

D2.6

Version	1.0
Author	REL
Dissemination	PU
Date	31-12-2019
Status	FINAL



D2.6 Data Management Plan

Project acronym	ElasTest
Project title	ElasTest: an elastic platform for testing complex distributed large software systems
Project duration	01-01-2017 to 31-12-2019
Project type	H2020-ICT-2016-1. Software Technologies
Project reference	731535
Project website	http://elastest.eu/
Work package	WP2 User-centered agile conception
WP leader	TUB
Deliverable nature	Report
Lead editor	Christos Roupas
Planned delivery date	31-12-2019
Actual delivery date	31-12-2019
Keywords	Open data, open access, data management



Funded by the European Union

License

This is a public deliverable that is provided to the community under a **Creative Commons Attribution-ShareAlike 4.0 International** License:

<http://creativecommons.org/licenses/by-sa/4.0/>

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

For a full description of the license legal terms, please refer to:

<http://creativecommons.org/licenses/by-sa/4.0/legalcode>



Contributors

Name	Affiliation
Christos Roupas	REL
Frances Pavlidi	REL
Guiomar Tuñón Hita	NAEVATEC
Guglielmo De Angelis	CNR
Francisco Gortázar	URJC
Varun Gowtham	TUB

Version history

Version	Date	Author(s)	Description of changes
0.1	13/03/2017	Christos Roupas Frances Pavlidis Andreas Afentakis	Initial ToC
0.2	13/05/2017	Christos Roupas	Executive Summary, Introduction, Conclusion
0.3	24/05/2017	Patxi Gortazaz	Revision of the document
0.4	22/06/2017	Christos Roupas Frances Pavlidis Andreas Afentakis	First draft version of Data Management Plan
0.5	25/12/2018	Christos Roupas Frances Pavlidis Andreas Afentakis	Second draft version of Data Management Plan
0.6	18/12/2019	Christos Roupas Frances Pavlidis Andreas Afentakis	Final draft version of Data Management Plan
0.7	26/11/2019	Guiomar Tuñón Hita	Revision of the document
0.8	31/12/2019	Francisco Gortázar	Fix some comments on backups

Table of contents

1	Executive summary	8
2	Introduction	8
2.1	General Data Protection Regulation (GDPR)	8
2.2	Why a Data Management Plan (DMP)?.....	9
2.3	What kind of data are considered in the DMP?	10
2.4	What can be expected from ElasTest DMP?	10
3	Data Management Plan Template (DMP)	11
3.1	FAIR Data Management.....	12
3.1.1	Making data findable, including provisions for metadata	12
3.1.2	Making data openly accessible.....	13
3.1.3	Making data interoperable	14
3.1.4	Increase data re-use (through clarifying licenses)	14
4	Data Set	14
4.1	Description of data	14
4.1.1	What data will we collect?	15
4.2	How will the data be managed, collected or created?	15
4.3	What documentation and metadata will accompany the data?	15
4.4	How will we manage copyright and Intellectual Property Rights (IPR) issues? ...	16
5	Register on numerical data sets generated or collected in ElasTest	17
6	Practical details of already known data sets	18
6.1	Project deliverables	18
6.2	Other internal documentation	20
6.3	Scientific publications	21
6.4	Other publications and outputs:	22
6.5	Contribution to standards	23
6.6	Software and applications	23
6.7	Datasets collected within the Various WP's.....	27
7	Conclusions	29
8	References	30

Table of figures

Figure 1 ElasTest open access to data and publications. 10

List of tables

Table 1: Information of dataset: Project deliverables.....	19
Table 2: Information of dataset: Other internal documentation.....	20
Table 3: Information of dataset: Scientific publications	22
Table 4: Information of dataset: Other publications and outputs.....	23
Table 5: Information of dataset: Project deliverables.....	27
Table 6. Data sets collected within WP2	28
Table 7. Data sets collected within WP3	28
Table 8. Data sets collected within WP5	28
Table 9. Data sets collected within WP6	28
Table 10. Data sets collected within WP7	29

Glossary of acronyms

Acronym	Description
DOI	Digital Object Identifier
DMP	Data Management Plan
EC	European Commission
GRIB	GRIdded Binary format, WHO
H2020	Horizon 2020, EU funding Strategy for 2014 - 2020
pdf	Portable Document Format
NetCDF	NETwork Common Data Format
ppt	Power Point
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language

1 Executive summary

Executive Summary This deliverable presents the final version of the Data Management Plan (DMP) and is an update of deliverable 2.6 Initial data management plan. It presents the data collected in the project and how the project will make this data findable, accessible, interoperable and reusable, in accordance with the concept of FAIR data management. It follows the template for DMP provided by the European Commission in the Guidelines on FAIR Data Management in Horizon 2020, version 3.0, July 2016, and provides technical details on the data collected, as well as purpose for data collection, data utility and where and how data can be accessed and reused.

The Data Management Plan (DMP) of ElasTest gives an overview of available research data, access and the data management and terms of use. The DMP reflects the current state of the discussions, plans and ambitions of the ElasTest partners, and will be updated as work progresses.

As specified in Section 2.2.1.2 on Part B of the DoA of this GA this project requires an open data management plan and this task is in charge of creating it and monitoring its implementation. This plan defines how to share and spread data generated by the project. This plan should not only contemplate the mechanisms for sharing ElasTest generated data, but also for enabling third parties to provide further testing data that might be helpful for training and improving ElasTest Machine Learning and Cognitive Q&A algorithms. Thus, the goal of this task is to investigate the appropriate methodologies, standards and technologies for data management so that the open access of all the data generated by the ElasTest is guaranteed and the ability to integrate new external data is granted. As a result of this task, a Data Management Plan deliverable shall have been generated at M6 and reviewed yearly.

Partner roles: RELATIONAL coordinates this task generating the plan and monitoring its appropriate implementation by the rest of partners on all data gathering and analysis activities.

- What data will be collected / generated in the course of the project?
- What data will be exploited? What data will be shared/made open?
- What standards will be used / how will metadata be generated?
- How will data be curated / preserved including after project completion

2 Introduction

2.1 General Data Protection Regulation (GDPR)

As of May 2018, the General Data Protection Regulation (GDPR) is applicable in all Member States in the European Union, as well as in the countries in the European Economic Area (EEA). GDPR updates and modernises existing laws on data protection to strengthen citizens' fundamental rights and guarantee their privacy in the digital age. The ElasTest project has reviewed the data collected through the project and how this has been processed and stored. Every partner has consulted its DPO (if appointed) and received confirmation that sensitive data handling, as well as its main data archiving

facility are operating in accordance and are in line with GDPR and sound research ethics. In addition, we have contacted all members of the project to get updated permission to store their contact data for involving them in project work and activities. All data collected from stakeholders in the project has been done in accordance with applicable ethical standards and requirements in the respective countries of the data collection, as well processed and handled secure and in line with applicable rules and regulations on privacy and data protection.

2.2 Why a Data Management Plan (DMP)?

It is a well-known phenomenon that the amount of data is increasing while the use and re-use of data to derive new scientific findings is more or less stable. This does not imply, that the data currently unused are useless - they can be of great value in future. The prerequisite for meaningful use, re-use or recombination of data is that they are well documented according to accepted and trusted standards. Those standards form a key pillar of science because they enable the recognition of suitable data.

To ensure this, agreements on standards, quality level and sharing practices have to be negotiated. Strategies have to be fixed to preserve and store the data over a defined period of time in order to ensure their availability and re-usability after the end of ElasTest.

As can be interpreted from article 29.3 in the Grant Agreement, the objectives of open access to data primarily concern two aspects: (1) to have raw-data available for post-validation of research results; and (2) to permit re-use in future research projects. Relatedly, as emphasized by the EC (2013a), open research data can help to accelerate innovation; foster collaboration and avoid duplication of efforts; build on previous research results; and increase the transparency of the scientific process. Open access to research data and to publications should however not represent a risk for compromising the privacy of informants participating in the different ElasTest case studies by openly publish datasets in which persons, households or families are identified. This DMP assesses when and how data can be shared within a sound research ethics framework, where directly or indirectly identifiable information is not disclosed at any stage in the research process. In addition to open access to research data, ElasTest will comply with the requirements for open access to scientific publications. We will return to this in section 4. In the below section 3, we describe the data sets to be gathered and processed in ElasTest, and the procedures followed to ensure open access to these data sets without violating the privacy of informants taking part in the ElasTest case-studies. Figure 1 illustrates the main points for how open access to research data and publications will be ensured in the project.

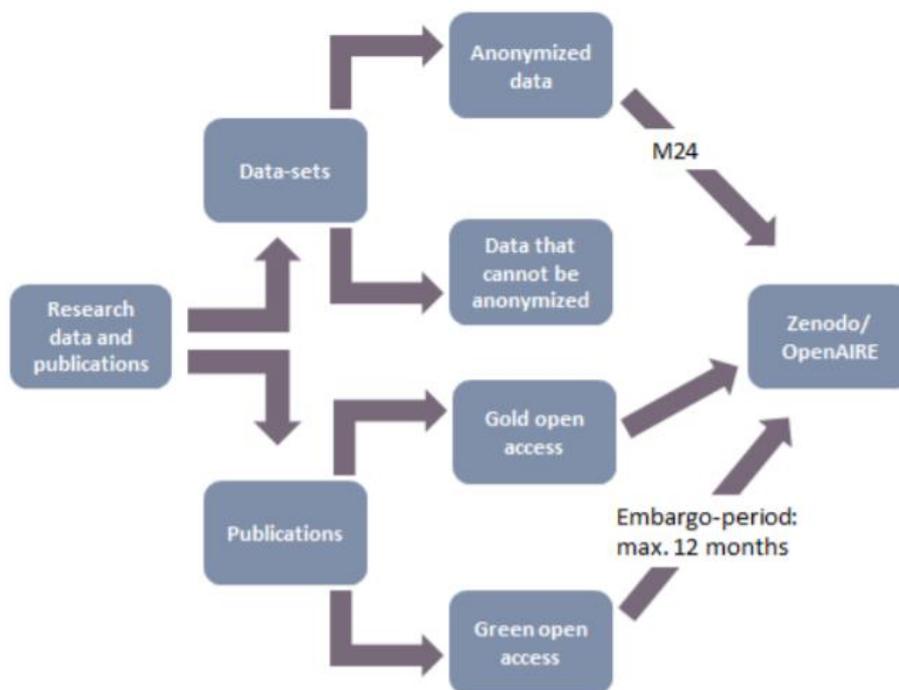


Figure 1 ElasTest open access to data and publications.

2.3 What kind of data are considered in the DMP?

The main purpose of a Data Management Plan (DMP) is to describe Research Data with the metadata attached to make them discoverable, accessible, assessable, usable beyond the original purpose and exchangeable between researchers.

According to the “Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020” (2015):

“Research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.”

2.4 What can be expected from ElasTest DMP?

Thought the document we will describe the lifecycle, responsibilities and review processes and data management policies of research data, produced in ElasTest. The DMP reflects the current status of discussion within the consortium about the data that will be produced. It is not a fixed document, but have evolved during the lifespan of the project.

The target audience of the DMP is all project members and research institutions using the data produced.

In deliverable D2.6, these were the expected data sets to be collected and generated in ElasTest project is presented below:

- Project deliverables.
- Scientific publications: the scientific publications, mainly scientific papers, created by the consortium members, will contain technical results from the ElasTest project
- Other publications and outputs: Besides the scientific publications mentioned above, e.g. in journals or conference proceedings, it is expected that the project will generate further publications and other project outcomes, such as:
 - Promotion material (brochures, flyers, posters, etc.).
 - Social network publications, through the official Twitter account of the project.
 - Press releases and further project announcements.
 - White papers created by the consortium on particular subjects.
 - Any further publication generated by the project.
- **Contribution to standards:** standardization constitutes an important dissemination activity in ElasTest project. It aims to contribute to the activities in major international standardization bodies, as defined in the ElasTest Description of Action.
- **Software and applications:** the ElasTest project developed and tested several applications. In addition to the source code and binaries, documentation of the developed applications, their specifications, and other related material will be available in the project deliverables.
- **Data collected for analysis and evaluation.**

In the next sub-section, we are including the results of the analysis required of each of those data-sets for this time of the project.

3 Data Management Plan Template (DMP)

The structure of the DMP will be according to the specifications set by the “Guidelines on FAIR Data Management in Horizon 2020” of the H2020 Programme.

Including the following sections:

- a. Data Summary
- b. FAIR Data
 - a. Making data findable, including provision for meta data
 - b. Making data openly accessible
 - c. Making data interoperable
 - d. Increase data reuse
- c. Allocation of resources
- d. Data security

- e. Ethical aspects
- f. Other

3.1 FAIR Data Management

3.1.1 Making data findable, including provisions for metadata

Metadata is structured information describing the characteristics of a resource; for example, the dates associated with a dataset or the title and author of a book. Metadata supports discovery, re-use and long-term preservation of resources. Metadata needs vary across scientific fields, but typically cover general descriptive and access metadata, data characteristics, and archive terms and access policies. A metadata record consists of a set of predefined elements that define specific attributes of a resource. Each element can have one or more values; for example, a dataset may have multiple creators. Documenting data enables other researchers to discover the data it-self. Metadata about the nature of your files is also critical to the proper management of digital resources over time. All the partners will agree on specific issues regarding for example:

- The way that the data will be organised or formatted so that everyone working on it now and in the future knows the origins of the data.
- The way that each file will be named (File Naming Conventions). The use of the following format is proposed for each file/document: "xxxx(year)yy(month)zz(day)_ElasTest_DX.X/RPX_V#.#":
 - xxxx(year)yy(month)zz(day): due date of the document (not actual delivery)
 - V#.#: . the number of the version starts with 0.X, until the first version is submitted to the WPL and Coordinator. Once the first draft has been revised, the version will be 1.X and for future updates the first number change in 2, 3 ... n.X.

Example: 20180430_ ElasTest _D8.1_V0.1

- Providing adequate metadata within the dataset (e.g. field labels or column headings) in order to be easy to interpret the data. Other examples of information that the data need to contain include: (i) reference period, (iii) Project funding information: European Union logo and information about Grant Agreement and the action/program that funds the project, (iv) Release policy including dissemination rules and purposes, (v) Information about data collection (source, frequency and adjustments), (vi) Keywords (Keywords or phrases describing the subject or content of the data), (vii) Geographic coverage of the dataset (if applicable), (viii) File formats and (ix) Comments.

Ways to identify different versions. It is proposed in each data set to include a versioning table, additionally to use the prefix ".v1" in each file/document name relevant to the versioning table. For versioning the rule that will be followed will be the use of a

sequentially numbered system: v1, v2, v3, etc. and “Final” for the final version. If changes need to be done in the final version then the name of the document will change including the relevant sequential version number, ensuring that the document with the “Final” prefix is indeed the final one. At a minimum, metadata records should be kept in a fielded form, such as a spreadsheet, CSV file, or tab-delimited file. Auxiliary information necessary to interpret the metadata - such as explanations of codes, abbreviations, or algorithms used - should be included as accompanying documentation.

3.1.2 Making data openly accessible

All the data related to Public Deliverables will be openly available as the default. The data related to IPR protection or to relevant provisions made in the consortium agreement will be eligible to be shared under the defined conditions. According to the type of data and to its confidentiality level, the data will be made accessible on the project’s communication channels or repositories (for raw data). This will be defined during the project once the data is better defined and known.

To ensure the safety of the data, the involved participants will use their available local file servers to periodically create backups of the relevant materials. Additionally, further level of storage and accessibility will be the members section in the ElasTest website (<http://elastest.eu//>).

In this section, the partners will be granted a username and password in order to access the shared platform where confidential documents and other relevant documentation created during the project implementation, such as deliverables, will be self-archived and preserved. The ElasTest Google Drive Folder allows users to store files in the cloud, share files, and edit documents, spreadsheets, and presentations with collaborators. The ElasTest Google Folder includes the following subfolders architecture:

- Folders and Subfolders related to each of the 8 Work Packages.
- Folder and Subfolder related to the communication & Marketing Material
- Folders and Subfolders related to the Deliverables (where all the deliverables will be stored– drafts in the beginning that will be substituted by the finals when submitted)
- Financial (including documents regarding financial issues of the project) - General (including general documents of ElasTest project e.g. Grant Agreement, Consortium Agreement, administrative templates etc.) Meeting (including folders for each meeting containing relevant documents and data)
- Periodic Report (Including data from the periodic reporting) -
- Other (to store and share any other document relevant to the ElasTest project)

All of the research data and material will be in place for at least the 2 years after the end of the project prescribed by the European Commission, as well as the foreseeable future following that according to the agreements reached by the consortium by the end of the project (if any additional is agreed). The Coordinator of the ElasTest project along with the Dissemination & Exploitation Manager will be in charge for data management and all the relevant issues.

3.1.3 Making data interoperable

The data produced in the project are interoperable because the Consortium activities should be based and strictly follow the Data Management Plan and the Project Manual. The guidelines provide the principles and requirements for technical and quality management system documentation in the regulatory context of CE marking the ElasTest multicomponent system, including product qualification and risk classification.

Attention has been paid to make the data and metadata interoperable taking into account all the standards and methodologies that were used. All the curated data produced during the project will be put into widely used formats such as Microsoft Office files, Adobe PDF, png/jpg images and avi videos. The data collected in this project is context specific, and the publicly available part of this data is at the highest level of detail that can be interpreted and understood by external readers. Including more data, as in the form of "raw data", could lead to misinterpretations of the data.

3.1.4 Increase data re-use (through clarifying licenses)

Re-use: All deliverables are available for download and re-use on the ElasTest project website as soon as possible after being submitted to the European Commission. All public/open deliverables include a description in section 1 of the intended readership of each deliverable. This outlines who the deliverable might be useful for outside the project consortium and provides guidance to external readers on whether the content of the deliverable is relevant and interesting for them to re-use. Non-public data from the project will remain available to the consortium partners only after the end of the project.

4 Data Set

4.1 Description of data

The data to be collected during the life time of the Project will fall into one of two general categories:

- Data, including metadata needed to validate the results presented in scientific publications
- Other data, including associated metadata which all participants of the project choose to make available in open access mode in addition to the data underlying publications.

Detailed data categories:

- Specification of software requirements, architectures and systems in different formats. These come from the different partners. They are useful for validation of architectural decisions.
- Specification of tests in different formats. These come from ATOS and NAEVATEC. They are used to train the recommendation engine.

- Machine readable representation of the software artifacts (e.g. code documentation, code comments, etc.) These come from the GitHub repositories. Each repository includes: code, tests and documentation. Information is used to recommend good practices when developing tests.
- Specification of Q&A provided by testers at different stages of the testing process. These come from CNR after anonymization and are used to validate the outcomes for the project.
- Runtime information of systems (e.g. logs, costs, energy, KPIs, etc.) These come from the EDM component. Useful to estimate costs, train machine-learning algorithms for early failure detection on tests, and build oracles based on the outcomes of that algorithms.
- Machine readable representation of testing knowledge (e.g. papers, books, guidelines, etc.)
- Anonymized information regarding tester and end-user satisfaction and subjective feelings on SuT and tests. These come from CNR after anonymization and are used to validate the outcomes for the project.
- Scientific publications.

4.1.1 What data will we collect?

Several types of data will be collected and analysed during the research in the Project. This data consists of end-user information obtained through the interaction with the ElasTest system and through the research with end users during the pilot and in other phases of the Project.

4.2 How will the data be managed, collected or created?

For the Management of data produced from the project, an online sub folder has been created to the project's Google Drive Folder offering specialised tools for managing data will be used. Access to data in the repository has been given depending on the specifications set by the partners.

Other repositories equipped with tools for the management of the accessibility and licensing of the data that could be used include EUDAT and Zenodo by OpenAIRE.

4.3 What documentation and metadata will accompany the data?

The documentation that will accompany the data will be coherent with the Commission Recommendation of 23rd June 2009 number 498. Thus, data will contain information about:

- Contacts: Individual or organizational contact details of persons in charge of data management.
- Metadata update: information indicating the date of last version of data (or

modifications occurred).

- Statistical presentation indicating: basic description of data, sector of coverage, statistical unit for which data have been collected, statistical population, reference area, time coverage.
- Unit of measure.
- Reference period.
- Project funding information: European Union logo and information about Grant Agreement and the action/program that funds the project.
- Institutional mandate including procedures for data sharing and coordination between data producing agencies.
- Confidentiality: property of data indicating the extent to which their unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.
- Release policy including dissemination rules and purposes.
- Information about data collection (source, frequency and adjustments)
- Comments

4.4 How will we manage copyright and Intellectual Property Rights (IPR) issues?

The IPR ownership is defined by the Consortium Agreement and Grant Agreement related to the Project. Such access will be provided by accepting the terms and conditions of use, as appropriate. Materials generated under the Project will be disseminated in accordance with Consortium Agreement. Those that use the data (as opposed to any resulting manuscripts) shall cite it as follows: The data created by the ElasTest project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 643566. For reuse of this data, please, contact ElasTest Consortium.

How will the data be stored and backed up during the research?

The data will be managed by collaborators of participants as well as other scientists interested in ElasTest relationships. The knowledge generated by the Project among partners, scientific community, target users and public at large during the Project are managed in two ways, depending on the data source:

1. The non-sensitive data will be organized into a repository that will contain all the knowledge produced by the Project partners. A restricted access is expected for the knowledge that will be used for exploitation purposes; open access for all the other knowledge. Specific attention must be paid to the creation of an open access to the data collected during the trials. To this end, raw data is organized in an exportable format to be used by the scientific community and practitioners for their own purposes. A registered access for data download will be the only request for their use, in order to understand which organization is interested in using them and for

which particular scope.

2. To manage and store the sensitive data obtained, all partners from ElasTest must comply with relevant European and national regulations as well as with the standards of practice defined by relevant professional boards and institutions.

The EU legislation standards for human studies include the following regulations as for the protection of individuals:

- The Declaration of Helsinki in its latest version (recommendation for conduct of clinical research).
- Directive 95/46/EC on the protection of individuals with regard to processing of personal data and on the free movement of such data Council Directive 83/570/EEC amending Directives 65/65/EEC, 75/318/EEC and 75/319/EEC dealing with proprietary medicinal products.
- Directive 95/46/EC (amendment 2003) of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the protection of privacy, storing of personal data and on the free movement of such data.
- The charter of fundamental rights of the EU (2000/C 364/01).
- EU Good Clinical Practice Directive (2001/20/EC)

5 Register on numerical data sets generated or collected in ElasTest

The register has to be understood as living document, which will be updated regularly during project lifetime. The intention of the DMP is to describe numerical model or observation datasets collected or created by ElasTest during the runtime of the project.

The information listed below reflects the conception and design of the individual work packages at the beginning of the project. Because the operational phase of the project started in January 2017, there is no dataset generated or collected until delivery date of this DMP.

The data register will deliver information according to Annex 1 of the Horizon 2020 guidelines (2015) (in *italics*):

- **Data set reference and name:** Identifier for the data set to be produced.
- **Data set description:** Descriptions of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
- **Standards and metadata:** Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

- **Data sharing:** Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).
 - In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, and commercial, privacy-related, security-related).
- **Archiving and preservation (including storage and backup):** Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

6 Practical details of already known data sets

This section aims at covering the committed statement: *During the project lifetime, additional information on the following aspects will be elaborated for all data sets on case by case basis before making consortium decision on handling of the particular data generated or collected:*

- Nature and scale of the data in consideration.
- To whom it could be useful, targeted audience and its size and level of interest.
- Information on the existence of similar data and possible synergies.
- Possibility for integration and reuse of the provided data by external users and researchers.
- Any further related issue.

These explanations aim at covering the Annex 1 of DMP template by the European Commission.

The internal management of the information and the datasets is being handled using the Project Management tools deployed for the project, however because of the participation in ORDP, this procedure is ready to be accommodated to other different repositories.

With that aim, we have created a template table in which we can embed the associated information in a handy and useful format.

6.1 Project deliverables

Reference/name	Deliverables
----------------	--------------

Data sharing method	Public website (section deliverables)		
Description	Public deliverables (according to GA) that will be generated during the execution of the project.		
Type	Written documentation.		
To whom it could be useful	Researchers, scientific community on the field		
Targeted audience	Size	5-10 for official review by EC.	
	Desc.	European Commission and technical reviewers	
Related WPs and tasks	All WPs		
Format, standards	<i>DX.Y – [Name]</i>		
Software	Generated after the writing of ElasTest partners. Software used can be text editors: MS Word, Google Drive or others.		
Possibility of integration and reuse of this data by external users and researchers	Deliverables are not prone to integration but the information contained in them can be used for further researches (with the proper referencing and IPR respect).	Possible synergies with similar data	N/A
Estimated size	At this moment (M36), the total size of public deliverables (aggregated) is: 96Mb. It is foreseen about 200-250 Mb.		
Storage	In our secure repository Google Drive.	Back up	A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included	Names of authors.	Means for personal data protection	Dealt with in CA.

Table 1: Information of dataset: Project deliverables

6.2 Other internal documentation

Reference/name		Other_internal_documentation	
Data sharing method		Not to be shared publicly.	
Description		Meetings minutes, agendas, internal documents of work, reporting docs...	
Type		Written documentation.	
To whom it could be useful		ElasTest partners.	
Targeted audience	Size	30-50	
	Desc.	People involved in the execution of the project.	
Related WPs and tasks		All WPs	
Format, standards		<i>PPD_UC[No. use case]_Entity</i> and the rest of naming procedures detailed in deliverable D2.2	
Software		Generated after the writing of ElasTest partners. Software used can be text editors: MS Word, Google Drive or others.	
Possibility of integration and reuse of this data by external users and researchers		N/A	Possible synergies with similar data N/A
Storage		In our secure repository Google Drive.	Back up A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included		Yes, names and emails of ElasTest partners people.	Means for personal data protection Dealt with in CA.

Table 2: Information of dataset: Other internal documentation

6.3 Scientific publications

Reference/name		Scientific_publications	
Data sharing method		Editorials/magazines/conferences proceedings, (URJC) Website Website (space “publications”), Researchgate, etc.	
Description		Scientific articles published at Scientific Journals, books or Conferences coming directly from technical work in ElasTest and authored by ElasTest participants	
Type		Written documentation.	
To whom it could be useful		Scientific Community	
Targeted audience	Size	-	
	Desc.	Followers in ResearchGate, current active researchers in the field of ICT and IoT	
Related WPs and tasks		WP2, WP4, WP5, WP6, WP7 and WP8	
Format, standards		Documents to be stored and shared in pdf. In the case of images, the default format should be svg if possible or png or jpg otherwise. Formats are depending on the journal/conference that they aim at being published. Namely, Springer LNCS or others.	
Software		Generated after the writing of ElasTest partners. Software used can be text editors: MS Word, Google Drive or others.	
Possibility of integration and reuse of this data by external users and researchers		Possible synergies with similar data	Publications can be integrated in new papers and scientific articles with the common tools of automated/classic referencing. Furthermore, the information contained in them can be leveraged for SotA studies for further researches (with the proper referencing
			Inner references in our scientific articles can be used for cascade literature research.

Estimated size	Up to 6, 8, 10, 12 (or other) limit of pages set by the editorial. Average size ~4Mb per scientific article.		
Storage	When open access is granted, a copy of the article is deposited in Zenodo.	Back up	Zeonodo's internal backup system. This open access repository guarantees access to publications even in the improbable event that the repository is closed.
Personal data included	Names, emails and affiliations of authors and co-authors.	Means for personal data protection	Informed consents and similar procedures established by the editorial prior to publication.

Table 3: Information of dataset: Scientific publications

6.4 Other publications and outputs:

Reference/name	Other_material	
Data sharing method	Public website, Social media channels, Email to newsletter members.	
Description	Poster, leaflet, supporting material for communication, newsletter.	
Type	Images, presentations, designs, full-size materials.	
To whom it could be useful	Followers, casual readers, ElasTest partners, event promoters, audience of certain events/fairs	
Targeted audience	Size	200/300 per year in conferences 8.000 in total through website 500 in total in social media
	Desc.	Attendees of events, visitors of our website, followers in social networks
Related WPs and tasks	WP8	

Format, standards	In the case of images, the default format should be svg if possible or png or jpg otherwise. The different files are expected to be below the 10 MB mark		
Software	These materials are produced by ElasTest partners using different software for edition: MS Visio, draw.io, GIMP, Photoshop and others.		
Possibility of integration and reuse of this data by external users and researchers	<p>This data is not prone to be integrated or reused by external users.</p> <p>It can be re-shared through social media by our followers.</p> <p>It can be used by event promoters to be included in publicity, promotion or event summaries</p>	Possible synergies with similar data	N/A
Estimated size	Average size of each material is ~6 Mb. Total size expected is 120-150 Mb.		
Storage	In our secure repository Google Drive	Back up	A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included	No	Means for personal data protection	N/A

Table 4: Information of dataset: Other publications and outputs

6.5 Contribution to standards

No contribution to standards have been initiated yet. Contacts with relevant entities and research of the current solutions and standardisation have been conducted but with no immediate action foreseen. If further actions happen, this information will be completed and detailed to make this activity compliant with the Data Management Plan.

6.6 Software and applications

Reference/name	Software_components
-----------------------	----------------------------

Data sharing method		<p>Website (private access)</p> <ul style="list-style-type: none"> • Nexus repository manager for private artifacts • AWS ECR for private docker images <p>The nexus repository is used as a private software Artifact repository where those components that shouldn't be published outside the consortium should be published.</p> <p>Activate the access to Nexus will create a user inside the nexus with the same username as GitHub, and a secure password will be sent to the provided email.</p> <p>Once the user has an active credential for the nexus the options to refresh and delete the credentials will be enabled.</p> <p>Public Repository:</p> <ul style="list-style-type: none"> • All the public artifacts (snapshots and releases) are hosted on Sonatype OSSRH (OSS Repository Hosting). • Releases are published to Maven Central from there. • The artifacts are available for everybody. • Publishing is only available from the CI server. • Maven settings will be provided in the following weeks. • GitHub • Zenodo (through GitHub) <p>Private Repository</p> <ul style="list-style-type: none"> • We host a Nexus Repository (OSS v3) in our own AWS EC2. • Publishing is only available from the CI server. • Nominative users will be created to access the repository. <p>Artifacts are available for all registered users from the consortium.</p>
Description		Software modules developed in the project and that can be made public.
Type		Software code with associated methodology
To whom it could be useful		EC, developers, open source community
Targeted audience	Size	-
	Desc.	EC, developers, open source community
Related WPs and tasks		WP2, WP3, WP4, WP5, WP6 and WP7

<p>Format, standards</p>	<p>Containerisation practices will be followed when possible. License Apache 2.0 will apply.</p>		
<p>Software</p>	<p>The software modules are being generated after ElasTest partners’ coding activities. The IDEs for development, frameworks and languages are diverse, such as Python, Java, Vue.js, IDE Eclipse, IDE VS Code, among others.</p> <p>The initial environment devised for running CI/CV tasks for the project started with a single instance that ran the main tools related to the software development process and from the first release of the ElasTest platform it has grown having now four instances, each one with a clear objective regarding the tasks that are meant to be executed over them:</p> <ol style="list-style-type: none"> I. Main Instance. Holds the main tools related to the software development process – CI server, repositories, credential generator, etc. – II. ElasTest Nightly Instance. Hosts the latest developed version of ElasTest (not necessarily stable). This instance main objective is to provide an ElasTest platform where latest changes on the code could be tested. III. ElasTest Stable Instance. Host the latest stable version of ElasTest. This instance will be used to test the ElasTest Nightly Instance with ElasTest, as specified in the DoA. IV. ElasTest Nightly K8s Cluster. This “instance” is a Kubernetes’ cluster with two nodes (one master, one slave with the objective of deploying a nightly version of ElasTest so it can be tested nightly and compare executions between Nightly and K8s Nightly in order to grant that both deployments are working, and ElasTest platform and each component works as expected. <p>The ElasTest Monitoring Platform (EMP has been integrated with the ElasTest GUI and the e2e-test is the principal mode to test the correctness of the UI integration. In the ElasTest CI server, Components e2e -> elastest-platform-monitoring -> emp-e2e-test performs this integration test. It uses JUnit together with Selenium to invoke various elements of the GUI to assess availability of known web-elements to assert whether the EMP + ElasTest GUI integration still remains valid at that point in time or not.</p>		
<p>Possibility of integration and reuse of this data</p>	<p>All components have been created in a modular way including</p>	<p>Possible synergies</p>	<p>Some synergies have been identified:</p>

<p>by external users and researchers</p>	<p>an interface for integration (API or others).</p> <p>All developments will be made public with associated markdown-type documentation (e.g. Readthedocs)</p>	<p>with similar data</p>	<p>1) Federation and SSO: if the target port includes some delegation support (e.g. SAML), there is no need to even store user profiles in the ElasTest platform.</p> <p>2) Storage: if the target port is also using Elasticsearch as database, both can be probably merged to reduce infrastructure costs</p>
<p>Estimated size</p>	<p>Average size of each material is ~6 Mb. Total size expected is 120-150 Mb.</p>		
<p>Storage</p>	<p>The Development platform is designed with the objective of providing the project with a complete set of tools and procedures that must grant the appropriate level of quality of each component and the right integration of all of them. The CI methodology comprises all the tasks that assure:</p> <ul style="list-style-type: none"> • High quality of each of the components from development to release. • High quality of integration between components. • High quality of ElasTest as a whole. • High quality of the CI methodology and CI 	<p>Back up</p>	<p>The use of Docker instances allows us to just focus on the data for back- up purposes.</p> <p>The IH allows to create clusters redundancy for scaling and safety purposes.</p> <p>It is also possible to dump/export data from Elasticsearch or other databases (E.g. Mongo)</p>

	environment., The specific configuration of the consortium and the diverse licenses (public/Apache 2.2 and Proprietary) of the components are managed within the CI tasks and tools to grant the appropriate access and dissemination of each component		
Personal data included	No	Means for personal data protection	N/A

Table 5: Information of dataset: Project deliverables

6.7 Datasets collected within the Various WP's

- Datasets collected within WP2**

Type of data	Format	Size of data	Public/Non-public
Text files: Recommender systems applied to software testing – results of a systematic survey Deliverable 2.2	PDF/A	15 pages	Public
Excel spreadsheets: <ul style="list-style-type: none"> ElasTest Requirements and end-to-end tests 	.xlsx	1 spreadsheet with 6 tabs	Non Public
Questionnaires <ul style="list-style-type: none"> Integration and system testing survey ElasTest user experience survey 	On Line Questionnaires		Non Public

- End-to-end testing survey

Table 6. Data sets collected within WP2

- **Datasets collected within WP3**

Type of data	Format	Size of data	Public/Non-public
Audio recordings	MP3, MP4	<ul style="list-style-type: none"> • 33MB • 15MB • 37MB 	Non-public
<ul style="list-style-type: none"> • EMP • ESM • EIM 			

Table 7. Data sets collected within WP3

- **Datasets collected within WP5**

Type of data	Format	Size of data	Public/Non-public
<ul style="list-style-type: none"> • WP5 Quatic Paper 	PDF/A	•	Public

Table 8. Data sets collected within WP5

- **Datasets collected within WP6**

Type of data	Format	Size of data	Public/Non-public
<ul style="list-style-type: none"> • Integrated E2E tests 	On Questionnaires	Line	Non-public
Excel spreadsheets:	.xlsx	1 spreadsheet with 28 tabs	Non-public
<ul style="list-style-type: none"> • ElasTest Requirements and end-to-end tests 			

Table 9. Data sets collected within WP6

- **Datasets collected within WP7**

Type of data	Format	Size of data	Public/Non-public
Excel spreadsheets:	.xlsx	1 spreadsheet with 9 tabs	Non-public
<ul style="list-style-type: none"> • ElasTest QE Data Collection v2 			

Table 10. Data sets collected within WP7

7 Conclusions

The objective of the document is to report the steps for data management which are to be followed during the execution of ElasTest project. The scope of the document is twofold: to define the detailed data management plan towards the dissemination of project outcomes and to report the detailed ethics management plan; specifying the ethics handling and preservation activities. The current document gives a preliminary information about the data types used and generated by the project consortium partners including focus on the means of sharing data captured by ElasTest framework and further specifies the methods of data storage thus providing general view over the complete data management life cycle.

8 References

- [1] ElasTest project Description of Action (DoA) – part B. Amendment 1. Reference Ares (2017)343382. 23 January 2017.
- [2] Guidelines on Data Management in Horizon 2020, Version 2.0, 30 October 2015: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [3] Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020, Version 2.0, 30 October 2015 :

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- [4] DMP Template: <https://dmponline.dcc.ac.uk/>
- [5] Open access & Data management: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm
- [6] Data Management. http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- [7] Guidelines: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [8] OpenAire: <https://www.openaire.eu/>
- [9] EUDAT: <https://eudat.eu/data-preservation>