



Deliverable D1.5

Data Management Plan

Editor(s):	Leire Orue-Echevarria (TECNALIA)
Responsible Partner:	TECNALIA
Status-Version:	Final – v1.0
Date:	31.05.2017
Distribution level (CO, PU):	PU

Project Number:	GA 731533
Project Title:	DECIDE

Title of Deliverable:	Data Management Plan
Due Date of Delivery to the EC:	31.05.2017

Workpackage responsible for the Deliverable:	WP1 – Project Management
Editor(s):	Leire Orue-Echevarria (TECNALIA)
Contributor(s):	AIMES, INNOVATI, ARSYS
Reviewer(s):	Pieter Gryffroy
Approved by:	All Partners
Recommended/mandatory readers:	All WPs

Abstract:	Plan detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.
Keyword List:	Data Management Plan
Licensing information:	This work is licensed under Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) http://creativecommons.org/licenses/by-sa/3.0/
Disclaimer	This document reflects only the author's views and the Commission is not responsible for any use that may be made of the information contained therein

Document Description

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	23.01.2017	TOC	Leire Orue-Echevarria (TECNALIA)
v0.2	11.05.2017	Sent to internal QA review	Leire Orue-Echevarria (TECNALIA)
V0.3	15.05.2017	Changes after internal QA review	Leire Orue-Echevarria (TECNALIA)
V1.0	31.05.2017	Version ready to submit	Leire Orue-Echevarria (TECNALIA)

Table of Contents

Table of Contents	4
Terms and abbreviations.....	5
Executive Summary	6
1 Introduction.....	7
1.1. About this deliverable	7
1.2. Document structure	7
2 Data Summary	8
2.1. Data related to the use cases	9
2.1.1 AIMES	9
2.1.2 ARSYS.....	9
2.1.3 INNOVATI.....	10
2.2. Data related to scientific publications.....	10
2.3. DECIDE public deliverables.....	10
2.4. DECIDE Open Source Software.....	10
3 Fair Data	11
3.1. Data related to the use cases	11
3.1.1 AIMES	11
3.1.2 ARSYS.....	11
3.1.3 INNOVATI.....	11
3.2. Data related to scientific publications.....	11
3.3. Data related to deliverables	11
3.4. Open Source Software.....	12
4 Allocation of resources.....	12
5 Data Security	12
6 Ethical Aspects.....	13
7 Conclusions.....	13

Terms and abbreviations

ACSml	Advanced Cloud Service (meta-) Intermediator
CD	Continuous Delivery
CI	Continuous Integration
CQ	Continuous Quality
DMP	Data Management Plan
EC	European Commission
GDPR	General Data Protection Regulation
KPI	Key Performance Indicators
KR	Key Results
OS	Open Source
WP	Work Package

Executive Summary

This deliverable aims to present a plan for the data management, collection, generation, storage and preservation related to DECIDE activities. In this action, we envision five different types of data: data related to the use cases, data related to the meta-analysis to be done in the social-sciences tasks, data coming from publications, public deliverables and open source software.

The document presents, following the EC template [1], how these different types of data will be collected, who the main beneficiaries are, and how DECIDE will store them, manage them, and if the project will make them accessible, findable and re-usable. The text continues with the foreseen resources needed for the openness and data to finalize with security and ethical aspects that will be taken into consideration in the context of DECIDE.

This plan is the first version of the data management plan, which will be updated in subsequent versions (M18 and M36) as part of the Technical Reports, having as input the work carried out in the use cases (WP6), the technical work packages (WP2 – WP5) and the dissemination activities (WP8).

1 Introduction

1.1. About this deliverable

This deliverable focuses on the management of the data in DECIDE. In DECIDE there will be two different data strands. The first strand relates to the publications generated as part of the research activities, and the second strand relates to the data collected from the use cases, that will be used as part of the implementation of the different key results established in the project.

1.2. Document structure

The document follows the established H2020 template for a Data Management Plan (DMP) [1]. Section 2 presents a summary of what the purpose of the data collection and generation is in the case of DECIDE. Section 3 explains how the data and metadata will be made fair, and thus accessible, findable and reusable. Section 4 briefly explains how the financial resources for this openness are envisioned at this stage to be allocated. Section 5 and 6 focus on the security and ethical aspects respectively. Section 7 presents the conclusions and future work.

2 Data Summary

DECIDE aims to create a set of tools to design, deploy and operate multi-cloud aware applications in an ecosystem of reliable, interoperable and legally compliant services. DECIDE will not generate any data beyond those of the public deliverables and open access publications. DECIDE will use data in the use cases. AIMES' use case with its clinical trials application may seem especially relevant but it should be noted that all data used shall be fictional.

To realize the main goal of the project, DECIDE will develop the following Key Results (KR):

- **KR1: Multi-cloud native applications DevOps framework:** This Key Result integrates KR2-KR5, explained next, in addition to the integration and extension of existing tools in the Open Source (OS) communities covering development, continuous integration (CI), continuous quality (CQ) and continuous delivery (CD).
- **KR2: DECIDE ARCHITECT:** ARCHITECT will provide architectural patterns and modelling practices for implementation, optimization and deployment of multi-cloud native applications. Apart from the theoretical description of the patterns and explanations of how they can be implemented, DECIDE ARCHITECT will provide a supporting tool, with suggestions on which pattern is to be applied and in which order.
- **KR3: DECIDE OPTIMUS:** DECIDE will address Multi-cloud deployment simulations as well. DECIDE OPTIMUS aims to simulate the behaviour, in stressful conditions, of the profiled and classified components of a multi-cloud native application deployed on multiple CSPs and using multiple cloud services, so as to provide the most adequate candidate deployment topologies for that application, based on a predefined prioritized set of requirements (e.g. for example, legal compliance, performance and cost, vs. cost, security, legal awareness) defined by the developer. The candidate topologies will be calculated by making use of big data optimization algorithms such as Dandelion codes, genetic algorithms, or Harmony search.
- **KR4: Advanced Cloud Service (meta-) intermediary (ACSml):** ACSml will provide means to assess continuously, through real time verification, the fulfillment by the cloud services of their non-functional properties fulfilment and legal compliance. ACSml will also provide a cloud services store where companies and developers across Europe can easily access centrally negotiated deals of compliant and accredited cloud services for applications developed by the software sector.
- **KR5: DECIDE ADAPT:** ADAPT is DECIDE's self-adaptation tool for the multi-cloud native applications. It is a Software tool to deploy, monitor and (semi-) automatically self-adapt multi-cloud native applications.

Out of all the Key Results envisioned for DECIDE, the one that is most data-related is the ACSml (KR4). In the timeframe of the project, the ACSml will use fictional user data to contract the services from the CSPs and it will provide encryption techniques for login-related information. The ACSml will be the DECIDE tool involved in the contracting of the CSP services, but the data relevant for the execution of the use cases will not be stored in the ACSml but rather in the services offered by the selected CSPs and therefore, external to DECIDE's tools. Following the vision and goal of DECIDE of contracting only legally compliant services, the ACSml will only contract only CSP services from CSPs that state that they comply with the laws and regulations relevant for the domain of DECIDE's use cases. This is especially relevant for the case of the AIMES study.

In DECIDE two distinct environments are envisioned, namely, an integration environment, and a production environment. The integration environment will be deployed at Innovati's DevOps infrastructure and will include the different components developed by the technological partners. For testing purposes, these partners will use synthetic, fictional, data or 'persona' data, but never

real data, anonymized or not, coming from the use cases. The production environment's final location is at this stage not yet decided.

Different types of data will be collected and generated in the context of the project. These can be summarized as follows:

1. Data related to the execution of the use cases, testing KR1 –KR5.
2. Data related to the ACSml
3. Data related to scientific publications.
4. DECIDE public deliverables.
5. DECIDE Open Source Software

2.1.Data related to the use cases

The aim of this section is to establish what the purpose is of the collection and generation of data within DECIDE, which the sources of this data are, if relevant, and how this is important to realize the Key Results of DECIDE.

Use cases are a key pillar of DECIDE as they will validate the work (in WP6) to be performed in the social and technological work packages (WP2-WP5).

Next we proceed to describe the data available at this stage of the project.

2.1.1 AIMES

With respect to the AIMES' use case, the following considerations with respect to data need to be stated:

- No personal or identifiable data will be used as part of the AIMES use case for the DECIDE project
- The streamline use case and other eHealth services will not use, host or process patient identifiable data as part of DECIDE
- Any eHealth related services which need to be accessed as part of demonstrating the viability and usability of tools will require the creation of a separate instance which is entirely separate to any service which hosts sensitive/personal data. In the event data is required to test the DECIDE tools, this data will be 'fake' and randomly generated.

While the AIMES use case does not involve the processing of personal data as a constitutive part, it can never be completely excluded that some accidental processing of personal data might take place, especially in relation to the use case partner's own employees. Any such processing shall however be minor and will take place in accordance with applicable law.

2.1.2 ARSYS

Data used and stored in the ARSYS use case are typically not personal, but rather non-personal data concerning the business continuity of ARSYS.

While the ARSYS use case does not involve the processing of personal data as a constitutive part, it can never be completely excluded that some accidental processing of personal data might take place, especially in relation to the use case partner's own employees. Any such processing shall however be minor and will take place in accordance with applicable law.

2.1.3 INNOVATI

INNOVATI's data needed for the execution of the use case is fictional, used only for testing purposes, aiming to simulate the behaviour of the system in production. INNOVATI stores only fictional data for the use case.

While the INNOVATI use case does not involve the processing of personal data as a constitutive part, it can never be completely excluded that some accidental processing of personal data might take place, especially in relation to the use case partner's own employees. Any such processing shall however be minor and will take place in accordance with applicable law.

2.2.Data related to scientific publications

DECIDE will publish scientific publications in conferences and journals as part of the planned dissemination activities. Following the EC Mandate on Open Access [2], DECIDE adheres to the Open Access policy, choosing the most appropriate route for each case. Whenever possible, DECIDE favours the 'green' open access route, in which the published article or the final peer-reviewed manuscript will be deposited in an online repository, before, at the same time as, or after publication, ensuring that the embargo period requested by certain publishers has elapsed.

Scientific publications data are often made available using accessible pdf files. The metadata to be used will be compliant with the format requested by OpenAire as well as the one requested by the repository where the papers are to be deposited.

The format in which the data related to the scientific publications will be accessible pdf files. The metadata to be used will be compliant with that of the repository where the paper is to be deposited and will be compliant with the format requested by Open Aire, as to ease the index.

DECIDE's partners will use Zenodo [3] for their joint publications. Some institutions, like TECNALIA, have developed their own Open Aire compliant repositories, where TECNALIA researchers have to upload their contributions to. When TECNALIA's researchers do so, these publications are automatically indexed in OpenAire.

2.3.DECIDE public deliverables

All information and material related to the public such as public deliverables, brochures, posters etc., will be freely available on the project website in the form of accessible pdf files. When IPR of foreground knowledge needs to be protected, the corresponding disclosures will be published.

All deliverables include a set of keywords and a brief description, which are meant to facilitate the indexing and search of the deliverables in search engines. The keywords in each deliverable aim to stress the main topics addressed in the document, be it a report or a software- related document.

The audience of the public deliverables of DECIDE range from general audiences, interested in the activities performed in the project, to more specialized audiences such as developers and operators of multi-cloud application or those who wish to learn about the benefits of DECIDE through the experiences gathered through the pilots.

2.4.DECIDE Open Source Software

DECIDE will develop the ICT tools mentioned above. The source code will be released, whenever the IPR of the partners is not breached, under a friendly open source licensing schema still to be decided as part the of the exploitation activities.

DECIDE tools will be developed in a variety of programming languages but deployed using a container-based approach following a micro-services [4] architecture. The size of the source code,

the readme files, the user manual and technical specifications as well as the docker scripts cannot be known at the moment.

The open source software is aimed at developers and operators of multi-cloud native applications.

3 Fair Data

This section focuses on the following aspects, namely, on the feasibility and appropriateness of making data findable, openly accessible, interoperable and reusable in the context of DECIDE.

3.1. Data related to the use cases

At this stage, for all use cases, both the collected and generated data, anonymized or fictional, are not envisioned to be made openly accessible. In principle, if needed, all data collected, stored and processed will be fictional and treated as strictly confidential, and kept for a specific period of time as stated on the consent form. This time period shall be no longer than necessary to achieve the aims of the scenario and to validate the project objectives, and after this point, the data will be destroyed as required.

3.1.1 AIMES

AIMES' use case data will not be made accessible, and thus not 'FAIR'¹.

3.1.2 ARSYS

ARSYS' use case data will not be made accessible, and thus not 'FAIR'.

3.1.3 INNOVATI

INNOVATI's use case data will not be made accessible, and thus not 'FAIR'.

3.2. Data related to scientific publications

The project will favour, whenever possible, the 'green' open access, in which the published article or the final peer-reviewed manuscript will be deposited in an online repository, before, at the same time as, or after publication, ensuring that the embargo period requested by certain publishers has elapsed. The Consortium will ensure open access to the publication within a maximum of six months. DECIDE partners have the liberty to choose the repository where they will deposit their publications, although Open Aire – compliant and Open Aire - indexed repositories such as Zenodo [3] will be favoured. The partner TECNALIA will use its own repository, already indexed by Open Aire.

For the case of the scientific publications, a persistent identification number will be provided when uploading the publications to the selected repository / repositories.

3.3. Data related to deliverables

For the project's publications on the website, the naming convention to be used will be <<Dx.y Deliverable name _ date in which the deliverable was submitted.pdf>>. All deliverables include a set of keywords and a brief description that are aimed to facilitate the indexing and search of the deliverables in search engines, in accordance to the defined template.

The deliverables will be stored at TECNALIA's hosting provider, and for three years beyond the duration time frame of the project.

¹ FAIR stands for Findable, Accessible, Interoperable, and Re-usable.

3.4. Open Source Software

DECIDE has envisioned a freemium business model for the project's Key Results which implies a free version of the software as well as a premium one. In addition to offering a free version of the software, these free versions of the open source components of the DECIDE KRs will be released as open source in a source code repository, namely GitHub, which will be stored at Innovati's premises.

The free versions of these components will be findable and reusable by any developer that is interested in the DECIDE Key Results, in agreement with the open source licenses of the components. Furthermore, DECIDE will explore the possibility of developers outside of the consortium to experience with such components, which will ensure the uptake and sustainability of the software results of the project.

For every software component, a readme file as well as the technical specification document will be released. Moreover, a docker script will also be released with the aim of facilitating the deployment of the DECIDE containerized components in any desired infrastructure.

4 Allocation of resources

DECIDE does not foresee additional needs for resources beyond the duration of the action to handle data or making the data fair. As expressed before, open access repositories will be favoured.

In the case of open source software, the partner TECNALIA will ensure that the github repository is available after the project duration, either by keeping it in its own premises or by transferring it to existing open source projects and communities.

5 Data Security

Out of the Key Results envisioned in DECIDE, the one that is most affected by security requirements is the ACSmI. To this respect, a Security management component has been included. The aim of this component is to be in charge of designing and developing the means to guarantee the secure operation of the ACSmI, including: Identity propagation and federated authentication and authorization. The sub-modules included in this component are:

- The *User management* is responsible for gathering information to create, delete and modify users.
- The *Policy Manager*, is responsible for creating, deleting and modifying the policies of the ACSmI and when a new user is created, is responsible for assigning the policies that apply to this new user and to properly update the user registry.
- The *Role Manager*, is responsible for creating, deleting and modifying the roles of the ACSmI. When a new user is created it is also responsible for assigning the roles to this user. This component will also properly update the user registry.
- The *Authentication Manager*, is responsible for the authentication of the users of the ACSmI (This activity will be carried out at DECIDE framework level, if the ACSmI is integrated). If the credentials are valid, the authentication manager will provide to the console the token of the identification for propagating the identity to the rest of the module when appropriate.
- The *User Registry*, is a database where all the information related to the users is stored.
- The *Data Encryption*, is responsible for encrypting the data within the ACSmI in order to keep these data secure in case of cyber-attacks.
- The *Back-up* is responsible for carrying out incremental back-ups in order to allow the recovery of data of the ACSmI should this be necessary.

- The *Communication security* is responsible for providing secure communication using SSL transport layer encryption both between client and platform and between platform and cloud infrastructures.

Moreover, DECIDE will ensure that the General Data Protection Regulation (GDPR), which will enter into force in May 2018, is complied with. The security components shown above will be implemented with that goal in mind.

6 Ethical Aspects

The basis of ethical research is the principle of informed consent. All participants in DECIDE use cases evaluation will be informed of all aspects of the research that might reasonably be expected to influence their willingness to participate. Moreover, project researchers will discuss before and after each practical exercise (e.g. interview, co-creation session, etc.) to maintain on-going consent. Participants will be recruited by each organization leading the use cases (AIMES, ARSYS and INNOVATI) to perform the planned qualitative assessment of the DECIDE Key Results. For instance, in the case of the evaluation of the usability of the tools this will be performed through questionnaires (see D6.1). This data will be anonymized and reported as aggregated data (when relevant) in the documents related to the evaluation of DECIDE outcomes. If participants wish to withdraw from the participation in the use cases at any time, they will be able to do it, and their data, even anonymized data, will be destroyed.

7 Conclusions

This deliverable has presented the plan for the management of data in the DECIDE project. In this action, mostly data related to scientific publications will be generated. Data coming from fictional users logging into the ACSml will be encrypted and they will not be made available compliant with the FAIR principles. Furthermore, data coming from the use cases will also be fictional, especially relevant in the case of Innovati and AIMES, and in principle also not disclosed as FAIR.

Data coming from publications will be stored in Open Aire indexed repositories favouring the green model whenever possible. Other publications such as deliverables will be stored at TECNALIA's hosting services.

This deliverable will be updated in subsequent releases, namely in M18 and M36 as part of the technical reports. It is envisioned that in those versions the aspects that at this stage are not fully clear will be clarified as work progresses in all the work packages.

References

- [1] European Commission,, "Data Management," July 2016. [Online]. Available: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template. [Accessed 9 January 2017].
- [2] European Commission,, "Guidelines on Open Access to Scientific Publications and Research Data," July 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. [Accessed December 2016].
- [3] Zendo, "Zenodo," [Online]. Available: www.zenodo.org. [Accessed 2016].
- [4] C. Richardson, "Microservice architecture patterns and best practices," 2017. [Online]. Available: <https://microservices.io/>. [Accessed March 2017].