HISTO-MRI Project

This project has received funding from the European Union's H2020 Programme under grant agreement no 737180— HISTO-MRI

D1.4: Data Management Plan

Version	Date	Reviewer
1.0	30th June 2017	José M. Benlloch

Table of contents

1 Deliverable description

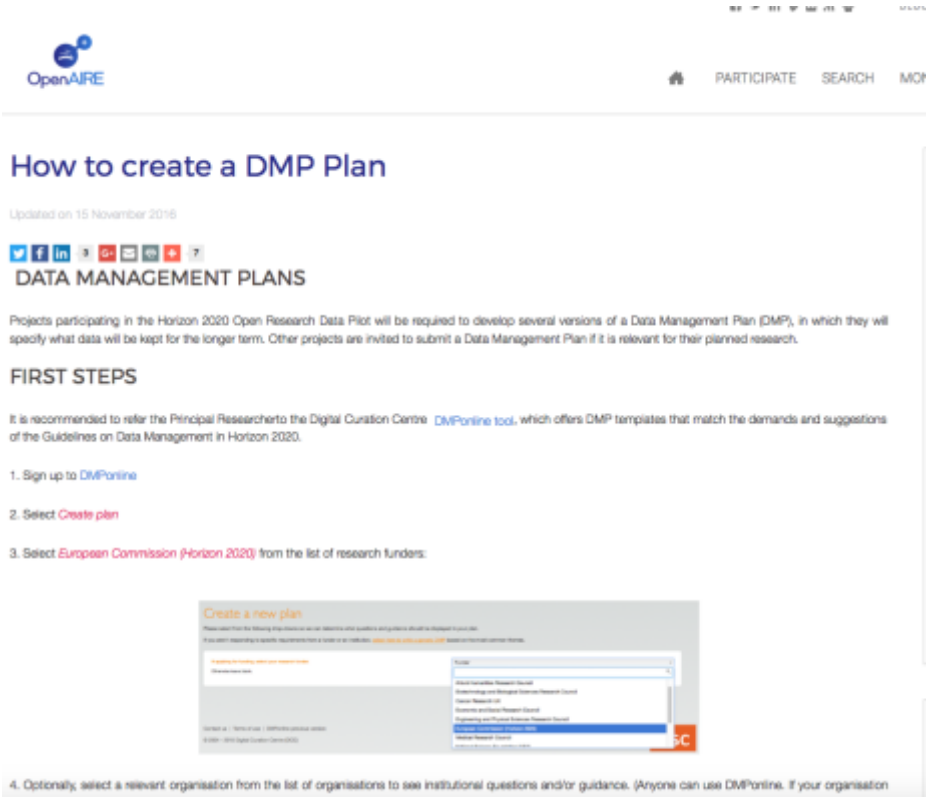
This deliverable is the first version of the Data Management Plan, that describes the strategy to manage all the data produced in the project. This first version provides an overview of the types of data that will be generated in the project and the means to store and preserve it. The next version will provide a more detailed description of the DMP, covering all its aspects.

This deliverable is the first result of *Task 5 Dissemination and Exploitation*, part of **WP1 – Management**.

The consortium will follow the Guidelines on FAIR Data Management in Horizon 2020. According to this document, Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

The consortium will make use of the online tool provided by OpenAire to create a DMP.



2 Introduction

The HISTO-MRI project participates in the Pilot on Open Research Data launched by the European Commission along with the H2020 programme. The use of a Data Management Plan is required for all participating projects.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data.

The DMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. It must be consistent with exploitation and IPR requirements.

Research data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is a H2020 obligation. The rest of research data will be deposited in an open access repository.

3 Data formats

The project will generate a big amount of information, mainly in the form of reports describing the specifications and performance of the sub-systems to be designed and developed in the scope of this project.

The formats to be used for these documents will be:

- Rich Text Format (.rtf): The Rich Text Format is a proprietary document file format with published specification developed by Microsoft Corporation for cross-platform document interchange with Microsoft products. Most word processors are able to read and write some versions of RTF. There are several different revisions of RTF specification and portability of files will depend on what version of RTF is being used. RTF specifications were changed and published with major Microsoft Word and Office versions.
- PDF (.pdf): The Portable Document Format is a file format used to present documents in a manner independent of application software, hardware, and operating systems. Each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, graphics, and other information needed to display
- OpenDocument Text (.odt): The OpenDocument Text Document Format is a format for editable textual documents. It is one of several subtypes in the ODF family for particular content categories. Designed to be a native format for word-processing applications, the format is sometimes called ODT after its usual file extension.
- MS Word (.doc/.docx): This is the text document format developed by Microsoft and used by the MS Word application, included in the Office suite. It is widely used and has become a de facto standard for text documents.

4 Data storage

Using the corresponding software tools and will be stored in an online repository protected with security mechanisms.

The consortium will provide “gold” open access to part of the information (report of the proof of concept results, including the outcomes of the preclinical validation), using its own repository, through the project web site, undert the PUBLIC RESULTS section and through public repositories like Zenodo.

Zenodo is a research data repository for the preservation and making available of research, educational and informational content. Access to Zenodo’s content is open to all, for non-military purposes only. It was created by OpenAIRE and CERN to provide a place for researchers to deposit datasets.

Content may be uploaded free of charge by those without ready access to an organized data centre.

The public documentation will be stored in the consortium own repository for five years after the end of the project.

The documents deposited in Zenodo will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for the at least next 20 years. Data files and metadata are backed up on a nightly basis, as well as replicated in multiple copies in the online system. All data files are stored along with a MD5 checksum of the file content. Regular checks of files against their checksums are made.

The rest of the information cannot be accessible, as the consortium (specially the industrial partners) will seek for its protection and commercialization.