Project title: Enforceable Security in the Cloud to Uphold Data Ownership
Project acronym: ESCUDO-CLOUD
Funding scheme: H2020-ICT-2014
Topic: ICT-07-2014
Project duration: January 2015 – December 2017

D6.4
Final Version of Data Management Plan

Editors: Pierangela Samarati (UNIMI)
Reviewers: Mark Shackleton (BT)
Ger Hallissey (EMC)
Christian Cachin (IBM)
Daniel Bernau (SAP)
Neeraj Suri (TUD)
Stefano Paraboschi (UNIBG)
Sabine Delaitre (WT)

Abstract

This deliverable describes the data management plan, that is, the policy regulating collection, management, sharing, archiving, and preservation of data in the ESCUDO-CLOUD project.

Type	Identifier	Dissemination	Date
Deliverable	D6.4	Public	2017.12.31

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644579. This work was supported in part by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract No 150087. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission or the Swiss Government.

ESCUDO-CLOUD Consortium

1.	Università degli Studi di Milano	UNIMI	Italy
2.	British Telecom	BT	United Kingdom
3.	EMC Corporation	EMC	Ireland
4.	IBM Research GmbH	IBM	Switzerland
5.	SAP SE	SAP	Germany
6.	Technische Universität Darmstadt	TUD	Germany
7.	Università degli Studi di Bergamo	UNIBG	Italy
8.	Wellness Telecom	WT	Spain

Versions

Version	Date	Description
0.1	2017.11.30	Initial Release
0.2	2017.12.14	Second Release
1.0	2017.12.31	Final Release

List of Contributors

This document contains contributions from different ESCUDO-CLOUD partners. Contributors for the chapters of this deliverable are presented in the following table.

Chapter	Author(s)
Executive Summary	UNIMI
Chapter 1: Data Management Plan	UNIMI
Chapter 2: Conclusions	UNIMI

Contents

Executive Summary

This deliverable describes the data management plan of the ESCUDO-CLOUD project. It summarizes the different kinds of data managed and produced by the project, describing their representation, management, sharing, archiving and preservation. Data regulated by the plan are all data that are either managed or produced by ESCUDO-CLOUD (i.e., documents and software tools).

We recall that, as already noted at the beginning of the project, ESCUDO-CLOUD does not use or manage any sensitive data, like Personally Identifiable Information (PII), confidential commercial information, or real-life data. In fact, the project has strictly built the technical framework enabling users (companies as well as individual users) to maintain control of their data in the cloud, but has neither collected nor used any personal data either directly or indirectly. No real-life data (and therefore no PII or sensitive information more specifically) have been collected or processed, or exposed to secondary use. Development, testing, validation, as well as demonstration of all the solutions developed in ESCUDO-CLOUD have been carried out only on purely synthetic data. Such data have been generated randomly, based only on structural information (such as table structure and data types) with no consideration of – or connection to – real-life data instances.

In summary, data produced by ESCUDO-CLOUD are technical solutions (in the form of documents or software tools). In its working the project has not acquired any data from other sources. The ESCUDO-CLOUD Consortium has been committed to timely and rapid distribution of the project's results, making them widely available and openly accessible. ESCUDO-CLOUD has pursued an open-access policy, making results and publications publicly available. Data needed for the working of the project (such as work communications and progresses of the technical work) have been restricted to the project participants.

In this deliverable, we describe the management plan that regulated the different kinds of data that have been handled or produced by ESCUDO-CLOUD.

1. Data Management Plan

Data produced by the project are technical solutions (in the form of documents or software tools). The project has not acquired or produced any specific data sets. Therefore, in the following we refer the data management plan to the different kinds of data managed or produced by the project distinguishing among the following kinds of data:

1. *data/information about the project*, that is, all static data describing the project that are produced for dissemination purposes (e.g., fact sheet, consortium, objectives, vision, planned work and its organization in work packages), as well as project progresses (e.g., news reporting related events, dissemination and exploitation relevant information);

2. *data for the working of the project*, that is, all information needed for the communication and interaction among the partners working in the project, (e.g., name and contact information of people working in the project, meetings, communication among partners, and intermediary progresses);

3. *documents produced by the project*, that is, scientific papers and deliverables/work documents presenting scientific and technical solutions produced by ESCUDO-CLOUD;

4. *software tools produced by the project*, that is, object and source code of the software implementing the scientific and technical solutions of the project, together with their associated metadata (e.g., unique identifier, creator/s, and versions).

There are three main servers where data on the project have been hosted. These have been dedicated to manage the following:

- Web site of the project (http://www.escudocloud.eu), with a public and a restricted area;
- Project document repository SVN (https://www.escudocloud.eu/svn/repos/escudocloud), for the project's internal working and communication;
- Mailing lists (escudocloud-...@filibusta.crema.unimi.it), for project working and communication.

All servers reside within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. They are backed up weekly.

Data about the project
Data set description
This kind of data refers to all the data describing the project that has had to be made publicly available (for dissemination purposes) such as: fact sheet, consortium, objectives, vision, planned work and its organization in work packages. In addition to these static data about the project, this kind of data includes information that has been continuously updated as the project progressed, such as news reporting related events, dissemination and exploitation relevant information.
Standards and metadata
Data are organized with a hierarchical structure allowing easy retrieval and navigation. Content management is handled with the Joomla content management structure.
Data sharing
Data are publicly accessible via the project web site: http://www.escudocloud.eu.
Archiving and preservation
Data are stored on a server residing within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. They will be maintained at least five years after the life of the project.
Example – ESCUDO-CLOUD home page

Menu

Home
The Project
Fact sheet
Consortium
Objectives
Vision
Research work
Work packages
Results
Deliverables
Publications
Innovation
Comprehensive requirements analysis
Data protection at rest
Object storage integrity and consistency
Over-Encryption in Swift
Mix&Slice
Query on encrypted data
Selective sharing
Query integrity
Access privacy
Multi/Federated cloud architecture
User requirements support
News and events
General news
Keynotes\Talks
Exploitation activities
Contact
Restricted area

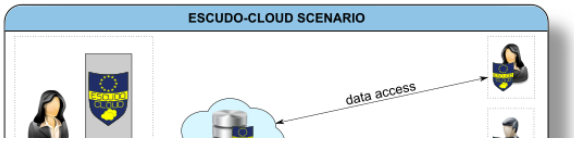
Welcome to ESCUDO-CLOUD

Background and motivation

Cloud computing is increasingly a necessary strategical ICT infrastructure component for European companies to successfully compete in the world-wide economy. The advantages of renting ICT infrastructures, platforms, and services, with easy access to scalability and elasticity, are driving an ever accelerating transfer toward the cloud of data and applications. Unfortunately, such a convenience comes at the price of the data owners losing control over their own data and any consequent misuse or security threats on them, which often limit the owner's adoption of the cloud's potential capabilities. On one hand, cloud providers can be assumed to employ basic security mechanisms for protecting data in storage, processing, and communication, devoting resources to ensure security that many medium and small companies may not be able to afford. On the other hand, data owners, when relying on the cloud, lose control over data and their processing, hence leaving them potentially exposed. Today data owners have to choose between having security but limited functionality or fully enjoying functionality but compromising on security and privacy guarantees. This situation has a strong detrimental impact on the adoption and acceptability of cloud services. Data owners may refrain from relying on the cloud for certain data, which they consider more sensitive or critical, or they use the cloud but remain exposed to the consequences of improper protection and control.

The goal of ESCUDO-CLOUD

> *ESCUDO-CLOUD aims at empowering data owners as first class citizens of the cloud. ESCUDO-CLOUD provides effective and deployable solutions allowing data owners to maintain control over their data when relying on Cloud Service Providers (CSPs) for data storage, processing, and management, without sacrificing on functionality.*



Data for the working of the project
Data set description
This kind of data refers to the information needed for the working of the project itself, such as name and contact information of people working in the project, meetings, communication among them, and intermediary progresses. As established by the Consortium Agreement (art 10.9), for people working on the project, only contact information has been acquired, and ESCUDO-CLOUD has not made available any personal data to other parties or processed any personal data on behalf of other parties.
Standards and metadata
Contact information is organized in different mailing lists to allow easy retrieval and references. Work in progress and documents for collaborative working are organized in a hierarchical structure with nested directories and files with self-explanatory names, and are accessed via an SVN service. Mailing lists and SVN information are also accessible via the restricted area of the project web site using individual credentials (login/password) assigned to each project participants.
Data sharing
Data sharing is restricted to project participants. Access to the SVN service is available to all project participants and is regulated with control of login and (randomly generated) password distributed to each individual participant by the administrator of the server. Access to the different mailing lists of the project is regulated with control of login and (randomly generated) password distributed to each individual participant by the administrator of the server.
Archiving and preservation
Data are stored on a server sitting within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. They will be maintained at least five years after the life of the project.
Example – Main mailing lists

`escudocloud@filibusta.crema.unimi.it`
`escudocloud-adpeople@filibusta.crema.unimi.it`
`escudocloud-tb@filibusta.crema.unimi.it`
`escudocloud-mb@filibusta.crema.unimi.it`
`escudocloud-pubs@filibusta.crema.unimi.it`
`escudocloud-wp1@filibusta.crema.unimi.it`
`escudocloud-wp2@filibusta.crema.unimi.it`
`escudocloud-wp3@filibusta.crema.unimi.it`
`escudocloud-wp4@filibusta.crema.unimi.it`
`escudocloud-wp5@filibusta.crema.unimi.it`
`escudocloud-wp6@filibusta.crema.unimi.it`

Project results - Documents
Data set description
This kind of data refers to all documents produced by the project. Among them we distinguish: deliverables/work documents and scientific papers.
Standards and metadata
All documents are made available in PDF format. Deliverables and work documents are identified following an easy-to-use notation with three parts: a letter that determines the kind of document (D for Deliverable, W for Work document); the number of the work package that coordinates its production; and a serial number that uniquely identifies the document within that work package. All documents have .bibtex metadata, easing the search, which can be exported for easy reference. These metadata include the ones prescribed in the Grant Agreement (art 29.2).
Data sharing
IPR and dissemination of project results are regulated by the Consortium Agreement (art 8). IPR of results remains with the project's party that generated them. Sharing and dissemination follow an open access policy. For sharing and dissemination, we distinguish work documents, deliverables, and scientific papers. Work documents, being internal deliverables produced for assessing the progress of work and for official communication of intermediate results among partners, are visible only to project's participants. All deliverables (apart from those reporting financial or exploitation information) are classified as public (PU). Consequently, they are made accessible to the general public following their final acceptance by the EC. This dissemination and sharing is made via the web site of the project (http://www.escudocloud.eu, link "Results/Deliverables"). ESCUDO-CLOUD embraces an open access policy, and also all scientific papers are made publicly available via the project website. Before undergoing public release, paper publication undergoes an internal process within the ESCUDO-CLOUD Consortium, regulated by the Grant Agreement (art. 29) and Consortium Agreement (art. 8.3) of the project. In particular, "prior notice of any planned publication shall be given to concerned Parties at least 21 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 15 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted." At this point, the paper is made publicly available via the project public web site (http://www.escudocloud.eu, link "Results/Publications").

Before papers are accepted as publication at scientific venue, this sharing and dissemination is made possible via the partners' institutional archives and other open archives (e.g., arXiv and Zenodo). For papers accepted for formal publication by publishers, this public sharing and dissemination access is made possible via "green" open-access publication (i.e., "self-archiving"), which is in line with the copyright policy of major institutions and associations of the most selective and recognized conferences and journals. Within this policy, the publishers allow authors to post the final versions of accepted papers in their personal web site, the web site of their employers, or selected pre-authorized institutional web sites. Project publications are then hosted (and publicly accessible) from the partners' website or open archives and linked from the web site of the project, thus ensuring broad visibility and easy access.

Archiving and preservation

Deliverables are stored on a server sitting within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. They will be maintained at least five years after the life of the project.

Scientific papers are linked from the project web site and are stored on partners institutional archives (also periodically backed up) and other open archives (like Zenodo), which provides guarantees of continuity of service preservation of access.

Example – Web page presenting deliverables

Menu

Home
The Project
Fact sheet
Consortium
Objectives
Vision
Research work
Work packages
Results
Deliverables
Publications
Innovation
Comprehensive requirements analysis
Data protection at rest
Object storage integrity and consistency
Over-Encryption in Swift
Mix&Slice
Query on encrypted data
Selective sharing
Query integrity
Access privacy
Multi/Federated cloud architecture
News and events
General news
Keynotes\Talks
Exploitation activities
Contact
Restricted area

Deliverables

D1.1 First version of requirements from the use cases (M6) [pdf] [bib]
It provides a first version of the security requirements coming from the use cases.

D6.1 Data management plan (M6) [pdf] [bib]
It provides a detailed description of the data management plan.

D1.2 Requirements from the use cases (M12) [pdf] [bib]
It provides the final version of the security requirements coming from the use cases.

D2.1 Report on data protection techniques (M12) [pdf] [bib]
It provides a first version of the data protection techniques designed in WP2.

D3.1 Report on techniques for selective access (M12) [pdf] [bib]
It provides a first version of the techniques for enforcing selective access on outsourced data.

D4.1 First report on security metrics and assessment (M12) [pdf] [bib]
It provides a formal definition of the security metrics that can be used to characterize the guarantees offered by different providers.

D5.1 First report on dissemination, communication, and exploitation (M12) [bib]
It provides a description of the activities performed to disseminate and exploit the project's results of the first year.

D6.2 First periodic report: Progress of work, use of resources, financial report (M12) [bib]
It provides an overview of the progress of work, the use of the resources, and a financial statement from each partner.

D2.2 Tools for protecting confidentiality and integrity of data and access (M18) [pdf] [bib]
It provides a first version of the tools developed for protecting the confidentiality and integrity of outsourced data and their accesses.

D3.2 Report on techniques for security testing (M21) [bib]
It provides a first version of the techniques for security testing.

Project results - Software tools

Data set description

This kind of data refers to all software tools produced by the project. All software tools developed within the project have also corresponding deliverables or work documents.

Standards and metadata

Code of software tools produced by the project are written using commonly-used programming languages (e.g., Java, C++, Python) or shell scripts. Every tool has structured metadata associated, specifying information such as unique identifier, creator/s, and versions, allowing for easy reference and retrieval.

Data sharing

IPR and dissemination of project results are regulated by the Consortium Agreement (art 8). IPR of results remains with the project's party that generated them.

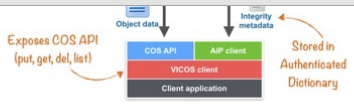
As for sharing, deliverables/work documents reporting on software tools, are managed according to the policy set for deliverables/work documents.
Object code of tools developed by academic partners and by industrial partners embracing an open access policy (i.e., IBM), subject to the applicable IPR clauses, is accessible to the general public. Public dissemination and sharing is also made possible via pointers to the software from the web site of the project.

Source code of software tools built by academic partners and by industrial partners embracing an open access policy (i.e., IBM), is provided as building blocks under an open source license and as applicable under the relevant IPR agreement.

Archiving and preservation

Object and source code of tools produced within ESCUDO-CLOUD is stored and managed by the software creator. Object code is also stored for dissemination in public repositories (e.g., GitHub), with pointers to it from the project's web site.

Example – Pointer to object code



The VICOS system consists of three components:

1. A *cloud object store* (*COS*) service, as offered by commercial providers. It maintains the object data (bulk data) stored by the clients using VICOS.
2. The *VICOS server* that runs remotely as a cloud services accessed by the VICOS client; it stores integrity-specific metadata of the object data being outsourced to the cloud storage service. The metadata is protected through the *Authenticated data structure Integrity Protocol* (*AIP*) for a simple key-value store.
3. The *VICOS client* enables clients to access the cloud-storage service and transparently protect the integrity and consistency of their object data. It exposes the cloud object store interface to a client application. During each operation, the VICOS client consults the cloud object store (using a COS API) for the object data itself and the VICOS server for integrity-specific metadata (through an AIP client). The integrity-specific metadata consists of a unique key of an object in the COS and its cryptographic hash.

The cloud object store and the VICOS server are *both in the untrusted domain*; they may, in fact, collude against the clients.

Related Publications

- Marcus Brandenburger, Christian Cachin, Nikola Knežević "Don't Trust the Cloud, Verify: Integrity and Consistency for Cloud Object Stores" in ACM Transactions on Privacy and Security (TOPS), Volume 20 Issue 3, August 2017, pp. 1-30 (article 8)
- Marcus Brandenburger, Christian Cachin, Nikola Knežević "Don't Trust the Cloud, Verify: Integrity and Consistency for Cloud Object Stores" in Proc. of the 8th ACM International Systems and Storage Conference (SYSTOR), Haifa, Israel, May 26-28, 2015

Software

- VICOS sofware is available at https://github.com/ibm-research/vicos

2. Conclusions

This deliverable provided a description of the Data Management Plan (DPM) for managing the data generated and collected during the project. Specifically, the DMP described the data management life cycle for all datasets produced by the project. It covered:

- data/information about the project;
- data for the working of the project itself;
- documents (scientific papers, and deliverables/work documents) produced by the project;
- software tools produced by the project.

For each dataset, this deliverable included a description and information about the methodology and standards applied, whether the dataset is shared/made open and how, and how the dataset is preserved.