

## A novel reconfigurable by design highly distributed applications development paradigm over programmable infrastructure












### D6.14 – Data Management Plan v1

<b>Editors:</b>	M. Volk (UL), P. Gouvas (UBITECH),
<b>Contributors:</b>	K. Tsagkaris(WINGS), S. Kovaci (TUB), M. Repetto (CNIT), A. Rossini (SINTEF), S. Siravo (MAGGIOLI), G. Kioumourtzis (ADITESS), L. Porwol (NUIG)
<b>Date:</b>	26/10/2015
<b>Version:</b>	1.00
<b>Status:</b>	Final
<b>Workpackage:</b>	WP6 –ARCADIA Framework Specifications
<b>Classification:</b>	Public

## ARCADIA Profile

<b>Grant Agreement No.:</b>	645372
<b>Acronym:</b>	ARCADIA
<b>Title:</b>	A NOVEL RECONFIGURABLE BY DESIGN HIGHLY DISTRIBUTED APPLICATIONS DEVELOPMENT PARADIGM OVER PROGRAMMABLE INFRASTRUCTURE
<b>URL:</b>	<a href="http://www.arcadia-framework.eu/">http://www.arcadia-framework.eu/</a>
<b>Start Date:</b>	01/01/2015
<b>Duration:</b>	36 months

## Partners

	<b>Insight Centre for Data Analytics, National University of Ireland, Galway</b>	<b>Ireland</b>
	<b>Stiftelsen SINTEF</b>	<b>Norway</b>
	<b>Technische Universität Berlin</b>	<b>Germany</b>
 <p style="font-size: small;">Univerza v Ljubljani</p>	<b>Consorzio Nazionale Interuniversitario per le Telecomunicazioni</b>	<b>Italy</b>
	<b>Univerza v Ljubljani</b>	<b>Slovenia</b>
	<b>UBITECH</b>	<b>Greece</b>
	<b>WINGS ICT Solutions Information &amp; Communication Technologies EPE</b>	<b>Greece</b>
	<b>MAGGIOLI SPA</b>	<b>Italy</b>
	<b>ADITESS Advanced Integrated Technology Solutions and Services Ltd</b>	<b>Cyprus</b>

## Document History

Version	Date	Author (Partner)	Remarks
0.1	1/9/2015	M. Volk (UL)	ToC formulated and chapter assignment performed
0.2	15/9/2015	P. Gouvas (UBITECH), M. Repetto (CNIT), K. Tsagkaris(WINGS), S. Kovaci (TUB),	Introduction, Additions to Chapter 2
0.3	29/9/2015	K. Tsagkaris(WINGS), S. Kovaci (TUB), M. Repetto (CNIT), A. Rossini (SINTEF), S. Siravo (MAGGIOLI), G. Kioumourtzis (ADITESS), L. Porwol (NUIG)	Additions to Chapter 3
0.6	13/10/2015	M. Volk (UL), P. Gouvas (UBITECH), K. Tsagkaris(WINGS), S. Kovaci (TUB), M. Repetto (CNIT), A. Rossini (SINTEF), S. Siravo (MAGGIOLI), G. Kioumourtzis (ADITESS), L. Porwol (NUIG)	Finalization of all Chapters
0.9	15/10/2015	M. Volk (UL)	Document submitted for internal review cycle
1.0	27/10/2015	M. Volk (UL), P. Gouvas (UBITECH)	Comments of internal reviewers addressed

## Executive Summary

This deliverable presents the data management plan for the ARCADIA project. This data management plan describes what kind of data is generated or collected in the ARCADIA project and how this data is published openly. A simple decision process is defined that either classifies a result as public or non-public. The publishing platforms used are the project website, the OwnCloud platform and GitHub for open-sourced code. All these platforms can be accessed openly.

The list of publications so far includes a set of deliverables, including the identification of the requirements for the development of the ARCADIA framework (D2.1) and the corresponding ARCADIA context model (D2.2). At the current phase of the project, there are no scientific publications, however contribution has been provided on behalf of the project to a set of EU-projects clustering activities, including the preparation of a white paper with regards to the description of software engineering evolution challenges.

The list of research data expected during the project consists of open-sourced, trust-related software components and a set of statistics and use cases' implementation results. Both datasets are expected to be collected during the implementation and evaluation phase of the project and are therefore subject to change.

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>6</b>
1.1	Scope of the Deliverable .....	6
1.2	Structure of the Deliverable .....	6
1.3	Terminology.....	6
<b>2</b>	<b>Publishing Infrastructure for Open Access.....</b>	<b>7</b>
2.1	Publishing Process.....	7
2.2	Publishing Platforms .....	8
<b>3</b>	<b>Project Results/Datasets.....</b>	<b>9</b>
3.1	Deliverables .....	10
3.2	Publications.....	11
3.3	Research Data.....	11
	<b>Annex I: References .....</b>	<b>12</b>

# 1 Introduction

---

The results of the ARCADIA project need to be published to communicate and spread the knowledge to all interested communities and stakeholders. Published results generate wider interest towards the improvements achieved by the project in order to facilitate and potentiate exploitation opportunities. The goal of this report is listing publishable results and research data and to investigate the appropriate methodologies and open repositories for data management and dissemination. The ARCADIA project partners try to offer as much information as possible generated by the ARCADIA project through open access. Such information includes scientific publications issued by the project consortium, white papers published, contribution to standardization bodies, open source code generated, datasets used or collected through the realization of the use cases and anonymous interview results.

In general, there are two types of project results which differ in the way they are published; namely publications and other research data. Our publication strategy follows the ideas of open access and open research data. Scientific publications and related research data is as far as possible published openly. On the other side not all data collected or generated can be published openly, as it may contain private information or interfere with legal aspects. This kind of data must be identified and protected accordingly.

## 1.1 Scope of the Deliverable

Each project in the EC's Horizon 2020 program has to define what kind of results are generated or collected during the project's runtime and when and how it is published openly. This document describes initially which results are published or expected to be published in the ARCADIA project after the first nine months. For all results generated or collected during the ARCADIA project a description is provided including the purpose of the document, the standards and metadata used for storage, the facility used for sharing openly, and for archiving them in the long term. This document is updated on a regular basis. However, it does not describe how the results are exploited, which is part of D6.15 and D6.16.

## 1.2 Structure of the Deliverable

The document is separated into three sections. The first section defines the purpose of the document, its structure, and terms that are necessary to understand it. In section two, we define a process that needs to be applied to all results collected or generated during the project. The process defines if a result has to be published or not. In addition, we provide a summary of all publishing platforms used in the project. In the third section, we list all publications and related data that is already or may be generated or collected during the project. For each result we provide - in accordance to the data management guideline [1] - a short description, the chosen way of open access, and a long-term storage solution.

## 1.3 Terminology

**Open Access:** Open access means unrestricted access to research results. Often the term open access is used for naming free online access to peer-reviewed publications. Open access is expected to enable others to:

- build on top of existing research results,
- avoid redundancy,
- participate in open innovation, and
- read about the results of a project or inform citizens.

All major publishers in computer science - like ACM, IEEE, or Springer - have participated in the idea of open access. Both green or gold open access levels are promoted. Green open access means that authors eventually are going to publish their accepted, peer-reviewed articles themselves, e.g. by depositing it to their own institutional repositories. Gold open access means that a publisher is paid (e.g. by the authors) to provide immediate access on the publishers website and without charging any further fees to the readers.

**Open Research Data:** Open research data is related to the long-term deposit of underlying or linked research data needed to validate the results presented in publications. Following the idea of open access, all open research data needs to be openly available, usually meaning online availability. In addition, standardized data formats and metadata has to be used to store and structure the data. Open research data is expected to enable others to

- understand and reconstruct scientific conclusions, and
- to build on top of existing research data.

**Metadata:** Metadata defines information about the features of other data. Usually metadata is used to structure larger sets of data in a descriptive way. Typical metadata are names, locations, dates, storage data type, and relations to other data sets. Metadata is very important when it comes to index and search larger data sets for a specific kind of information. Sometimes metadata can be retrieved automatically from a dataset, but often it needs some manual classification also.

## 2 Publishing Infrastructure for Open Access

---

The ARCADIA publication infrastructure consists of a process and several web-based publication platforms that together provide long-term open access to all publishable, generated or collected results in the project. Both the process and the used web-based platforms are described in the following subsections.

### 2.1 Publishing Process

The project partners defined a simple, deterministic process that defines if a result in the project must be published or not. The term result is used for all kind of artefacts collected or generated during the project like white papers, scientific publications, and anonymous usage data. By following the process, each result is either classified public or non-public. Public means that the result must be published under the open access policy. Non-public means that it must not be published. A non-public classification always prevails a public classification. For each result generated or collected during the project runtime, the following questions must be answered to classify it:

1. Does a result provide significant value to others or is it necessary to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public. If this question is answered with no, the result is classified as non-public. For example, code that is very specific to the

ARCADIA platform is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published.

3. Does a result allow the identification of individuals even without the name?

If this question is answered with yes, the result is classified as non-public. The information must be reduced to a level where single individuals can not be identified, usually by using abstraction techniques. Sometimes data inference can be used to superimpose different user data and reveal indirectly a single user's identity. In ARCADIA, such datasets are non-public. ARCADIA will use established anonymization techniques to conceal a single user's identity, e.g. abstraction, dummy users, or non-intersecting features.

4. Does a result include business or trade secrets of one or more partners of the project?

If this question is answered with yes, the result is classified as non-public. Business or trade secrets needs to be removed in accordance to all partners' requirements before it can be published.

5. Does a result name technologies that are part of an ongoing, project-related patent application?

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

6. Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and the project's ethics?

If this question is answered with yes, the result is classified as non-public.

7. Does a result break national security interests for any project partner?

If this question is answered with yes, the result is classified as non-public.

## 2.2 Publishing Platforms

In ARCADIA, we use several platforms to publish our results openly. The following list presents the platforms used during the project and describes their concepts for publishing, storage, and backup.

### 2.2.1 OwnCloud

OwnCloud is a suite of client-server software for creating file hosting services and using them. OwnCloud is functionally very similar to the widely used Dropbox, with the primary functional difference being that OwnCloud is free and open-source, and thereby allowing anyone to install and operate it without charge on a private server, with no limits on storage space (except for disk capacity or account quota) or the number of connected clients.

In order for desktop machines to synchronize files with their OwnCloud server, desktop clients are available for PCs running Windows, OS X, FreeBSD or Linux. Mobile clients exist for iOS and Android devices. Files and other data (such as calendars, contacts or bookmarks) can also be accessed using a web browser without any additional software. Any updates to files are pushed between all computers or mobile devices connected to a user's account.

The OwnCloud platform for ARCADIA is hosted by UBITECH and runs on a server at UBITECH's premises, therefore keeping all data on an own managed server. The OwnCloud platform is securely backed in the UBITECH system infrastructure and holds all project-related data. This includes data



about the ARCADIA members, the ARCADIA projects, deliverables, and publications. Information and data from services or platforms of ARCADIA project partners will also be stored on the OwnCloud platform. The ARCADIA OwnCloud platform will not duplicate any project-related information to external servers, such as issues, requirements, product code, or deployment information. The ARCADIA OwnCloud platform will be available during the project runtime, and will still be available for at least one year after the official project end.

Web link: <https://euprojects.net/owncloud/>

### 2.2.2 ARCADIA Website

The partners in the ARCADIA consortium decided early to setup its own project-related webpage. This webpage describes the mission and the general approach of the project and its development status. A blog informs about news on a regular basis. Later in the project the developed ARCADIA software components and software development paradigm will be announced. A dedicated area for downloads is used to publish reports and white papers. All documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title and the type of the document. The webpage is hosted by partner ADITESS at its own infrastructure. All webpage-related data is backed on a regular basis. All information on the ARCADIA website can be accessed without creating an account. The information is indexed by web search engines like Microsoft Bing or Google Search. The webpage is backed manually once per month.

Web link: <http://www.arcadia-framework.eu/>

### 2.2.3 GitHub

GitHub is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides also some facilities to work on documentation and to track issues. GitHub provides paid and free service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over provided material. For textual metadata items, English is preferred. The service is hosted by GitHub Inc. in the United States. GitHub using a rented Rackspace hardware infrastructure where data is backed continuously to different locations.

All source-code components that are implemented during the project and decided to be public will be uploaded to an open access GitHub repository.

Web link: <https://github.com/>

## 3 Project Results/Datasets

---

In this section, a list of all existing or foreseeable results is presented, separated into public deliverables, publications and open research data. For each result and in accordance to the data management guideline [1], we provide a description, name the standards used for storage and metadata, and define which open access platform is chosen.

## 3.1 Deliverables

### 3.1.1 Description of Highly Distributed Applications and Programmable Infrastructure Requirements

Data set reference and name

Description of Highly Distributed Applications and Programmable Infrastructure Requirements (report D2.1).

Data set description

This document is the first technical deliverable of the project and is committed to depict the general context, to identify relevant actors and technologies, and to set the main group of requirements that will drive the design of the ARCADIA framework. The document is a result of the ARCADIA Task 2.1.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title and the partner organizations and members that classify this report.

Data sharing

The document was published openly on the ARCADIA webpage. The access is free for everyone and without restrictions.

Web link: <http://www.arcadia-framework.eu/wp/documentation/deliverables/>

Archiving and preservation (including storage and backup)

The document was published on the ARCADIA webpage. All earlier versions of the document are archived on the project-internal OwnCloud repository. The repository is backed on a regular basis by UBITECH.

### 3.1.2 Definition of the ARCADIA context model

Data set reference and name

Definition of the ARCADIA context model (report D2.2).

Data set description

This document introduces the first version of the facets included in the ARCADIA context model and elaborate on their usage. A standalone version of the ARCADIA Context Model is presented by providing all appropriate information needed by the end-user prior to the explanation of the modeling artifacts.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title and the partner organizations and members that classify this report.

Data sharing

The document was published openly on the ARCADIA webpage. The access is free for everyone and without restrictions.

Web link: <http://www.arcadia-framework.eu/wp/documentation/deliverables/>

Archiving and preservation (including storage and backup)

The document was published on the ARCADIA webpage. All earlier versions of the document are archived on the project-internal OwnCloud repository. The repository is backed on a regular basis by UBITECH.

It should be noted that all the public deliverables of ARCADIA, upon their finalization, are going to be published and made available in the project's website.

## 3.2 Publications

### 3.2.1 Cluster “Software Engineering for Services and Applications” White Paper

Data set reference and name

ARCADIA white paper 1: Cluster “Software Engineering for Services and Applications” White Paper.

Data set description

This is a white paper that is still under preparation within the cluster “Software Engineering for Services and Applications”. The white paper aims at collecting the challenges that have to be faced towards the design and development of novel software engineering approaches. It is a collaborative work among all the projects that participate to the cluster. The work is coordinated by UBITECH and is going to include the challenges identified on behalf of the ARCADIA project.

Standards and metadata

The document is under preparation. The final version is going to be stored in the cross-platform portable document format (PDF). Metadata is going to be added manually and include the title, the partner organizations, and keywords that classify this research paper.

Data sharing

This research paper will be published within the webpage of the Cluster “Software Engineering for Services and Applications”. It will be freely available worldwide.

Web link: <https://eucloudclusters.wordpress.com/software-engineering-for-services-and-applications/>

Archiving and preservation (including storage and backup)

The research paper will be made available by the the Cluster “Software Engineering for Services and Applications” will therefore be persistently available to the public.

## 3.3 Research Data

With regards to research data that is going to be used in the project, the ARCADIA Consortium does not envisage the need to collect raw data, nor to produce huge amount of data during the Project lifespan. At the time of writing, the Consortium plans to put at disposal of the “Pilot on Open Research Data in Horizon 2020” the input data set that will be used in the final demonstrations of the ARCADIA use cases, with the aim to facilitate the reproducibility of the experiments and confirm the correctness of the obtained results. The final demonstrations has been planned to be organized within the works of WP5, in the dedicated Tasks 5.1, 5.2, 5.3 and 5.4, and will constitute the core of the deliverables D5.1, D5.2 and D5.3. All ARCADIA partners participate in WP5 and will contribute in defining and gathering the demonstration data set that will be available in the use cases.

## Annex I: References

---

- [1] European Commission: "Guidelines on Data Management in Horizon 2020", Version 1.0, December 2013.