



Artificial intelligence Supporting CAncer Patients across Europe

Project Title	Artificial intelligence Supporting CAncer Patients across Europe
Project Acronym	ASCAPE
Grant Agreement No	875351
Instrument	Research and Innovation action
Call / Topic	H2020-SC1-DTH-2019 / Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment
Start Date of Project	01/01/2020
Duration of Project	36 months

D5.8 – Open Research Data Pilot Contribution

Work Package	WP5 – Dissemination, Exploitation and Legal monitoring
Deliverable Leader (Org)	ALBV
Contributor(s) (Org)	UBITECH, ALBV
Due Date	30.06.2020
Actual Date of Submission	30.06.2020
Version	1.0

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public (*on-line platform)
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)



The work described in this document has been conducted within the project ASCAPE. This project has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme under the Grant Agreement no 875351. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Versioning and contribution history

Version	Date	Author	Notes
0.1	01/05/2020	Dimitra Stefanatou (ALBV)	ToC
0.2	12/05/2020	Christina Stratigaki (UBITECH)	Section 4
0.3	15/05/2020	Dimitra Stefanatou (ALBV)	Contribution under section 1.
0.4	30/05/2020	Dimitra Stefanatou (ALBV), Christina Stratigaki (UBITECH)	Integration of partners' input under Section 3 and initial input under sections 2, 3, 5, 6 and section 7.
0.5	01/06/2020	Dimitra Stefanatou (ALBV)	First draft complete version
0.6	18/06/2020	Dimitra Stefanatou (ALBV)	Updates under section 3 and input under section 8.
0.7	19/06/2020	Christina Stratigaki (UBITECH)	Formatting changes across the document.
0.8	19/06/2020	Dimitra Stefanatou (ALBV)	Submission for internal review.
0.9	29/06/2020	Dimitra Stefanatou (ALBV), Christina Stratigaki (UBITECH)	Integration of reviewers' comments.
1.0	30/06/2020	Dimitra Stefanatou (ALBV), Christina Stratigaki (UBITECH)	Final editing and release of Final Version

Disclaimer

This document contains material and information that is proprietary and confidential to the ASCAPE Consortium and may not be copied, reproduced or modified in whole or in part for any purpose without the prior written consent of the ASCAPE Consortium

Despite the material and information contained in this document is considered to be precise and accurate, neither the Project Coordinator, nor any partner of the ASCAPE Consortium nor any individual acting on behalf of any of the partners of the ASCAPE Consortium make any warranty or representation whatsoever, express or implied, with respect to the use of the material, information, method or process disclosed in this document, including merchantability and fitness for a particular purpose or that such use does not infringe or interfere with privately owned rights.

In addition, neither the Project Coordinator, nor any partner of the ASCAPE Consortium nor any individual acting on behalf of any of the partners of the ASCAPE Consortium shall be liable for any direct, indirect or consequential loss, damage, claim or expense arising out of or in connection with any information, material, advice, inaccuracy or omission contained in this document.

Table of Contents

Executive Summary	5
1 Introduction	6
1.1 Purpose of the document.....	6
1.2 Target Audience	7
2 Data Management in Horizon 2020.....	8
3 ASCAPE Data Management Overview	10
3.1 Types and formats of artefacts generated/collected	10
3.2 ASCAPE Artefacts and Access Rights	10
3.3 Data Protection.....	12
3.4 Existing data re-used (if any).....	13
3.5 Expected size of the data (if known)	13
4 ASCAPE ORDP Participation	14
4.1 Publishing Infrastructure for Open Access	14
4.1.1 Publishing Process.....	15
4.1.2 Publishing Platforms.....	16
4.1.3 Access and Sharing	17
5 FAIR data.....	19
5.1 Making data findable, including provisions for metadata.....	19
5.1.1 Discoverability of the data	20
5.1.2 Data identification mechanisms	20
5.1.3 Naming conventions used.....	20
5.1.4 Approach towards search keywords	21
5.1.5 Clear versioning of the documents.....	21
5.1.6 Standards for metadata creation (if any).....	22
5.2 Making data openly accessible.....	22
5.2.1 Methods or software needed to access the data	24
5.2.2 Deposit of data, associated metadata, documentation and code	24
5.3 Making data interoperable.....	24
5.3.1 Interoperability of data assessment	24
5.3.2 Vocabulary use	25
5.4 Making data re-usable	25
5.4.1 Increase data re-use through clarifying licenses.....	25
5.4.2 Data quality assurance process	25
5.4.3 Length of time for which the data will remain re-usable	26
5.5 Artefact Template.....	26
6 Allocation of resources	28
6.1 Data management responsibilities.....	28
6.2 Cost of potential value of long-term preservation	28
7 Data Security	29

8	<i>Ethical Aspects</i>	30
9	<i>Concluding remarks and next steps</i>	31

List of tables

Table 2-1- Clarifications of terms	8
Table 3-1 – Artefacts overview	10
Table 3-2 Research Item Accessibility	11
Table 3-3 Software Accessibility	11
Table 3-4 Pilots’ Dataset Description and Access Rights	12
Table 5-1 – Proposed Document History Table overview	21
Table 5-2 – Artefacts’ Types	24
Table 5-3 Making Data Findable Template	26
Table 5-4 Making data Accessible Template	26
Table 5-5 Making data Interoperable Template	27
Table 5-6 Making data Re-usable Template	27

List of figures

Figure 1 ASCAPE Artefacts’ Type.....	10
Figure 2 Artefacts Access Rights	14
Figure 3 Template to be used for project documentation metadata overview.....	22
Figure 4 Open access to scientific publication and research data in the wider context of dissemination and exploitation	23

Executive Summary

ASCAPE is a H2020 research and innovation action European project and aims at take advantage of the recent ICT advances in Big Data, Artificial Intelligence and Machine Learning to support cancer patients' quality of life and health status. To achieve its objective, ASCAPE will create an open AI infrastructure that will enable health stakeholders (hospitals, research institutions, companies, etc.) to deploy and execute its AI algorithms locally on their private data.

Based on European Commission's Guidelines for Horizon 2020, to submit a DMP (*Data Management Plan* as a deliverable) within the first six (6) months of the project, the present report forms the Data Management Plan (DMP) of ASCAPE project. Any occurring further updates in this respect will be provided in the course of the project duration, on the basis of the actual developments of the technical work. More specifically, in line with the F.A.I.R Principles, the deliverable provides for how making the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. The DMP is, thus, a key document presenting the data management practices employed by consortium partners, describing -among other- standards and methodology for data collection and generation will be followed, and whether and how data will be shared. To this end, partners were requested to provide input with respect to the artefacts that they all will produce/provide in the ASCAPE project and the level of dissemination (Public/Private).

It is intended that the present DMP is updated over the course of the project, whenever significant changes arise either with respect to the data management practices of individual partners or at project level. Note that a detailed description of the data management practices concerning -in particular- processing of personal data in the context of the piloting activities is provided under "[D7.1 – Human - Requirement No. 2](#)" and "[D7.2 – Protection of Personal Data – Requirement No.5](#)", also, due in M6 of ASCAPE project.

1 Introduction

Research and innovation projects such as ASCAPE usually produce large sets of data. Depending on the discipline, the data could come for example from social science research, laboratory testing, field studies or observations. However, it often remains unclear and uncertain, what will happen with the data after they have been analysed and the project is finished. Furthermore, a lot of data sets are potentially interesting also for other researchers, but due to the fact that they are either stored on a local server or miss crucial meta-data (or both), their potential value cannot be exploited. Hence, consortium partners need to think about the data that they will produce at the beginning of the research and how they will handle them throughout and after the project duration, in view of ultimately increasing transparency, to the extent of course, feasible on the basis of the progress of the technical work; this is exactly the purpose of the Data Management Plan (DMP).

This deliverable presents the ASCAPE data management plan, as captured in M6 of the project. In accordance with European Commission's Guidelines for Horizon 2020 Programme¹, to submit a DMP within the first six (6) months of the project, the present report forms the DMP of ASCAPE project reflecting the technical progress at the moment of the drafting of the present document. Further updates in this respect will be provided in the course of the project duration, on the basis of the actual developments of the technical work. More specifically, in line with the F.A.I.R Principles, the deliverable provides for how making the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. To this end, the deliverable outlines -among other- how the research data collected and/or generated will be handled during and after the ASCAPE project, describes which standards and methodology for data collection and generation will be followed, and whether and how data will be shared. Note that the document is largely based on the related template provided by the European Commission.

The delivery of the present document falls under "Task 5.4. Legal framework, monitoring and ethical aspects", that extends until M36 of ASCAPE project. In the context of the related activities, it is, thus, intended that the DMP is a living document, subject to updates -to the extent necessary- on the basis of the progress of the project activities.

1.1 Purpose of the document

This Data Management Plan (DMP) outlines how data collected or generated by the ASCAPE project will be organised, stored, and shared. It specifies the types of research data that will be generated or collected during the project, the standards that

¹ H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0, 26 July 2016, available at:

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

will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse.

1.2 Target Audience

The present report forms a public deliverable -primarily- addressed to;

- European Commission
- Partners and Advisory Group in the ASCAPE project
- EU Parliament
- Horizon 2020 projects and other healthcare related projects (clustering activities)
- Organizations and experts involved in the ASCAPE case studies.
- Other relevant organizations both public and private, including associations of relevant stakeholders

2 Data Management in Horizon 2020

According to the European Commission (EC) all project proposals submitted to "Research and Innovation actions", "Innovation actions" and "Coordination support actions" have to include a section on research data management which is evaluated under the criterion 'Impact'. Projects participating in the pilot action on open access to research data have to develop a DMP) to specify what data will be open.²

The DMP is defined as:

“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. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research.”

The purpose of a DMP is to provide a discussion of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

Table 2-1- Clarifications of terms

Research data	Research data is the evidence that underpins all research conclusions (except those which are purely theoretical) and includes data that have been collected, observed, generated, created or obtained from commercial, government or other sources, for subsequent analysis and synthesis to produce original research results. These results are then used to produce research papers and submitted for publication.
Open research data	Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated, free of charge for the user.
Secondary data	Secondary data are data that already exist, regardless of the research to be conducted.
Open access	Open access is understood as the principle that research data should be accessible to relevant users, on equal terms, and at the lowest possible cost. Access should be easy, user-friendly and, if possible, Internet-based.
Metadata	Metadata is data used to describe other data. It summarizes basic information about data, which can make finding and working with instances of data easier.
Research data repositories	Research data repositories are online archives for research data. They can be subject based/thematic, institutional or centralized.

Overall, having taken into account all relevant principles regarding lawful processing of personal data, scientific research data should be easily discoverable, accessible, assessable and intelligible, useable beyond the original purpose for which it was collected and interoperable to specific quality standards.

²http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

The ASCAPE Data Management also follows the Guidelines on FAIR Data Management in Horizon 2020, released by the European Commission Directorate – General for Research & Innovation. This Horizon 2020 FAIR DMP template³ has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. According to these guidelines the management and organization of data should be based on four basic principles, which determine how research outputs should be processed so that they can be more easily accessed, understood, exchanged and reused. This means that data must be findable, accessible, interoperable and re-useable, for example by researchers interested in using the data in further research in the field.

These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. EC provides a Template with the FAIR principle. This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept. The template represents the set of questions that someone should answer with a level of detail appropriate to the project.

It is possible to develop a single DMP for any project to cover overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), someone should clearly spell this out.

The template proposes the following issues to be addressed:

- Data Summary
- FAIR data
- Allocation of resources
- Data security
- Ethical aspects
- Other issues
- Further support in developing your DMP

Each of the previously defined has its own set of questions that has to be addressed. The proposed template states that it is not required to provide detailed answers to all the questions of the DMP that needs to be submitted by month 6 of the project, subject -also- to potential future updates. Rather, the DMP is intended to be a living document -to the extent necessary- in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

³ H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0, 26 July 2016, available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3 ASCAPE Data Management Overview

As described in the Guidelines on FAIR Data Management in Horizon 2020 a Data Management Plan is a key element to ensure data is well managed. For this reason, in this section we will firstly identify the type of artefacts that will be generated and collected in the framework of the project. During the lifetime of the ASCAPE project, several artefacts will be produced. The artefacts that will be collected/generated are listed below in Section 3.2. As the project evolves, this list may require modifications (addition or removal of artefacts) with respect to the project developments.

3.1 Types and formats of artefacts generated/collected

In order to provide an overview of the different data sets that are currently and will be produced in the ASCAPE project, the following table shows the data type, the related WP number and the format, in which the data will be presumably stored.

Table 3-1 – Artefacts overview

#	Artefact type	Explanation	WP#	Format (indicative)
1	Research Item	Models and Meta models, Policies, Questionnaires, Deliverables, Papers	1-6	.xls, .csv, .txt, .docx, .pdf
2	Software	Code, APIs, microservices, libraries, dashboard	2,3	.xls, .csv, .txt, .docx, .pdf
3	Dataset	Retrospective, Prospective, Synthetic	1-4	.xls, .csv, .txt, .docx, .pdf

3.2 ASCAPE Artefacts and Access Rights

In the survey conducted during the first months of the project, the input collected from most of the partners is depicted in the following chart. The types of ASCAPE Artefacts are distributed as 23% being datasets, 45% research items and the remaining 32% is of a software artefact type.

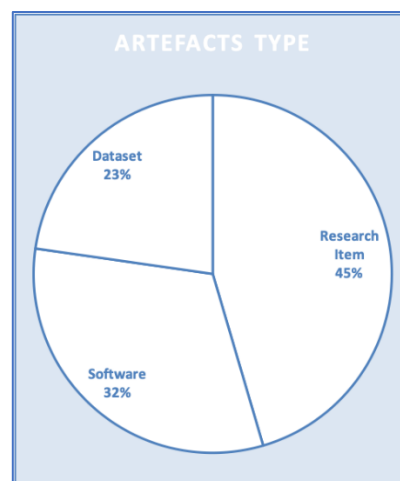


Figure 1 ASCAPE Artefacts' Type

The following tables present the current status and consensus within the Consortium with regards to identified artefacts and their access rights. It is provisioned that those tables are a recurring exercise and all future updates and additions will be documented under the relevant deliverables.

Table 3-2 Research Item Accessibility

Partners' Research Item provision		
Partner Name	Artefact	Access
SIE	HE AI Model	Non-Publishable
INTRA	Policies, Questionnaires	Non-Publishable
FCRB	Models and Meta models	Non-Publishable
	Questionnaires	Non-Publishable
UNSPMF	FL models for cancer-care predictions; Papers; Deliverables	Publishable
ATOS	ASCAPE Economic Model	Publishable
	Business Model for ASCAPE AI based platform	Non-Publishable
CC	Questionnaires	Non-Publishable
	Data Dictionaries	Non-Publishable
	Correlations	Non-Publishable
ALL	Deliverables, Papers, White Papers	Publishable

Table 3-3 Software Accessibility

Partners' Software provision		
Partner Name	Artefact	Access
SIE	software code	Non-Publishable
INTRA	Dashboard (server-side & client-side)	Publishable
FORTH	SGX enabled code	Non-Publishable
	AI&DL models parallelisation for GPU performance enhancement	Non-Publishable
FCRB	Code, APIs, microservices, libraries	Non-Publishable
STS	Software as a Service for Security auditing	Non-Publishable
FundiSYS	Xemio	Non-Publishable
UNSPMF	FL framework for cancer-care applications - code, DP components for data protection - code	Publishable
ATOS	HL7-FHIR model for QoL of cancer patients	Non-Publishable
	ASCAPE Web Page	Publishable
DFKI	Software for risk prediction explanations and intervention suggestions justification	Publishable

Partners' Software provision		
Partner Name	Artefact	Access
	Software for personalized what-if simulations and statistical assessment of the effect of suggested interventions and QoL related outcomes.	Publishable
UBITECH	Data Anonymizer	Non-Publishable
	Data Cleaner	Non-Publishable

Table 3-4 Pilots' Dataset Description and Access Rights

Pilots' Dataset		
Partner Name	Artefact	Access
FCRB	Medical dataset	Non-Publishable
ORB	Medical dataset	Non-Publishable
CC	Breast Cancer retrospective data	Non-Publishable
	Prostate Cancer retrospective data	
	Breast Cancer prospective data	
	Prostate Cancer prospective data	
	Fabricated data	
	Statistics	
Technical Partners	Synthetic anonymized Dataset and statistics	Publishable

Note that the above information provided by partners is currently available at the project's repository.

3.3 Data Protection

A detailed description of the data management practices concerning -in particular- processing of personal data in the context of the piloting activities is provided under "[D7.1 – Human - Requirement No. 2](#)" and, especially, "[D7.2 – Protection of Personal Data – Requirement No.5](#)", also, due in M6 of ASCAPE project.

The above documents capture key aspects of the lawful processing of health related data, such as the necessity to acquire an explicit consent from the individuals whose personal data will be processed for the purpose of ASCAPE and to conduct a Data Protection Impact Assessment (DPIA), as part of the technical and organizational measures in place. The individuals will be informed about the use of the information collected by them and will have to agree to the data collection while providing their approval in the form of written consent. The identity of any individual interviewed or in any other way initially engaged in the project (e.g. by email, correspondence, newsletter) will be further protected by this anonymization of his personal data. The relation of real names and codes will only be known to the partners (hospitals), who will keep the records in secure place. All copies of approvals/notifications regarding the processing of personal data by the competent authorities and Ethics Committee

that will be available by M6, are annexed under "[D7.1 – Human - Requirement No. 2](#)", also, due in M6.

Overall, any processing of personal data for the purpose of ASCAPE will be conducted-primarily- in accordance with the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, as well as in accordance with the policies and -where relevant- the terms of service of the ASCAPE pilots.

3.4 Existing data re-used (if any)

Data will be collected throughout the project. Regarding the usage of retrospective datasets collected from the Pilots' specific protocols will be documented and approved by their ethic committees and/ or competent authorities in order to align their usage with the terms and conditions agreed for the prospective datasets.

Likewise, it is aimed that sharing data with project partners outside the EU is avoided and that the respective decisions regarding technical implementation are taken accordingly.

3.5 Expected size of the data (if known)

It is expected that as a research outcome will generate research datasets (i.e. results of the technologies, services of the demos, etc.), publications, new services proposal through the Open Call, dissemination material, etc. Due to size of the project, scope of work and complexity, the expected size cannot be estimated at the moment.

4 ASCAPE ORDP Participation

The Open Research Data Pilot (ORDP) of the European Commission enables open access and reuse of research data generated by Horizon 2020 projects. There are two main pillars to the Pilot: a) developing a Data Management Plan (DMP) and b) providing open access to research data.

A project that opts-in ORDP have to adhere to the following conditions:

- Develop (and keep up-to-date) DMP.
- Deposit the data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate this data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research.

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that are specified in the DMP.

From the current consensus within the consortium most of ASCAPE Artefacts will not be publicly available as depicted in the graphic below (Figure 2).

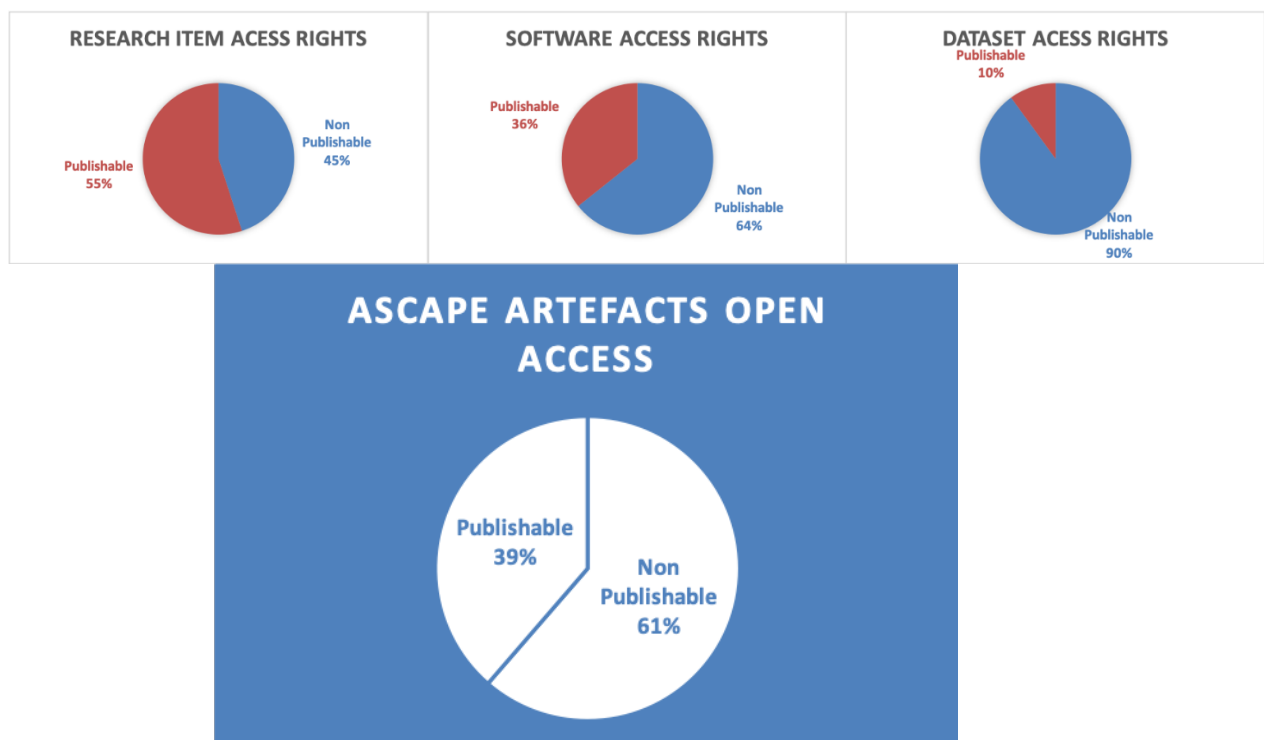


Figure 2 Artefacts Access Rights

4.1 Publishing Infrastructure for Open Access

The ASCAPE 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 of the project. The implementation of the project will be done in accordance with the applicable regulations in national and EU level and, especially, with the General Data Protection Regulation (GDPR)⁴ protection of personal data.

More specifically, *-apart from the Pilots' personal data collection and further processing-* there are not cases where personal data information or sensitive information of internet users is collected (IP addresses, email addresses or other personal information) or further processed. For the whole duration of the project, from the beginning to its end, the appointed project's Data Protection Officer (DPO), also, in the context of the activities of Task 5.4 mentioned above, will provide guidance on compliance aspects of the ASCAPE activities and tools, especially, with the GDPR.

. In the potential future case where the ASCAPE consortium will collect and/or further process personal data, this will be done in accordance with GPDR. Overall, it is aimed that ASCAPE only collects and/or further processes personal data are necessary for the attainment of the project objectives.

Both the process and the used web-based platforms are described in the following subsections.

4.1.1 Publishing Process

ASCAPE partners defined a simple, deterministic process that decides if a result in ASCAPE must be published or not. The term result is used for all kind of artefacts generated during ASCAPE like white papers, scientific publications, and anonymous usage data. By following this 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.

For each result generated or collected during ASCAPE runtime, the following questions must be answered to classify it:

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. Such a result could be code that is very specific to ASCAPE platform (e.g., a database initialization) which is usually of no scientific interest to anyone, nor does it add any significant contribution.

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. This also bares witness on the repetitive nature of the publishing process, where results which

⁴ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

are deemed in the beginning as non-publishable can become publishable once privacy-related information is removed from them.

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. Sometimes data inference can be used to superimpose different user data and reveal indirectly a single user's identity. As such, in order to make a result publishable, the included information must be reduced to a level where single individuals cannot be identified. This can be performed by using established anonymization techniques to conceal a single user's identity, e.g., abstraction, dummy users, or non-intersecting features.

Does a result include business or trade secrets of one or more partners of ASCAPE?

If this question is answered with yes, the result is classified as non-public, except if the opposite is explicitly stated by the involved partners. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

Does a result name technology that is 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.

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

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

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.

4.1.2 Publishing Platforms

In ASCAPE, 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.

The project Website

The partners in the project consortium decided early to setup a project-related website. This website describes the mission and the general approach of ASCAPE and its development status. A blog informs about news on a regular basis. Later in the project the developed ASCAPE platform will be announced. A dedicated area for downloads is used to publish reports and white papers as well as scientific publications (in pre-camera ready form, or through links to the publisher's websites in case these are not open access). All documents are published using the portable document format

(PDF)⁵. All downloads are enriched by using simple metadata information, such as the title and the type of the document. The website is hosted by partner UBITECH. All webpage-related data is backed up on a regular basis. All information on the project website can be accessed without creating an account. The website is backed up once per month.

Zenodo

Zenodo is a research data archive / online repository which helps researchers to share research results in a wide variety of formats for all fields of science. It was created through EC's OpenAIRE+ project and is now hosted at CERN using one of Europe's most reliably hardware infrastructures. Data is backed nightly and replicated to different locations. Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). Zenodo provides a connector to GitLab that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC0 license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results generated or collected during the project lifetime will be uploaded to Zenodo for long-term storage and open access.

GitLab

GitLab 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. GitLab 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 GitLab 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 GitLab over provided material. For textual metadata items, English is preferred.

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

4.1.3 Access and Sharing

The accessing and sharing of data is firstly ruled by two documents: the non-disclosure agreement, which stipulates under which conditions transmitted information between the project partners is deemed confidential and must not be further disseminated; and the Description of Action (DoA) which stipulates the dissemination level of each deliverable. Moreover, the project consortium will comply with the FAIR (findable,

⁵ Note that the site will not host spreadsheets. It exclusively host PDFs

accessible, interoperable and reusable) (European Commission, 2016) guidelines of the H2020 programme.

The data necessary to successfully complete the project Work Packages (WPs) will be shared without any restrictions amongst the WP partners either via internal repositories or direct communication. Public data will be made available at the project's website or other repositories, as appropriate. Users will be made aware of this data primarily through research publications, patent applications, dissemination activities, invited talks, social networks and the project website. Data will be made available to the project consortium as soon as it is available; to standardization bodies when required; and to the public at the due date of the derivable, and, in case a research publication is based on that, as soon as the paper is submitted (if submission is anonymous, this will be postponed). If access to confidential data is necessary by the public, restrictive measures will be put in place.

5 FAIR data

ASCAPE project supports the reuse of research data and follows FAIR principles⁶. FAIR represents a set of guiding principles to make data **Findable, Accessible, Interoperable, and Reusable**.

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable and reusable. Research data must be of quality that makes them accessible, findable and reusable.

- **Findable:** data has a unique, persistent ID, located in a searchable resource, and documented with meaningful metadata.
- **Accessible:** data is readily and freely retrievable using common methods and protocols, metadata is accessible even if the data is not.
- **Interoperable:** data is presented in broadly recognized standard formats, vocabularies, and languages.
- **Re-useable:** data has clear licenses, and accurate meaningful metadata conformity to relevant community standards and identifying its content and provenance.

5.1 Making data findable, including provisions for metadata

This document launches the data management plan to support the effective collection and integration of the ASCAPE data. Storage, processing and sharing (among project participants) will occur via data exchange platforms (such as Nextcloud), whereas interaction with the wider public will be achieved through the official project website, as well as the business platform that will be developed for the purposes of the Open Call procedure. Also, data will be stored at the coordinator's repository and will be kept for minimum 5 years after the end of the project. Where requested, data will be kept for 2 more years.

A naming convention will include a concise description of contents, the host institution collecting the data and the month of publication.

Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename.

No specific standards or metadata have been identified for the time being for the proposed datasets.

Data will be anonymized meaning that data will not identify any individuals and therefore real names of participants will NOT be distributed.

Data will be shared only in relation to publications (deliverables and papers). As such, the publication will serve as the main piece of metadata for the shared data. When this

⁶ Force11 (2016) The FAIR Data Principles, <https://www.force11.org/group/fairgroup/fairprinciples>

is not seen as being adequate for the comprehension of the raw data, a report will be shared along with the data explaining their meaning and methods of acquisition.

5.1.1 Discoverability of the data

In order to be able to use the data generated by the project it is essential to integrate data from the participants in the open calls and the activities undertaken by project partners. Taking into account the FAIR data principles (Wilkinson et al., 2016⁷) (meta)data should:

- Be assigned to a globally unique and persistent identifier;
- contain enough metadata to fully interpret the data, and;
- be indexed in a searchable source.

By applying these principles data become retrievable and include their authentication and authorization details.

5.1.2 Data identification mechanisms

All documents associated project will be identified with a project name and unique and persistent document type designator and number that will be given to the coordinator for the submission to the EC. Versioning of the document should be part of the document name and title.

As per the documents related to project activities and/or deliverables, the tasks or deliverables number will be used to identify the document followed by a brief title of the activity or deliverable.

Examples

- ASCAPE-D5.8- Open Research Data Pilot Contribution-v1.0.pdf

5.1.3 Naming conventions used

Each set of data produced (dataset, deliverables, etc.) will be named in a uniform way and will include a table with a version control.

The recommendations to name documents of the project are as follows⁸:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid Language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

For deliverables: **ASCAPE_[Deliverable Code]-[Deliverable Title]_[Partner]-vA.BB** i.e.: ASCAPE_D6.1-Project Management Handbook-v1.00 (*for submission to the Commission*)

⁷ Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data3:160018 doi: 10.1038/sdata.2016.18 (2016).

⁸ <https://www.ukdataservice.ac.uk/manage-data/format/organising>

For datasets: **WP [Work Package number] P [Pilot number; pilot activity number] - [description of the activity]** i.e.: WP4 P1.3 Results of demonstration performance.

5.1.4 Approach towards search keywords

Documents related to the activities of the project will be done following the templates agreed by the consortium, these templates include a keywords section to make documents findable.

The information submitted by the applicants to the open calls will use keywords related to the topics covered by the ASCAPE such as:

- Big Data,
- Data Driven Analytics,
- Artificial Intelligence,
- Machine Learning,
- Cancer Patients,
- Quality of Life,
- Health Status,
- Health Economic Data,
- Health Care metadata,
- Cancer care prediction and
- ASCAPE architecture

The keywords used to easily identify documents related to a project will be the ones used throughout the submission process, where applicant will have to select the characteristics of their projects selecting descriptors from a dropdown menu.

5.1.5 Clear versioning of the documents

Only documents created by the consortium will be versioned, for this purpose templates include 3 descriptors to identify the versions and status of the documents:

Table 5-1 – Proposed Document History Table overview

Version	Date	Comments	Author
1	xx	xx	xx
2			
3			
4			

Moreover, partners, following the recommendations included in section “Naming conventions” will identify the different versions by using a two-digit number following the descriptor Draft. A document reviewed by another partner should be returned to the principal author by including **rev + acronym** of the organisation. Only the principal author will change the draft number and will add the word FINAL to documents ready to be sent to the EC or those to be used as final versions.

The document history included in the document template should be filled in as follows:


Version	Date	Comments	Author
1	XX/XX/2019	Section 2.1 needs to be completed	ABC

2	XX/XX/2019	Section 2,1 completed. Comments added to the document.	CDE
3	XX/XX/2019	Added suggestions by EMAX	ABC
4	XX/XX/2019	Included some topics on section 2.1	XYZ
	XX/XX/2019	Final version with partners contribution	ABC

5.1.6 Standards for metadata creation (if any)

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators and contribute to easily find the information requested. To this end, all documents related to the project have to include in the front-page information about author(s) & editor(s), WP, dissemination level and version.

Figure 3 Template to be used for project documentation metadata overview



Artificial intelligence Supporting CAncer Patients across Europe

Project Title Artificial intelligence Supporting CAncer Patients across Europe

Project Acronym ASCAPE

Grant Agreement No 875351

Instrument Research and Innovation action

Call / Topic H2020-SC1-DTH-2019 / Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment

Start Date of Project 01/01/2020

Duration of Project 36 months

Dx.y – Deliverable title

Work Package	WPx – Work Package Title
Lead Author (Org)	Name Surname (Organization)
Contributing Author(s) (Org)	
Due Date	dd.mm.yyyy
Actual Date of Submission	dd.mm.yyyy
Version	

Dissemination Level

PU: Public (*on-line platform)

PP: Restricted to other programme participants (including the Commission)

RE: Restricted to a group specified by the consortium (including the Commission)

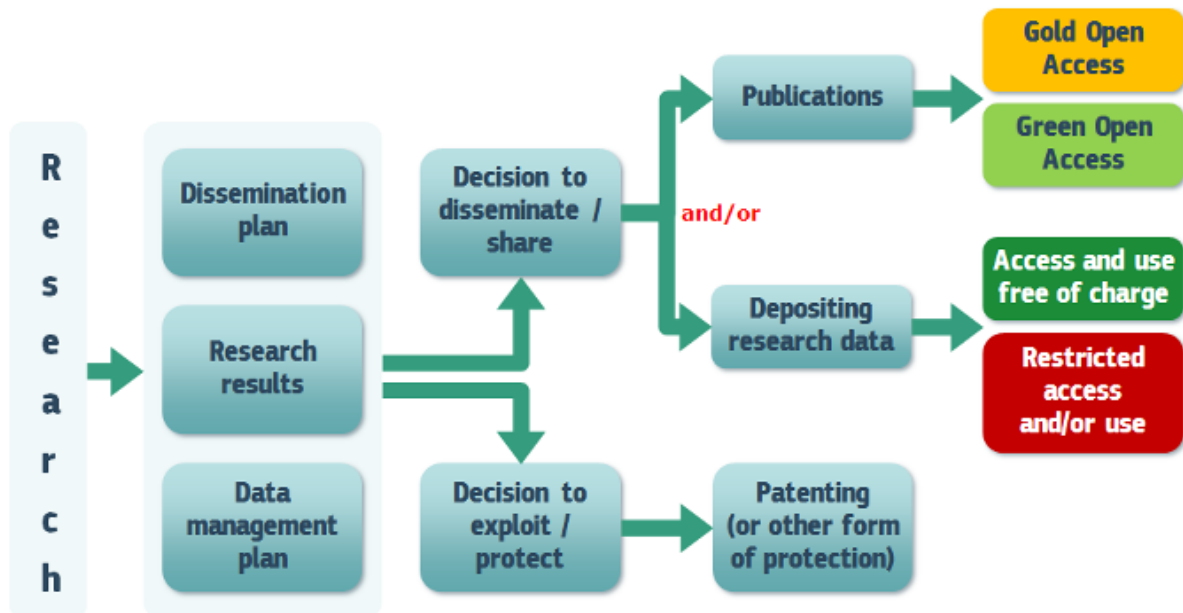
CO: Confidential, only for members of the consortium (including the Commission)

The work described in this document has been conducted within the project ASCAPE. This project has received funding from the European Union's Horizon 2020 (H4520) research and innovation programme under the Grant Agreement no 875351. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

5.2 Making data openly accessible

Where possible data will be made available subject to Ethics and participant agreement. However, the personally-identifiable nature of the data collected within ASCAPE means that in most instances it would be difficult to release collected data. Where data is made available, we will do so using the Project's file repository hosted in coordinator's premises.

Figure 4 Open access to scientific publication and research data in the wider context of dissemination and exploitation⁹



Prior to release, a requesting party will need to contact the Project Coordinator describing their intended use of a dataset. The Project Coordinator will send a terms and conditions document for them to sign and return. Upon return, the dataset will be released. Documentation will be included with the release of the data.

As regards the research data and its management and taking into account the scope of ASCAPE the consortium had originally opted out from the Pilot on Open Research Data Pilot (ORDP). . With regards to the dissemination of the scientific results, and in alignment with the EC Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, the ASCAPE consortium will also follow a combination of Gold and Green Open Access strategy to its scientific publications, with a potential embargo period for peer-reviewed publications that will be agreed during the first months of project execution. Gold Access will be encouraged for high-impact journal publications while the self-archiving, Green Access will be granted for the rest of the publications.

⁹ European Commission Directorate-General for Research & Innovation (2017) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

Table 5-2 – Artefacts' Types

#	Artefact Type	Explanation	Accessibility
1	Research Item	Models and Meta models, Policies, Questionnaires, Deliverables, Papers	Publishable/ Non-publishable
2	Software	Code, APIs, microservices, libraries, dashboard	Publishable/ Non-publishable
3	Dataset	Retrospective, Prospective, Synthetic	Publishable/ Non-publishable

5.2.1 Methods or software needed to access the data

No specific software tools will be needed to access the data, since anonymous data sets will be saved and stored in word, pdf or excel to facilitate its exploitation and guarantee their long-term accessibility.

Pilots' dataset access

In general¹⁰, infrastructures need to be setup at the pilot sites that will have an edge node. A physical hardware needs to be deployed inside the pilot site premises, where data provided by the pilot site is made available and can be accessed from the central ASCAPE model training serve, either as part of the federated learning or the homomorphic encryption scheme.

5.2.2 Deposit of data, associated metadata, documentation and code

Data will be deposited and secured on Nextcloud file repository and additional instance of all data on coordinator's account.

5.3 Making data interoperable

The concept interoperable demands that both data and metadata must be machine-readable and that a consistent terminology is used.

5.3.1 Interoperability of data assessment

Partners will be responsible of storing the data in a comprehensive format and adapted to the real and current needs of the possible practitioners interested in using, merging or exploiting the data generated throughout the project. The assessment of data interoperability will be updated in future reviews in order to guarantee the ASCAPE data fits the needs of a specific scenario (such as data infrastructures, interests or purpose of data).

¹⁰ The infrastructures to be set up at each pilot site will be discussed in depth under the related pilot specific deliverables.

5.3.2 Vocabulary use

The vocabulary used in the project is a very standard and common language within the business creation culture and the logistics. Vocabulary won't represent any barrier for data interoperability a re-use.

5.4 Making data re-usable

For data to be re-usable, it is -generally- considered that meta(data) have a plurality of accurate and relevant attributes and that they are released with a clear and accessible data usage license. Moreover, it is considered that (meta)data are associated with their provenance and that they meet domain-relevant community standards¹¹.

Note that the overall management of knowledge and the provisioning for the establishment of the related Intellectual Property Rights is dictated in detail under ASCAPE's Grant Agreement and the consortium agreement stipulating -among other- for the ownership of the background and the foreground knowledge, as well as for the commercial exploitation of the project's results.

5.4.1 Increase data re-use through clarifying licenses

Due to the sensitive nature of the data they will only be available on application/Nextcloud platform/share portal and their use will be restricted to the research use of the licensee and colleagues on a need-to-know basis. This non-commercial licence is renewable after 2 years, data may not be copied or distributed and must be referenced if used in publications. These arrangements will be formalised in a User Access Management licence which describes in detail the permitted use of the data.

5.4.2 Data quality assurance process

The project coordinator will be responsible of assuring the quality of the data by making sure dataset follow the FAIR principles included in this plan, and that data is updated.

Personal data processing will be done following the EU, national and international laws taking into account the "data quality" principles listed below¹²:

Data processing is adequate, relevant and non-excessive;

- Accurate and kept up to date;
- Processed fairly and lawfully;
- Processed in line with data subjects' rights;

¹¹ See, also, FAIR data principles (FORCE11 discussion forum) available at: <https://www.force11.org/group/fairgroup/fairprinciples>

¹² Wilms, G. Guide on Good Data Protection Practice in Research of the European University Institute. (March 2017). Retrieved from <http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

- Processed in a secure manner;
- Kept for no longer that necessary and for the sole purpose of the project.

Data quality assurance process will be led in accordance with the REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

5.4.3 Length of time for which the data will remain re-usable

The Consortium will contribute to maintain data re-usable as long as possible after the end of the project. A first period of 4 years has been established; however, this time can be extended under partners agreement. This period can vary depending on the value of the data after the end of the project.

5.5 Artefact Template

The following tables try to capture the description of the data that will be produced in the context of ASCAPE. Every use case will fill in such a template and subsequently all the templates will be collected with the beginning of WP4, the demonstration applications work package of the project.

Table 5-3 Making Data Findable Template

Making data Findable	
Name of data set	<i>Univocal identifier of the considered data [ASCAPE_Wx_Tz_01] Please, provide one sentence description.</i>
Data types	<i>[Real time data stream, unstructured like tweets, synthetic data stream, log data of IDS, etc.]</i>
Data generation and/or collection	<i>Description of the type of input used to generate the data and the complete methodology and tools used for data collection</i>
Purpose	<i>What are the data collected/generated specifically used for?</i>
Data origin	<i>[Where applicable, information from applications to be developed by the partner.]</i>

Table 5-4 Making data Accessible Template

Making data Accesible	
Accessibility	<i>Open/Confidential</i>
Repository	<i>Description/location of the available data.</i>

Making data Accessible	
Shareability restrictions / related Information	<i>[Where applicable, information from applications to be developed by the partner.]</i>

Table 5-5 Making data Interoperable Template

Making data Interoperable	
Format	<i>Data format, measuring unit, typical order of magnitude [JSON-like, CSV]</i>
Expected size of the data	<i>[To be defined, 3 TB/Day or 12 GB/day when compressed etc.]</i>
Standards and metadata¹³	<i>[The metadata attributes list. The used methodologies.]</i>
Standard software Interfaces	<i>List of the standards used to promote results replicability.</i>
Extensions to standard interfaces	<i>Extensions to the above standards as developed during the project.</i>

Table 5-6 Making data Re-usable Template

Making data Reusable	
RE-use of existing data	<i>[No reuse of existing data, for the generation of synthetic datasets, it will be essential to create a recipe, reusing the existing data in logs etc.]</i>
Data backup	<i>Consistent location of the data, including previous releases</i>
Quality Consistency	<i>Constraints determining the quality/currency of the collected data.</i>
Emulation tools	<i>Description/location of possible emulation tools useful for replicating the data</i>

¹³ Note that the fields pertinent to standards are, also, relevant for reusability purposes.

6 Allocation of resources

6.1 Data management responsibilities

Data will be stored at the Collaboration file repository (Nextcloud), set by the Coordinator as the project's repository, and will be kept for 5 years after the end of the project. Where requested, data will be kept for 2 more years. The handling of the repository on behalf of ASCAPE as well as all data management issues related to the project fall in the responsibility of the coordinator.

As for the publications, where the analyses of the empirical research data will be presented, the consortium will publish them in scientific journals that allow open access. The costs related to open access will be claimed as part of the Horizon 2020 grant.

Regarding the data resulting from the activities of the project, each WP leader will be responsible for the storage and compliance of the data and then for uploading in the ASCAPE Nextcloud web portal, or other storage systems to share the information of the project.

The ASCAPE coordinator assisted by the WP leaders will be responsible for updating this document and develop a strategy to encourage:

- the identification of the most-suitable data-sharing and preservation methods;
- the efficient use of data assuring clear rules on its accessibility;
- the quality of the data stored and
- the storage in a secured in a user-friendly interface.

6.2 Cost of potential value of long-term preservation

As stated in previous section, the costs of data storage and maintenance are not going to require extra funding once the project ends. As per the value of the data, it is important to take into account that the topics covered by the project respond to a current need of the logistics sector and customers' needs. Therefore, data coming out of this project will have a direct impact in the coming years but might not be of relevance as the challenges are being tackled or replaced by other priorities.

7 Data Security

ASCAPE data exchange platform (Nextcloud) applies technological and organizational measures to secure processing of personal data against publishing to unauthorized persons, processing in violation of the law and change, loss, damage or destruction.

- **Information security:** SSL (Secure Socket Layer) certificates are applied. In order to ensure the appropriate level of security, the password for the account will exist on the platform only in a coded encrypted form.
- **Options for reading data:** the platform offers the possibility to make data available in a read-only or downloadable format, hindering the access to information by unauthorized users.
- **Back-up policy:** complete and redundant back-ups are done every week. Moreover, every time a modification is done an older version is saved.
- **Accidental deletion or modifications:** in case of a catastrophic event that implies the partial or complete deletion of the data sets, the data from the most recent back up will be automatically restored (back-up won't be older than 60 minutes). In case of accidental deletion or modification only the most recent document will be restored, so in case of accidental changes or deletion data can be easily recovered.
- **Deletion or modification of data by users:** only administrators have the rights to delete or modify the information included in the datasets.
- **Terms and conditions:** the Nextcloud platform have specific terms of use and conditions that have to be accepted by all users of the platform.

8 Ethical Aspects

The ASCAPE consortium is aware of the ethical aspects pertinent to the scope of ASCAPE, which are addressed under the Work Package 7 on "Ethics Requirements".

In particular, Work Package 7 puts particular emphasis on two aspects pertinent to the scope of ASCAPE: a) the involvement of research participants and b) the protection of the personal data to be collected and further processed for the purpose of ASCAPE project. Work Package 7, thus, focuses on the piloting activities by capturing, for instance, the procedures for approvals by the Ethics Committees for the performance of the relevant clinical trials and by producing an overview of the technical and organizational measures for the protection of personal data of the research participants.

Note that the content of the earlier stated deliverables on Ethics due in M6 will be updated in the context of the activities under Task 5.4 on "Task 5.4. Legal framework, monitoring and ethical aspects", that extends until M36.

9 Concluding remarks and next steps

The present document captures the DMP of ASCAPE project, based on the progress of the technical work by M6. It did so, by providing information resulting by the inputs of the technical partners, by putting focus on related aspects of the piloting activities and by -ultimately- producing an overview of the data management practices at consortium level. Due to the scope of ASCAPE- including-the specific piloting activities envisioned, it is expected that further updates pertinent to the contents of the present document will be, also, provided -to the extent relevant- through the rest of the public deliverables, due under the earlier stated "Task 5.4. Legal framework, monitoring and ethical aspects".