Published in "Cyber Security and Privacy", Communications in Computer and Information Science Volume 182, 2013, pp 28-40
http://link.springer.com/chapter/10.1007/978-3-642-41205-9_3#

Bringing Accountability to the Cloud: Addressing Emerging Threats and Legal Perspectives

Massimo Felici[1], Martin Gilje Jaatun[2], Eleni Kosta[3], Nick Wainwright[1*]

[1] Hewlett-Packard Laboratories, Long Down Avenue, Bristol BS34 8QZ, UK
[2] SINTEF ICT, NO
[3] Tilburg University, NL

Abstract. This paper is concerned with accountability in cloud ecosystems. The separation between data and data subjects as well as the exchange of data between cloud consumers and providers increases the complexity of data governance in cloud ecosystems, a problem which is exacerbated by emerging threats and vulnerabilities. This paper discusses how accountability addresses emerging issues and legal perspectives in cloud ecosystems. In particular, it introduces an accountability model tailored to the cloud. It presents on-going work within the Cloud Accountability Project, highlighting both legal and technical aspects of accountability.

Keywords: Accountability, Data Governance, Cloud Computing

1 Introduction

Cloud computing has emerged as a new paradigm used across industries for deploying technological resources. Economic forecasts show that cloud computing will enable efficient, competitive and cost-effective deployments of computational resources in order to accommodate emerging user needs [1]. Alongside many business benefits both for consumers and providers of information services, cloud computing presents new challenges in terms of security and trust. Personal data is duplicated across cloud resources making them more accessible and less subject to loss. Unfortunately, personal data is also exposed to security threats and lack of trust across cloud supply chains [2].

Cloud consumers and providers are exposed to various problems. For instance, from a resource viewpoint, it is necessary to improve data management processes and to a certain extent to automate them. The increasing amount of data and resources requires new mechanisms enabling cost-effective management while guaranteeing critical features like security and privacy. One of the essential characteristics of cloud computing is rapid elasticity – *"Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand"* [3]. This involves horizontal (connecting different resources such that they work as a single logical unit) and vertical (increasing the

* Corresponding author: nick.wainwright@hp.com

capacity of a single unit by adding additional resources to it) scalability of cloud computing. Unfortunately, scalability of code and data still remains among the main challenges affecting quality of services and interactions among customers and providers [1]. From security and trustworthiness perspectives, some of the issues that consumers and regulators are mostly concerned about are things like lack of transparency and control in cloud service provision. The international dimension of some situations (for instance, foreign government surveillance) may involve dealing with further complexities from a legal perspective. Other challenges in cloud computing relate in particular to multi-tenancy, which *"raises multiple concerns that implicit impact on the quality of the cloud systems and in how far the respective characteristics can be fulfilled"* [1]. Such challenges are perceived as barriers and are limiting the adoption of cloud computing.

Accountability has emerged as a critical aspect of data protection [4]. Recent research on accountability [5] has identified its essential elements (e.g. organizational commitments, mechanisms for privacy policies and assurance reviews). Unfortunately, a generally accepted definition of accountability is still beyond any consensus [6]. Without a well-defined concept of accountability, it is difficult to interpret accountability from an operational viewpoint of analysis in cloud ecosystems. However, within the on-going debate on a definition of accountability, this paper is concerned with how accountability information enables data governance in cloud ecosystems. The relationship between accountability and information is also referred to as information accountability: *"information usage should be transparent so it is possible to determine whether a use is appropriate under a given set of rules"* [7]. Accountability provides a means to unlock the cloud potential by addressing relevant problems of data protection emerging in cloud ecosystems [8].

This paper discusses emerging issues (focusing on data governance and protection) in cloud ecosystems and presents relevant legal perspectives. The separation between data and data subjects as well as the exchange of data between cloud consumers and providers increases the complexity of data governance in cloud ecosystems. This problem is worsened by emerging threats and vulnerabilities in cloud ecosystems. This paper discusses how accountability addresses emerging issues and legal perspectives in cloud ecosystems. In particular, it introduces a model of accountability addressing both technical and legal perspectives in cloud ecosystems. This paper is organized as follows. Section 2 discusses the problem of data governance in cloud ecosystems. Section 3 describes some emerging data protection problems in the cloud. It highlights how cloud computing requires new information mechanisms orchestrating data governance and relationships among stakeholders. Section 4 introduces our model of accountability in the cloud. Section 5 discusses relevant data protection issues drawn from legal perspectives. Section 6 highlights some remarks.

2 Data Governance in Cloud Ecosystems

Cloud computing has transformed the way information technology is delivered, promising rapid, efficient, and cost-effective deployment of computational resources across different industries, geographic areas and application domains. More recently,

the scope of cloud computing has expanded to include 'big data', the increasingly large amounts of data held by cloud service providers that is the raw material on which new and innovative cloud services are founded. However, alongside its numerous business benefits for consumers and providers of information services alike, cloud computing presents new challenges in terms of security, privacy and trust. The transfer of personal or confidential data into the cloud may provide the opportunity for innovators to create new services, offering operational advantages such as improved accessibility and reducing the probability of catastrophic loss. At the same time this may make data more vulnerable to unauthorised access or modification. The broader issue is essentially one of loss of transparency and control in what happens to data once moved to the cloud. As stewardship of data becomes shared between users and potentially complex chains of cloud providers, the former have to place trust on the cloud ecosystem and its governance (see Figure 1). This has proven to be a significant barrier limiting the adoption of cloud computing – one that can be lifted by ensuring that there is accountability throughout the cloud ecosystem.

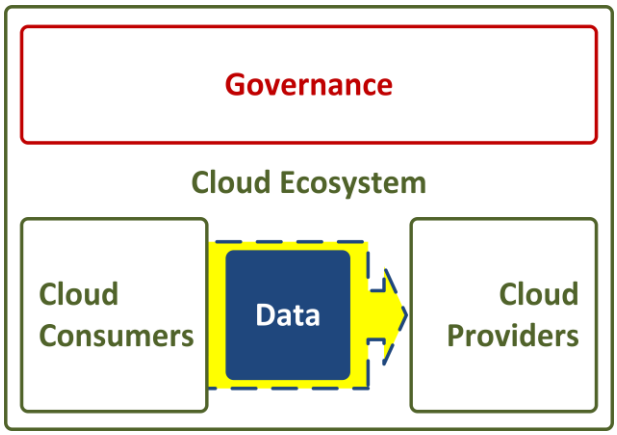

Figure 1 Cloud Ecosystem

Accountability is emerging not only as an essential aspect of data protection (for several decades it has been regarded as a privacy principle), but also in particular within the deployment of cloud computing as argued by the Article 29 Data Protection Working Party[1] – *"In IT accountability can be defined as the ability to establish what an entity did at a certain point in time in the past and how. In the field of data protection it often takes a broader meaning and describes the ability of parties to demonstrate that they took appropriate steps to ensure that data protection principles have been implemented."* [18].

[1] Under Article 29 of the Data Protection Directive, a Working Party on the Protection of Individuals with regard to the Processing of Personal Data is established, made up of the Data Protection Commissioners from the Member States together with a representative of the European Commission. The Working Party is independent and acts in an advisory capacity. The Working Party seeks to harmonize the application of data protection rules throughout the EU, and publishes opinions and recommendations on various data protection topics.

Accountability, if implemented by cloud service providers, can unlock further potential cloud services by addressing relevant problems of data stewardship and data protection in emerging in cloud ecosystems. Governance is the process by which accountability is implemented in the cloud. All the actors involved in the cloud – service providers, consumers of cloud services (whether individual end-users, businesses, public organisations and even other cloud service providers), and those directly involved in IT governance have a role to play in making cloud services accountable for how data is used and managed in the cloud.

3 Emerging Challenges in Cloud Ecosystems

This section discusses by means of examples emerging challenges and issues in cloud ecosystems. Cloud services are not isolated, they exist in an ecosystem where all the parts interact and rely on each other. Figure 2 illustrates the main challenges and threats that we will discuss in the remainder of this section.

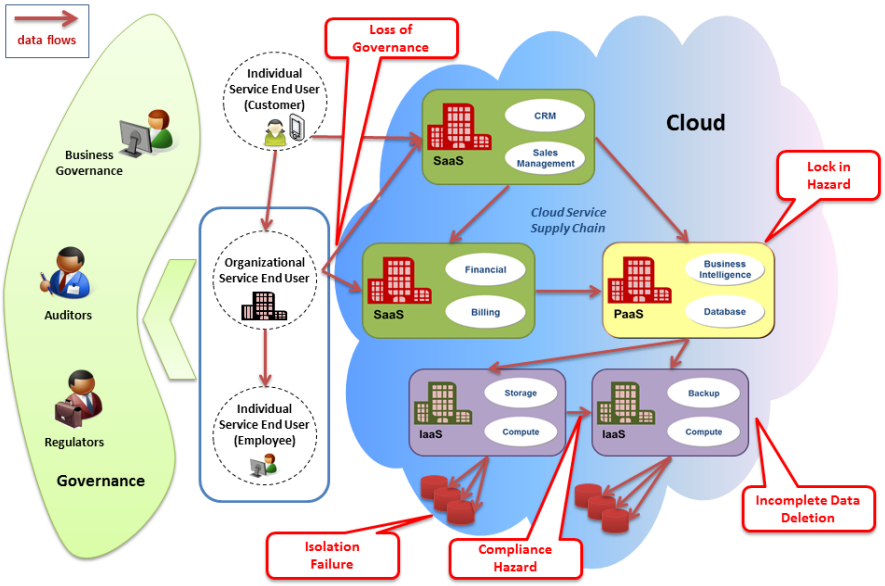

Figure 2 Threats in a Cloud Ecosystem

The governance challenges in cloud computing are in part related to the complex provider supply chains in such ecosystems, for instance, where the Software-as-a-Service (SaaS) application that a user interacts with may be based on another provider's Platform-as-a-Service (PaaS) solution, which in turn may be running on yet another provider's Infrastructure-as-a-Service (IaaS) offering. To complicate things even further, services and data may be replicated horizontally among multiple providers, making it extremely difficult to determine where your data is at any one time. As if the complexity of the Cloud ecosystem supply chains was not enough, the scale of cloud operations is daunting. Economy of scale is one of the most often

quoted arguments for the viability of cloud computing, and the major cloud provides operate data centres of a size which is downright intimidating. Finally, the vast amounts of data on individuals available to cloud providers enable them to perform sophisticated data mining operations, revealing things about us that we may not even know ourselves.

Isolation Failure. Multi-tenancy means data from several customers are stored and processed on the same infrastructure, and improper protection may expose confidential data. An example of isolation failure would be if it was possible to log into anyone's account without a password. Isolation failures do not always arise due to provider errors. Cloud consumers might misconfigure their services, effectively making data which should have been private publicly available on the internet. A specific case of the data isolation challenge comes when a given employee (say, of Acme Inc.) uses an Enterprise SaaS application for work-related transactions, but at the same time also uses an eGoverment/eHealth app in the capacity of a private citizen (see Figure 3). This is challenging on several levels; not only is there a risk of mixing personal information and sensitive business information on the same device (which is strictly speaking not a cloud challenge), but the eGovernment SaaS service and the Enterprise SaaS service may actually be based on the same IaaS service (as Figure 3 illustrates).

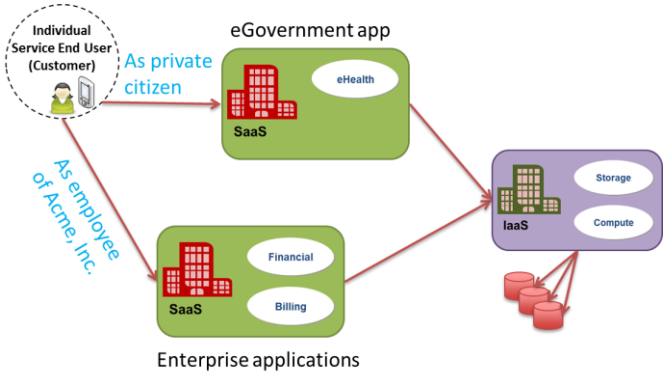

Figure 3 Example of data isolation problem

This highlights that isolation failures could both have consequences for personal data (individuals accessing data belonging to other individuals) and sensitive company data (one company accessing another company's trade secrets) and even combinations of the two (Acme Inc. accessing the personal eHealth data of their employees).

Compliance Hazard. Data protection laws prohibit transferring personal data from EU to jurisdictional domains without sufficient protection. Determining what constitutes *"sufficient protection"* might not be something the average cloud user should be expected to manage, but even if a local provider is chosen, there is a high likelihood that data is transferred across borders. Data flows in the cloud ecosystem are dynamic, and may go both horizontally (e.g. between IaaS providers) and vertically (e.g. from Saas to PaaS to IaaS).

Figure 4 illustrates an example where an (imagined) SaaS application in cloud X gathers medical sensor data from home-based patients, passes partially processed data on to another SaaS application in cloud Z, and uses IaaS storage services from cloud Y for long-term storage and backup. Cloud Z may be using the same IaaS services as Cloud X, but it could just as easily be using a completely different cloud. Cloud Y could, in turn, use yet another cloud service to back up the data in their data centres.

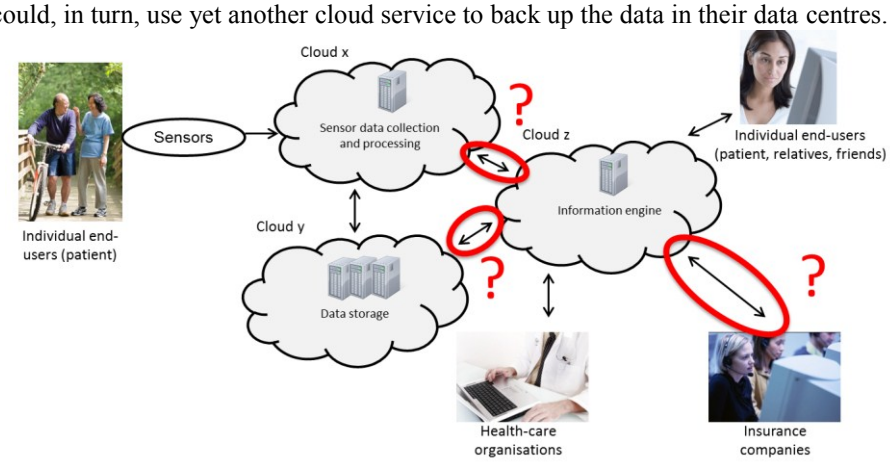

Figure 4 Example of data flows in a cloud ecosystem

Figure 4 also shows that some statistical data is transferred to insurance companies; here it will be important to ensure that the data is properly and fully anonymized, or that the patients have given consent for this use of their data (note that consent may not be required if data are really anonymous, however this could be subject to ethical practices described in professional codes of conduct).

Incomplete Data Deletion. Redundant data storage and data migration may lead to multiple copies being stored on multiple physical infrastructures, and a command to delete a particular piece of data may take a month to take effect (due to data being duplicated across different data centres). Furthermore, some providers state that data may remain in backup logs for 90 days or more, or even indefinitely (i.e. stored forever) [10].

Lock-in Hazard. Proprietary formats may make moving from one cloud provider to another difficult, if not impossible. Data transfer costs may also be prohibitive, and serve as a lock-in feature in itself. There are examples where it is cheap to upload data to a given service, but comparatively expensive to download. This can be the case where a customer is allowed to upload a small amount of data for free every month; a small trickle of data over the years translates to a data deluge when it has to be moved all at once. It may be even more dramatic when a cloud service provider goes out of business – when the Megaupload file sharing service was shut down due to copyright infringement, a large number of innocent customers lost access to their files and images without warning [9]. Some standard terms offered by cloud providers can also leave consumers with little control on how to migrate their data or their accounts, e.g. allow providers to terminate consumers' accounts for any reason at any time, with no advance notice.

Loss of Governance. The challenge related to loss of governance may be seen a combination of all issues discussed in this section, but maybe in particular for a business user of cloud services. When a business places its data and processes in the cloud, control is necessarily ceded to the cloud provider, but the SLA for this service does not necessarily also cover governance services from the CSP, unless this has been specifically negotiated. Furthermore, control of what happens further down the provider chain is not something that automatically can be assumed; this should be clear even from the simple examples we have provided.

4 Accountability in Cloud Ecosystems

Recent research has identified basic features of an accountability-based approach [5], and has highlighted the complexity of accountability [6]. Different definitions of accountability have been proposed (e.g. see [5], [6], [7], [8] for relevant discussions on accountability). This paper is concerned with the problem of supporting and achieving accountability in practice. The following definition captures a shared understanding of accountability based on reviewing previous related work and discussion within the Cloud Accountability Project:

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

Governance here is the processes which devise ways of achieving accountability. The conceptual definition of accountability encompasses different understandings drawn from different disciplines. It is intentionally generally applicable across different domains. Further to this generic definition, we tailor the conceptual definition of accountability to data protection in the cloud. Thus, the following definition contextualises the notion of accountability (i.e. Conceptual Definition of Accountability) and makes it relevant to the problem of data governance in the cloud.

Definition of Accountability for Data Stewardship by Cloud Services: *Accountability for an organisation 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 is 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.*

The definitions highlight the main conceptual aspects of accountability. They characterise the necessary practices emerging in organisations that take an accountability-based approach (with respect to specific attributes of accountability). An analysis that deconstructs the accountability definitions highlights a model

consisting of *accountability attributes*, *practices*, *mechanisms and tools*. Figure 5 shows how they form together a model of accountability for cloud ecosystems.

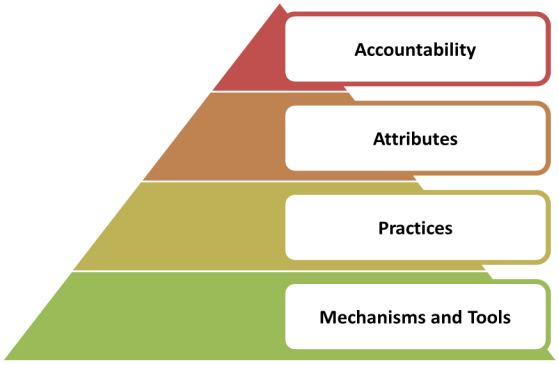

Figure 5 Accountability Attributes, Practices, Mechanisms and Tools

The central elements of this model are:

- **Accountability attributes** – conceptual elements of accountability as used across different domains (that is, the conceptual basis for our definition, and related taxonomic analysis)
- **Accountability practices** – emergent behaviour characterising accountable organisations (that is, how organisations operationalize accountability or put accountability into practices)
- **Accountability mechanisms and tools** – diverse mechanisms and tools that support accountability practices (that is, accountability practices use them).

5 Legal Perspectives of Data Protection in Cloud Ecosystems

The Data Protection Directive [12] lays down rules for the processing of personal data and recognizes specific rights of individuals on their personal data, while ensuring that such data can move freely within the internal EU market. When data can be linked directly or indirectly to an individual (the so-called data subject) they qualify as personal data. Only data that are truly anonymous are excluded from the provisions of the Directive [13]. Great amounts of information can be found in cloud ecosystems, some of which qualifies as personal data, while some does not. When information in cloud ecosystems refers to an identified or identifiable natural person, then the Data Protection Directive may apply. Important territoriality issues arise, which will not be further discussed within this work.

The European data protection legal framework distinguishes between two principal actors (besides the data subjects): the data controller and the data processor. The data controller is the one who defines the means and the purposes for the processing of personal data, while the data processor carries out the processing on behalf of the controller. This distinction is of great importance as the data controller (and not the data processor) is the party who carries the obligations described in the

Data Protection Directive and also the party required to define the details of the data processing. Cloud computing raises significant challenges in identifying who are the responsible entities, in order to assign accountability obligations, since usually multiple actors are involved. Figure 6 shows an example of accountability relationships for cloud ecosystems drawn from a data protection viewpoint (note that other relationships creating different governance models are possible too).

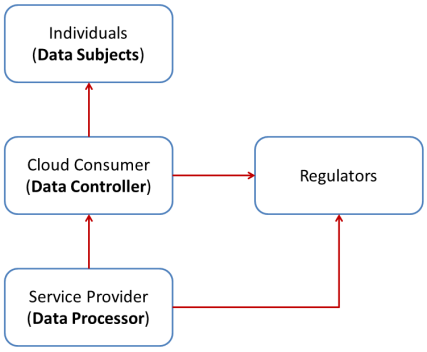

Figure 6 Data Subject, Controller, Processor and Regulator

The Directive contains general principles for processing personal data that have to be respected, which balance the interests both of data controllers and of data subjects [14]. These principles include fairness and data quality (data should be correct and up-to-date), purpose specification and use limitation (data may only be processed for previously specified purposes), and legitimate ground (processing must be based for instance on user consent, a contract, a legal obligation, or vital interest of the data subject). Of special importance is the principle that the design of data-processing systems be aimed at processing either no personal data at all or as little as possible (data avoidance and data minimisation) [15]. Consent will often be the legal basis for processing personal data in cloud computing, but special attention must be paid to the processing of health and medical data: the processing of these is in principle prohibited, unless special grounds apply. Figure 4 (in section 3) shows an example in which the cloud providers collect and process personal data of persons, who in many cases are also patients. In such cases, the rules on the protection of sensitive data have to be taken into account, with attention for the special requirements that relate to health and medical data [11].

The Data Protection Directive also addresses the issue of data security, imposing a statutory obligation on data controllers to ensure that personal data are processed in a secure environment. Moreover, the Directive contains rules for the transfer of personal data to third countries, an issue of great importance in all trans-border applications. The transfers of data between cloud computing providers, which are located in different countries, raise issues on the trans-border flows of personal data. Specifically, questions arise on the entities that have to take into account data protection from the design of the system and on who is responsible for the integrity and security of the data. Especially when several providers are involved in applications that transfer personal data of users in a way that leaves the control of the developer of the system, the identification of the responsible parties becomes difficult

[17]. The European Commission has proposed the replacement of the Directive with a Regulation, which aims at ensuring a consistent level of protection for individuals among the 27 European Member States and at providing legal certainty and transparency for economic operators, including micro, small and medium-sized enterprises [16]. Although the draft General Data Protection Regulation will take several years before coming into effect, developers of cloud computing systems and applications have to take into account the envisaged amendments and changes in order to make sure that they will comply with the future legislation as well. The draft Regulation introduces the concept of joint controllers and creates stricter accountability obligations for data processors. Of particular relevance are requirements to implement 'data protection by design and by default' and to execute Data Protection Impact Assessments for all operations that present specific risks.

6 Discussion and Concluding Remarks

This paper highlights accountability as an enabler for cloud ecosystems. In order to make accountability meaningful, it is useful to distinguish between accountability attributes, practices, mechanisms and tools. Figure 7 illustrates the contextual relevance of the accountability (in terms of the model basics, i.e. attributes, practices, mechanisms and tools) between cloud ecosystems with respect to regulatory regimes (e.g. Data Protection Directive).

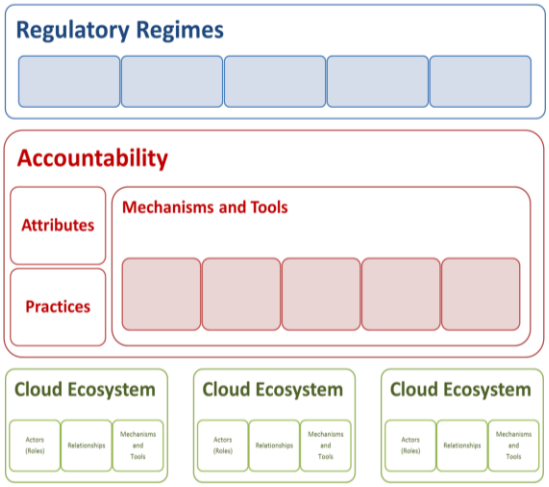

Figure 7 Accountability context

The attributes of accountability identify relevant concepts, e.g. responsibility, liability and transparency, that support accountability (other attributes can be observability, verifiability and attributability). Accountability practices concern different operational aspects of cloud ecosystems. That is, despite the supporting elements, accountability practices differ across cloud ecosystems. Emerging relationships among attributes and practices of accountability, supported by specific

mechanisms and tools (e.g. technical tools like software implementing security controls and policies as well as legal mechanisms like sanctions), enable cloud ecosystems to position and comply with relevant regulatory regimes. This stresses an operational interpretation (that is, what work practices and other domain-specific factors show being accountable) of accountability in cloud ecosystems. The different attributes of accountability and their contextual practices enable various mechanisms and tools in cloud ecosystems.

The emerging accountability model enables the analysis of accountability relationships among cloud actors. The attributes of accountability (e.g. responsibility, liability and transparency) highlight accountability relationships among cloud actors. The analysis of such accountability attributes enables us to understand how accountability relationships emerge in cloud ecosystems. For instance, let us consider responsibility in order to analyse emerging relationships among cloud actors. A Cloud Service Provider (CSP) is responsible to its customers, as specified in contracts between them, for the way personal data are stored and managed. Similarly, the CSP is responsible to Data Protection Authorities (DPAs) to comply with data protection legislation – the extent of such responsibility varies depending on the CSP's role in managing personal data. Each CSP's employee is responsible to the provider (employer), but not directly to Customers and DPAs. This is an example of how the elements of accountability enable us to analyse emerging relationships among cloud actors. Different accountability relationships emerge among actors in cloud ecosystems. Chains of accountability consist of the sets of relationships existing between any two actors in a cloud ecosystem. The characterization of accountability and the analysis of the emerging relationships among actors allow us to identify opportunities (in terms of mechanisms and tools) to support accountability in cloud ecosystems. Our accountability characterization of cloud ecosystems involves the identification of the main actors and the analysis of their relationships with respect to the attributes of accountability. The emerging accountability model enables cloud ecosystems that need to comply with regulatory regimes constraining their application domains. It supports diverse mechanisms and tools throughout chains of accountability relating actors one another.

In conclusion, this paper has discussed the problem of data governance and protection in cloud ecosystems. The problem has been explained from two different viewpoints: a technical one discussing emerging threats affecting data governance in cloud ecosystems, and a legal one highlighting the complexity of the EU Data Protection Directive constraining the provision of cloud services. This paper is concerned with addressing both perspectives by supporting accountability in cloud ecosystems. The paper has introduced a model of accountability in the cloud that enables diverse mechanisms and tools, which support organisational practices for being accountable.

Acknowledgments. This work has been partly funded from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no: 317550 (A4CLOUD – http://www.a4cloud.eu/) Cloud Accountability Project. Figure 2 Threats in a Cloud Ecosystem is taken from a presentation by Siani Pearson. Figure 4 Example of data flows in a cloud ecosystem is based on original by Karin Bernsmed. We also would like to thank the contributions to the accountability

conceptual framework of our partners within the Cloud Accountability Project: Daniele Catteddu, Giles Hogben, Amy Holcroft, Theofrastos Koulouris, Ronald Leenes, Christopher Millard, Maartje Niezen, David Nuñez, Nick Papanikolaou, Siani Pearson, Daniel Pradelles, Chris Reed, Chunming Rong, Jean-Claude Royer, Dimitra Stefanatou, Vasilis Tountopoulos, Tomasz Wiktor Wlodarczyk.

References

1. European Commission, Advances in Clouds – Research in future cloud computing, Expert Group Report, Public version 1.0. European Union (2012).
2. ENISA, Cloud Computing: Benefits, risks and recommendations for information security. European Network and Information Security Agency (2009).
3. Mell, P., Grance, T.: The NIST Definition of Cloud Computing, NIST Special Publication 800-145 (2011).
4. Article 29 Data Protection Working Party, Opinion 3/2010 on the principle of accountability, 00062/10/EN WP 173 (2010).
5. The Galway Project, Accountability: A compendium for stakeholders. The Centre for Information Policy Leadership (2011).
6. Guagnin, D., et al. (eds.): Managing privacy through accountability. Palgrave Macmillan (2012).
7. Weitzner, D. J., et al.: Information Accountability. Commun. ACM 51(6):82-87 (2008).
8. Pearson, S.: Toward Accountability in the Cloud. IEEE Internet Computing, 15(4):64-69 (2011).
9. Stilgherrian: Collateral damage in the copyright wars. Accessed online (June 2013): http://www.abc.net.au/unleashed/3787384.html
10. Bennett, C., Molnar, A., Parsons, C. A.: Forgetting, Non-Forgetting and Quasi-Forgetting in Social Networking: Canadian Policy and Corporate Practice. Accessed online (January 28, 2013): http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2208098
11. Dumortier, J., Goemans, C.: Legal Challenges for Privacy Protection and Identity Management. In: Jerman-Blažič, B., Schneider, W., Klobučar, T. (eds.) Security and Privacy in Advanced Networking Technologies, NATO Science Series, III: Computer and Systems Science, vol. 193, pp. 191-212, IOS Press, (2004).
12. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281, 23/11/1995, p. 0031-0050 (1995).
13. Kuner, C., European Data Protection Law – Corporate Compliance and Regulation, Oxford: Oxford University Press, 2008, p. 51.
14. Walden, I.: Privacy and Data Protection. In: Reed, C., Angel, J. (eds.) Computer Law: The Law and Regulation of Information Technology, 7th edition, Oxford University Press (2011).
15. Holznagel, B., Sonntag, M.: A Case Study: The JANUS Project. In: Nicoll, C., Prins, J. E. J., van Dellen, M. J. M. (eds.) Digital Anonymity and the Law – Tensions and Dimensions, Information Technology and Law (No. 2), The Hague (2003).
16. Proposal for a General Data Protection Regulation, COM(2012) 11 final, 25.01.2012.
17. Löhr, H., Sadeghi, A.-R., Winandy, M.: Securing the e-health cloud. In: Veinot, T. (ed.) Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10), pp. 220-229, ACM (2010).
18. Article 29 Data Protection Working Party, Opinion 05/2012 on Cloud Computing, 01037/12/EN, WP196 (2012).