

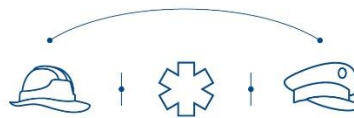
ASSISTANCE

Adapted situation awareneSS tools and tallored training curricula for increaSing capabiliTie and enhANcing the proteCtion of first respondErs



European Commission

Project co-funded by the European Union within the Horizon 2020 Programme



assistance

Project Ref. N°	ASSISTANCE H2020 - 832576
Start Date / Duration	May 1, 2019 (36 months)
Dissemination Level¹	PU (Public)
Author / Organisation	Jordi Arias/ETRA

Deliverable D9.4

Research data management, Open Data and Open Access strategy

¹ PU: Public; PP: Restricted to other programme participants (including the EC services); RE: Restricted to a group specified by the Consortium (including the EC services); CO: Confidential, only for members of the Consortium (including the EC services).

ASSISTANCE

Nowadays different first responder (FR) organizations cooperate together to face large and complex disasters that in some cases can be amplified due to new threats such as climate change in case of natural disasters (e.g. larger and more frequent floods and wild fires, etc) or the increase of radicalization in case of man-made disasters (e.g. arsonists that burn European forests, terrorist attacks coordinated across multiple European cities).

The impact of large disasters like these could have disastrous consequences for the European Member States and affect social well-being on a global level. Each type of FR organization (e.g. medical emergency services, fire and rescue services, law enforcement teams, civil protection professionals, etc.) that mitigate these kinds of events are exposed to unexpected dangers and new threats that can severely affect their personal safety.

ASSISTANCE proposes a holistic solution that will adapt a well-tested situation awareness (SA) application as the core of a wider SA platform. The new ASSISTANCE platform is capable of offering different configuration modes for providing the tailored information needed by each FR organization while they work together to mitigate the disaster (e.g. real time video and resources location for firefighters, evacuation route status for emergency health services and so on).

With this solution ASSISTANCE will enhance the SA of the responding organisations during their mitigation activities through the integration of new paradigms, tools and technologies (e.g. drones/robots equipped with a range of sensors, robust communications capabilities, etc.) with the main objective of increasing both their protection and their efficiency.

ASSISTANCE will also improve the skills and capabilities of the FRs through the establishment of a European advanced training network that will provide tailored training based on new learning approaches (e.g. virtual, mixed and/or augmented reality) adapted to each type of FR organizational need and the possibility of sharing virtual training environments, exchanging experiences and actuation procedures.

ASSISTANCE is funded by the Horizon 2020 Programme of the European Commission, in the topic of Critical Infrastructure Protection, grant agreement 832576.

Disclaimer

This document contains material, which is the copyright of certain ASSISTANCE consortium parties, and may not be reproduced or copied without permission.

The information contained in this document is the proprietary confidential information of the ASSISTANCE consortium (including the Commission Services) and may not be disclosed except in accordance with the consortium agreement.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the project consortium as a whole nor a certain party of the consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

The information in this document is subject to change without notice.

Executive Summary

This document *D9.4: Research data management, Open Data and Open Access strategy* is a living document that aims at providing an analysis of the main elements of the data management policy that will be used by the ASSISTANCE Consortium regarding the project research data.

This document will evolve during the development of the project, when the project progresses and when significant changes occur, in order to keep an updated version of the guidelines and recommendations for making the research data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (FAIR) and therefore contribute to knowledge discovery and innovation.

The current version is the second iteration, after the deliverable *D1.2 Data Management Plan* was presented at M3, in which the envisioned data management strategy is provided and a first effort is done to define 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.

List of Authors

Organisation	Authors
UPVLC	Federico Carvajal, Manuel Esteve, Israel Pérez
ETRA	Jordi Arias, Luisa Perez, Eva Muñoz

Change control datasheet

Version	Changes	Chapters	Pages	Date
0.1	First draft	All	12	25/06/2020
0.2	Internal version updated	All	13	30/06/2020
0.3	Content ready	All	25	07/08/2020
0.4	UPV review	All	25	10/08/2020
1.0	Updates and changes after UPV considerations	All	25	12/08/2020
1.1	Final version after peer reviewers' and partners comments	All	25	17/08/2020
1.2	Final version	All	30	16/10/2020
1.3	UPV review	All	31	21/10/2020
1.4	Final version ready to submit	All	31	27/10/2020

Content

Executive Summary	3
List of Authors	4
Change control datasheet	5
Content.....	6
List of Figures.....	7
List of Tables	7
Acronyms.....	8
1. INTRODUCTION	9
1.1. Purpose of the document	9
1.2. Scope of the document.....	9
1.3. Structure of the Document	9
2. DATA SUMMARY	11
2.1. Purpose of research data management and relation to the project	11
2.2. Data set types, formats, and standards	11
2.3. Re-use of data	13
2.4. Timetable for updates of the Research data management, Open Data and Open Access strategy	13
3. FAIR research data.....	14
3.1. Making data findable, including provisions for metadata	15
3.2. Making data openly accessible.....	17
3.2.1. Open research data repository	18
3.2.2. Open Access publications	19
3.2.3. ASSISTANCE internal repository	21
3.2.4. Public deliverables	23
3.3. Making data interoperable	24
3.4. Increase data re-use	25
4. Allocation of resources	25
4.1. Responsibilities and decision making	26
5. Data security	27
6. Ethical aspects	28
7. References.....	29

List of Figures

Figure 1 - Research Data Management according to the FAIR principles [8]	15
Figure 2 – Summary of Metadata Best Practices [9].	17
Figure 3 – Open access to scientific publication and research data in the wider context of dissemination and exploitation. Source H2020 Open Access portal	21
Figure 4 – ASSISTANCE repository overview	22
Figure 5 – Example of folders and files in ASSISTANCE repository	23
Figure 6 - Public deliverables in ASSISTANCE Website	24

List of Tables

Table 1 – Timetable for updates of the DMP	14
--	----

Acronyms

ASSISTANCE	Adapted situation awareness tools and tailored training curricula for increasing capabilities and enhancing the protection of first responders
D#.#	Deliverable number #.# (D1.1 deliverable 1 of work package 1)
DDP	Distributed Data Protocol
DOI	Digital Object Identifier
DMP	Data Management Plan
EC	European Commission
EU	European Union
GA	Grant Agreement
H2020	Horizon 2020 Programme for Research and Innovation
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
ISRC	International Standard Recording Code
M#	#th month of the project (M1=May 2017)
MQTT	Message Queuing Telemetry Transport
PC	Project Coordinator
URL	Uniform Resource Locator
WP	Work Package

1. INTRODUCTION

1.1. Purpose of the document

The purpose of this document is to provide an analysis of the main elements of the data management policy that needs to be considered by the ASSISTANCE Consortium with regards to the project research data. It presents the knowledge management strategy and the basic rules to guarantee an open access to the research data retrieved by the project.

This document aims to outline how the ASSISTANCE project will make the research data Findable, Accessible, Interoperable and Reusable (FAIR) and therefore contribute to knowledge discovery and innovation. The objective is to lay the foundations for creating an effective data management strategy covering the complete research data life cycle.

The *D9.4: Research data management, Open Data and Open Access strategy* is not a fixed document; on the contrary, it will evolve during the lifespan of the project. This deliverable will be a living document in which information will be available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur [1].

1.2. Scope of the document

The focus of this deliverable is the covering of the complete research data life cycle. The knowledge management strategy will be used by the ASSISTANCE consortium as a guideline when handling the research data during and after the end of the project. In this document, a new iteration of the Data Management Plan is presented to provide further details about the definition of the types of research data that are 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.

1.3. Structure of the Document

This deliverable follows the template provided by the services of the European Commission (EC) on Data Management [2]. The structure of the template has just been adapted to follow ASSISTANCE document procedures –e.g. section 1 “Introduction” is mandatory in all the project documents.

D9.4: Research data management, Open Data and Open Access strategy

In section 2, the data sets that will be used in the project are introduced and briefly described. Then, in section 3 the document focuses on making ASSISTANCE research data findable, accessible, interoperable and re-usable (FAIR). Section 4 provides an overview of the allocation of resources needed to implement this strategy. Finally, section 5 deals with data security aspects.

Respecting the guidelines provided by the Commission [1], this document will be updated in time with the periodic evaluation/assessment of the project at month 24, and at the end of the project, at month 36. During these reviews and whenever significant changes arise, the current document will be updated with newly available information.

2. DATA SUMMARY

2.1. Purpose of research data management and relation to the project

ASSISTANCE is a project that enhances cooperation and coordination among different kinds of FRs' organizations for helping and protect them in a large scale disasters, enhancing their capabilities for facing complex situations by providing them with advanced training, tailored to their actual needs and based on novel approaches like Virtual (VR), Mixed (MR) and Augmented Reality (AR) in order to increase FRs Situation Awareness (SA).

ASSISTANCE will comply with European recommendations regarding Data Management Plans², as is explained in the *D1.2 Data Management Plan* providing clear procedure for findable, accessible, interoperable and re-usable (FAIR) data and updating the current document along with the development of the project.

The purpose of the D9.4 is to provide a better analysis of the main elements of the research data management policy, explained in the first version of the Data Management Plan -D1.2, that will be used by the Consortium with regards to the project research data. This document reflects consortium data management policies, systems, and procedures - which will be implemented and embedded in research procedures and regularly reviewed throughout the research cycle.

The project will strive to make data open but cannot overrule limitations that partner institutions put on data that they contribute (as specified in their Background included in the ASSISTANCE Consortium Agreement). Moreover, an ethical approach will be adopted and maintained throughout the fieldwork process. The responsible partners will assure that the EU standards regarding ethics and Data Management are fulfilled, as stated in WP10.

2.2. Data set types, formats, and standards

Throughout the duration of the project, ASSISTANCE will gather – and also generate – a variety of research data. Research data refers to information collected to be examined and considered 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 [3].

² http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

All of them have to be stored in a way such that they are easily accessible by both humans and software, as appropriate.

Data in ASSISTANCE can be classified into three main categories:

- Organizational data: these are those data relevant to the implementation of the Research and Innovation Action, such as project deliverables, internal reports, agendas and minutes of meetings, etc.
- Technical data: they include raw and processed data. The Situation Awareness Platform (SAP) to be developed in ASSISTANCE will receive information from different sensors and systems and will integrate all this information, showing it in a tailored manner through its main Human Machine Interface (HMI), to the different FRs that will use the platform. Moreover, some other modules and systems will also receive information from sensors and different sources for performing their internal process and will send these processes result to the SAP.
- Scientific data: these includes scientific analyses/publications, papers and all kind of dissemination material, as well as software code and algorithms. Data for analysis and evaluations, such as interviews and questionnaires also fall within this category.

The detailed definition of the different types, formats and standards of the data to be collected, processed and/or generated during the project was done in the framework of WP1 and WP2, more specifically in task T2.4 “*System and Network Architecture Design*” and the associated deliverable D2.4 “*ASSISTANCE System and Network Architecture Design*” where technical data is analysed, and Task 1.1 “*Project Management*” and the deliverable D1.2 “*Data Management Plan*”, which describes the main types of documentation that will be generated or collected during the project.

D2.4, submitted in M9 of the project, provides the description of the open, secure and flexible architecture including details of the types and formats of data to be handled. The deliverable also provides the description of the security features required by the ASSISTANCE architecture and solutions as well as the identification of relevant standards for the project development.

The starting point of Task T2.4 was to list the types of data to be handled in the framework of ASSISTANCE and to associate to each type of data the specific details, which will be defined in the upcoming deliverables, the sensitivity of data in terms of: a) privacy; b) data protection due to privacy or other reasons, such as commercial interest or critical infrastructure data.

In any case, as already stated at the ethical deliverables, the project is not supposed to employ private data related to individuals. ASSISTANCE will exclusively manage anonymised or aggregated data in project workshops and events, as well as the pilot and training exercises.

In addition, no sensitive personal data, neither personal data itself, is being processed in the ICT systems that are relevant for ASSISTANCE demonstration activities.

Regarding scientific data, the following template will be used for the management of ASSISTANCE scientific publications:

Publication reference and name:	The name and the ID of the publication (e.g. DOI).
Publication Abstract:	A short description of the content of the publication.
Standards and metadata:	The type of the document format and any type of metadata associated with the content of the document.
Publication sharing:	Where the final version of the scientific publication is stored.
Archiving and preservation (including storage and backup):	How the scientific publication will be preserved, archived and preserved

2.3. Re-use of data

ASSISTANCE will use and exploit some of the preliminary solutions and technologies developed in two H2020 projects: AF3 Advanced Forest Fire Fighting [3] and CAMELOT [4] projects. These two projects will be the basis of some of the ASSISTANCE Innovation activities, complementing and going beyond the work done in these projects, providing new and more advanced services.

Moreover, during the project lifetime, available results from other research activities, publications, and further relevant information available will be analysed. This information will be mainly used for internal project studies and will be provided in respective project deliverables with appropriate references to origins of the gathered information.

2.4. Timetable for updates of the Research data management, Open Data and Open Access strategy

This document is a living document that will be updated over the course of the project whenever significant changes arise, such as new data, modification in consortium policies, changes in consortium composition, external factors, etc.

As a minimum, the *D9.4 Research data management, Open Data and Open Access strategy plan* will be updated according to the following timetable:

DMP Version	Delivery Month	Description
V1.0	M3	D1.2 Data Management Plan First version

V2.0	M18	D9.4: Research data management, Open Data and Open Access strategy including WP3 inputs and data updates
V3.0	M24	DMP including any refinement coming from the first developing activities
V4.0	M36	DMP Final version

Table 1 – Timetable for updates of the DMP

3. FAIR research data

Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results [5]. According to the EC H2020 guidelines [1], good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. In general terms, the FAIR principles [6] describe four key concepts in research data management. Research data should be:

- **Findable** – Easy to find by both humans and computer systems and based on a mandatory description of the metadata that allows the discovery of interesting datasets;
- **Accessible** – Long term storage so data can be easily accessed and/or downloaded with well-defined license and access conditions, whether at the level of metadata, or at the level of the actual data content;
- **Interoperable** – Ready to be combined with other datasets by humans, as well as computer systems;
- **Reusable** – Ready to be used for future research and to be processed further using computational methods.

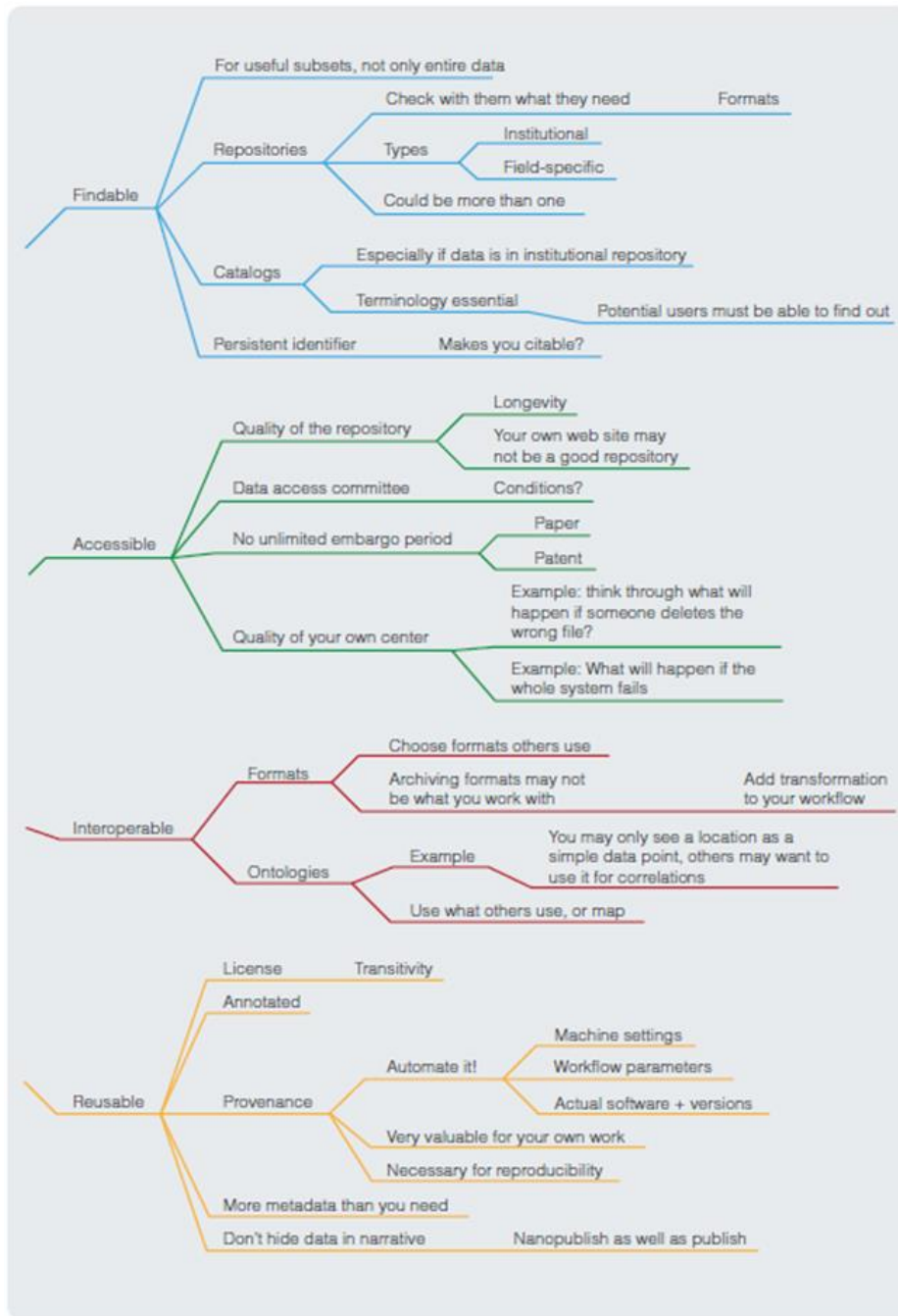


Figure 1 - Research Data Management according to the FAIR principles [8]

3.1. Making data findable, including provisions for metadata

Identification and localisation means will be used for some specific data to be generated during ASSISTANCE project.

The data to be generated in ASSISTANCE project will be identifiable and locatable by means of unique identification mechanisms. Files will be uniquely identifiable by using standardised name conventions and clear versioning. These conventions for the documents – and data sets - are already provided in D1.1 Project Management Handbook and D1.2 Data Management Plan.

According to those documents, the following conditions apply regarding identification and location of data:

- Code development: rules will be specified within WP3, WP4, WP5 and WP6. The repository server will be able to include links to SVN repositories for versioning management as required and defined by WPs.
- Deliverables can be classified as Confidential (CO) or Public (PU) in ASSISTANCE. All these documents will be named and numbered according to the following rules, in order to facilitate the quick identification and indexing:

ASSISTANCE_<dnum>_<dname>_v<ver>.pdf

All the documents names start with the word “ASSISTANCE” in order to facilitate the identification with other projects documents, and to raise the awareness of the project within a number of people that will download the documents from the public website. The fields <dnum> represents the code of the deliverable, <dname> the name of the deliverable according to the title, and <ver> is the version of the document.

Versions 0.X will indicate that the document is still a draft not approved by the internal reviewers. The official document to be sent to the EC will be numbered as v1.0. Further revisions or new issues of a deliverable will make use of the following format: v1.X, vY.X.

In order to facilitate the work and localisation of the documents, all the documents will be posted in the repository as soon as possible.

- Public deliverables will be available for the public at the ASSISTANCE project website <https://assistance-project.eu> when they had been officially approved by the EC. Confidential deliverables will not be available on the website. In case of request from any external party, the Security Advisory Board may decide to disseminate the corresponding deliverables or specific parts of the deliverables to particular external parties.

ASSISTANCE project research data will be inventoried and annotated with metadata always that this data can be made publicly available in a repository, as in the case of data supporting scientific publications, which will be provided as datasets as detailed in section 3.2. At that end, metadata must describe the dataset itself (e.g. date of creation, title, content, author, type, size). This information about the data needs to be added to the catalogues to help discover the data. Metadata needs to be both human understandable and machine readable [9].

The metadata will be published with the data using a machine-readable format and standard terms to define the metadata; the overall features of the dataset will be described with information about local parameters, licence, origin, and quality.

The figure below summarizes the best practices outlined by the European Data Portal which will be followed the ASSISTANCE partners when they have to publish datasets linked to scientific publications.

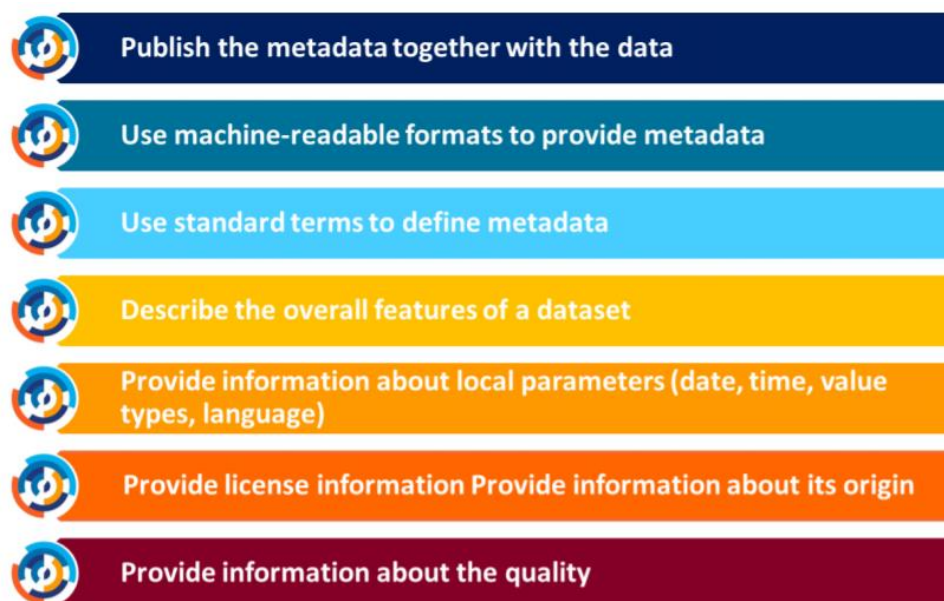


Figure 2 – Summary of Metadata Best Practices [9].

3.2. Making data openly accessible

An analysis of which ASSISTANCE research data will be made openly accessible and which data will be kept closed will be done while the project is progressing.

According to the classification provided in section 2.2, **Organizational data** are susceptible to be published openly only if they are categorised as “Public”, as it is the case of the Public Deliverables. On the other hand, **Technical data** will be never open access: research data – especially considering the pilot clusters – is sensitive due to security issues and therefore will be kept confidential. Finally, **Scientific data** will be open under specific circumstances, to be assessed in the framework of the WP9 Exploitation and Dissemination.

In general terms, ASSISTANCE research data will be made available when possible without compromising privacy, ethical or commercial sustainability, to parties with a legitimate research interest.

The outcome of this analysis shall be implemented in the coming versions of the DMP.

During the project lifetime, information on the following aspects will be elaborated for data sets on a case by case base, 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 / 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 / researchers, and
- Any further related issue **Error! Marcador no definido.**

3.2.1. Open research data repository

For implementing open access to the research data sets, ASSISTANCE will use Zenodo [9], an established online European scientific repository that is fully integrated with OpenAIRE. The decision on whether a research data set will be uploaded to and opened for access in Zenodo will be made on a case-by-case basis between the coordinator (UPVLC), the Innovation Manager (ETRA) and the partner(s) that have ownership of the data. This will be specifically for the case of datasets that are linked to scientific publications, and not for those datasets that are classified as Technical Data, as explained previously.

When a data package study has been marked as public, it will be made openly available. Data gathered by partners outside of the project work plan and protected by IPR, or inside the work plan but containing confidential information (e.g. related to personal data), will be kept closed for privacy reasons.

Zenodo offers a simple online service that enables researchers, scientists, EU projects and institutions to share, preserve and showcase multidisciplinary research results (data and publications), that are not part of the existing institutional or subject-based repositories of the research communities. It provides service hosting according to industry best practices in CERN's professional data centres. A detailed description of Zenodo's policies regarding the handling of the data and usage of the service is found at <https://zenodo.org/policies>



To share datasets in the publicly accessible disciplinary repository Zenodo, descriptive metadata will be used as required/provided by that repository. Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. (It is NOT possible to edit a Zenodo DOI once it has been registered). Zenodo further supports harvesting of all content via the OAI-PMH protocol³.

A digital object identifier (DOI) is a character string (a "digital identifier") used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found.

³ Open Archive Initiative - Protocol for Metadata Harvesting (OAI-PMH)

The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the metadata for the DOI to link to the new URL. A DOI name differs from standard identifier registries such as the ISBN (ISO 2108:2005) and ISRC (ISO 3901:2001). The purpose of an identifier registry is to manage a given collection of identifiers, whereas the primary purpose of the DOI system is to make a collection of identifiers actionable and interoperable.

3.2.2. Open Access publications

The rules and principles of the European Commission's Horizon 2020 Framework Programme clearly establish that scientific results generated within H2020 projects will be made available as open access publications, i.e. freely available online to any user. Following these rules, enforced in the ASSISTANCE Grant Agreement – Article 29.2, open access will be ensured to all peer-reviewed scientific publications related to ASSISTANCE and its composite solutions. Furthermore, other project outputs and results may be provided in the Open Access upon respective decision of the project consortium.

According to the “Open access to publications and data in Horizon 2020 Fact sheet” [10], two main routes exist for open access to scientific peer-reviewed publications:

- Self-archiving (also called ‘Green’ open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed (‘embargo period’ of six months of publication) as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period
- Open access publishing (also called ‘Gold’ open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead charged to the research institute to which the researcher is affiliated, or to the funding agency supporting the research.

Since these two routes described above are not mutually exclusive in an EU-funded action, within the ASSISTANCE project each beneficiary will be able to choose the most suitable approach for each publication concerned. In any case, this specific aspect will be further discussed and described in an updated version of the Data Management Plan.

D9.4: Research data management, Open Data and Open Access strategy

For the moment, parallel publishing has been chosen as the primary strategy for providing open access. In addition, a budget 10.000 € was included in the dissemination activities leader (PIAP) budget under other goods and services in order to cover the open access publications cost that the partners perform.

This allows consortium members to publish their results in the scientific fora and journals of their choice for maximum impact and still ensure optimal dissemination of the results by open access. It is worth to mention that the overwhelming majority of academic journals support either the gold, the green or a hybrid open access route. It means that ASSISTANCE beneficiaries have freedom to publish where they feel it is the most appropriate. All publications will have a DOI making them easily findable and citable. Project partners will be able to publish in open access until this reserved budget (10k€) was over.

It is important to note that the open access requirement does not mean an obligation to publish results. As stated by the guidelines: “The decision to publish is entirely up to the grant beneficiaries. Open access becomes an issue only if publication is chosen as a means of dissemination”. The graph below (see Figure 3) illustrates the decision related to research results and possible path for publication or other options.

It is worth to mention that project partners have reported a number of events and initiatives they intended to participate in or undertake (see *Deliverable D9.2 Updated Exploitation and Dissemination Plan*). Nevertheless, it is important to underline that the current COVID-19 outbreak situation and the uncertainty of the situation in the next months and years have heavily affected some of the already planned communication and publications activities (many congresses and events have been cancelled or postponed, like e.g. INTERSCHUTZ 2020 event or ISCRAM 2020 conference) as well as any potential future planning. Therefore, the outline of planned activities related to publications or events, should be treated as only an initial plan for further actions that will be updated in the course of project implementation and adjusted to the evolving situation.

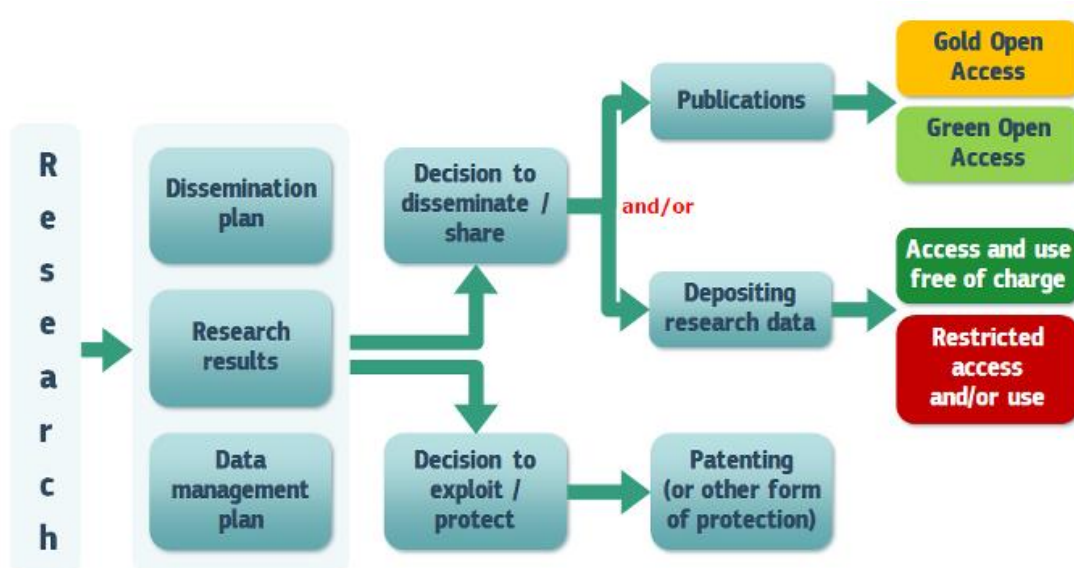


Figure 3 – Open access to scientific publication and research data in the wider context of dissemination and exploitation. Source H2020 Open Access portal

ASSISTANCE will create a space (a “Community”) in Zenodo, where all public research data can be made open access, citeable and discoverable. Examples research outputs that can be uploaded are: publications (book, book section, conference paper, journal article, patent, preprint, report, thesis, technical note, working paper, etc.), posters, presentations, datasets linked to scientific publications, images (figures, plots, drawings, diagrams, photos), software, videos/audio and interactive materials such as lessons.

3.2.3. ASSISTANCE internal repository

During the life cycle of ASSISTANCE, data collected or generated by the project is being stored and systematically organised in the official project repository on GitLab.

GitLab⁴ is a secure and flexible content management web application developed using Java technology. In the framework of the ASSISTANCE project, it is used mainly as a repository to securely store and share files, making data available to the whole Consortium. The repository (documents section) consists of an internal project area, not possible to be accessed by external users.

⁴ <https://gitlab.com>

D9.4: Research data management, Open Data and Open Access strategy

ASSISTANCE / Repository

Archivos Incidencias Pull Requests Wiki

Sin descripción

331 Commits 2 Ramas 0 Releases

Rama: master Repository Nuevo archivo Subir archivo HTTPS https://git.services.satrd.es

Folder Name	Commit Hash	Description	Time
fecarro	5994141be2	Subir archivos a 'Submitted Deliverables'	hace 1 semana
Admin	b1a4b899db	Repository structure	hace 1 año
Approved Deliverables	b1a4b899db	Repository structure	hace 1 año
Contractual Documents	b629015873	Subir archivos a 'Contractual Documents'	hace 1 año
Coordination with other rel...	a5f161b9ad	Eliminar 'Coordination with other related and relevant projec...	hace 1 año
Coordination with standar...	da562a8ac9	Subir archivos a 'Coordination with standardization organiza...	hace 1 año
Dissemination material	95f2cc9ba8	Prześlij pliki do 'Dissemination material'	hace 1 mes
Events	b1a4b899db	Repository structure	hace 1 año
Meetings	6e99758489	D6.2 telco minutes	hace 6 meses
Reviews	161a6d1cb4	Prześlij pliki do 'Reviews/First Review'	hace 4 meses
Submitted Deliverables	5994141be2	Subir archivos a 'Submitted Deliverables'	hace 1 semana
Templates	4517d1cc5f	Usuj 'Templates/ASSISTANCE_PPT_internal_template_F.ppt'	hace 11 meses
WPs	b58363a662	Upload files to 'WPs/WP1/T1.4'	hace 1 mes
ASSISTANCE_Communica...	38c1832e77	Prześlij pliki do ''	hace 1 mes

Figure 4 – ASSISTANCE repository overview

As shown in the snapshots below, folders are organized in a hierarchical and clear structure and files are uniquely identifiable and versioned by using a name convention.

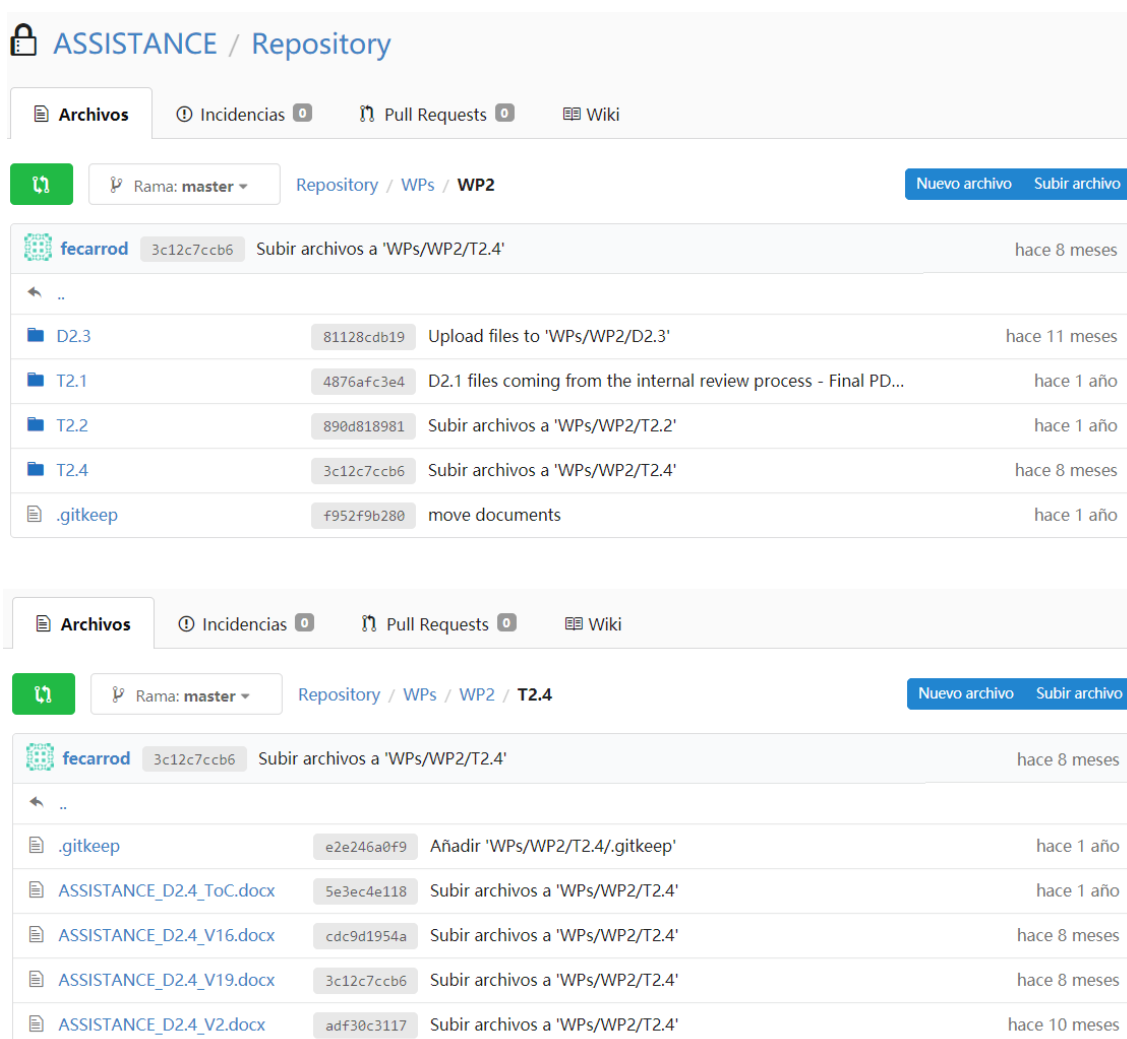


Figure 5 – Example of folders and files in ASSISTANCE repository

3.2.4. Public deliverables

The project public deliverables will be available for download on the website and even in Zenodo after their submission and approval from the EC. Confidential deliverables will be stored in the ASSISTANCE private repository only accessible for restricted users. Confidential deliverables might be requested by external parties, in which case the Consortium might make decision to disseminate corresponding deliverables or specific parts of the deliverables to particular external parties. The project deliverables on the website will be provided in widely adopted PDF format.

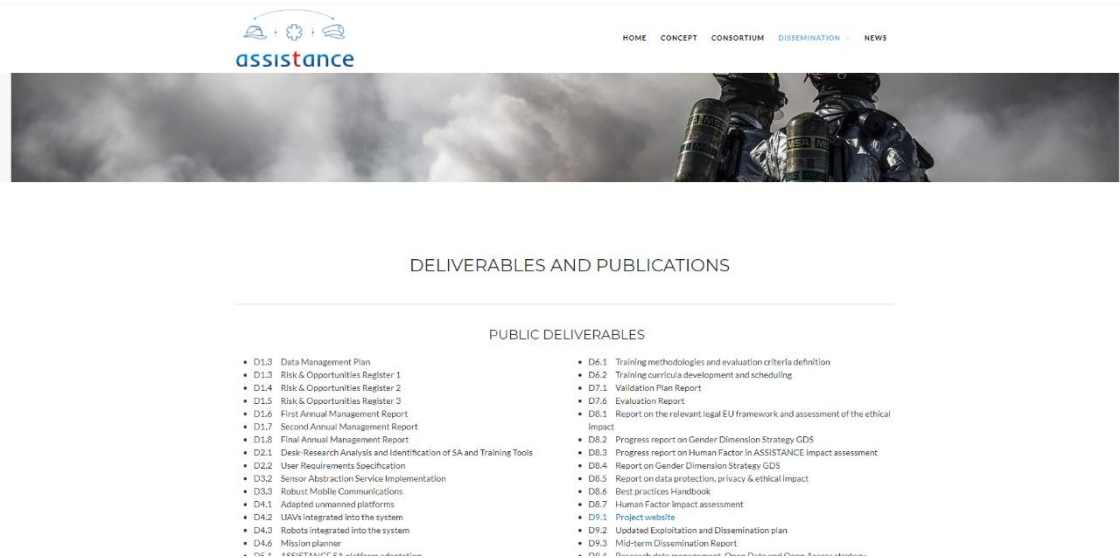


Figure 6 - Public deliverables in ASSISTANCE Website

3.3. Making data interoperable

Information and data models as well as the specification of system interfaces have been defined in WP3 to get overall interoperability. This interoperability is ensured under three main aspects:

- Communications and data exchange among the actors identified in the ASSISTANCE architecture
- Communications and data exchange between the services proposed by the project and current infrastructure
- Communications and data exchange between the services proposed by the project and new applications.

The use of open and standard interface definition languages allows getting the easiest and most diffused adoption by all parties. The identification of the most relevant standards is key to achieve interoperability: ASSISTANCE has analysed the relevant protocols to be used accordingly, being the following ones adopted in the implementations:

- MQTT, which is one of the most recognised standards for IoT messaging [14].
- NATS, enables the exchange of data in messages among applications and services through a simple, secure and high performance open source messaging system [15].
- DDP, as a protocol between a client and the server to handle Remote Procedure Calls (RCP) and to manage data in real time [16].

Complete information about standards and specifications allowing interoperability can be found in the following deliverables: D3.1 “Sensor Abstraction Service Adapted Interfaces Definition” and D2.4 “ASSISTANCE System and network Architecture Design”.

3.4. Increase data re-use

As indicated in section 3.2, data will be treated on a case study basis during the project. Once a data set is marked as public, and, therefore, can be made publicly available on Zenodo, it will be fully reusable (with the possibility of specifying embargo period or with controlled access to whitelist of persons; see Zenodo policies [11]).

When possible, and as recommended by European Commission’s guidelines [12], data will be made available with Creative Commons Licences (CC BY or CC0). The data sets may be given different licenses according to their specificities⁵.

The Zenodo repository ensures sustainable archiving of the final research data. Items 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 at least the next 20 years. All publicly available uploads on Zenodo will be stored safely for the future in the same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software INVENIO, which is used by some of the world's largest repositories such as INSPIRE HEP and CERN Document Server.

The data will remain re-usable at least until Zenodo discontinues the dataset(s) (i.e. warranted for a minimum of 20 years).

4. Allocation of resources

As this deliverable is currently based on the use of free resources and open source software, the only costs that will be incurred are related to the server(s) (hardware) required to run them and the working time needed to setup, maintain and evolve the different tools (efforts measured by person-months). In addition, a budget 10.000 € is allocated to the dissemination activities leader (PIAP) budget under other goods and services in order to cover the open access publications costs.

⁵ The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data (<https://eudat.eu/services/userdoc/license-selector>)

4.1. Responsibilities and decision making

As indicated in previous sections, the Data Management, Open Data and Open Access strategy presented in this deliverable is just the first iteration-after being presented D1.2 DMP, and the related Consortium discussions will be continuously carried out, to identify the relevant project outputs as well as to decide on way and means of their open access (if applicable). To ensure it, a dedicated time slot will be reserved at each project plenary meetings and, if needed, at selected Consortium audio conferences. EC and project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Project Coordinator – to prepare and lead related discussions at the relevant project meetings and to maintain the project document repository GitLab;
- Scientific and Technical Project Manager – to identify data collected by the project and technical project outcomes eventually suitable for publication;
- Dissemination (WP9) Leader – to identify publications suitable for publication in the considered repositories and maintain ASSISTANCE inputs for the Open Access;
- Each individual partner – to identify own project results suitable for publication and to share the published scientific articles in advance with project coordinator and dissemination manager.

Moreover, each ASSISTANCE partner has to respect the policies set out in this deliverable. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

The Project Coordinator and the Dissemination Manager have a particular responsibility to ensure that data shared through the ASSISTANCE website are easily available, but also that backups are performed and that proprietary data are secured.

The Project Steering Committee, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP. Metadata constitutes an underlying definition or description of the datasets and facilitate finding and working with particular instances of data.

Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.

Quality control of these data is the responsibility of the relevant WP Leader, supported by the Project Coordinator.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

5. Data security

The GitLab and Zenodo repositories will ensure secure and safe storage of both public and non-public data.

Zenodo provides clear security guaranties. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files and metadata are backed up on a nightly basis. Files are regularly checked against their checksums (using MD5 algorithm) to assure that file content remains constant. In case of closure of the repository, Zenodo ensures that efforts will be made to integrate all content into suitable alternatives [11].

GitLab is hosted on a private internal server with local backup mechanism (managed by the project coordinator UPVLC).

The servers hosting the research data will be accessible only by authorized system administrators. Files containing confidential data should be protected by owners using local encryption tools (i.e. password-protected archives) before being uploaded to shared repositories. Interaction through web user interfaces will use https protocol (i.e. secure). Also, a secure file transfer protocol (sftp) will be provided as the need arises.

To assure data privacy, all data will be anonymised, encrypted and stored on servers to which only the relevant staff have access. More specifically the servers onto which the data will be stored will have server side encryption. This means that the server's administration personnel will be able to generate public keys for specific personnel who will have access to the data but will not be able to access the data themselves (since the private keys required for this access will be generated on the machine of the person with access to the data). This means that only the required personnel will have access to the data and, even in the remote case of a possible data leak or server hack, the data stolen will be fully encrypted and thus virtually fully non-accessible.

Finally, and after a retention period (to be defined), a secure deletion software will be used to destroy data, i.e. using Gutman algorithm (35-pass overwrite technique).

If deemed necessary, a full format can be used in conjunction with overwriting, to provide further assurance that data cannot be recovered, guaranteeing the destruction of the project personal data.

The following guidelines will be used in order to ensure the security of the data:

- use anonymised and aggregated data instead of individual data;
- encrypt data by the local researchers;
- store data in at least two separate locations to avoid loss of data;
- limit the use of USB flash drives;
- label files in a systematically structured way in order to ensure the coherence of the final dataset.

6. Ethical aspects

As described in D10.1 and D1.2, no sensitive information personal data (e.g. health, sexual, lifestyle, ethnicity, political opinion, religious, or philosophical conviction) is envisaged to be handled by the project. ASSISTANCE will prepare an information sheet and an informed consent form for the collection of data during the pilot phase (name/surname of the participants, organization type and images of the participants to the pilot scenarios).

The main ethical aspects to be considered within ASSISTANCE regards to the participation of non-member states in ASSISTANCE. The partners in the project – independently on their nationality - have confirmed the commitment to comply with: (a) ethical principles (including the highest standards of research integrity) and (b) applicable international, EU and national law.

7. References

- [1] "Guidelines on FAIR Data Management in Horizon 2020," [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- [2] European Commission, "TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)," [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx.
- [3] European Commission, "H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020," [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2.
- [4] CORDIS, "AF3 - Advanced Forest Fire Fighting," [Online]. Available: <https://cordis.europa.eu/project/id/607276>.
- [5] CORDIS, "CAMELOT - C2 Advanced Multi-domain Environment and Live Observation Technologies," [Online]. Available: <https://cordis.europa.eu/project/id/740736>.
- [6] University of Leicester, "What is Research Data Management," [Online]. Available: <https://www2.le.ac.uk/services/research-data/old-2019-12-11/rdm/what-is-rdm>. [Accessed 2020].
- [7] GO-FAIR, "FAIR principles," [Online]. Available: <https://www.go-fair.org/fair-principles/>.
- [8] K. Russell, "FAIR data principles and Data Management Plans," 31 October 2017. [Online]. Available: <https://www.slideshare.net/AustralianNationalDataService/fair-data-principles-and-data-management-plans-31-oct-2017>.
- [9] European Data Portal, "Europe moves towards a common metadata language," [Online]. Available: <https://www.europeandataportal.eu/en/news/europe-moves-towards-common-metadata-language>.

D9.4: Research data management, Open Data and Open Access strategy

- [10] “Dublin Core Metadata Initiative,” [Online]. Available: <http://dublincore.org/metadata-basics/>.
- [11] DublinCore Metadata Element Set , [Online]. Available: <http://dublincore.org/documents/dces/>.
- [12] Zenodo Data Repository, [Online]. Available: <https://zenodo.org/>.
- [13] “Fact sheet open access to publications and data in Horizon 2020 - Fact sheet,” [Online]. Available: https://www.iprhelphdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf.
- [14] “MQTT,” [Online]. Available: <https://mqtt.org/>.
- [15] “NATS,” [Online]. Available: <https://nats.io/>.
- [16] GitHub, “DDP,” [Online]. Available: <https://github.com/meteor/meteor/blob/devel/packages/ddp/DDP.md>.
- [17] Zenodo, “Zenodo policies,” [Online]. Available: <https://www.zenodo.org/policies>.