



Strategic partnership for the co-design of an innovative and scalable eportfolio ecosystem to improve the quality and visibility of skills

D1.4 – Data Management Plan

Jean-Baptiste Milon¹, Andrea Contino², Mirjam Hillenius²,

¹ICDE, International Council for Open and Distance Education (Norway), ²MYD - MyDocumenta (Spain)

| | |
|---|--------------------------------|
| Dissemination Level | PU - Public |
| Due Date of Deliverable | (30/06/2018) |
| Actual Submission Date | (30/06/2018) |
| Work Package | WP1 - Project management |
| Type | ORDP: Open Research Data Pilot |
| Abstract: | |
| <p>This document is the first of four that will be periodically delivered at month 6, 12 18 and 36. Whereas future versions will comprises guidelines, templates and process descriptions, the following versions will focus on monitoring and reporting and the usage of data.</p> | |
| <p>The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.</p> | |

Project funded by the EU Horizon 2020 Research and Innovation Programme under the agreement No. 780435



History

| Version | Date | Reason | Revised by |
|---------|------------|------------------------------------|---------------------|
| 0.1 | 15/06/2018 | First Draft by the project manager | MYD |
| 0.2 | 22/06/2018 | 2nd draft from the project manager | MYD |
| 0.3 | 28/06/2018 | 3rd Draft form the project Manager | MYD, ICWE, UOC, AVU |
| | 30/06/2018 | Submission | |

EXECUTIVE SUMMARY

This living document presents the Data Management Plan (DMP) describing how the data generated within the project are processed and preserved during the EPICA lifetime. As the project will handle and process sensitive data from different sources, the need for appropriate data handling is becoming even more important.

The research data management procedures will be further refined in the future releases of the deliverable as we will have more concrete information about EPICA research findings and our plans will be more specific.

This DMP is based on the EC guidelines[1] and the principles of making the research data Findable, Accessible, Interoperable and Re-usable (FAIR). This document will be updated during the project lifetime as the data to be generated, stored and managed will be further specified, new data arise and in case of changes in consortium policies.

The DMP is highly related and aligned to deliverable D1.2 - "Ethics, Data Protection and Privacy Management Plan" especially regarding the procedures and policies of collecting and protecting personal and sensitive data.

This document is based on the terms and conditions established in the Grant Agreement and its Annexes, as well as in the Consortium Agreement specifications and requirements.

The use of the present guidelines can ensure better collaboration among the consortium members, individuals and groups. They can also facilitate the entire consortium to get engaged in the work that is produced by the project.

The Data Management Plan is a deliverable which is intended to be used by all the project partners, to ensure quality assurance of project processes and outputs and prevent possible deviations from the project work plan as described in Description of Action, GA Annex 1.-

Abbreviations and Acronyms

| | |
|-------------|--|
| DMP | Data Management Plan |
| FAIR | findable, accessible, interoperable and reusable |
| DOA | Description of Action |
| ISO | International Organization for Standardization |
| IEC | International Electrotechnical Commission |
| AES | Advanced Encryption Standard |
| | |

TABLE OF CONTENTS

| | |
|--------------------------------|----------|
| Executive Summary | 2 |
| 1. Data Set Description | 5 |
| 2. FAIR Data | 5 |
| 4. Conclusion | 7 |
| 5. References | 7 |

1. DATA SET DESCRIPTION

The purpose of the collected and generated data is to use them for carrying out the tasks as described in DOA, as well as to be able to validate and assess the EPICA eportfolio and component. Thus the data is crucial for the project.

The data can be divided into personal and anonymised data. At this stage of the project, it is not yet known which formats the data will have, but it will at least contain .docx, .pdf, .jpeg and geopositional data. Furthermore, the participants will have to fill out surveys and the answers will be another data source for the researcher within the project.

Data from the pilots are being reused in order to optimize the final version accordingly.

The data will be generated and/or provided by the participants, consortium and from the teachers.

The researcher, the consortium and teacher will find the collected data useful.

2. FAIR DATA

Due to the early stage of the project and the insecurity about the data formats, its quality and quantity, it is not possible to plan how to make the data findable, accessible, interoperable and reusable (FAIR). The consortium is striving to make the project data FAIR as described in the guidelines, provided by the European Commission and their directorate-general for Research & Innovation, for H2020 programs but an exact plan will not be made before a later stage of the data management plan. This is due in month 18 and 36 of the project.

3. DATA SECURITY

Considering that EPICA deals with sensitive data, to ensure data confidentiality updates on standards concerning encryption systems are analysed (ISO/IEC 18033-2:2006 Information technology -- Security techniques -- Encryption algorithms -- Part 2: Asymmetric ciphers). Both data at rest and data in flight will be encrypted using the AES symmetric cipher while secure key exchanges will be performed using an asymmetrical cipher of at least 2048 bits in length.

EPICA has an explicit plan on how the consortium will handle data generated or provided by the participants (Fig. 1). The figure also illustrates how different data types are handled and stored with a special emphasis on the different approaches between anonymised and pseudo-anonymised data samples. These differences are especially important due to the different ethical problems these respective data types uphold, as described in the Chapter 2 of the deliverable 1.2 Ethics, Data Protection and Privacy Management Plan.

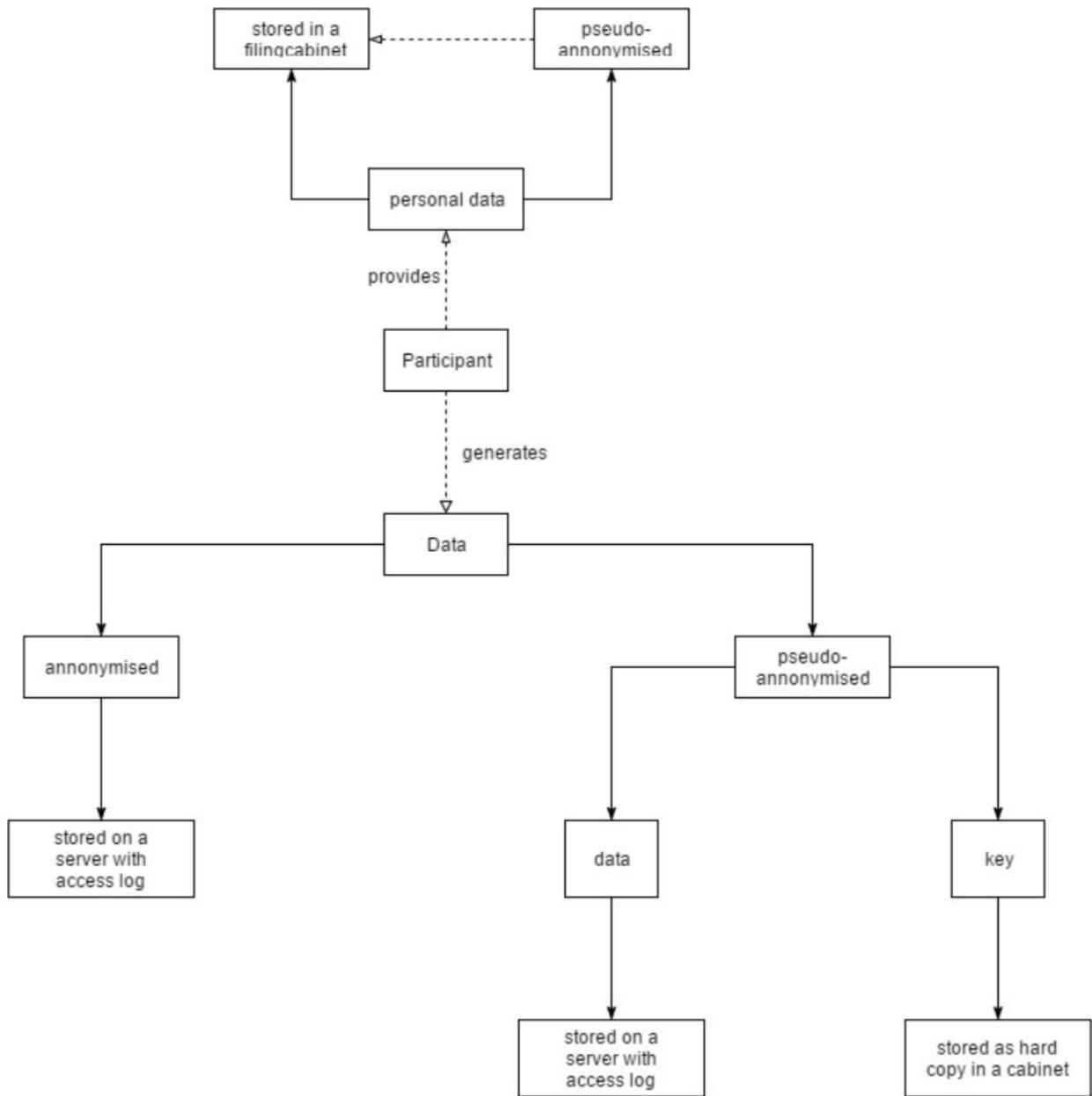


Figure 1 : Data Storage Plan

4. CONCLUSION

This Data Management Plan provides a framework on how the data will be handled in EPICA. Moreover it describes the expected sources, from where the data will be stemmed from and the principles of making them FAIR. This framework is dynamic and expected to be enriched within the project lifetime. As the project is in its first 6-months of its life no estimations can be made yet regarding the volume of data that will finally be preserved in the developed eportfolio as well as the costs required for their preservations.

The work performed in the next months is expected to generate all the required details in order to provide more accurate information. The document will keep being updated along with the EPICA progress. Two more versions of D1.4 are going to be created based on the available results during its corresponding period. These versions will be delivered:

1. Second version: M12
2. Third version: M18
3. Fourth version: M36

5. REFERENCES

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf