anywhere.

2.4.10. Incident Response Tool (IRT)

The Incident Response Tool (IRT) informs both individual end users, such as Alice and Bob, as well as business users/cloud customers, such as the Business Enterprise in this use case, about security and privacy incidents in relation to their personal data that were outsourced to the Cloud and provides appropriate options to respond to the incidents. Messages informing about incidents may contain sensitive personal information about the data subjects concerned and their personal data that were compromised. For instance, the fact that Alice's Social Network or bank account was hacked can have negative implications on her financial credibility or social trustworthiness. Hence, both the content of these messages as well as the identity of the victims need to be protected. In particular, if the IRT tool directly informs Alice and Alice has acted so far pseudonymously in regard to the service that caused the incident, recipient anonymity should be provided towards this service.

2.4.11. Plug-in for Assessment of Policy Violation

The plug-in is a plug-in for the Data Track to assess the severity of policy violations locally. Beyond the risk of providing data subjects with inaccurate assessments, which may lead data subjects into taking inappropriate actions, the plug-in has a limited impact on privacy in BUC3.

2.4.12. Remediation & Redress Tool (R&RT)

The Remediation & Redress Tool can be used by an individual user, e.g. Alice,, or business user, e.g. the Business Enterprise in our use case, to respond to incidents that the user was informed about by the IRT or that the user suspects or heard about from other sources (e.g., via news reports). R&RT can either be activated by the user herself, or be triggered by the IRT. In the latter case, the user is already informed about the incidents that occurred and proposed actions. In this first case, R&T should help the user to contact the responsible service provider that outsourced the data (the Health Care Service - Company C) or to contact the cloud provider to which she outsourced data (e.g., Company F). In case that a service broker was involved, the tool should direct the user to the service provider that is behind the service broker. In order to determine which service provider needs to be contacted, the tool has to obtain personal data about the services used. For contacting the respective provider, Alice has to prove that it was her data which were potentially compromised. Besides, she has to describe details of the conflict that arouse, and may thus have to reveal personal context information. Hence, the following data, which could include personal data, is provided and processed by R&T:

- Authentication proofs that the party contacting the cloud provider is subject of the incident and thus may have the right to obtain redress;
- Information describing (cloud) services that were used and the disputes with those respective services sides/cloud services sides, which can be sensitive information, e.g. about Alice's health data that were illegally processed in the Cloud.

Potentially, less privacy issues arise if the tool is user controlled (i.e. if Alice runs it on her local machine) and not a central service provided by a third party. When Alice has to prove that her personal data were compromised, she should optimally not reveal more identifying data than the service or cloud service provider already knows about her. In particular, if she has used a service or cloud service anonymously or pseudonymously, it should not be required that she is now requested to identify herself.

2.4.13. Transparency Log (TL)

TL would presumably be deployed at every company as part of "the cloud" and the business enterprise. We start by looking at the impact of TL from the point of view of the customer Bob. When Bob discloses his personal data to the business enterprise, he will participate in the TL setup. This involves an exchange of some cryptographic material (ephemeral keys), that could take place, for example, when Bob register's his account with the business enterprise. From Figure 2 we can see that the business enterprise uses several cloud services composed by

Company A, B, E and F. As Bob's personal data is processed at the different companies, transparency logging should continue[4]. This is possible by extending the TL setup from one entity to another, which means that the business enterprise can enable other companies (and companies can in turn enable other companies as well) to send data through TL to Bob. We assume that each company (and the business enterprise) run their own TL server for log storage.

As soon as data is written to the TL API (running at the sender, so at the business enterprise and each involved company) until it is read by Bob, the design of TL provides strong privacy protection [PPW13]: data is encrypted, tamper-evident, the construction and reconstruction of the log enables unlinkable log entries, and the foundations are in place for recipient unobservability. Key concerns for TL involves what TL enables *before* data is written to TL and the impact of Bob's device being compromised:

- The intended use of TL is enable data to be sent from the business enterprise (and the cloud services it uses) to data subjects like Bob. This data may not have been shared at all if it was not for TL. This means that the introduction of TL may lead previously closed systems to open up their processing of personal data, which inherently increases the risk of this personal data being leaked. The design of TL can only hope to minimise this increase in risk, never fully eliminate it. For example, TL may normalise the flow of personal data from core software performing data processing (like a sales management system) to an additional system (the TL server) that has to be accessible anonymously over the Internet.

- With the ability of data subjects to retrieve personal data about them from TL, the impact of the compromise of a data subject's device increases. Using the ephemeral key material generated as part of a TL setup, an attacker can fully reconstruct (anonymously) all data logged to TL for the particular data subject (and associated ephemeral key, if no one-to-one mapping). As before, the design of TL can only strive to minimise this risk, not eliminate it.

We now shift our focus to Alice (Employee A), who like Bob (the customer) is a data subject in BUC3. When Alice setup her account at the healthcare service (Company C) and the social network service (Company D), she also presumably performed a TL setup for each respective account. Beyond the potential privacy threats introduced for Bob by TL, Alice has one additional risk: she is an employee of the business enterprise and uses her own device for processing business confidential data. This means that the risk that her device will be "compromised" due to legal or technical obligations related to her employment increases, which is exceptionally privacy invasive in BUC3 in conjunction with TL. For example, Alice's device could be seized by authorities as part of an investigation related to business confidential data Alice processed using her device, or compromised as a consequence of being forced to run vulnerable software due to enterprise IT policy[5]. The consequences of device compromise for Alice entails compromise of all (potentially sensitive) personal data sent through TL by the healthcare and social network services.

[4] At what granularity depends on other tools in the project and is at time of writing not yet apparent.

[5] A prime example would be the mandated use of a particular anti-virus software, or the client for accessing the sales management platform provided by Company A.

2.5. Summary

In the initial privacy analysis of the proposed A4Cloud tools and for the BUC3 analysis, we ignored common sources of security and privacy issues, such as misconfiguration or excessive application level logging. When it comes to tool compromise, our analysis differs for different tools. We took this approach simply because of the lack of information about tools, even at the time of writing this deliverable. Our initial analysis, provided to the respective tool owners as part of the work in WP C-7, raised privacy issues early in the design of tools. Further analysis with respect to potential privacy issues is surely needed as the tools mature.

The analysis of tools in regard to BUC3, as done in Section 2.4, highlighted the increased risk of tool "compromise" as a consequence of the data subject Alice bringing her own device for her role as an employee. Such compromise may have severe negative impact on her privacy, potentially revealing personal hence sensitive data with respect to Alice's health. Earlier in WP C-7, we also performed a privacy analysis of the tools in regard to BUC1 and BUC2, available in Appendix C. BUC1, like BUC3, involves a data subject whose sensitive personal data is processed by cloud providers. The sheer number of entities that may have access to personal data is daunting, but does not per se affect our analysis of the tools. BUC2 concerns primarily the processing of data subjects' shopping histories, which while being personal data, is not necessarily sensitive. The analysis of the tools in regard to the BUCs serve as a way to illustrate privacy issues with tools. Further analysis of the different BUCs and tools are surely needed, especially since the interplay between tools may as a whole introduce new privacy issues unforeseen at this relatively early stage of tool development and integration efforts.

3. Privacy by Design Approach

This section discusses privacy by design guidelines and proposes and exemplifies a privacy by design engineering approach based on previous related work. Within the lifetime of the A4Cloud project, the proposed approach should guide how to design and implement A4Cloud tools developed in stream D, privacy friendly. Besides, it will also provide guidance to other developers of related accountability or transparency tools beyond the A4Cloud project.

The following foundational privacy by design guiding principles have been defined and promoted by Ann Cavoukian, the Information and Privacy Commissioner of Ontario [Cav09]:

1. Proactive not Reactive; Preventative not Remedial
2. Privacy as the Default Setting
3. Privacy Embedded into Design
4. Full Functionality – Positive-Sum, not Zero-Sum
5. End-to-End Security – Full Lifecycle Protection
6. Visibility and Transparency – Keep it Open
7. Respect for User Privacy – Keep it User-Centric

While these foundational principles explain well the basic philosophy of privacy by design, Gürses et al. [GTD11] point out that it remains unclear how these principles can be practically guiding the engineering of privacy by design. In their paper "Engineering Privacy by Design" [GTD11], Gürses et al. are therefore proposing guidelines for an engineering approach to privacy by design, which consist of the following steps:

- ***Functional Requirements Analysis:*** As a first step, the functionality of the system, objectives and purposes of personal data processing need to be clearly defined. Vague descriptions with no clear purpose limitation could result in systems that collect more data to accommodate possible future uses for a broader range of purposes.
- ***Selecting PETs for Data Minimisation:*** At this step the data that is absolutely necessary for implementing the described functional requirements need to be analysed. This step also requires a survey of state of the art and a selection of PETs and alternative architectures that should be used for the specific system to technically enforce data minimisation.
- ***Modelling Attackers, Threats and Risks:*** At this step, possible attackers and the types of threats that they could conduct should be analysed. Hence, a security model should be defined and a corresponding risk analysis can be conducted afterwards.

- ***Multilateral Security Requirements Analysis:*** This step deals with the analysis and resolution of conflicting security requirements for achieving multilateral security. Optimally, a "positive sum" should be reached, which accommodates both security and privacy. Typically, conflicts may arise between the privacy and security requirements of anonymity and accountability, or confidentiality and availability.

- ***Implementation and Testing of the Design:*** In this final step, the system is implemented, tested for potential vulnerabilities and the functioning of the system according to its functional requirements is validated.

As Gürses et al. point out, while the functional requirements analysis clearly has to be the first step taken, there is no definitive order for the other activities. Rather, reiterations of steps may be needed at different points, e.g. functionality may be revised based on the results of the risk analysis, or data minimisation and the security requirement analysis may have to be iterated for achieving multilateral security.

The privacy by design guidelines that we recommend for A4Cloud are comprising the privacy by design engineering approach and steps defined by Gürses et al. In addition, we propose an additional step for

- ***Selecting PETs for Data Protection:*** Selecting PETs that enforce that personal data are processed in compliance with legal privacy requirements, such as informed consent, purpose binding, transparency and data subject rights and data security. This step also requires a survey of existing PETs for protecting personal data.

Typically, after PETs have been analysed and an architectural design for achieving data minimisation has been conducted, privacy risks should be re-analysed and PETs for data protection needs to be chosen for protecting the remaining personal data that could not be minimised.

As we pointed out in Section 2.1, the step "Functional Requirements Analysis" comprises the first two steps of our initial privacy analysis ("light PIA") of A4Cloud tools, namely the summary of the tool specification (specifying the objectives and purposes of data processing) and the summary of the information flow during the data's life cycle. The "Functional requirement analysis" as a PbD step should preferably include a more detailed functional description, which was however mostly not available for most of the A4Cloud tools in project year 1.

The step "Modelling Attackers, Threats and Risks" comprises the third and fourth step of our initial privacy analysis of A4Cloud tools, i.e. the analysis of privacy issues and of how far basic legal privacy principles are at stake.

As mentioned above, the selection of appropriate PETs for data minimisation and data protection first requires a survey of PETs. In Section 4 we will provide an overview of PETs that are suitable for A4Cloud technologies.

For the privacy by design of accountability tools, the multilateral security requirements analysis will be an important activity due to the inherent conflict between privacy and accountability. While accountability is an important privacy principle, accountability mechanisms may require additional personal data to be collected and retained (e.g., log data). In Section 5, we will therefore discuss in more detail how the balancing of privacy, accountability and other security protection goals for achieving multilateral security can be performed. Then, in Section 6

we will exemplify how the proposed privacy by design approach can be followed for a privacy-enhanced design of the A4Cloud Audit Agent System.

4. Suitable Privacy-Enhancing Mechanisms

In this section, we suggest a number of privacy-enhancing technologies (PETs) for A4Cloud technologies. As discussed in the previous section, our recommended privacy by design approach requires a survey of state of the art PETs for data minimisation and data protection. PETs can be defined as technologies that are enforcing legal privacy principles in order to protect and enhance privacy of users of information technology and/or data subjects [FH09]. Different PETs can be selected for enforcing legal privacy principles when taking a privacy by design approach for accountability tools (c.f. Section 3).

There are different classification schemes for PETs. For instance, recently [Hoe14] presented "Privacy Design Strategies" derived from data protection legislation, that provide a useful classification of privacy design patterns and the underlying PETs. However, it is important to keep in mind that A4Cloud tools can be seen as PETs by themselves, as they are enforcing legal privacy principles of transparency, user control and accountability. Nevertheless, as discussed in the previous sections, A4Cloud tools can require the processing of personal data, and may thus create new privacy risks. The list of PETs in this section is not a complete list of all prominent PETs, but is rather a selection of PETs that we specifically identified as being useful for a privacy-friendly design of A4Cloud tools during our analysis work.

We classify PETs into two categories; *application PETs* and *communication PETs*. Application PETs are protecting privacy on the application layer, while communication PETs are focused on providing communication privacy over a network. Furthermore, we distinct between *practical PETs*, for which broadly used open source implementation exist, and *research PETs*, for which only research papers or prototypes, but no directly usable practical implementation exist, to our knowledge. While the practical PETs could be easily deployed by A4Cloud tool developers within the project's lifetime, the research-types of PETs probably may only be considered in the long term.

4.1. Application PETs

Practical PETs

Encryption Encryption of application data is enforcing the legal privacy requirements of security of data processing[6] (c.f. Art. 17 EU Data Protection Directive). For most use-cases when the core provided functionality is enough, we *strongly recommend* finding an implementation of NaCl: Networking and Cryptography library[7]. Other practical encryption tools are for instance provided by the Legion of the Bouncy Castle[8] which is a collection of cryptographic APIs for Java and C# available under the MIT license. Bouncy Castle makes it easy to add support for different cryptographic standards, ranging from signing email following the OpenPGP standard to authenticated encryption of data following ISO/IEC 19772:2009.

[6] That is, processing in the legal sense. Homomorphic encryption, discussed later, protects data that is processed in the technical sense.

[7] `http://nacl.cr.yp.to/`, last accessed 2014-04-22

[8] `http://bouncycastle.org/`, last accessed 2013-12-09

Key management The use of encryption mechanisms to assure secrecy of course requires a dedicated key management mechanism. Only authorized entities should be able to receive relevant keys. The underlying key distribution mechanism may also ensure back and/or forward secrecy. For example, an end-customer may opt for the revocation of an earlier authorized entity. In this case, the encryption key should be updated and new keys should be distributed accordingly. SKS[9] is an open source PGP keysever, enabling setting up a web of trust. Support for user credential management, including key management, exists in all modern operating systems, such as in Windows with its Windows Credential Manager[10], in Linux with GNOME Keyring[11], and in OS X with Apple Keychain[12].

Identity Management User controlled identity management technologies are effective means for enforcing the legal privacy principle of data minimisation at application level. The Identity Mixer (idemix[13]) and U-Prove[14] technologies provides support for *anonymous credentials*, which enables cryptographically strong authentication with support for minimal disclosure of personal data (revealing only what is required), proving predicates over attributes (e.g., prove that age is over 18, without revealing actual age), and conditional anonymity (e.g., reveal the name on a drivers license credential if an accident happens). Implementations are freely available under permissive licenses in Java for idemix and C# for U-Prove. Anonymous credentials can for instance also be used to prove the ownership of different types of pseudonyms, such as transaction pseudonyms. Further technical means for securely implementing pseudonyms, which could for instance be used for anonymous/pseudonymous authentication, include self-generated transaction pseudonyms (e.g., in form of the public keys of users).

Research PETs

Homomorphic encryption Traditional encryption mechanisms, be they symmetric (such as AES) or asymmetric (such as RSA) fall short in a scenario where there is a need to process data in encrypted form. Homomorphic encryption mechanisms [Gen09, GH11, SV10, BV11] can be considered as candidates to achieve such a goal. They allow a third party to perform meaningful computation over encrypted data without access to the decrypted data. Although fully homomorphic encryption solutions may be considered as very expensive, some partially homomorphic encryption mechanisms also named as ”somewhat homomorphic” mechanisms which are homomorphic in a restricted set of operations can still be considered as a potential candidate for near-term use.

[9] `https://bitbucket.org/skskeyserver/sks-keyserver/wiki/Home`, last accessed 2014-01-20
[10] `http://windows.microsoft.com/en-us/windows7/what-is-credential-manager`, last accessed 2014-01-20
[11] `http://live.gnome.org/GnomeKeyring`, last accessed 2014-01-20
[12] `http://www.opensource.apple.com/source/libsecurity\_keychain/libsecurity\_keychain-27723/`, last accessed 2014-01-20
[13] `http://www.zurich.ibm.com/idemix/details.html`, last accessed 2013-12-09
[14] `https://uprovecsharp.codeplex.com/`, last accessed 2013-12-09

Privacy preserving word search Using specific efficient homomorphic encryption mechanisms it is possible to perform word search over encrypted data. The word search primitive enables verifying whether a certain word is part of an encrypted dataset or not. Such a primitive assures the search for expressions while preserving the privacy and unlinkability of the data, of the search queries and even of the result to the queries. In classical primitives [BDOP04, BKO07, CM05, CGKO06, OK04, SWP00, BdPMO12] only the owner of the data (the entity which encrypts the data) can generate queries correctly. Today's applications may also request the search by authorized delegated third parties such as auditors. As part of our work in C-7, we have defined a new delegated word search mechanism [EOM13] which allows authorised users to perform lookup operations and includes proper revocation operations.

Proxy re-encryption From a high-level viewpoint, a proxy re-encryption scheme is an asymmetric encryption scheme that permits a proxy to transform ciphertexts under a public key into ciphertexts encrypted under another public key. In order to do this, the owner of the data gives a re-encryption key to the proxy that makes this process possible, without learning anything about the underlying plaintexts and the private keys. Besides defining encryption and decryption functions, like any traditional public key scheme, a proxy re-encryption scheme also defines a re-encryption function for executing the transformation. Proxy re-encryption schemes [AFGH06, GA07, LV11, WDLC10] are of particular interest in cloud computing, specially when one considers scenarios that deal with the sharing of confidential or sensitive data. Data in the cloud can always be in an encrypted form, and the possibility of re-encrypting it to different actors facilitates the sharing.

Anonymisation and Pseudonymisation Anonymisation or pseudonymisation of data is a means for supporting data minimisation, and can for instance be used as an additional level of protection for personal data in log or audit data. If data is "anonymised", the Directive 95/46/EC does not apply, as its Recital 26 states that the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable. However, in practice, the question whether data is anonymous or not is very difficult to answer. This particularly applies to statistical data, where despite the fact that the information may be presented as aggregated data, the original sample is not sufficiently large and other pieces of information may enable the identification of individuals. For instance, data sets published by AOL, a media company, and by Netflix, a provider of on-demand streaming media, in 2006 that were claimed to be anonymised were later proven not to be since a number of individuals could be re-identified from the data set (see [NS09]). The Recital 26 demands that for deciding whether data is "anonymous" all the means likely reasonably to be used either by the controller or by any other person should be taken into account. Still, as an additional measure, means for anonymisation or pseudonymisation can further enhance privacy. Schemes for the pseudonymisation of log data were for instance previously proposed by [SFHR97], [Fle05], and [WM11].

Policy User Interfaces for obtaining consent For obtaining explicit consent from users, usable user interfaces for informing end users about the legally required policy information including, pursuant Art. 10 EU Data Protection Directive, at least about the identity of

the controller and data processing purposes. Research on usable user interfaces for the PrimeLife policy language (PPL) was for instance presented in [AFHWP12].

Usage Control Access control is a basic security mechanism for protecting the confidentiality of personal data. In addition, usage control schemes exist for enforcing legal privacy requirements, such as purpose binding, see e.g. the work presented in [FH01], [KS02], and PPL.

4.2. Communication PETs

The following list includes the most broadly used practical PETs for encryption and anonymisation of communication traffic. Further PETs for providing anonymous communication, such as DC-nets or Crowds, have been theoretically explored, but have hardly been used in practice due to several limitations, and as such they are currently not directly available for use. Therefore, we restrict the list of communication PETs to the following types of practical PETs:

Encryption of Communication Transport Layer Security (TLS) is a protocol for secure communication with a focus on data confidentiality, authenticity, and integrity. It is the de facto standard for securing communication on the Internet — the "S" for secure in "HTTPS" — responsible for providing a secure tunnel for HTTP traffic. Support for TLS is widespread, from web servers such as the Apache HTTP server[15] to client and server support in Java. TLS is a bare minimum for securing communication and should be used unless other means are in place for at least equivalent protection.

Anonymous Communication Anonymous communication technologies are important means for enforcing data minimisation by hiding the IP addresses of communication partners. Tor is a low-latency anonymity network run by the Tor project[16]. There are two actively maintained libraries for interacting with Tor: Stem and Netlib. Stem[17] is a library for controlling the Tor client using Python. Netlib from Silvertunnel[18] is a Java library that fully implements a Tor client. Further practical PETs for anonymous communication include for instance JAP[19] operated by TU Dresden and its project partners.

4.3. Selecting PETs for A4Cloud Tools

In a milestone report in project month 15, we have selected PETs for data minimisation and data protection for A4Cloud tools. This initial suggestion of PETs was communicated through this milestone report to all A4Cloud tool developers and is presented in Appendix D. The selection of PETs per tool is also summarised in the next section in Table 6.

[15] `https://httpd.apache.org/`, last accessed 2013-12-09
[16] `https://www.torproject.org/,lastaccessed2013-12-09`
[17] `https://stem.torproject.org/`, last accessed 2013-12-09
[18] `https://silvertunnel.org/netlib.html`, last accessed 2013-12-09
[19] `http://anon.inf.tu-dresden.de`, last accessed 2014-04-23

5. Balancing Privacy and Security Protection Goals

In the original methodology by Gürses et al., the multilateral security analysis is devoted to studying the effect of design choices, in particular those that orient privacy towards other important security objectives such as integrity, availability, etc. and vice versa. However, in the context of accountability, we are dealing with additional objectives, such as the ones related to the attributes of accountability [A4C14b]. Thus, the goal of this multilateral analysis is to assess the impact of the proposed privacy measures, in the form of privacy-enhancing technologies (PETs), with respect to the basic accountability objectives, which are represented by the attributes of accountability.

In Section 4, we proposed several PETs for privacy measures for accountability tools:

- Encrypted/Authenticated communications (M1), which comprehends mechanisms for achieving confidentiality and authentication of communications, such as the use of TLS/SSL (see Section 4.2).
- Anonymous communications (M2), which are proposed for protecting the anonymity of senders and receivers in a network. Examples of these mechanisms are Tor and JAP (see Section 4.2).
- Encryption of data at rest (M3), used for protecting confidentiality of stored data (see Section 4.1).
- Key Management (M4), which is intimately related to the use of encryption technologies.
- Strong Access Control (M5), proposed for restricting access to sensitive data only to authorized parties and for valid purposes.
- Multifactor Authentication (M6), which mandates that the authentication procedure uses two or more authentication factors, namely knowledge, possession and inherent factors.
- Anonymisation and Pseudonymisation of Data (M7), which is proposed as an additional measure for protecting stored personal data (see Section 4.1).
- Privacy-Preserving Word Search (M8), which can also support privacy-respectful delegated search to third parties such as auditors (see Section 4.1).
- Proxy Re-Encryption (M9), which enables sharing encrypted information to authorized parties through delegation of decryption rights (see Section 4.1).

After reviewing the description of each A4Cloud tool and the initial privacy suggestions presented in Appendix B, we summarize in Table 6 which are the privacy measures applicable to each A4Cloud tool:

Tool	M1	M2	M3	M4	M5	M6	M7	M8	M9
A-PPLE	X		X	X	X				
Audit Agent System	X		X	X	X	X	X	X	X
Data Track	X	X	X						
Data Transfer Monitor Tool	X		X		X				
Redress tool	X	X	X	X	X				
Transparency Log	X	X	X			X			X

Table 6: Summary of proposed privacy measures for A4Cloud tools

Once we have identified the proposed privacy measures, we have to perform an impact analysis of them with respect to the accountability attributes, namely Transparency, Verifiability, Attributability, Observability, Responsibility, Remediability and Liability. However, it is not necessary to analyse each potential combination (9 privacy measures $\times$ 7 accountability attributes $=$ 63 combinations). Instead, we will perform an *a-priori* accountability impact analysis of these privacy measures. In Table 7, we identify which accountability attributes are *a-priori* affected by each privacy measure.

Proposed privacy measure	Transp.	Verif.	Attrib.	Observ.	Respons.	Remed.	Liab.
M1: Encrypted/Authenticated communications	X	X	X	X			X
M2: Anonymous communications				X			
M3: Encryption of data at rest		X		X			
M4: Key Management		X		X			
M5: Strong Access Control			X	X			X
M6: Multifactor Authentication		X	X	X	X		X
M7: Anonymisation of Data	X	X		X		X	
M8: Privacy-Preserving Word Search		X	X	X			
M9: Proxy Re-Encryption		X		X			

Table 7: A-priori accountability impact analysis of proposed privacy measures

A first result of this analysis is that the attributes which seem more impacted are Observability, Attributability and Verifiability, which is understandable since the goal of these privacy measures is to minimise the knowledge gain of sensitive information by third parties. Next, we provide a more detailed justification of the analysis in Table 7, for each proposed privacy measure:

M1: Encrypted/Authenticated communications Encrypted communications between A4Cloud tools' components may negatively affect logging of the communications, rendering the job of auditors more difficult. Hence, verifiability, attributability and observability could be affected. At first sight, Transparency may seem to be affected, but it is worth recalling that in the accountability context defined by the Conceptual Framework [A4C14b], Transparency is oriented towards the data controller being open regarding its policies and procedures, so the use of encrypted communications does not impact on this.

With regard to the use of authenticated communication channels, it is clear that this is a necessary feature for assuring the identities of the communication endpoints. Therefore, insufficient assurance regarding the authenticity of communications could have a negative impact in important aspects of accountability, such as Verifiability, Liability and Attributability. Note that although authentication and confidentiality are different security properties, certain mechanisms, such as public-key cryptography, achieve both of them. In order to ensure proper levels of confidentiality and authentication, the use of standard mechanisms such as TLS is advised.

M2: Anonymous communications The purpose of using anonymous communication technologies in some communication channels within the A4Cloud tools is to disallow the tracing of communications of the participating entities. In particular, they are proposed for use in the communication between users and service providers (in the case of the Data Track and the Redress Tool) and between the Transparency Log and service providers. The a-priori analysis reveals that observability may be affected by the use of such technologies, since from a strict point of view, observability is related to the ability of gaining knowledge about the internal state of a system by analysing its outputs. If the ability of tracing communications is hindered, then the capacity of an observer for knowing the internal state of the communicating entities is decreased. However, considering the discussion on the accountability attributes in [A4C14b], we can argue that the use of anonymous communications is reasonable in the aforementioned situations, since the purpose of using anonymous communications is to protect communicating entities from, possibly malicious, third parties that try to trace their traffic over the Internet.

It may be useful for accountability that previously anonymous communications may be "de-anonymized". This would imply the use of specific anonymisation schemes that permit to perform this process. In this case, the anonymisation mechanisms could, under some circumstances, support Observability.

M3: Encryption of data at rest This privacy measure is a common practice for all the A4Cloud tools; examples of this are the Audit Agent System, which handles sensitive information in the form of pieces of evidence, and the Transparency Log client, which deals with encrypted log entries and credentials. Therefore, we must ensure that the encryption mechanisms in place are appropriate and do not negatively impact accountability. The main properties potentially affected by the use of encryption are Observability and Verifiability, since encrypted data cannot be interpreted nor analysed. However, in the same fashion as for the use of encrypted communications and anonymous communications, the use of encryption for data at rest is not only reasonable but necessary in some cases.

With respect to this privacy measure, one aspect that could be measured is the scope of the encryption process. For instance, when encrypting a log, we could encrypt the whole file, or encrypt each entry separately keeping the timestamp in clear text. It is clear that the second option would make the system more observable (and therefore more accountable), as it facilitates the interpretation of the log. On the other hand, the disclosure of the timestamps by itself, could indirectly leak useful information to an attacker in case the log file is compromised. It is very difficult to provide a specific metric for this aspect, since

discerning which pieces of data should be encrypted highly depends on the context of application. This aspect is also intimately related to other privacy measures, such as the use of advanced encryption schemes that enable certain processing on encrypted data, e.g. Privacy-Preserving Word Search (M8) and Proxy Re-Encryption (M9).

The quality of encryption is also intimately related to the quality of key management, which we analyse below.

M4: Key Management Observability and Verifiability require access to data by different parties. Depending on how the system choose to manage encryption keys, some parties may be unable to access data. Limiting the exposure of the cryptographic keys to the maximum level will clearly benefit privacy, but at the same time, might negatively impact Observability and Verifiability.

As described in [A4C13], when encryption is used to protect the confidentiality of data at rest in the cloud, there are many approaches to the distribution and the use of cryptographic secrets that ultimately protect the data in the cloud. The cloud client can encrypt the data before it even reaches the cloud, with a key that is only known to the client. More often, however, the cloud provider will encrypt the data with a key that is under its control. We can identify several levels of key exposure that reflect the degree of confidentiality afforded to cryptographic secrets, from a cloud client point of view:

Level	Description
1	Access to decrypted data or cryptographic secrets by the CSP is necessary to provide some functionalities of the service.
2	Access to decrypted data or cryptographic secrets is available to specific personnel of the CSP, for administrative or debugging purposes only.
3	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.
4	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.
5	Cryptographic secrets needed to decrypt the data are known to the cloud client only.

Table 8: Key Exposure Level

In most of the situations where encryption of data at rest is advised, such as for protecting logs, levels 2, 3 and 4 of key exposure are appropriate. In these cases, access by the CSP to cryptographic keys is permitted only under special circumstances, which could be used for example during audit tasks, thus supporting Observability and Verifiability. However, there are some cases where the maximum level is required, such as for protecting the database of disclosures in the Data Track.

Another aspect to worth considering is the problem of key revocation. When there is a key exposure, there should be a revocation and rekeying mechanism and this usually is part of the key management solution. Another interesting aspect related to this measure is key recovery mechanisms.

M5: Strong Access Control It is clear that proper access control is needed to support accountability, as it is a key mechanism for protecting the data entrusted by users and the systems and tools that handle sensitive information. Strong access control would have a beneficial impact on accountability attributes such as Attributability and Liability, since it provides proper traceability of actions to the actors responsible of them. With regard to negative impact, we observe that Observability is affected, since access to information from the system is restricted, for example for monitoring and auditing purposes. However, in the same way as for other privacy measures, if the boundaries of access control are reasonable, then the benefits outweigh the disadvantages. Thus, we must guarantee that the implemented access control measure is of good quality and that it leaves reasonable room for supporting Observability.

In order to assess the quality of access control systems, there already exist well-established criteria, such as the *Guidelines for Access Control System Evaluation Metrics* from NIST [HS12]. These guidelines are extensive, so we should focus only in the most relevant aspects for our purposes.

M6: Multifactor Authentication The analysis in Table 7 indicates that the use of authentication mechanisms may have an impact in several accountability attributes, namely Verifiability, Attributability, Observability, Responsibility and Liability. This is due to the fact that the quality of the authentication mechanisms in place have an effect on how well the true identity of actors can be linked to actions in the system; although this is mostly related to the concept of Attributability, it also impacts the aforementioned attributes. Thus, high-quality authentication mechanisms should be used. A suitable metric for measuring the quality of these mechanisms is the *levels of assurance* provided by the NIST standard 800-63-1 [BDN$^+$11]. This metric is a well-defined set of levels to measure the quality of authentication mechanisms, to which we can add a level 0 describing no authentication at all. At level 1, simple challenge response mechanisms are allowed and no identity proofing is required. At 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 correspond to multifactor authentication, where proofs of control of the authentication token are done through a cryptographic protocol. Level 4 represents the strongest degree of assurance, where multi-factor 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.

In the case of A4Cloud tools, multifactor authentication is advised in the Transparency Log and in the Audit Agent System, as described in Appendix B. The use of authentication mechanisms with a lower assurance level, namely levels 0, 1 and 2, will negatively impact accountability, and therefore, must be avoided, since the potential impact of impersonating end-users (in the case of the Transparency Log) and auditors (in the case of the Audit Agent System) is very high. For this reason, in these cases is required at least level 3, where multifactor authentication and cryptographic protocols are in place.

M7: Anonymisation and Pseudonymisation of Data As already noted in Section 4, anonymi-

sation and pseudonymisation are important privacy mechanisms towards achieving data minimisation. However, it is clear that full anonymisation would render data of little use to any kind of auditing task, as already pointed out in Appendix D. It may also affect Remediability. Thus, in this point there is a trade-off between data anonymisation and effectiveness. Nonetheless, some specific anonymisation tools are proposed, such as the rsyslog module mmanon and the Apache module mod_anonstats, which are intended for minimizing data relative to IP addresses. The first one provides anonymisation by removing the last bits of the IP addresses that are logged; the number of bits to be removed are configurable, but the default value is 16, which is said in their website to be compliant to German laws regarding data protection. In this case, one could assess the degree of anonymisation provided by this tool by measuring the size of the resulting anonymity set, using some of the privacy metrics described in [A4C13]; in order to further analyse the effect of anonymisation one should study the purpose of auditing tasks and its requirements regarding identification of senders and receivers. The second tool provides two types of anonymisation, namely, full and hash. Full anonymisation replace all IP addresses by 127.0.0.1, whereas hash anonymisation replaces each IP with a salted hash, and the salt changes every few minutes. In this case, full anonymisation would not be recommended for use in A4Cloud as it completely negates the potential of audit tasks.

A possible side-effect of the use of anonymisation and pseudonymisation of users' data is that it may difficult users' requests for access or modification of their personal data, which is an important aspect of Transparency, as understood by A4Cloud; however, if this data is properly minimised, it cannot be considered personal anymore.

M8: Privacy-Preserving Word Search This PET allows auditors to query a database manager (e.g. the evidence store of the AAS), without having access to the entire database and without the database manager being able to access neither to the data, nor to the queries and the responses. This can be a useful tool for AAS if evidence is outsourced and the server storing data (evidence) should not have access to the data or to the queries.

It is clear then that by limiting the scope of auditing, one is potentially diminishing the levels of Observability, Verifiability and Attributability. However, as in the case of other proposed privacy measures, the key aspect here is to identify if the limitation is within reasonable boundaries, which would make an optimal impact on accountability by decreasing privacy risks while enabling functionality. In addition to this, since we are dealing here with encrypted data, the same analysis than for encryption of data (M3) holds.

In order to do so, we would need that the requirements regarding the audit tasks are fully established, so there are clear limits to the scope of auditing. Nevertheless, as it is pointed out in the next section, there is currently not much information to this matter.

M9: Proxy re-encryption Being a kind of encryption scheme, in principle proxy re-encryption has a similar impact as general encryption (M3), as it happens with the previous privacy measure. However, proxy re-encryption facilitates the sharing of encrypted information through the delegation of decryption rights to authorized parties. That is, encrypted data

is no longer visible just for the owner of the original decryption key but also to other entities. Thus, there is a potential of having a lesser effect on accountability.

Another important aspect of proxy re-encryption, if it is finally used in any A4Cloud tool, is that in order to minimise risks one should choose proxy re-encryption schemes that are collusion-resistant, such as [AFGH06] and [LV11]. This kind of proxy re-encryption schemes protect the secret key of the delegator in case of collusion of the proxy with the delegatee, a situation which we cannot rule out in our scenarios.

6. Example: Privacy by Design of the Audit Agent System

Here we briefly perform one iteration of our privacy by design guidelines for the Audit Agent System. We start with the (understandably) high-level functional requirements of the tool and briefly perform each activity once. We end our example with some reflections.

6.1. Functional Requirements Analysis

We identify three key functional requirements of AAS based on the second WP D-2 survey:

R1 Given an audit task and A-PPL policy, configure software agents to determine what data to collect.

R2 Agents store collected data in an evidence store.

R3 Based on data in the evidence store, perform analysis (based on the audit task and A-PPL policy) and produce a report.

Another way to state R1 is that there should be dynamic and configurable data collection from an environment, which suggests that high system privileges may be needed in the environment. Since one of the main points of AAS appear to be possibility to construct different agents for different tasks, we cannot make R1 more precise in terms of what should be possible to collect (and therefore also not how it should be collected).

Once data has been collected, it is to be stored in an evidence store (R2). Based on the data in the evidence store, a component should produce a report (R3). The report is the main output of AAS, based on the main input: an audit task and a A-PPL policy. The evidence store is just an intermediate component.

Our primary privacy goals are to:

1. Minimise the impact of agents as far as possible, in particular by requiring strong authentication for defining audit tasks and for where they can execute.
2. Minimise the amount of data stored in the evidence store, protect such information from unauthorized access and delete it as soon as possible.
3. Only enable access to data in the evidence store that is strictly needed.
4. Keep reports private. Only include raw data from the evidence store when it is in the interest of the data subject.
5. Ensure that all actions in the system can be traced in case of conflict.

6.2. Multilateral Security Requirements Analysis

We have the following stakeholders:

CSP The cloud service provider running an instance of AAS. The CSP has a set of agents it can deploy throughout its own environment.

Auditor An auditor is an individual tasked with performing an audit. He or she will use the AAS instance at the CSP to perform the audit.

Other CSPs Other CSPs than the primary CSP, that may either be using the service offered by the primary CSP or are used by the primary CSP, as part of a service delivery chain.

Other Agents Other agents than the primary agent tasked to audit the primary CSP. They may also audit the primary CSP, or audit other CSPs.

Conflicting goals appear first when one considers the interests of a malicious CSP whose audit may reveal wrongdoing or a dishonest auditor that wishes to forge an audit report. When it comes to accountability attributes, we have little impact towards our main stakeholders since we do not (at this point in time) limit the collection of data by agents. This may change towards other CSPs if the process of producing reports aggregates data in such a way that it destroys essential information (which it should not, but may). For third-parties, the impact on attributes such as observability is significant (as it should be) due to our use of strong PETs.

6.3. Modelling Attackers, Threats and Risks

CSPs are considered as being honest-but-curious, which means that CSPs perform all operations in a correct way regarding the deployment and processing of their local agents; however, the CSP may be interested on discovering what is audited and the content of evidence store at others' premises in the service chain.

Auditors, on the other hand, are also considered as being honest-but-curious; the main difference between CSPs and auditors is that Auditors should only be able to access to the report corresponding to their task and should not be able to retrieve any additional information.

6.4. Defining PETs for Data Protection

In order to cope with the honest-but-curious model for CSPs and the semi-trusted model for Auditors, we propose to include privacy preserving primitives for the evidence store: Indeed, all stored information in the evidence store need to be encrypted and only accessed by authorized entities such as Auditors or primary CSPs. The underlying encryption mechanism should still allow Auditors to perform some operations over the data; thus the encryption solution should implement homomorphic operations. Moreover, AAS may implement privacy preserving delegated word search solutions to allow Auditors to query some parts of the data stored at the evidence store given their credentials; thanks to this primitive neither the evidence store nor an external entity will discover the content and the result of queries.

Additionally, since Auditors are not fully trusted, we propose to use the Transparency Logging tool to store an audit trail of:

- All audit tasks issued by auditors for a particular policy.
- All "credentials" issued by the CSP to auditors to query the evidence store and to which audit task this relates.

- The resulting audit reports.
- All transfers of audit reports from the CSP to other CSPs or auditors.

The audit trail serves as a transcript of all events in AAS and may resolve disputes between CSPs and Auditors. Additionally, TL may also implement proxy re-encryption solutions in order to authorize other Auditors or the primary CSP to access audit reports generated by other CSPs. Thanks to such solutions the TL will never be able to access the content of the audit reports while still allowing the primary CSP or other Auditors to decrypt the dedicated content.

6.5. Data Minimisation

We have one evidence store per CSP, with no outside access. The only data that leaves a CSP is in the form of audit reports. We assume that reports actually aggregate data, and therefore acts as a mechanism for data minimisation. While the agents at the CSP may collect excessive data, the CSP only gives auditors the ability to do specific queries over the collected data at the evidence store.

6.6. Implementation and Testing of the Design

Our design after one iteration is as follows. A CSP has agents it fully controls scattered around its environment. With strong authentication of auditors, auditors can issue audit tasks related to privacy policies. All actions by auditors are logged. Based on the audit tasks, auditors get credentials for querying the evidence store for data. Based on the data from queries, reports are generated and stored in TL. The CSP can share data in TL by proxy re-encryption with other CSPs or other auditors. Figure 3 gives an overview of the high-level design.

The data flow in Figure 3 is as follows:

Continuous logging The CSP's AAS controller continuously logs all its actions to TL. It uses itself as the recipient. This ensures a non-repudiable and tamper-evident trace of all its processing related to AAS.

1. **Setup** The auditor and CSP mutually authenticate using TLS. On successful authentication, the auditor specifies which audit tasks he or she would like to perform. The CSP logs the audit tasks and creates appropriate key material that enables the auditor to query the evidence store for the appropriate words.
2. **Audit task** The CSP instructs its agents to perform the audit tasks.
3. **Store evidence** The agents store the resulting evidence from the audit tasks in the evidence store in encrypted form. Ideally, the agents minimise the collected data to the bare minimum to achieve the audit task.
4. **Query** The auditor queries the evidence store using the credentials from step 1.
5. **Report** Based on the results from the evidence in the evidence store, the auditor creates a report. The report is stored in TL with the CSP as the recipient.

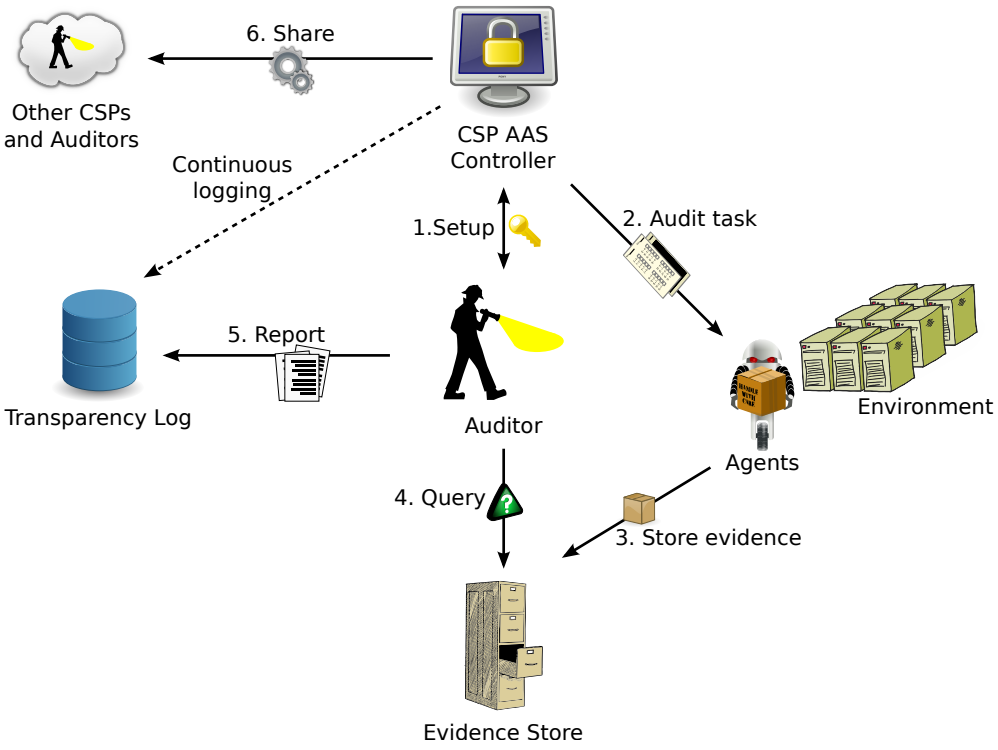

Figure 3: High-level design of AAS after one iteration

6. Share The CSP can share reports stored in TL with other CSPs or auditors by using proxy re-encryption of the encrypted report stored in TL.

We have not implemented or tested our design. We can however identify several shortcomings after one iteration. For starters, we know too little about what an "audit task" is, what the agents are supposed to detect, and what is meant by "evidence". The result in our design is that agents are still extremely powerful and we do not know how to minimise the collected data. Another consequence of knowing too little about what the goal of an audit task is, is that we cannot determine if (even excessive) data collection may be in the interest of the data subject. As it is now, the software agents may pose a big security risk to the CSP with little to no identified gain for data subjects (or auditors). Yet another example is the potential impact of reports being spread to other CSPs and auditors. Without exact knowledge of what are in the reports and how it is generated, the negative impact on privacy may be significant. The main problem, in terms of our guidelines, is that we cannot make functional requirements detailed enough. From vague requirements come overly broad solutions, and our design iteration is a prime example of this. It may be that it is impossible to make a "privacy friendly" agent system with fully flexible agents, and as such, it makes no sense to talk about "privacy by design" for too invasive tools. Furthermore, one of the stakeholders which seems to miss in this design is the Data Subject. We believe that Data Subjects (DS) should also have the right to access Audit reports. Such a capability/permission may also raise additional privacy problems like accessing information about other Data Subjects for example.

7. Conclusions

This deliverable first provided a concise analysis of ten A4Cloud tools among which some were already implemented while others were at the very early stage of design. Each tool is evaluated with respect to different privacy issues and the underlying business use cases, and a number of privacy-enhancing technologies (PETs) are suggested accordingly. The first privacy analysis showed that each tool presents different privacy issues. The proposed PETs are classified into two main categories: application and communication layer PETs. While the deliverable suggests some practical solutions, it also enumerates some "'research-type"' technologies that seem very appropriate although not ready to be implemented. Finally, the deliverable defines some guidelines any tool developer could follow in order to end up with a privacy friendly solution. This approach is illustrated with the application to the Audit Agent System and show that even with one iteration, some issues already raise and should be taken into consideration.

Providing privacy by design guidelines for the specific A4Cloud tools at an early project stage was in practice not so easy, because the tool specification were either changing very frequently or partly not available in detail in the beginning of the project, as the D stream only started in year 2. However, we believe that the analysis and guidelines provided in this deliverable, which were largely already communicated to the A4Cloud tool developers in a milestone report in project month 15, could and can in the future still influence the work on the design and implementation of any accountability tool, including relevant A4Cloud tools within stream D.

References

[A4C13] A4Cloud Project. D:C-5.1 – Metrics for accountability. Technical report, A4Cloud, 2013.

[A4C14a] A4Cloud Project. Glossary — accountability for the cloud. `http://www.a4cloud.eu/lexicon/glossary/`, April 2014.

[A4C14b] A4Cloud Project. MSC-2.3 – Conceptual framework. Technical report, A4Cloud, 2014.

[AFGH06] G. Ateniese, K. Fu, M. Green, and S. Hohenberger. Improved proxy re-encryption schemes with applications to secure distributed storage. *ACM Transactions on Information and System Security (TISSEC)*, 9(1):1–30, 2006.

[AFHWP12] Julio Angulo, Simone Fischer-Hübner, Erik Wästlund, and Tobias Pulls. Towards usable privacy policy display & management for primelife. *Inf. Manag. Comput. Security*, 20(1):4–17, 2012.

[BDN+11] William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, and Emad A. Nabbus. Electronic authentication guideline. SP 800-63-1, NIST, 2011.

[BDOP04] D. Boneh, G. DiCrescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In *Proceedings of Eurocrypt 2004*, pages 506–522, Barcelona, Spain, 2004. LNCS 3027. ISBN 978-3-540-72539-8.

[BdPMO12] E.-O. Blass, R. di Pietro, R. Molva, and M. Önen. PRISM - Privacy-Preserving Search in MapReduce. In *Proceedings of the 12th Privacy Enhancing Technologies Symposium (PETS 2012)*. LNCS, July 2012.

[BKO07] D. Boneh, E. Kushilevitz, and R. Ostrovsky. Publickey encryption that allows pir queries. In *Proceedings of Crypto 2007*, pages 50–67, Santa Barbara, USA, 2007. LNCS 4622. ISBN 978-3-540-74142-8.

[BV11] Z. Brakerski and V. Vaikuntanathan. Efficient Fully Homomorphic Encryption from (Standard) LWE. In *Proceedings of the 52th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011.

[Cav09] Ann Cavoukian. Privacy by design: The 7 foundational principles. 2009.

[CGKO06] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In *Proceedings of ACM Conference on Computer and Communications Security, CCS*, pages 79–88, Alexandria, USA, 2006.

[CM05] Y.-C. Chang and M. Mitzenmacher. Privacy preserving keyword searches on remote encrypted data. In *Proceedings of Applied Cryptography and Network Security (ACNS)*, volume 3531, pages 442–455. LNCS, 2005. ISBN 3-540-26223-7.

[EOM13] K. Elkhiyaoui, M. Önen, and R. Molva. Privacy preserving delegated word search in the cloud. In *Workshop on Trustworthy Clouds, in connection with ESORICS*, September 2013.

[FH01] Simone Fischer-Hübner. *IT-Security and Privacy - Design and Use of Privacy-Enhancing Security Mechanisms*, volume 1958 of *Lecture Notes in Computer Science*. Springer, 2001.

[FH09] Simone Fischer-Hübner. Privacy-enhancing technologies. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 2142–2147. Springer US, 2009.

[Fle05] Ulrich Flegel. *Pseudonymizing audit data for privacy respecting misuse detection*. PhD thesis, 2005.

[GA07] M. Green and G. Ateniese. Identity-based proxy re-encryption. In *Applied Cryptography and Network Security*, pages 288–306. Springer, 2007.

[Gen09] C. Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of 41st ACM Symposium on Theory of Computing (STOC)*, pages 169–178, 2009.

[GH11] C. Gentry and S. Halevi. Implementing Gentry's Fully-Homomorphic Encryption Scheme. In *Proceedings of EUROCRYPT*, pages 129–141, 2011.

[GTD11] Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering Privacy by Design. In *Conference on Computers, Privacy & Data Protection, 25-28 January 2011*, January 2011.

[HKW12] P. De Hert, D. Kloza, and D. Wright. A privacy impact assessment framework for data protection and privacy rights. 2012.

[Hoe14] Jaap-Henk Hoepman. Privacy design strategies. In *Proceedings of the 29th International Information Secruity and Privacy Conference (IFIP SEC)*. Springer, 2014.

[HS12] Vincent C. Hu and Karen Scarfone. Guidelines for access control system evaluation metrics. IR 7874, NIST, 2012.

[KS02] Günter Karjoth and Matthias Schunter. A privacy policy model for enterprises. In *Proceedings of 15th IEEE Computer Security Foundations Workshop (CSFW)*, pages 271–281, 2002.

[LV11] B. Libert and D. Vergnaud. Unidirectional chosen-ciphertext secure proxy re-encryption. *IEEE Transactions on Information Theory*, 57(3):1786–1802, 2011.

[NS09] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Proceedings of the 30th IEEE Symposium on Security and Privacy*, pages 173–187. IEEE, 2009.

[OK04] W. Ogata and K. Kurosawa. Oblivious keyword search. volume 20, pages 356–371, April 2004. ISSN 0885-064X.

[PPW13] Tobias Pulls, Roel Peeters, and Karel Wouters. Distributed privacy-preserving transparency logging. In Ahmad-Reza Sadeghi and Sara Foresti, editors, *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 83–94. ACM, 2013.

[SFHR97] Michael Sobirey, Simone Fischer-Hübner, and Kai Rannenberg. Pseudonymous audit for privacy enhanced intrusion detection. In *International Information Security and Privacy Conference (SEC)*, pages 151–163, 1997.

[SV10] N. P. Smart and F. Vercauteren. Fully homomorphic encryption with relatively small key and ciphertext sizes. In *Proceedings of the 13th International Conference on Practice and Theory in Public Key Cryptography (PKC)*, 2010.

[SWP00] D.X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 44–55, Berkeley, California, 2000.

[WDLC10] Jian Weng, Robert H Deng, Shengli Liu, and Kefei Chen. Chosen-ciphertext secure bidirectional proxy re-encryption schemes without pairings. *Information Sciences*, 180(24):5077–5089, 2010.

[Wes70] Alan F. Westin. *Privacy and freedom*. Atheneum, New York, 1970.

[WM11] Stefan G. Weber and Max Mühlhäuser. Multilaterally secure ubiquitous auditing. In *Intelligent Networking, Collaborative Systems and Applications*, pages 207–233. 2011.

A. Template for Initial Privacy Analysis of Tools

A.1. Tool Specification

to be filled in for each tool

A4Cloud tool specification summary	
WP	*WP delivering the tool*
Contact	*A4Cloud participant in charge of tool development*
Type of Personal Data Processed	*e.g., content data, traffic data, location data,*
Data Subjects	*Individuals about whom the tool collects data*
Purpose of Data Processing	*For what purposes are the data collected and processed?*
Data Controller	*Party which alone or jointly with others determines the purposes and means of the processing of personal data (EU Directive 95/46/EC)*
Data Processors	*Parties which process personal data on behalf of the controller*
Stakeholders (tool users)	*e.g., Cloud Consumers, Cloud Providers, Cloud Auditors, and Cloud Carriers.*
Retention Period	*How long are the data retained?*

A.2. Information Flow during the Datas Lifecycle

By whom are the data collected? How are personal data transferred / flowing between different parties? Where/when are the data destroyed or anonymised?

A.3. Initial Privacy concerns

What privacy issues arise by the tool that need to be addressed?

A.4. Privacy and Data Protection Principles

How far are basic legal privacy requirements an issue or how far are they already considered? What means of protection are already taken by the tool design?

Privacy and Data Protetcion Compliance Check	
Informed Consent	*Is informed consent obtained from the data subjects? How?*
User Control	*Do data subjects have access to their data? Can they decide/configure who has what type of access to their data?*
Purpose Binding	*Are there controls in place to limit the use of the data for specified and legitimate purposes only?*
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	*Is the collection and use of personal data minimised? Are the data adequate, relevant and not excessive? Are they kept in a form which permits identification of data subjects for no longer than is necessary?*
Transparency & Data Subject Rights	*How far are data subjects informed about the data collection and processing?*
Confidentiality, Integrity & Availability	*Are appropriate technical security measures taken to protect the data?*

B. Initial Privacy Analysis of Tools

Table 9 summarises all the A4Cloud tools identified for analysis during the first year of the project. This list is provided as-is as documentation of the work in WP C-7. Some tool ideas have been scrapped (like the Cloud Control Tool), and others are now more mature (like the WP D-3 tools). A new privacy analysis may yield significantly different results. The initial analysis of the Audit Agent System and the Transparency Log was presented in Section 2.

The initial privacy analysis of tools were conducted by sending out a survey to tool owners that were filled in together with a member of WP C-7.

This survey was divided into three parts: The initial part includes questions about the owner of the tool, its purpose and the different stakeholders playing a role within it; the survey was further analysing the information flow during the data's life-cycle, the initial privacy concerns the tool owners were already aware; finally, the owners were asked on whether some privacy or data protection principles such as data minimisation, data confidentiality or purpose binding were already taken into consideration.

After the initial C-7 survey, WP D-2 conducted their own survey of tools. For each tool below we include the summary from the D-2 survey that describes the purpose of each tool in the tool owners' own words.

B.1. Accountability and Privacy Enforcement Tool

From the D-2 survey:

> Enable enforcement of policies in ways that can be verified and audited. The tools that are developed require a secure architecture facilitating verification of security and privacy requirements by external auditors as well as can provide notifications to end users.

Accountability and Privacy Enforcement Tool	
WP	D-3
Contact	Jakub Sendor
Type of Personal Data Processed	all kind of data including business sensitive data
Data Subjects	End-users, employees of DC, depends on the use case
Purpose of Data Processing	Enforcing accountability policies (agreed between the cloud provider and consumers) Example: data deletion, event interception
Data Controller	Depends on the use case.
Data Processors	
Stakeholders (tool users)	All
Retention Period	Depends on, defined by the policy.

Table 10: Summary for the Accountability and Privacy Enforcement Tool.

B.1.1. Information Flow During the Data's Life-Cycle

No answer yet. Subject to discussion among partners.

Tool	WPs	Stakeholders	Purpose
Accountability and Privacy Enforcement Tool	D-3	CCs, CPs, CAs	Enable enforcement of policies in ways that can be verified and audited
Audit Agent System	C-8	CCs, CPs and CAs	Enable stakeholders to perform audits of their cloud infrastructures
Cloud Control Tool (CCT)	C-4/D-6	CPs	The declarative definition of implementation techniques and tools for the addition of new functionalities to existing Cloud applications and infrastructures
Data Protection Impact Assessment Tool	C-6	CCs (organisational users) and CPs (within the service provisioning chain)	Selecting service providers by assessing the impact of a particular choice before sensitive information is transferred or shared
Data Track	D-5	CCs (end-users)	Enabling data subjects to exercise their rights online to access, correct and request removal of their data at data controllers
Data Transfer Control Tool	D-3	CCs, CPs, CAs	Accountable data localization and transfer across cloud software, platform and infrastructure services, usually run by data processors
Incident Response Tool	D-4	CCs (end-users)	Provides data subject access rights and to file complaints regarding the processing of personal data by the controller and/or an incident
Plug-in for Assessment of Policy Violation	C-5	CCs (end-users)	Provide an assessment on the relevance of the violation of policies, so that stakeholders receive information about the most relevant policy violations
Redress Tool	D-4	CCs	Advises end users and organisations on actions that can be taken in case of violations
Tool for Cloud Contracts	D-4	CCs (organisational users)	A simple expert system to help with drafting and managing custom contracts
Transparency Log	D-5	CCs (end-users) and CAs	Transfer of data from CPs to CCs/CAs in a secure and privacy-friendly way

Table 9: All tools initially (put together around M6) planned for development in A4Cloud.

B.1.2. Initial Privacy Concerns

Hard to evaluate the potential impact but issues may arise as the work progresses.
Examples:

- The policy stressed that users data have to be deleted in 2 years. To enforce this deletion, the cloud has to store some evidence about this deletion which still may have some information about the user: this of course induces a privacy problem.
- Data transfer from one provider to another : need for anonymization and confidentiality techniques because some information such as IP addresses can be considered as business sensitive data.
- System level audits (logs about VM). No privacy concerns because they do not use personal data.
- Event interception. The goal of an adversary could be to observe some data transfer operations and perform some verification to retrieve some correlations.

B.1.3. Privacy and Data Protection Principles

Accountability and Privacy Enforcement Tool	
Informed Consent	this will be defined using the policy language only.
User Control	It is not sure we will be able to include this functionality in our prototypes, but this is feasible. For the system level operations, the user will not have control over its data
Purpose Binding	The main aim of the tool is to limit the use of the data for specified and legitimate purposes only defined either by the Data Subject or the Cloud Customer within the policy.
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	No answer yet. Data minimization must be respected during the policy definition phase. In the case of data transfer, the data (logs) has to be anonymized.
Transparency & Data Subject Rights	They won't be informed for the example on system level enforcement.
Confidentiality, Integrity & Availability	Research theme in C4-D3, for example encryption of transferred data can be one requirement.

Table 11: Privacy and data protection principles for the Accountability and Privacy Enforcement Tool.

B.2. Data Protection Impact Assessment Tool

From the D-2 survey:

> The goal of this task is to create a decision support tool building on the risk and trust models that would identify what the risks are for a given configuration and environment and highlight this to organizations (including cloud service providers). The tool will be used as part of the process of selecting service providers by assessing the impact of a particular choice before sensitive information is transferred/shared. We are not going to target individual end users, but organisational end users, and this may include cloud service providers within the service provision chain that are using other cloud service providers.

Since the tool is aimed at being used before sensitive information is tranferred or shared, there is little risk for negative consequences with respect to end-user privacy. The only issue is that potentially confidential data about a configuration or environment would have to be shared with the tool user, for the tool to be able to produce its assessment. This is however out of the scope of this deliverable.

B.3. Data Track

From the D-2 survey:

> The Data Track is a tool used by data subjects (individual end-users tracking their disclosures of their own personal data) that provides an overview of all personal data the data subject has disclosed and in particular who they have disclosed the data to (data controller). From this overview, working in conjunction with several other tools in A4Cloud (the Accountability and privacy enforcement tool from D-3, redress from D-4, assessment of policy violation from C-5, Transparency Log from D-5), data subjects can (dependency in parenthesis):
>
> - request access to their data stored at a data controller (D-3),
> - request to correct their data, or manage their consent (D-3),
> - request that their data should be deleted or blocked (D-3),
> - detect policy violation by matching log data (C-5, transparency log from D-5),
> - seek redress (D-4),
> - view detailed information about how their data has been shared and used by data controllers (D-3, D-5),
> - and view previous data disclosures and associated metadata like privacy policy, time, etc.

Data Track	
WP	D-5
Contact	Tobias Pulls
Type of Personal Data Processed	All data disclosed by users to CPs, potentially all additional data the CP derives or collects about the data subject
Data Subjects	Individual end-users, i.e., the per definition data subjects
Purpose of Data Processing	Empowering data-subjects to exercise several of their rights wrt. their personal data
Data Controller	end-users themselves (depending on interpretation)
Data Processors	The Transparency Log servers will store encrypted data needed by the Data Track
Stakeholders (tool users)	individual end-users
Retention Period	The Transparency Log data is stored for a configurable duration, the data subject always has the ability to delete a particular encryption key, the user can delete local data whenever

Table 12: Summary for the Data Track.

B.3.1. Information Flow During the Data's Life-Cycle

The Transparency Log (TL) ensures secure and privacy-friendly delivery of data from the CPs to end-users, see Section 2.3. The data from TL will be available for a known period of time. After that, the end-user keeps data locally in an encrypted form if the user wishes to keep the data longer than it is available in the TL.

Requests made by the DT to CPs, for example to access or correct data, goes directly between the DT and the CP(s) in question. Such a request (should) result in one or more messages sent to the user through the TL.

B.3.2. Initial Privacy Concerns

The Data Track, conceptually, stores a copy of all data a user discloses to any and all service providers. Depending on what the user wants, this data may be kept indefinitely. Assuming that the TL functions properly, the secrecy of the encryption keys for accessing the DT data is paramount.

Requests made by the DT to CPs need to be secured to prevent information leaks.

B.3.3. Privacy and Data Protection Principles

Data Track	
Informed Consent	Data subjects need to explicitly participate for the Data Track to function. What data is sent to the user (through the Transparency Log) by the CP(s) are hopefully specified in the privacy policy.
User Control	The Data Track enables data subject access to their data. For all data sent by the CPs through the TL to the users, the tool provides *forward secrecy* when deleting keying material.
Purpose Binding	Yes, the user is in control, with the exception of any personal data derived from when, how, and what type of requests are made by the user, which the CP(s) control.
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	When it comes to receiving data, the TL is the key component, see the TL analysis in Section 2.3. Requests to CPs could be sent over Tor (and using TLS), since Tor is already (planned for) used by the TL. Of course, the recipient CP(s) can determine the type of request and deduce information based on it, but at least this prevents some leakage to third parties.
Transparency & Data Subject Rights	Hopefully always, depends on CP behaviour.
Confidentiality, Integrity & Availability	Confidentiality through encryption (locally, or as part of the TL), no local integrity protection (beyond the use of authenticated encryption), TL is the key component impacting availability.

Table 13: Privacy and data protection principles for the Data Track.

B.4. Data Transfer Control Tool

From the D-2 survey:

> We address the lack of tools to support accountable data localization and transfer across cloud software, platform and infrastructure services, usually run by data processors. We designed a framework for automating the collection of evidence that obligations with respect to personal data handling are being carried out in what concerns personal data transfers.

Data Transfer Control Tool	
WP	D3
Contact	Jakub Sendor and Anderson Santana de Oliveira (SAP)
Type of Personal Data Processed	no PII on Data Subjects but system logs, information on data location which can be considered as business sensitive data
Data Subjects	cloud consumers, data controller
Purpose of Data Processing	collection of evidences for personal data transfers, tracing and auditing
Data Controller	the party which outsources the tool and the logs
Data Processors	the party that runs the tool and stores the logs
Stakeholders (tool users)	All: Service Providers, Auditors, Regulators
Retention Period	depends on the policy

Table 14: Summary for the Data Transfer Control Tool.

B.4.1. Information Flow During the Data's Life-Cycle

Data are collected by the Data Controller. There is no transfer of data except inside the cloud infrastructure. The environment is considered as being a closed system. Data are destroyed at the end of the retention period and they are not anonymized.

B.4.2. Initial Privacy Concerns

It is possible to infer some information from logs. Location information may need to be anonymized. There might be a need for access control in order to assure that only authorized parties query the logs. The communication channel should also be "secured".

B.4.3. Privacy and Data Protection Principles

Data Transfer Control Tool	
Informed Consent	No
User Control	They can have access to their data but cannot decide on anything. There might be a problem but the tool owners did not think about such an issue.
Purpose Binding	The goal is that only Auditors and the Data controller have access to these logs.
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	The tool does not deliver any solution for anonymity, pseudonymity or unlinkability. However, it may address the ”unobservability” property thanks to the integration of a proper access control mechanism: only authorized parties can access information on locaiton.
Transparency & Data Subject Rights	Data Subjects are not informed about the data collection and processing.
Confidentiality, Integrity & Availability	Nothing defined yet

Table 15: Privacy and data protection principles for the Data Transfer Control Tool.

B.5. Incident Response Tool

From the D-2 survey:

> The tool will allow individuals to respond to privacy and security incidents. It allows them to effectuate their data subject access rights (inspection, request for correction or deletion of the personal data held by the controller (DSAR)) and to file complaints regarding the processing of personal data by the controller and/or the incident that may have triggered the user to use the tool.

The Incident Response tool was at the very early stage of its design and therefore was hard to be evaluated with respect to privacy.

B.6. Plug-in for Assessment of Policy Violation

From the D-2 survey:

> Plug-in for assessment of policy violation: This plug-in will provide an assessment on the relevance of the violation of policies, so that stakeholders receive information about the most relevant policy violations. That is, it takes as inputs a policy

and an instance of a policy violation, and provides a quantitative assessment that enables an ordering of instances of violations with respect to their importance. This plug-in will feed part of the tools delivered in WP:D-5. Note that the plug-in for assessment of policy violations is not a tool as such, but a plug-in to be used by another tool. In particular, this plug-in will be used by the tools developed in D-5.2 (Data Track tool).

Since the tool is nothing more than a plug-in for the Data Track that provides a sorting function, there are very few if any privacy considerations with the tool itself.

Plug-in for Assessment of Policy Violation	
WP	C-4
Contact	Carmen Fernandez Gago
Type of Personal Data Processed	Business sensitive data and personal data
Data Subjects	Only the data subject who is using the plug-in
Purpose of Data Processing	helping users assess the importance of policy violations
Data Controller	Not applicable / end-user
Data Processors	Not applicable / end-user
Stakeholders (tool users)	individual end-users
Retention Period	Not applicable

Table 16: Summary for the Plug-in for Assessment of Policy Violation.

B.6.1. Information Flow During the Data's Life-Cycle

All input to the plug-in is provided by the Data Track, and all output is consumed by the Data Track.

B.6.2. Initial Privacy Concerns

None, any potential privacy concerns are related to how the input is provided or consumed (see Section B.3).

B.6.3. Privacy and Data Protection Principles

Plug-in for Assessment of Policy Violation	
Informed Consent	As part of using the Data Track.
User Control	Not applicable
Purpose Binding	Not applicable
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	Not applicable
Transparency & Data Subject Rights	Not-applicable
Confidentiality, Integrity & Availability	The result of the tool is only available to the user who runs the plug-in.

Table 17: Privacy and data protection principles for the Plug-in for Assessment of Policy Violation.

B.7. Redress Tool

From the D-2 survey:

> This tool advises end users and organisations on actions that can be taken in case of violations (consent violations, contractual violations, SLA violations and potentially data breaches). Potentially, the tool can automatically act on certain violations (file complaints, notify entities).

Redress Tool	
WP	WP D4
Contact	Ronald Leenes
Type of Personal Data Processed	The following information, which could include personal information, is provided and processed by this tool: • Authentication proofs that the party contacting the cloud provider is subject of the incident and thus may have the right to obtain redress; • Information describing (cloud) services that were used and the disputes with those respective services sides / cloud services side
Data Subjects	Individuals who have outsourced data to the cloud, or about whom an organisation has outsourced data to the cloud, and whose data were compromised.
Purpose of Data Processing	This tool processes data describing conflicts that individuals or organisations have with their cloud provider with purpose of settling this conflict and providing redress, e.g. by recovering from incidents and/or providing (financial) compensation if appropriate.
Data Controller	The redress tool itself processes personal data in regard to incidents. Hence, it is important to distinct cases where the tool is user controlled (i.e. running locally on the users machine) or provided as a central service? In the latter case, the central service is the data controller of the personal dispute-related data. The tool sends personal data to the following parties (which may in turn be data controllers of the personal data that caused the disputes): • The service provider that outsources personal data to a cloud provider and/or • the cloud provider, to whom individuals directly outsource their data. (is however a matter of legal controversy whether especially in the latter case the cloud provider is really determining the purposes of data processing and can thus be assumed to play the role of a data controller). The party that the user contacted in order to get a service is in many cases just a service broker/mediator for another service provider that takes the role of a data controller (e.g., Techcrunch is a mediator for Facebook).
Data Processors	Service providers, cloud providers and sub cloud providers may all process the type of personal data specified above.
Stakeholders (tool users)	End users and service providers that outsource data. (Also cloud providers that outsource data to sub cloud providers?)
Retention Period	Until the dispute/conflict is settled (maybe there are legal requirements for storing the data even longer?)

Table 18: Summary for the Redress Tool.

B.7.1. Information Flow During the Data's Life-Cycle

The tool should help the individual concerned to contact the responsible service provider that outsourced his data, or to contact the cloud provider to which he outsourced data. In case that

a service broker was involved, the tool should direct the user to the service provider that is behind the service broker. In order to decide which service provider to contact, the tool has to obtain personal data about the services used.

For contacting the respective provider, the individual has to prove that it was his data which were potentially compromised. Besides, he has to describe details of the conflict that arouse. The service provider needs to forward this complaint to the respective cloud provider. Similarly, the cloud provider may have to send or forward the complaints to a sub cloud provider that was responsible for the incident.

When the conflict is settled and all sides involved are satisfied with the outcome, the data should be deleted (unless there are legal provisions requiring retaining the data for a longer period of time).

B.7.2. Initial Privacy Concerns

Potentially, less privacy issues arise if the tool is user controlled and not a central service provided by a third party. When an individual has to prove that his personal data which were compromised, he should not reveal more identifying data than the service or cloud service provider already knows about him. In particular, if he has used a service or cloud service anonymously or pseudonymously, it should not be necessary that he is now identifying himself. The information describing the conflict can be sensitive and should be well protected and kept confidential.

B.7.3. Privacy and Data Protection Principles

Redress Tool	
Informed Consent	When the data subject takes the initiative to request to get redress, the redress should ask him to provide consent to sending and forwarding the personal data needed. When a service provider requests to get redress from a cloud provider on behalf of an individual, it should not disclose any personal data that the cloud provider does not have yet without the individuals consent
User Control	In principle, if the tool is running on the user side, the user can keep better control over his data. Users should preferably have access to information at the cloud service providers side revealing the status of conflict resolution of their cases and how their data describing the conflict case have been processed.
Purpose Binding	The data should only be processed for the purpose of settling the conflict and providing redress
Data Minimisation • Anonymity of Content Data, Communication or Location Data • Pseudonymity of Content Data, Communication or Location Data • Unlinkability of Data Items • Unobservability of Tool Usage	Not clear yet. However, it is in principle possible to authenticate individuals as data subjects of compromised data without identifying the individuals.
Transparency & Data Subject Rights	Not clear yet. However, it should be possible to grant data subject access to dispute data
Confidentiality, Integrity & Availability	It is not clear yet how the data will be technically secured, but there are technical standard security solutions that can be implemented.

Table 19: Privacy and data protection principles for the Redress Tool.

B.8. Tool for Cloud Contracts

From the D-2 survey:

> This is a simple expert system to help with drafting and managing custom contracts. This would be suitable for specialised situations where both the cloud provider and the customer can negotiate cloud contracts. We might also look to generalise the tool slightly to cover outsourcing scenarios, or usage by entities that are not in a position to negotiate cloud contracts, just for tracking which ones are in place and possibly highlighting some meta-level information in relation to those. The tool will take into account, while drafting the contracts, organizational risk and trust aspects as part of the contract terms. The output would be a suggestion for contractual terms in the given situation (with some reference to the cloud supply chain).

Since the main aim of this tool is to help its users to write contracts, we believe that the tool will not use or output any personal data or privacy sensitive information. Therefore, the privacy analysis did not seem appropriate.

C. Initial Privacy Analysis in Relation to BUCs

C.1. BUC1: Healthcare

The first business use case, illustrated in Figure 4, deals with the flow of information from medical sensors attached to or in the proximity of elderly patients (individuals that are cloud consumers and data subjects) to three cloud providers; Cloud x, Cloud y, and Cloud z. Cloud x collects and processes the sensor data, which in turn is stored at Cloud y. Data is fed into Cloud z, that provides an "information engine" used by a number of different cloud consumers: the end-users themselves, their relatives and friends, health-care organisations, insurance companies, researchers, pharmacies, and data analysts.

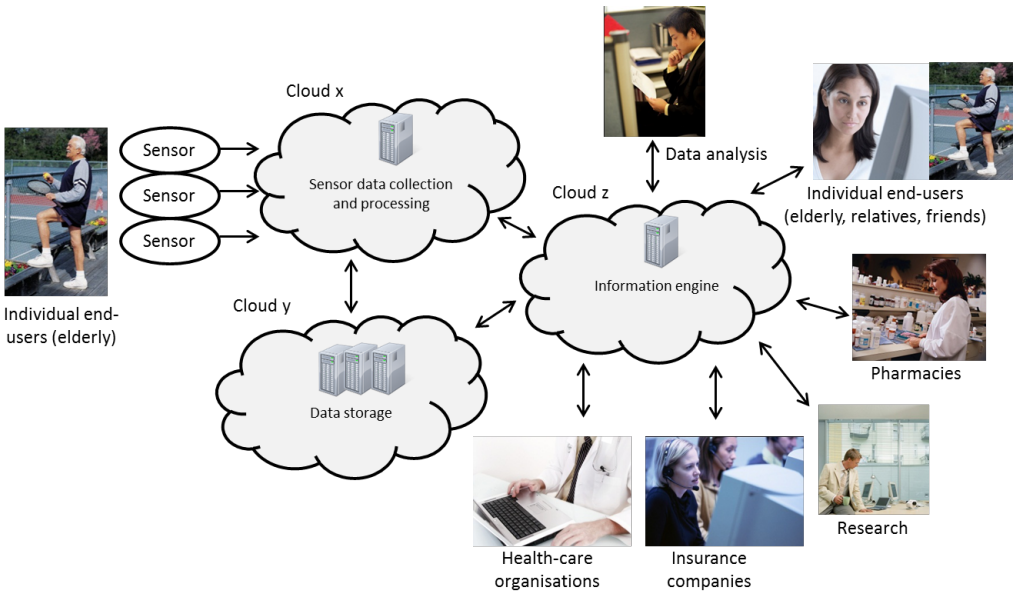

Figure 4: The setting of BUC1 (from DB-3.1)

Starting from the perspective of the individual end-users (the elderly patients), data collected about them from sensors will primarily be *sensitive* personal data concerning their health. This, in conjunction with their position (which is one of the recorded sensory inputs, see p19 of DB-3.1) will reveal significantly more about a data subject's personal life, such as potentially whom they are interacting with, religious affiliation, sex life, and so on.

From the point of view of the other cloud consumers, namely the businesses and organisations using the information engine provided by Cloud z, what data they are interested in, and when, is apparent to Cloud z. This may or may not be considered (business) confidential data. For example, observing that an insurance company, a health-care organisation, and a pharmacy are scrutinizing the data of a particular patient may serve as an indicator of an adverse event.

There are some legal hand-waving going on in the description of the BUC, such as implying that the storage provider (Cloud y) learns no personal data because of the fact that the data is encrypted before storage, such that Cloud y does not even need to be classified as a data

processor as defined by the EU Data Protection Directive. Legal consideration aside[20], let us consider the following scenario: Cloud y is storing all sensor data collected by Cloud x by sensors sensing elderly people. Cloud x performs perfect encryption of the data before storing it at Cloud y, to such a degree that it is not apparent for Cloud y which person what (encrypted) data belongs to. An employee of Cloud y now picks up a rock, and throws it at an unsuspecting sleeping elderly person whose data is being stored at Cloud y. The impact of the rock, naturally, startles the elderly patient, which in turn triggers a flurry of data to be generated by the observing sensors. This spike of new data goes from the sensors, to Cloud x, and then in turn results in a spike at Cloud y. Now Cloud y can associate a blob of encrypted data with the elderly person who was just attacked. In other words, some side-channel information allowed Cloud y to link the encrypted data to a particular individual. By simply observing the usage of this encrypted blob of data over time, Cloud y can deduce personal data. Encrypting the data at Cloud y is a necessary but far from sufficient criteria for preventing Cloud y from being able to deduce personal data. In the spirit of accountability, especially with regards to committing to being responsible stewards of personal data, taking into consideration *ethical* obligations (even when legal obligations are inadequate), for this analysis we consider the case where *all* cloud providers in the BUC are running accountability tools.

C.1.1. Accountability and Privacy Enforcement Tool

The tool will be in the centre of all access to personal and confidential data, presumably with instances running at each CP and CC, including individual end-users[21]. Who is in control of each instance of the tool is of paramount importance, since it (per description) should enable enforcement of privacy policies, which among other things dictates how data can be used, enabling the entity in control of the tool to observe all of these enforcement decisions.

Any data generated as a result of an access request for personal or confidential data may be personal or confidential data as well, since it is *related* to the personal or confidential data (the conservative approach is to treat all such data as persona or confidential data). This includes, for example, something as simple as a transaction log of all requests and the outcome of the request (access granted for request for attribute A under purpose P for person X by system Y). The tool's purpose is to facilitate *verification* of enforcement of policies, the key questions are then: by whom and how? If the " how" is by keeping, for example, a verifiable log of transactions then all data in the log is also most likely personal data and depending on "who" is able to do the verification, potentially extremely invasive since we are dealing with sensitive medical data that is generated by sensors. A similar argument as for personal data can be made for confidential data. If the tool, for example, is verified by administrators from the healthcare organisations then they may learn of how frequent insurance companies look at recorded patient data.

To illustrate the vital role of the tool: when a sensor wants to store some data about an individual end-user, the data goes from the sensor to Cloud x[22]. At Cloud x, an instance

[20] It may very well be the case that Cloud y falls outside the scope of the current EU Data Protection Directive, but we are performing a privacy analysis, not a legal analysis.

[21] This was the case in PrimeLife for the PrimeLife Policy Engine, which is noted as a candidate for being extended by C-4 in their answer to the D-2 survey.

[22] There might be an interaction with the Accountability and Privacy Enforcement Tool at the individual end-user as

of the Accountability and Privacy Enforcement Tool is running which stores the sensor data associated with the prior agreed to data handling policy (sticky policy in PPL speak) in its data store as part of the policy engine of the tool. But, as we know, Cloud y is used for the actual encrypted storage so presumably the data is encrypted (by the tool of Cloud x) and stored at Cloud y, where in turn the data is associated to a data handling policy and stored. When the sensor data needs to be analysed by a CC associated to Cloud z, then the information engine (part of which is a policy engine, as part of the tool) needs to get the data, either by directly requesting it from Cloud y or by first getting permission from Cloud x to retrieve the data. Of course, before sending a request for getting the data, the instance of the tool at Cloud z needs to evaluate the request, just as each instance of the tool and Cloud x and y need to do the same. Finally, the information engine returns the data to the requesting CC which, depending on which CC it is, may or may not also run an instance of the tool. At all of these steps the different instances of the tool are potentially processing, or able to deduce information about, sensitive personal data of the individual end-users. For example, if the engine leaks (to any third or unauthorised party) that a particular individual's sensor data was evaluated by, say, a doctor specialised in a particular type of disease, this is sensitive personal data of the individual. Note that this is sensitive personal data regardless of what the sensor data actually was, since it was simply the existence of a *link* between the individual and doctor that constitutes the sensitive personal data.

C.1.2. Audit Agent System

Presumably, there will be at least software agents running at each CP, that in some form or another is in the control of the individual end-users (depending on if each end-user has or her own agent, or if an agent is split among a group or all individual end-users). Key considerations are then:

1. How is the evidence collected by software agents? How is personal or confidential data separated for each user, i.e., what prevents the agent of one user to look at the personal or confidential data of another user?

2. How is evidence stored by agents? How is it stored by individual end-users? For how long?

3. How does agents and individual end-users communicate?

Regarding point 1, naturally, if an agent for one end-user can access the sensitive personal data of another end-user this is bad for privacy.It results in a *weakest link* scenario, where the (direct or indirect) compromise of one agent leads to the compromise of all data accessible by agents at a system. Agents should only have the capability to capture evidence that relates to *only* the individual end-user for which it serves. Furthermore, there is a need to protect confidential data, perhaps belonging to other CCs of Cloud z, such as when and what for they are querying the information engine.

well, depending on how tightly the model from PrimeLife is followed or not

For point 2, any evidence (which can be considered personal or confidential data, see working draft of MSC8.1) created and stored by the tool needs to be protected. If the tool leaks (to any third or unauthorised party) how much data is stored for a particular end-user this data may, over time, reveal sensitive information about the end-user, such as when they were under a lot of stress, their health status changed, and so forth.

Last, but not least, point 3 highlight an important consideration in the setting of the tool; agents are communicating with individual end-users and other agents. How is this communication protected, and what information does it leak? For example, CPs may be able to observe *when* end-users are querying their agents, if they are updating their agents with new instructions or not, and *how much* evidences are sent to the user. End-users in BUC1 are more likely to worry about their privacy in situations when they feel vulnerable, such as after a visit with their doctor when discussing their health.

C.1.3. Data Protection Impact Assessment Tool

In BUC1, potentially *confidential* data in the form of the configuration and environments of all CCs, such as the insurance companies and healthcare organisations, would have to known to the CCs.

C.1.4. Data Track

The tool, conceptually, contains a copy of *all* (sensitive) personal data an individual end-user has disclosed to the involved CPs. Protecting this data is paramount, and its existence increases the value for third-parties when compromising systems running the tool. Going through the list in Section B.3:

- For bullets 1–3, see Section C.1.6.
- For bullet 4, log data may reveal (sensitive) personal data of other end-users, or confidential data of the CPs or other CCs. For example, log data may reveal the operational procedures of an insurance company, without being relevant as policy violations. The matching takes places *locally*, see Section B.6.
- Redress, in bullet 5, may require the end-user to reveal personal data to a third party to prove the right to redress. See Section C.1.8 for further analysis.
- For bullet 6, a record of how (sensitive) personal data has been processed is also (sensitive) personal data. Furthermore, the record may reveal confidential data about the entities performing the processing. In BUC1, this is the other CCs of Cloud z. See Sections 2.3 and C.1.10 for further details on how this data is recorded, stored, and communicated.
- Regarding bullet 7, this data is stored as part the Transparency Log, see C.1.10.

C.1.5. Data Transfer Control Tool

As for the Accountability and Privacy Enforcement Tool, the Data Transfer Control Tool needs to be in a privileged position to monitor all processing to be able to collect evidences for when personal or confidential data is being transferred. Since the tool in particular deals with data transfer, the example from Section C.1.1 regarding leaks of information flow is relevant for this tool as well; how personal or confidential data is transferred may in and of itself be personal or confidential data.

Just as in the case for the analysis of the Audit Agent System in Section C.1.2, how evidences are stored and communicated are key considerations, together with *to whom* the evidence is available.

C.1.6. Incident Response Tool

Requests by individual end-users may reveal *when* end-users are making requests, both to one or more CPs and any observing third party. The fact that CPs have to support electronic requests from CCs may also pose a threat if not properly secured. For example, if requests are not confidential, then any third party observing the request may learn of the contents of the requests, which is potentially (sensitive) personal data. Furthermore, if a third-party can observe a particular end-user making a request to a particular CP or CC, this may in and of itself be sensitive personal data. For example, observing that an end-user in BUC1 sends a request to a particular healthcare organisation or pharmacy may reveal data about the end-users medical condition.

Regarding filing of complaints regarding processing of personal data, key considerations are (i) what is contained in the complaint, (ii) and how is it filed? Similar to access requests, if the filing of complaints can be observed, for example such as who is the sender and recipient, this may leak (sensitive) personal data.

C.1.7. Plug-in for Assessment of Policy Violation

Since the tool is run locally by an individual end-user, and input is provided by the Data Track, primary considerations for the tool is regarding the use of its input and output, which is out of scope for the tool itself.

C.1.8. Redress Tool

(Reservation: the description of the tool is very fluffy at this point in time, making an analysis difficult.) Key considerations are how violations are detected, stored, and communicated. Furthermore, as with filing complaints for the Incident Response Tool analysed in Section C.1.6, how the tool acts (potentially automatically) on certain violations may reveal (sensitive) personal or confidential data.

C.1.9. Tool for Cloud Contracts

As with the Data Protection Impact Assessment Tool analysed in Section C.1.3, since the tool is only used to assist in drafting and managing custom contracts, there are no obvious issues with end-user privacy. The only considerations are related to potentially confidential data, both of the contracts themselves and any data needed for the expert system to function.

C.1.10. Transparency Log

The purpose of the tool in BUC1 would be to transfer log data from CPs and some of the CCs, to the individual end-users and potential CAs. Support for CAs are optional, and would of course involve a breach of privacy of the end-users if parts of their (sensitive) personal data were shared with third-party CAs. A similar argument can be made for confidential data. Taking CAs outside of the equation, one key consideration is ensuring that only log data that concerns one individual end-user is only made transparent to that end-user, and no (sensitive) personal data concerning one end-user is leaked to another end-user. Another consideration is that data shared with end-users may reveal confidential data about the CPs or CCs whose processing is made transparent in the log data. Other key considerations are:

- If the log data is not *confidential*, then a third party can learn potentially (sensitive) personal or confidential data.
- If the log data is not integrity-protected, then a third party could modify or delete the log data.
- If each individual log entry could be linked to an individual end-user, then this may also leak (sensitive) personal or confidential data about the end-user or the CPs/CCs processing.
- If the data kept for the transparency log on different CPs and/or CCs can be linked, then this may reveal at which CPs and/or CCs an individual end-user's data has been or are being processed. This may also be (sensitive) personal data in this BUC1, similar to the example in Section C.1.1.

C.2. BUC2: Enterprise Resource Planning

The second use case tackles the problem of storage and processing of personal and business confidential data for cloud based enterprise resource planning (ERP) software. As illustrated in Figure 5, a SaaS named **MarchéAzur** offers its customers a loyalty program and collects and analyzes their personal information in order to propose dedicated offers. **MarchéAzur** uses **PaasPort**, a PaaS which allows the connection of **MarchéAzur** customers through a mobile application and **InfraRed** as a IaaS to provide virtualized infrastructure. **MarchéAzur** also connects to a third party software service that provides customized extensions (mobile payment in this use case) and that may be involved in personal data handling.

As in any of the three use cases, outsourcing of data storage immediately raises the problem of data confidentiality and privacy. In this particular scenario, data being outsourced to the

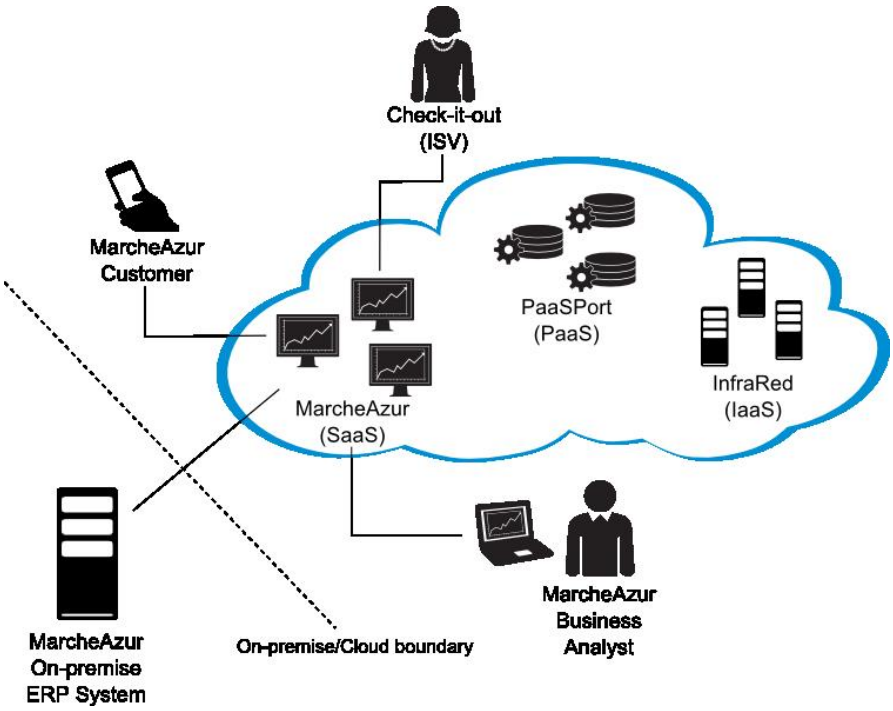

Figure 5: Overview of BUC2 (from DB-3.1)

cloud regroup customers' personal data which are considered as being highly privacy sensitive. Furthermore, customers share different data with different services which may use the same cloud service. In this case, an adversary may link data from different sources to a single customer.

Additionally, access to this data by end-customers may also raise privacy concerns. For example, the update of privacy preferences of a certain customer can be considered as privacy sensitive and may even reveal some information on the location of his personal data on the outsourced database.

The access of the data by business analysts may also be considered as business confidential. InfraRed should not discover which Business Analyst accesses some data and how frequently.

C.2.1. Accountability and Privacy Enforcement Tool

As already mentioned in section C.1.1, this tool will provide and enforce the access to personal and business confidential data. Therefore the control of the access to and the usage of this specific tool should definitely be the very major privacy requirement. Furthermore, as previously mentioned in section C.2, the access to the data by either the end-customer or an authorized business analyst may result on the creation of additional privacy sensitive data such as access logs. Even if the original personal data is protected through encryption mechanisms for example, an adversary may discover that some data was requested several times. Although there is no information on the content of such data, the fact that the same data was accessed

several times may imply some leakage.

C.2.2. Audit Agent System

When an end-customer requests some evidence on what information the supermarket **MarchéAzur** has collected (Scenario 7.1.1a in deliverable D:B-3.1) or when the business analyst requests a report regarding the operations performed during some predefined period (Scenario 8.1.1a), some personal identifiable information (PII) may be part of the evidence. Therefore, the storage of such evidences by the tools should of course be protected. The storage of evidences should not reveal who this evidence is related to. Even the time at which the evidence has been generated can be considered as being sensitive data. Finally, the communication channel between the tool and the requester should of course be secured as well.

C.2.3. Data Protection Impact Assessment Tool

C.2.4. Data Track

Since Data Track basically helps the end-customer to track all data directly or indirectly related to him/her, all such information should be protected while being stored or communicated to the end-customer. The additional operations performed upon customers' requests (such as the update of their privacy preferences) should not reveal any additional information. Both the request and the response during this exchange should be privacy protected. Information on policy violations even a simple notification may be considered as being highly sensitive by businesses and thus should be considered as confidential. Only the relevant end-customer some authorized entities and optionally the Auditor should discover the existence of such an event.

C.2.5. Data Transfer Control Tool

As in the case of BUC1, the information on location and flow of data is considered as sensitive and their storage and processing should automatically be protected against unauthorized access and analysis. In particular in BUC2, since different services may use the same storage service, information on location and flow of data can be of interest for different curious or malicious services.

C.2.6. Incident Response Tool

During the usage of this tool, requests from end-customers may reveal some information on the history of such actions. For example, the tool may reveal when or how frequently a specific customer has sent a request. As in BUC1, the filling of complaints should also be protected with privacy preserving mechanisms in order not to reveal any privacy sensitive information.

C.2.7. Plug-in for Assessment of Policy Violation

The tool is considered as a plug-in for the Data Track tool and therefore does not imply any additional privacy issues than existing ones.

C.2.8. Redress Tool

C.2.9. Tool for Cloud Contracts

C.2.10. Transparency Log

In BUC2, this tool would store log data or customers' privacy preferences. The tool should assure that only authorized entities, i.e. the customer whose information and actions were logged can have access to the such information. The tool should also assure unlinkability because even if logs are being encrypted, they may reveal some information on similarity for example. Furthermore, once a customer has accessed his/her respective data, he/she should not be able to discover any other information with respect to other end-customers' data.

D. Initial Privacy-Enhancing Mechanisms Suggestions

This appendix contains initial suggestions of PETs for some A4Cloud tools identified initially in the project. Several tools were too immature to make any suggestions for at the time of writing. Some tools and suggestions are outdated at this time of writing, but are included here for sake of completeness.

D.1. Audit Agent System

In terms of general PETs, communication between components over any network should be protected by TLS with mutual authentication. For authenticating auditors, strong multi-factor authentication should be enforced. For example, this could be done with client certificate authentication in TLS together with a password for each auditor. Passwords could be protected using scrypt, with a strict policy in place of locking accounts after failed authentications.

Figure 6 shows the high-level architecture of the Audit Agent System, as envisioned in the end of 2013 by the developers. For each component:

APM Beyond strong access control, as described earlier for general considerations in terms of authentication, all audit tasks should be logged in a secure and irrefutable way. As part of the Accountability and Privacy Enforcement Tool described in Section B.1, there will be secure logging mechanisms in place, which may be used. Another option is using the Transparency Log described in Section 2.3. Another avenue for access control is to use the policy engine as part of the Accountability and Privacy Enforcement Tool.

AAC As for the APM, strong access control is needed together with all actions being logged.

ES The evidence store is the “cookie jar” of the AAS and therefore deserves extra attention. The component could be based on the “privacy preserving third-party word search” mechanism that EURECOM is working on. Another option is to at least encrypt the data at rest such that only the EPP can read the data in the store.

EPP Securing reports and other representations of evidence are paramount. Data should be encrypted at rest and the distribution of data logged.

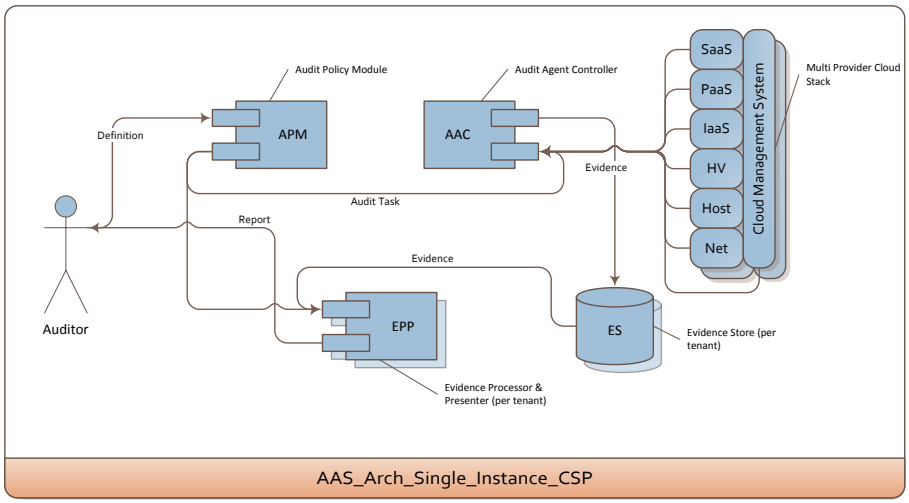

Figure 6: The Audit Agent System architecture

Figure 7 provides a layered view of the architecture of AAS and its components. We highlight two components of AAS Agents:

Agent Collectors The capability of collectors should be minimal. Presumably, Agents will contain some framework/plug-in system for collectors, enabling collectors to be swapped in and out as needed. Perhaps it makes sense to have collectors sandboxed with limited access to the surrounding system, such that all access to the system can be logged.

Agent Minimizers There exists a number of tools for "anonymising" log data, such as the rsyslog module mmanon[23] and the Apache module mod_anonstats[24]. If evidence collection is triggered by audit tasks, then presumably the collection of data by collectors are already focused, which lowers the value of minimizers. In fact, minimizers in a strict "collect only what is needed" mode, may do more harm than good even from a privacy perspective (accurate assessments by auditors of accountability obligations are presumably in the interest of data subjects).

[23] `http://www.rsyslog.com/doc/mmanon.html`, last accessed 2013-12-10
[24] `http://bug.st/mod_anonstats`, last accessed 2013-12-10

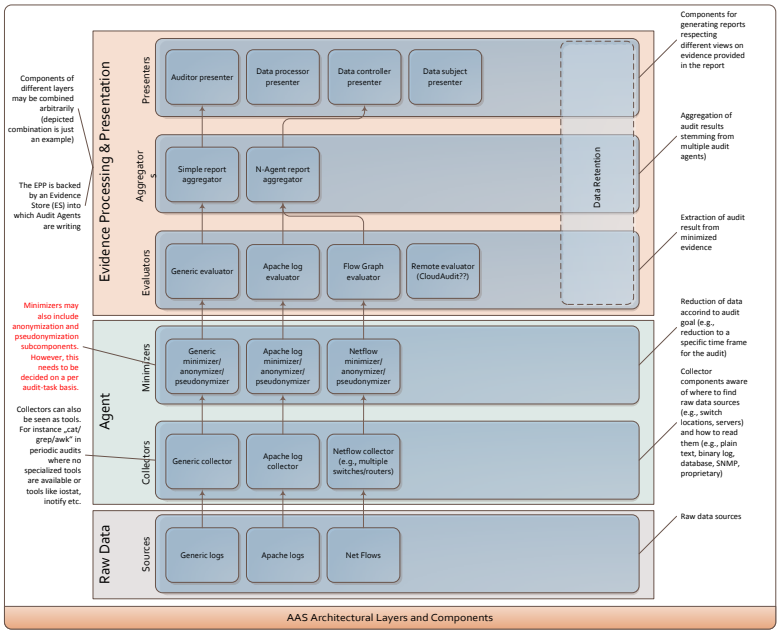

Figure 7: The Audit Agent System layers

Figure 8 shows the runtime environment for AAS. We highlight two components:

Agent Runtime (owned by provider) Similar to Agent Collectors, the runtime could probably benefit from being heavily sandboxed. If the runtime used for agents auditing a provider is provided by the provider being audited, one important consideration is how trustworthy the actual data from the agent can be? A similar setting can be found in the secure logging area, which as mentioned earlier, is something both the Accountability and Privacy Enforcement Tool and Transparency Log tool may help address.

Agent Multi-Tenant Collection Environment While we cannot offer any solution, it is worth highlightening the potential risk posed by other agents run by potentially other entities in the environment. Thread carefully.

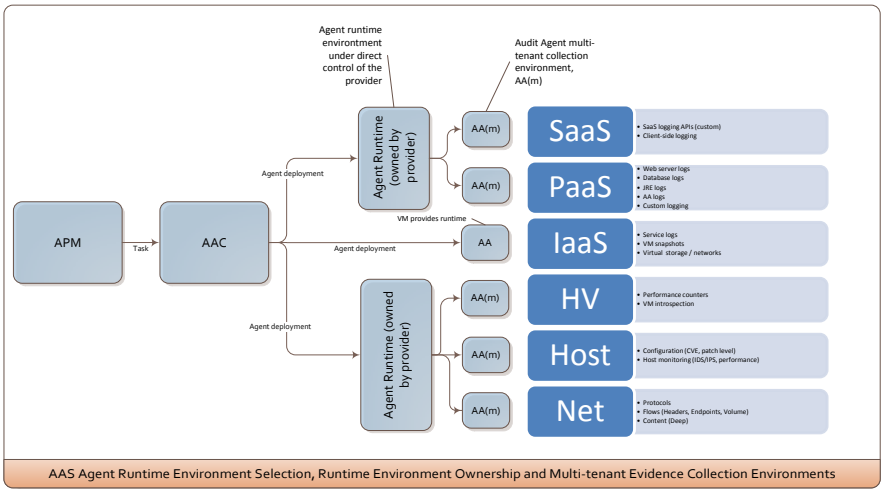

Figure 8: The Audit Agent System runtime environment

D.2. Data Track

Figure 9 shows an overview of the Data Track tool. As can be seen from the figure, the Data Track interacts with three tools: (i) the Plug-in for Assessment of Policy Violations, (ii) the Transparency Log, and (iii) the Accountability and Privacy Enforcement Tool. While we note that the interactions and result of combining the different tools need to be analysed together, we leave the "big picture" analysis for when more details about each tool are available. Most prominently for the Data Tack are two components:

Data disclosures The database of data disclosures need be protected. This could be done by encrypting the database, where the key is derived from a password using scrypt.

Remote access When users request remote access to their disclosed data, for retrieving data, correcting data, or revoking their consent for processing, they should be able to do this without revealing more information to the service provider than it already knows. Assuming that the authentication using the Transaction Log allows for anonymous (pseudonymous) authentication, all communication between users and service providers should be using TLS over Tor.

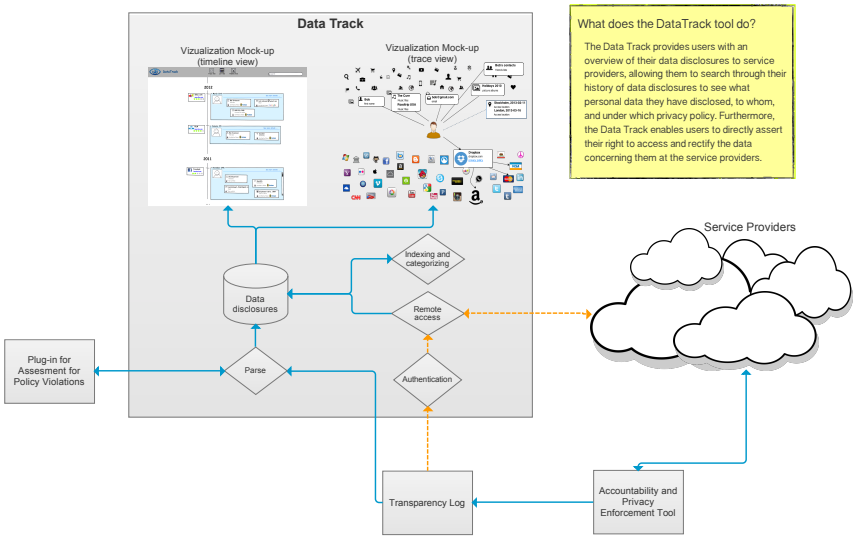

Figure 9: An overview of the Data Track from the second D-2 survey

D.3. Data Transfer Control Tool

As opposed to some other tools (like the Audit Agent System for example), the Data Transfer Control Tool is not entirely defined yet. We therefore provide some suggestions on PETS and may revise this section whenever the tools will be ready. Since location information is considered as privacy sensitive, the tool requires to implement some encryption mechanism to assure confidentiality. The access to such information should be controlled via a proper access control mechanism. The communication channel between different entities can be protected via well-known mechanisms such as TLS.

D.4. Redress Tool

As discussed above, PETs are needed to protect personal information describing the conflict as well as allowing the user to authenticate himself as the data subject of compromised data (i.e. subject of the incident that has the right to obtain redress) without revealing more identifying information than needed. If the redress tool runs as a plugin for the Data Track tool, means for anonymous (pseudonymous) authentication available for the Data Tack can be used for anonymous/pseudonymous authentication of the data subject of compromised data, i.e. of the subject of an incident. This requires however also that Tor and TLS are used for anonymising and protecting the communication. For protecting the personal information describing the conflict, encryption and access control at the service provider/cloud provider side should be used.

D.5. Accountability and Privacy Enforcement Tool (A-PPLE)

Similarly to most of the other tools, since the Accountability and Privacy Enforcement Tool is at its early phase of design, we provide some general suggestions based on its current

status. As it is the case for any A4Cloud tool, the communication channel between entities should be protected via encryption and authentication mechanisms such as TLS. Additionally, since the Accountability and Privacy Enforcement Tool provides storage for PII and logs, such information should be encrypted either through traditional encryption mechanisms. The chosen encryption mechanisms may be semantically secure in order to assure unlinkability: if a same data segment is encrypted twice, the output at each encryption operation becomes different. The encryption keys should be distributed following a dedicated key management protocol. Only authorized entities should be able to decrypt relevant information. The tool may also have the right to revoke some access at any time.

D.6. Transparency Log

Figure 10 shows an overview of the Transparency Log tool. As can be seen from the figure, the Transparency Log interacts with two tools: (i) the Data Track and (ii) the Accountability and Privacy Enforcement Tool. While we note that the interactions and result of combining the different tools need to be analysed together, we leave the "big picture" analysis for when more details about each tool are available. Most prominently for the Transparency Log are two components:

Communication links Both for setup and reconstructions the client should use TLS over Tor. Between the service provider and the server, TLS is a bare minimum. Depending on the service offered by the service provider, traffic analysis of its communication with the TL server may reveal information about the service's users and its own operations worth protecting. If so, there is added benefit in using Tor.

Log entry and credential storage At the client, the secure storage of log entries and credentials are of utmost importance. The data should be encrypted on disk and the credentials properly managed. Creating a central repository of all credentials of a client may pose a too big risk in case of compromise. Solutions where, for example, a smart card in addition to the key material on disk is needed for reconstruction of log data may be a good idea. Basic secret sharing between multiple devices of a client is another option.

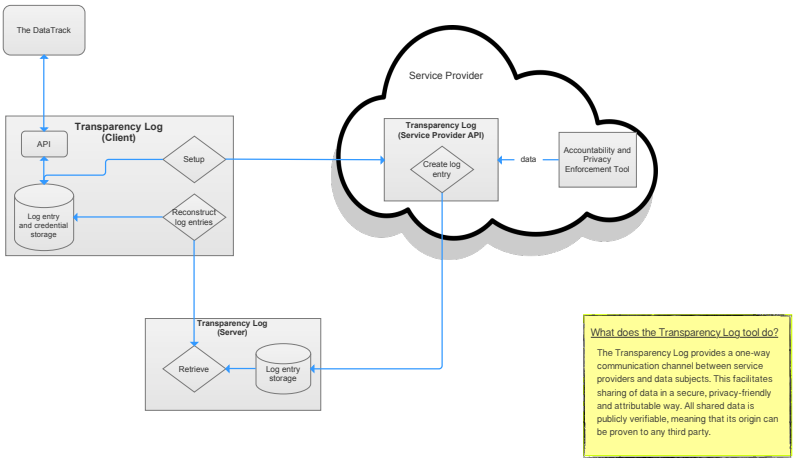

Figure 10: An overview of the Transparency Log from the second D-2 survey M15