Project Title	Artificial Intelligence in Secure PRIvacy-preserving computing coNTinuum
Project Acronym	AI-SPRINT
Project Number	101016577
Type of project	RIA - Research and Innovation action
Topics	ICT-40-2020 - Cloud Computing: towards a smart cloud computing continuum (RIA)
Starting date of Project	01 January 2021
Duration of the project	36 months
Website	www.ai-sprint-project.eu/

D7.3 - Data Management Plan

Work Package	WP7 | Project Management
Task	T7.5 |Data and Ethics Management
Lead author	Politecnico di Milano (POLIMI)
Contributors	All Partners
Peer reviewers	C. Pepato (IDC), R. Badia (BSC), S. Parker (TRUST-IT)
Version	V1.0
Due Date	30/06/2021
Submission Date	30/06/2021

Dissemination Level

X	PU: Public
	CO: Confidential, only for members of the consortium (including the Commission)
	EU-RES. Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON. Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC. Classified Information: SECRET UE (Commission Decision 2005/444/EC)

AI-SPRINT - Artificial Intelligence in Secure PRIvacy-preserving computing coNTinuum, has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 101016577.

Versioning History

Revision	Date	Editors	Comments
0.1	28/04/2021	POLIMI/FPM	DMP structure and table of contents, including tables for collecting research data info from the partners and examples to fill in the tables
0.1	21/05/2021	Cristina Pepato (IDC), Rosa Badia (BSC)	D7.3 v0.1 revision
0.2	24/05/2021	POLIMI/FPM	Updating D7.3 based on revision; adding Institutional Requirements section
0.2.1	10/06/2021	Giulio Fontana (POLIMI)	Filled in POLIMI parts of sections 2.2, 2.3, 2.9
0.2.2	14/06/2021	Matteo Matteucci, Danilo Ardagna	POLIMI remaining sections finalised
0.2.2	23/06/2021	Andrei Popa	BECK's input in section 2.1, 2.2, 2.4, 2.6, 2.7, 2.8, 2.9, 2.10
0.3	24/06/2021	POLIMI/FPM	Finalisation of DMP draft for internal revision before submission
0.3	29/06/2021	Rosa Badia (BSC)	D7.3 v0.3 revision
0.3	29/06/2021	Stephanie Parker (TRUST)	Internal review
1.0	30/06/2021	POLIMI/FPM	Editing DMP based on the reviewers' feedback; finalisation of D7.3 v1.0

Glossary of terms

Item	Description
DMP	Data Management Plan
FAIR (data)	Findable, Accessible, Interoperable, and Re-usable data
GDPR	General Data Protection Regulation
WP	Work Package

Keywords

Data Management Plan; FAIR data

Disclaimer

Executive Summary

This document is “D7.3 Data Management Plan” (DMP) of the H2020 AI-SPRINT Project aims to set the life-cycle management plan for handling research data that will be collected, generated, and/or processed within the project according to the “FAIR data” and “as open as possible, as closed as necessary” principles.

In chapter 1, the overall aim and scope of the DMP are clarified, as well as its audience, methodology, revision process, and structure. In chapter 2, all issues associated with research data management in the context of the AI-SPRINT project are addressed by following the *Guidelines on FAIR Data Management in Horizon 2020* (version 3.0; 26 July 2016) as a general framework. This starts with a data summary and covers the FAIR principles of findability, accessibility, interoperability, reusability, as well as the allocation of resources, data security and ethical issues.

The current version 1.0 is the initial DMP, due by month 6 (June 2021) according to the H2020 Grant Agreement. It will be constantly updated during project implementation whenever relevant changes occur and, in any case, for the Interim and Final Reports due by month 18 (June 2022) and 36 (December 2023) respectively.

Table of Contents

1. Introduction

The Data Management Plan (DMP) is a document describing the management of the research data to be collected, generated, and/or processed in the context of a Horizon 2020 project. Sound research data management is instrumental to ensuring that research data are handled according to the FAIR data principle, i.e. research data are made Findable, Accessible, Interoperable, and Reusable so as to foster research data exploitation and further knowledge development and innovation.

Some research data, however, may also be kept undisclosed for well-justified reasons, e.g. protection of personal data according to the General Data Protection Regulation (GDPR), future industrial exploitation, etc. In these cases, the DMP should clearly point out any limitation to research data accessibility in line with the "as open as possible, as closed as necessary" principle, provided that research data needed to validate scientific publications should be fully disclosed.

As clarified in the *Guidelines on FAIR Data Management in Horizon 2020* (European Commission, 2016), which is used as a framework to address all issues associated with research data management in the context of the AI-SPRINT project, the DMP should therefore include the following information:

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

The DMP is addressed to the AI-SPRINT consortium, the European Commission and its Officers, and any stakeholder (e.g. technology suppliers, the research community, etc.) who will interact with the project on several levels and whose data will be collected by the project partners.

For the preparation of the DMP, the following methodology was used by the consortium:

- The objectives and the general structure of the DMP were illustrated and shared with partners;
- A template for the collection of the partners' feedback was defined;
- Each partner was invited to contribute to the DMP and clarify their involvement in research data management through the guided template;
- Partners' contributions were integrated and revised.

The current version 1.0 is the result of this process and the initial AI-SPRINT Project DMP, due by month 6 (June 2021) according to the H2020 Grant Agreement. However, in this early stage of the project, some issues associated with research data management are still not fully clear and/or hard to foresee. As suggested in the above mentioned H2020 Guidelines on FAIR data management, the DMP will therefore be a "living document", i.e. it will be constantly updated as soon as relevant changes in the project occur (e.g. new data, changes in the consortium policies and/or composition, etc.) and according to the revision timeline detailed below (also to be kept updated) with a finer level of detail and accuracy. In any case, a revision process will take place at least on the occasion of the Interim and Final Reports due by month 18 (June 2022) and 36 (December 2023) respectively.

Table 1 - Timeline for D7.3 DMP revision

version	month	date	milestone
1.0	M6	30 June 2021	Initial DMP
2.0	M18	30 June 2022	Interim Report
3.0	M36	31 December 2023	Final Report

In line with the H2020 DMP Guidelines, for each identified data set the DMP provides a description of the following elements:

- Institutional requirements, clarifying if the partner organisations in the consortium already have procedures and tools relevant with reference to research data management;
- Data summary, describing the types of data, the origin of the data (generated internally or collected from external sources), how they fit into the project (where they are produced or collected and what contribution is provided to project objectives);
- Details on how to make data findable, including provisions for metadata;
- Details on how to access data and how data will be findable and accessible;
- Details on how to make data interoperable;
- Policies to support re-use and sharing of data;
- Resources necessary to support the collection and maintenance of data;
- Policies to guarantee secure management of data;
- Ethical aspects and other issues.

In chapter 2, each section is dedicated to one of the research-data-related aspect listed above. Information on each aspect is provided by each partner.

2. Data Management Plan

2.1 Institutional Requirements

As the background for the AI-SPRINT DMP, this section aims at providing information about existing policies within the consortium partner organisations with reference to the broader context of Open Access to Scientific Publication and Research Data. Overall, the majority of the partners do not have specific policies on the matter in place. However, some organisations, particularly Higher Education Institutions (POLIMI, TUD, UPV) and research centres (BSC) already have measures in place to address it. Information for each partner is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - Is there any Open Access Policy for publications in your organisation? If yes, please clarify. - Does your organisation require using a specific Open Research Data repository?
1. POLIMI	- Accepted preprints published on RE.PUBLIC@POLIMI repository. - Zenodo
2. BSC	The BSC Open Access policy requires that an electronic copy of the accepted version (either author final manuscript or publisher version) of all peer-reviewed articles, books, conference proceedings etc. is deposited in the institutional repository UPCommons (https://upcommons.upc.edu/). The full text (not just the abstract) of all publications must be made openly available at the time of deposit or as soon as possible thereafter. In the case of publications that cannot be made immediately openly available because of publisher restrictions, the deposit remains mandatory, but the access will be set to closed until publisher embargo elapses. As for Open Research Data repository, BSC recommends the use of Zenodo.
3. TUD	- TUD has created an Open Access infrastructure as a repository for scientific publications. The infrastructure and open access service is hosted and provided by the SLUB, the Universities and State's library. - The high performance computing center (ZIH) provides a repository to host scientific data that is linked to publications and research projects.
4. UPV	- The UPV has an institutional repository for publications, supporting the deposit of Open Access articles and license-free preprints (RiuNet). The UPV also started providing guidelines for the storage of research data. - The I3M (Institute of Instrumentation for Molecular Imaging), which is the unit in charge of the work in AI-SPRINT, is a joint centre from CSIC-UPV. The I3M also uses DIGITAL-CSIC, the institutional open access repository of the CSIC, which also supports the deposit of research data.
5. GREG	GREG does not have any Open Access policies nor any requirements regarding the use of specific repositories.
6. BECK	BECK does not have any Open Access policies nor any requirements regarding the use of specific repositories.
7. C&H	C&H does not have any Open Access policies nor any requirements regarding the use of specific repositories.

8. 7BULLS	7Bulls does not have any Open Access policies nor any requirements regarding the use of specific repositories.
9. AF	AF does not have any Open Access policies nor any requirements regarding the use of specific repositories.
10. TRUST	Trust-IT does not have any Open Access policies nor any requirements regarding the use of specific repositories.
11. IDC	IDC does not have any Open Access policies nor any requirements regarding the use of specific repositories.

2.2 Data Summary

The main aim of the AI-SPRINT project is to develop a platform composed of design and runtime management tools to seamlessly design, partition, and operate Artificial Intelligence (AI) applications among the variety of cloud-based solutions and AI-based sensor devices (i.e. devices with intelligence and data processing capabilities) in order to provide resource efficiency, performance, data privacy, and security guarantees. For this reason, the main data that will be collected and generated in the AI-SPRINT project are related to the development of the platform which will be implemented through Work Packages 2: Design Time Tools, WP3: Runtime Environment, and WP4: Security Tools and validated within WP5 through their application in 3 use cases, namely a) Farming 4.0, b) Maintenance and Inspection, and c) Personalised Healthcare.

In this section, issues about the type, format, origin, and size of the data are addressed as well as on aspects of data re-use and utility. Detailed information on the different data is provided by each partner in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - What is the purpose of the data collection/generation and its relation to the objectives of the project? - What types and formats of data will the project generate/collect? - Will you re-use any existing data and how? - What is the origin of the data? - What is the expected size of the data? - To whom might it be useful ('data utility')?
1. POLIMI	About POLIMI, data collection and generation will take place in the context of Work Package 5 (Validation on Use Cases), 2, 3, and 4. In particular, POLIMI is directly involved in the collection/generation of data concerning the Farming 4.0 use case, where POLIMI's expertise in AI and Robotics will be leveraged for two consecutive data-relevant operations: (i) to set up the sensors and data collection systems installed on the spraying machine provided by partner GREG and to use them to capture datasets; (ii) to analyse and process collected datasets to build the AI modules that will subsequently be tested as governors of the machine's spraying system and the yeld monitoring and prediction module. For what concerns operation (i), data consists of raw sensor streams, i.e. data produced by the sensors onboard the machine (such as RGB, RGBD and NIR cameras, LiDAR, GPS), possibly integrated online or subsequently with data from external sources (such as weather data or data from sensors deployed in the field). For what concerns these datasets, the plan is to make them available not only to AI-SPRINT partners, but also to external actors for scientific work. If the data will allow this (something that will be known only after its collection), POLIMI will extract from them as well-formatted and FAIR-compliant datasets to be shared with the scientific community. However, some of the datasets/streams may include information that - if published - can be damaging to GREG's intellectual property and technology assets. Therefore, the data will be subjected to a careful analysis together with partner GREG before disclosure. For what concerns the outcomes of operation (ii), i.e. the experimental AI modules that WP5 will build and install on GREG's machine and use to govern spraying, there are no plans to make them public (except possibly in simplified form). These models, in fact, are an exploitable asset and one of the results of the project having potential industrial

	relevance. If such relevance will be proved, exploitation will take place according to the provisions of the Consortium Agreement. For the yield prediction and estimation task, public images datasets will be used to train AI models. These data are already public or they are already distributed according to a licence thus the requirements of such licence will be enforced. Additional data shall be used from field campaigns performed as part of external collaborations of POLIMI with research centers which are not part of the AI-SPRINT consortium; in this case, data will be used under permission of the institutions which have collected the data, and it will not be redistributed unless there is an agreement among all the data owners. In WP2, 3, and 4, POLIMI will generate data related to the performance metrics of the applications under study. If the applications are open benchmarks, those data will be made available to the scientific community. If the data are related to the other use cases (managed by the partners AF and BSC), the data will be made publicly available only under the permission of the partners involved.
2. BSC	In the context of the Personalized Healthcare use case in WP5, BSC will use anonymized personal data of human subjects, namely stroke sufferers and healthy individuals, who will be monitored during an observation time. The data consist of lifestyle information based on a questionnaire, biochemical measurements (blood tests), if available, and digital data from the sensors of wearable devices. The stroke foundation "Freno al Ictus" will collect and anonymize the personal information of the recruited subjects. To comply with privacy preservation, a federated learning approach will be implemented so that no personal data will be stored or shared but only the parameters of the local models at the edge and the global model in the cloud are collected. Data utility of the Personalized Healthcare use case resides in the implementation of a smart and secure automated system to deliver personalized assessment of individual health aiming at increasing the quality of stroke healthcare as well serving as a demonstrator for the AI-SPRINT technology. For the operation of programming and runtime tools, data consist of the input and output data of the application components implemented using such tools and adopted by the use cases. These data will not be made public. Performance data will be produced in order to analyse the behaviour of the applications on the resources and will be processed internally by the BSC team to provide the proper adaptation to the code. Annotated code and the execution graph of the applications will be shared with POLIMI to be processed to generate performance models and resource allocation.
3. TUD	TUD's Palaemon service will collect data about application launches, attestation procedures as well as secure connections established between processes that make use of trusted execution environments. The data are mainly log data that contain solely IP and mac addresses of the different computers the processes are running. The data are used to debug errors in the system as well as to detect unauthorized accesses. Certain parts of the data will also include personal information such as the names of people and IP addresses in case they are part of the so-called policy board which provides the possibility to modify security policies. These data are confidentiality protected and are associated with so called policies that will be automatically purged after removal/turn down of a certain service. Hence, personal information (identifiers, user credentials, and e-mail) is solely used for operation purposes of Palaemon and will not be included in the research data. The data will be neither shared with partners nor third-party organisations. The size of the log data spans tens of MBs.
4. UPV	UPV will collect data from the operation of the infrastructure services. We will consider in this document two sources: (i) virtual infrastructure data provided by Infrastructure Manager (IM - https://imdocs.readthedocs.io/en/latest/) and (ii) horizontal elasticity data provided by CLUES. IM data will consist of plain text logs and semi-structured files that

	gather information about the infrastructures deployed, including the virtual infrastructure topology in TOSCA, the logs from the operation (including date and time of each operation and infrastructure deployment id). The data about the power-on and off of the cloud resources provided by CLUES (https://github.com/grycap/clues/blob/master/docs/reports.md) include information about the changes of state of the back-end resources. The interest of this data is mainly for internal operational purposes and eventually to support the development of performance models. The data will be kept to guarantee the traceability of the operations and to provide support in case of errors. Information about the power-on and power-off and aggregated information of the deployments (such as number of them, size of the deployments, type of resources, number of invocations) will be kept for statistical purposes. UPV will not use any pre-existing data and will not release personal data to third parties. Personal information (identifiers, user credentials, and e-mail) will be used for operation purposes only and will not be included in the research data. The amount of data to be kept is in the order of a few hundreds of Megabytes, and it will be useful to application and system developers in the consortium who could need it to understand issues, tune up configuration, evaluate scalability and other similar performance evaluations. The aggregated data will be released to any partner in the consortium that will require it. Non-aggregated data will not be disclosed and will be deleted timely.
5. GREG	For what concerns GREG, data collection and generation will take place in the context of WP5 Farming 4.0 use case with POLIMI and BECK, for two consecutive data-relevant operations: (i) to set up the sensors and data collection systems installed on the spraying machine provided by GREG and to use them to capture datasets; (ii) to analyse and process collected datasets to build the AI modules that will subsequently be tested as governors of the machine's spraying system and the yeld monitoring and prediction module. For what concerns operation (i), data consists of raw sensor streams, i.e., data produced by the sensors onboard the machine (such as RGB, RGBD and NIR cameras, LiDAR, GPS), possibly integrated online or subsequently with data from external sources (such as weather data or data from sensors deployed in the field). For what concerns these datasets, the plan is to make them available not only to AI-SPRINT partners, but also to external actors for scientific work. If the data will allow this (something that will be known only after its collection) POLIMI will extract from them well-formatted and FAIR-compliant datasets to be shared with the scientific community. However, some of the datasets/streams may include information that - if published - can be damaging to GREG's intellectual property and technology assets. Therefore, the data will be subjected to a careful analysis together with partner GREG before disclosure. For what concerns the outcomes of operation (ii), i.e., the experimental AI modules that WP5 will build and install on GREG machine and use to govern spraying, there are no plans to make them public (except possibly in simplified form). These models, in fact, are an exploitable asset and one of the results of the project having potential industrial relevance. If such relevance will be proved, exploitation will take place according to the provisions of the Consortium Agreement. For the yield prediction and estimation task, public images datasets will be used to train AI models. These data are already public or they are already distributed according to a licence thus the requirements of such licence will be enforced. Additional data shall be used from field campaigns performed as part of external collaborations with research centers which are not part of the AI-SPRINT consortium; in this case, data will be used under permission of the institutions which have collected the data, and it will not be redistributed unless there is an agreement among all the data owners.

6. BECK	BECK will collect data related to Farming 4.0 use-case in collaboration with POLIMI and GREG. Storing of this data will be only on provided and agreed systems with all the security controls imposed by the project coordinator. Data will be used during the implementation and testing of the use case solution. The types of data collected and stored include but might not be limited to: - Sensor data: images from RGB, RGBD and NIR cameras, LiDAR, GPS Location, other sensors and nozzle data, environmental metrics (i.e. weather conditions); - Application and infrastructure monitoring data: events, logs and metrics regarding availability, performance, and functionality of the in-scope solution; - Specific use-case information, i.e. yield estimates, disease infestation, spraying substance usage. A public set of images will be used initially for training the AI-model. Re-training of the model might use images collected during the project. Data will be used to achieve the Use-case main goals which are: - Estimate foliage and adjust for the most effective and optimized spraying; - Estimate yield in different season phases; - Identify potential disease in the vineyard. Collection and usage of the data will be used mostly during the WP3, 4, and 5. All data collected and used by BECK will remain internal and will not be made public without all partners agreement and approval.
7. C&H	The data which will be collected and stored during the project on Cloud&Heat side are mostly (i) log data which will be used for development purposes, (ii) user account data which will be used for authorization purposes of users in Openstack accounts and (iii) Openstack instance metadata and temporary config data which is used to set up Krake. In the future, however, there may be a new feature for Krake called "monitoring," where metrics data is collected on-site and stored permanently. The data will mostly be in JSON format or at least in plain text. A re-use of the data is not planned at the moment, but especially log data could be used for development and debugging purposes. The log-data which is generated on the Cloud&Heat servers can contain IP-addresses. Due to the textual nature and short form in general, the data generated will be quite small, perhaps a few kb and at most a few mb, more is unlikely. The Openstack authorization data and Krake configuration files will most likely only be used by the AI-SPRINT project partners, but the log data could also be used by the Cloud&Heat development teams. The authorization data will only include names and email addresses and will be saved in a MariaDB. The temporary Krake resource data are saved in an etcd database. Certificates, used for authentication processes of connections in Krake, are stored client-side. For logging purposes of server instances, Cloud&Heat plans to use the ELK stack, an elastic search service. Therefore, log data would be gathered with Logstash.
8. 7BULLS	7BULLS plans to collect time series data which will describe the state (performance, health status, resource usage) of monitored systems. Monitoring sub-system will also generate, manage, and provide credentials used to connect, query and store data. Collected data (metrics) will be used by other systems which will be responsible for resource management, configuration adjustments, or performing additional tasks in case of detected anomalies. Metrics will contain timestamp (number in nanosecond precision), value (number or short text), and set of labels (text key-value). Credentials will be provided either as tokens (text) used by monitored systems to connect or as login-password pairs used by administrators.

	We plan to use data provided by monitored systems. The data will be used to generate various statistics and to control if the whole system works properly. We expect data volume to be huge: 100s of gigabytes. If monitored systems will use local metric data buffering, they should also provide space to store temporary monitoring data (10s gigabytes).
9. AF	AF will collect a dataset of JPG images of wind turbine blades in the context of WP5 Maintenance & Inspection use case validation. The main purpose of data collection is building an automated inspection pipeline. Images will be taken from a UAV and saved together with metadata, i.e. GPS coordinates of UAV and camera specification. The data will be collected from multiple turbines and multiple wind farms located in diverse environments. Blade damages will be marked and used to train object detection and classification models. In some of the experiments, models pre-trained on existing computer vision datasets (e.g. imagenet, pascal voc) will be used. Annotations will contain information about position, shape, type, and severity of the damage and will be stored as a set of JSON files. The amount of data to be kept is in the order of a few hundreds of Gigabytes. The dataset will not be made public nor shared with partners. The only data that can be shared with partners are the results of the experiments in form of tensorboard logs and .csv files, and a limited sample of the dataset in size of a few Gigabytes.
10. TRUST	Processing of personal data for communication, dissemination and stakeholder engagement activities is performed by Trust-IT that has full responsibility regarding the purpose and systems used for such activity. In this context, Trust-IT takes the GDPR role of "Data Controller". The origin of the data is linked to the activities on the website by stakeholders (registration, newsletter subscriptions, event registration, etc.) and through direct interaction with stakeholders at physical events. Full list of activities and details about the personal data collected can be found in the AI-SPRINT Privacy Policy publicly available on the project website: https://ai-sprint-project.eu/privacy-statement. Personal data are mainly collected through a web form on the project website. Data are also collected during direct contacts with the data subjects (in physical meetings or events). Other personal data are collected through cookies and analysed in an anonymous way in order to generate statistics about the website usage. Personal data, logs, credentials, and navigation data are available to Trust-IT employees involved in the AI-SPRINT project only. Personal data, logs, credentials, and navigation data are not transferred to third parties. The expected size of the data collected is in the order of a few Gigabytes.
11. IDC	IDC is a market intelligence company. IDC collects data through primary research regarding all market sectors and technology. The data are collected through more than 400,000/yr focused interviews and surveys across the globe. All data collection is anonymously and no data about individuals is collected or stored. Contact information about respondents is managed by external companies (not IDC) that ensure proper data management. In the scope of the project, no data will be collected regarding project activities or participants. On the other hand, IDC data will provide information regarding the exploitation potential of the technologies in comparison to real market offerings, products, and market outlook. All data will be available to project partners for their own exploitation planning and marketing plans and activities. All data will be made available to the project for the scope of the project and in all cases they will include no information about individuals.

2.3 Data findability

In this section, issues of making research data findable are addressed. Aspects related to the production of metadata and their standard, naming conventions, file versioning, etc. are detailed for each partner organisation in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? - What naming conventions do you follow? - Will search keywords be provided that optimize possibilities for re-use? - Do you provide clear version numbers? - What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
1. POLIMI	As explained in the POLIMI subsection of Section 2.2 above, the data that POLIMI plans to share take the form of structured datasets comprising multiple and time-synchronized sensor data streams. The actual structure and contents of the datasets will depend on the quality and features of the actual collected sensor data, which can only be evaluated after collection; so it is not possible to provide precise information in the early stage of the project. Collection activities will benefit from internal iteration governed by such evaluation, in order to maximise final data quality. POLIMI will leverage its long-standing experience in designing, building, and publishing high-quality multisensor datasets (starting from FP6 project RAWSEEDS) to ensure the quality of the final product. The datasets will have the form of archives composed of multiple files (one per stream or, for some types of sensors such as cameras, possibly one per data item: e.g. frame) and one or more TOC documents explaining their internal organisation and formatting. Suitable metadata will describe the features of the sensors and devices used to collect the data. Storage will rely on Zenodo, while dataset findability will be ensured by linking to them from Politecnico di Milano's web pages, by publicizing their existence and location on AI and Robotics and Cloud mailing lists, and by providing links to the datasets in research papers and dissemination material.
2. BSC	BSC plans to establish a naming convention with clear version numbers that will be applied to the different file sets that will be generated. For the purpose of the Personalized Healthcare use case in WP5, we do not expect to require search keywords associated with the generated metadata. The types of metadata that will be created are related to the characterization of the files generated during the different stages of the use case execution and their content, such as the type of data, the timestamps, anonymous sample identifiers, etc. Log data produced by PyCOMPSs and dislib tools is stored in plain text files using log4j format. Configuration files are written in xml format
3. TUD	The log data produced by SCONE and Palaemon are stored in plain text and follow the standard log4j/log4rust logging format. The information does not allow to identify individuals as only internal object identifiers are used. The log data file names will include the timestamp of the time the data were produced.

4. UPV	Data produced will be named using a structured format related to the experiments. The experiments will be defined in plain text. The IM data will follow the syslog format. There is no standard metadata for such format, so in our case, the metadata of the file will include a JSON file with the name and purpose of the experiment, the date, and time. The CLUES data will be coded in JSON and made available also as JavaScript reports containing the sequence of events. A similar metadata file will be generated. Experiments will be unique, so no versioning will be provided.
5. GREG	As for POLIMI, the datasets will have the form of archives composed of multiple files (one per stream or, for some types of sensors such as cameras, possibly one per data item: e.g. frame) and one or more TOC documents explaining their internal organisation and formatting. Suitable metadata will describe the features of the sensors and devices used to collect the data. Storage will rely on Zenodo, while dataset findability will be ensured by linking to them from Politecnico di Milano's web pages, by publicizing their existence and location on AI and Robotics and Cloud mailing lists, and by providing links to the datasets in research papers and dissemination material.
6. BECK	BECK will collect data together with GREG and POLIMI as part of the farming 4.0 use case. Data will be mostly time-synchronized data streams (RGB, 3d-camera etc.), in some cases enriched with GPS information. Data structure will be as illustrated above for POLIMI and GREG. BECK will also rely on Zenodo and findability will be assured through links to the POLIMI web page and by providing links in any dissemination material.
7. C&H	For Krake logging, the standard Python logging library is used. If Cloud&Heat implements the ELK stack with Logstash in the near future, server logging would follow the Logstash standard, i.e. the logs would be stored in Elasticsearch. If Elasticsearch will be used, server log-data would be searchable, but for sure not publicly.
8. 7BULLS	Data will be stored in the central clustered instance of the Influx DB database. Each collected metric will contain a timestamp and set of labels (in the form of key-value pairs) which will provide uniqueness and identify the origin of each data entry. We do not plan to provide versioning of collected metrics.
9. AF	Data will be stored on a private server. Images and damage annotations will be stored in separate directories. Filenames of images will represent their sha256 hash codes ([sha256].jpg). Corresponding files containing annotations and information about turbines and wind farms can be matched by sha code in filename ([sha256].json). GPS coordinates and camera details will be stored in Exifs info embedded in JPG files. It will be possible to search the data using JSON annotations.
10. TRUST	Data processed by TRUST-IT are only related to communication, dissemination, and stakeholder engagement project activities and they are not meant to be shared or found publicly. Data will be stored on a private server and will be accessible to TRUST-IT employees involved in the AI-SPRINT project only. Issues of metadata and data findability are therefore not relevant for the data processed by TRUST-IT.
11. IDC	Data is all stored on IDC servers and provided in reports that are normally in .pdf, .doc, and excel formats. The metadata is irrelevant for project uses as it pertains to internal IDC classification and storage.

2.4 Data accessibility

This section aims at clarifying which AI-SPRINT research data are made fully accessible and/or at justifying if any restriction to data accessibility is deemed necessary, including the rationale behind it. It describes the data that are made accessible, the methods or software tools needed to access them, where to find the data, and other relevant issues. Information for each partner organisation is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Please note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out. - How will the data be made accessible (e.g. by deposition in a repository)? - What methods or software tools are needed to access the data? - Is documentation about the software needed to access the data included? - Is it possible to include the relevant software (e.g. in open source code)? - Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories that support open access where possible. - Have you explored appropriate arrangements with the identified repository? - If there are restrictions on use, how will access be provided? - Is there a need for a data access committee? - Are there well-described conditions for access (i.e. a machine-readable license)? - How will the identity of the person accessing the data be ascertained?
1. POLIMI	Data that POLIMI plans to share take the form of structured datasets comprising multiple and time-synchronized sensor data streams. Data will be accessible from Zenodo and will be in the standard formats of compressed archives composed of multiple files (one per stream or, for some types of sensors such as cameras, possibly one per data item, e.g. frame) and one or more TOC documents explaining their internal organisation and formatting. Suitable metadata will describe the features of the sensors and devices used to collect the data. It will be possible to use available tools to access the data and no licenced software will be required. It will be possible to access all data from project partners upon request while, in general, data will be distributed to the public once cleared of their exploitability, i.e. after having published a paper describing their use, their collection procedure, or simply their availability.
2. BSC	The anonymized personal data produced and used in the context of the Personalized Healthcare use case in WP5 will not be made publicly available for privacy and security reasons as it consists of personal information of individuals. To comply with privacy preservation in the model development, a federated learning approach will be employed using dedicated software tools with available documentation. Research findings and software code will be made accessible through certified repositories that support open access such as UP Commons and Zenodo, in agreement with BSC policies on this matter

	(see section 2.1 above). As for this specific material, we do not expect restricted access and the need of a data access committee. Data computed from programming and runtime tools (PyCOMPSs, dislib) will be stored in the organization where the computation takes place. Log data are available in flat text files and do not contain any application specific details. Information and documentation are available in BSC tools web pages.
3. TUD	The log data that SCONE/Palaemon produces will not be made publicly available. The data will only be shared with partners in very rare circumstances where for debugging purposes this information is required.
4. UPV	Given the size and purpose of the data, UPV envisages that the data will be stored within the organization. Log data from the IM is coded into semi-structured text files and any standard log processing software could be used. CLUES data will be exported in JSON format. If necessary, data will be deposited in a repository such as DIGITAL.CSIC. Documentation about the data is available in the pages of the different tools.
5. GREG	As for POLIMI, data will be accessible from Zenodo, a web-based interface and will be in the standard formats of compressed archives composed of multiple files (one per stream or, for some types of sensors such as cameras, possibly one per data item, e.g. frame) and one or more TOC documents explaining their internal organisation and formatting. Suitable metadata will describe the features of the sensors and devices used to collect the data. It will be possible to use available tools to access the data and no licenced software will be required. It will be possible to access all data from project partners upon request while in general data will be distributed to the public once cleared of their exploitability, i.e. after having published a paper describing their use, their collection procedure, or simply their availability.
6. BECK	As mentioned above, for data storage BECK will use the systems and tools provided by partners involved in the Farming 4.0 use case within WP5. From BECK's perspective, data should be accessible to all involved parties in the Farming 4.0 use case that might need it. It will be possible to access all data from project partners upon request while in general data will be distributed to the public once cleared of their exploitability, i.e. after having published a paper describing their use, their collection procedure, or simply their availability.
7. C&H	It is not planned to share log-data. Config data, however, may be shared with project partners to ensure proper Krake deploying. Authentication data for Openstack users will not be publicly shared. This data will only be given to the specific partner who will use Openstack and Krake.
8. 7BULLS	All collected metrics will be openly available as the default. Registered users' data (logins, passwords) or credentials used by monitored sub-systems will not be publicly shared. Metrics will be accessible via REST API or via provided web user interface. REST API is well documented in Influx DB's documentation.
9. AF	The data handled by AF cannot be shared with partners due to formal restrictions. The data is property of AirFusion clients. Only a small portion of the dataset can be shared with partners through cloud storage for exclusive project use only.
10. TRUST	Data related to communication, dissemination, and stakeholder engagement activities in the context of WP6 Dissemination & Impact are not meant to be shared or found publicly. These data are accessible to TRUST-IT staff members involved in the AI-SPRINT project only for privacy reasons, in compliance with the GDPR. They may only be made available in aggregated and anonymized form for reporting and/or publications on the project dissemination activities and impact.
11. IDC	Data produced in the context of the project will be made openly available as needed for exploitation reasons. Data will be made accessible in reports and presentations to be

	archived in the project document repository. No specific tools are needed to access the data. Data are all human readable and not machine readable in most cases.

2.5 Data interoperability

This section aims at describing how the data produced in the project are made interoperable, i.e. how data exchange and re-use between researchers, organisations, countries, etc. is enabled, for instance, by using standard formats and/or data compliance with open software applications. In the table below, information about the data processed by each partner organisation is detailed.

Partner	
	For each partner, please provide in the cells below information on the following: - Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)? - What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable? - Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability? - In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
1. POLIMI	Data streams from sensors and performance data which will be distributed will have an accompanying document describing the format and the content of the data. Libraries for loading and processing each stream of data will be publicly available and all the details required to open and visualize the data will be provided as well. We do not see any interoperability issue on the data as open formats will be used and documented.
2. BSC	We do not expect specific issues of interoperability and/or the need to create specific ontologies or vocabularies.
3. TUD	The log files SCONE/Palaemon produces contain a timestamp and a custom message which are stored in plaintext. Hence, the data does not follow some standard or parsable format except for the timestamp at the beginning of each line.

4. UPV	The log files from IM contain structured data (date / time / service / type of message / message text / infrastructure id) and parsers will be provided. Logs will be stored in plain format to reduce storage footprint.
	The horizontal elasticity from CLUES will be exported as JSON with the following structure:
	`{ maxtime_avail": float number` timestamp of the last measure,
	`"requests": [` collection of LRMS requests `],`
	`"mintime_avail": float number` timestamp of the first measure,
	`"hostevents": {`
	`"nodeDNS": [`
	`{ "memory_used": float number,`
	`"state": integer` from a state list: idle, used, off, powering on, powering off, on (err), off (err), error, unknown,
	`"t": long integer,`
	`"memory": float number,`
	`"slots": float number` VCPUs available,
	`"slots_used": float number` VCPUs used `}]`
	`}`

	}
5. GREG	As for POLIMI, datastreams from sensors and performance data which will be distributed will have an accompanying document describing the format and the content of the data. Libraries for loading and processing each stream of data will be publicly available and all the details required to open and visualize the data will be provided as well. We do not see any interoperability issue on the data as open formats will be used and documented.
6. BECK	As for POLIMI, datastreams from sensors and performance data will be distributed with documentation describing the format and the content of the data. Libraries for loading and processing each stream of data will be publicly available and all the details required to open and visualize the data will be provided as well. We do not see any interoperability issue on the data as open formats will be used and documented.
7. C&H	The data generated by Cloud&Heat is not interoperable and will not be shared. However, there can be issues where it is necessary to share data with AI-SPRINT project partners for debugging purposes. There is no need to specify explicit issues for interoperability.
8. 7BULLS	Collected metrics can be fetched from Influx DB in JSON format. Format of data records is described in Influx DB's documentation. The label vocabulary (key-value pairs) will be defined for each instance of the system.

9. AF	The data is well described with metadata, but due to formal restrictions it will not be shared. Exceptions might apply to some selected subsets of data where data owners provided their consent.
	The structure of JSON metadata will be as follows:
	{
	"lastUpdateTime": string date of last update,
	"meta": {
	"BD": string blade id,
	"BE": string blade side,
	"WFO": string wind farm id,
	"WTG": string wind turbine id
	},
	"object": [
	{
	"bndbox": {
	"xmax": int,
	"xmin": int,
	"ymax": int,
	"ymin": int
	},
	"name": string damage type name,
	"severity": string severity level,
	}
	],
	"sha256": string,
	"size": {
	"depth": int,
	"height": int,
	"width": int
	},
	"version": int
	}.

10. TRUST	Not applicable to TRUST-IT data
11. IDC	Not applicable to IDC data and reports

2.6 Data reusability

This section addresses how the data will be licensed to enable the widest re-use possible during and after the project life cycle, and if any restriction to data reusability is deemed necessary. Information for each partner organisation is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - How will the data be licensed to permit the widest re-use possible? - When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible. - Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why. - How long is it intended that the data remains re-usable? - Are data quality assurance processes described?
1. POLIMI	Data will be released as soon as a publication exploiting their content has been accepted by either a conference or a journal. Non-exploitable data, deemed as interesting for the scientific community, will be released immediately after the accompanying document has been drafted. Data with industrial value will not be released until the end of the project; then, if deemed of public interest and no industrial exploitation process is already started, they will be made publicly available. A Common Creative licence which will impose the citing of the source and will not allow the redistribution without permission will be chosen for all the data which will be made available.
2. BSC	To prevent the risk of re-identification, the anonymized personal data generated in the context of the Personalized Healthcare use case will not be made available for re-use by third parties during the duration of the project and after the end of it.
3. TUD	The only data that will be released will be performance related, such as throughput and latency data. The data will be publicly accessible, however, in plotted/graph form rather than rare data. Due to the size of the data spanning only a few MBs, we intend to keep the data accessible permanently using TUDs/ZIH research data platform.
4. UPV	Data provided will be associated with publications which will describe the experiment conditions in detail. The data will be openly released under creative commons license. The usability of the data by third parties could be used for reproducibility or the development of other models. Considering the size of the data, we foresee keeping it available for a long time.
5. GREG	As for POLIMI, datastreams from sensors and performance data which will be distributed will have an accompanying document describing the format and the content of the data. Libraries for loading and processing each stream of data will be publicly available and all the details required to open and visualize the data will be provided as well. We do not see any interoperability issue on the data as open formats will be used and documented.
6. BECK	Use of data should be freely available to all partners following the access rights regulated by the Cooperation Agreement. We rely on the project's Coordinator to decide when and how the data can be made available to any third party.
7. C&H	Due to the nature of the data (see section 2.2), no specific re-use is planned at the moment. Log data and authentication data for Openstack users will not be made publicly available,

	as it constitutes personal data and contains no information that could be licitly relevant to third parties.
8. 7BULLS	Use of data will be freely available to all partners following the access rights regulated by the Cooperation Agreement.
9. AF	All the data processed by AF is considered confidential. No re-usability outside the project is foreseen. The only foreseen exception is aggregated metrics data on models performance that has been developed and trained in the course of the project. Such data can be published and used as reference for future work.
10. TRUST	Re-usability outside the project is not applicable.
11. IDC	Data can be reused and republished as long as IDC is given credit and appropriately cited.

2.7 Allocation of resources

This section aims at clarifying which are the costs, if any, for making research data FAIR in the AI-SPRINT project, provided that such costs are eligible as part of the Horizon 2020 grant if compliant with the conditions set in the Grant Agreement. In the table below, information for each partner organisation is detailed.

Partner	
	For each partner, please provide in the cells below information on the following: - What are the costs for making data FAIR in your project? - How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). - Who will be responsible for data management in your project? - Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?
1. POLIMI	Cost of open access publications will be covered by the project funds. Cost of the hosting of the data, if any, will be covered by the hosting institution for the duration of the project, until the end of the project. After the end of the project, data will remain available via public open repositories, e.g. Zenodo, or via a partner institution. However, in the latter case, we cannot guarantee hosting data longer than a year, in which an open public repository will be searched to move the data (note that datasets from the FP6 RAWSEEDS project, concluded in 2009, are still hosted today by POLIMI and publicly accessible via ftp).
2. BSC	Costs for open access publications will be covered with the funds of the project. Costs for the maintenance of the institutional repository of the open access publications are covered by UPC (Universitat Politècnica de Catalunya).
3. TUD	Costs of maintaining the data repository are covered by the hosting institution's ZIH and will not be charged on the project. Costs for open access publications will be covered with the funds of the project during the duration of the project.
4. UPV	Costs for making data FAIR will include activities for the production of the metadata, data storage, curation, encoding into machine readable formats, and extensive descriptions of the experiment conditions in the publications for their reusability. The fulfilment of the FAIR principles will be intrinsic to the publication of the scientific results derived from the experiments. Therefore, the cost cannot be easily isolated from other tasks related to the sharing of the data within the project partners. Due to its reduced size, the preservation costs will be small. Costs related to open publications will be covered with the funds of the project during its lifetime and with the budget of the institution beyond the end of the project. The UPV has negotiated a free quota for open publications along with the yearly contracts with the publishers, and we will explore this opportunity.
5. GREG	GREG seems not concerned by these topics.
6. BECK	Costs of maintaining the repository of the data are covered by the hosting institution and will not be charged on the project. Responsible for the maintenance of the infrastructure is the repository owner (Coordinator)

7. C&H	C&H foresees no significant costs associated with data storage and management. The limited data generated (see section 2.2 above) can be stored and managed as part of C&H process with minimal effort.
8. 7BULLS	7bulls seems to not be impacted with this topic.
9. AF	Costs of maintaining the data repository are covered by AF and will not be charged on the project.
10. TRUST	FAIR is not an applicable principle to the data TRUST-IT is managing in the context of AI-SPRINT. However, any eventual costs for managing the data will be covered with the funds of the project.
11. IDC	Not applicable: all data and reports are provided free of charge and maintained on IDC servers with no cost to the project.

2.8 Data security

In this section, provisions for data storage, transfer, and recovery are described, including aspects of long-term preservation and curation. Information for each partner organisation is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? - Is the data safely stored in certified repositories for long term preservation and curation?
1. POLIMI	Data collected by POLIMI will be replicated on physical HDDs stored in different buildings of the institution and on a RAID 6 Network Attached Storage at POLIMI. This is in addition to the publicly accessible dataset (this is the approach followed for RASWEEDS project datasets).
2. BSC	The Personalized Healthcare use case will apply the strategies for transparent encryption of files and network traffic adopted by the AI-SPRINT. All documents related to the personal data of the individuals involved in the pilot study will be stored in-house by the Foundation “Freno al Ictus” and not shared with the consortium or other parties.
3. TUD	Palaemon uses a file protection shield to securely store its log data in case it contains personal data such as the names of the policy board members.
4. UPV	Neither IM’s nor CLUES’s data include personal information.
5. GREG	Data security will be managed by POLIMI
6. BECK	During the project, BECK will adopt the tools and align to security policies provided by the project coordinator and partners.
7. C&H	Data security will be managed according to C&H’s standard processes, in compliance with the ISO 27001 standard.
8. 7BULLS	Data collected and managed by the Monitoring Sub-system will be stored in the clustered Influx DB instance run in the cloud. Cloud provider and cluster management will be responsible for data security.
9. AF	Data is stored on an isolated private server. Copy of the dataset is stored in cloud storage, which guarantees backups of data. AF is data administrator. Some of the data might be shared with partners for the purposes of the project. In such a case, dedicated shared folder administered by the project coordinator will be used.
10. TRUST	The AI-SPRINT website is based on Drupal version 7. All the open-source modules are kept up to date and all security patches are applied as soon as they are released. The infrastructure runs on a VM or docker container in a cloud environment with a firewall and restricted access to specific port and IP. All the operating systems based on Linux are maintained and updated. Databases are based on MySQL/MariaDB and PostgreSQL with daily backups stored in two different locations.
11. IDC	Data and reports on IDC servers are maintained according to IDC security department arrangements and according to company policy.

2.9 Ethical aspects

This section aims at clarifying if any ethical or legal issue may impact data management and sharing in the project. In this regard, the main issue in AI-SPRINT is associated with the collection of personal health-related data in the context of the Personalised Healthcare use case within WP5, which, as further clarified below, is dealt with according to the Ethics Summary Report linked to the Description of the Action. Other personal data, if any, will be kept undisclosed in compliance with the GDPR. Information on ethical aspects for each partner organisation is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA). - Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?
1. POLIMI	On POLIMI's side, data collection will take place in the context already described in Section 2.2 above. Such data is completely plant-oriented or application performance oriented, and will not include sensitive data about people. Any appearance of people in datasets (e.g. an AI-SPRINT researcher setting up a camera before data collection) will be accidental and highly occasional, if occurring at all. Each of such appearances will be checked against the requirements of the GDPR, and (if needed) managed either by collecting suitable authorizations or by excising the relevant segment from the datasets prior to publication.
2. BSC	The Personalized Healthcare use case, which concerns health monitoring via wearable sensors, questionnaires, and health data collection, envisages the participation of human subjects. According to the Ethics Summary Report associated with the DoA, the participant recruitment and management will be subcontracted to an external party, namely the Foundation "Freno al Ictus". The patient and caregiver informed consent forms have been provided with the DoA.
3. TUD	No sensitive data is collected, hence, there are no ethical issues associated with this data.
4. UPV	As no sensitive or personally identifiable data is stored, there are no ethical issues associated with this data.
5. GREG	GREG data collection will not include sensitive data about people. Any appearance of people in datasets (e.g. an AI-SPRINT researcher setting up a camera before data collection) will be accidental and highly occasional, if occurring at all. Each of such appearances will be checked against the requirements of the GDPR, and (if needed) managed either by collecting suitable authorizations or by excising the relevant segment from the datasets prior to publication.
6. BECK	BECK does not expect that any data collected and used to be sensitive or personal data under GDPR regulations. Also, we do not foresee ethical issues associated with the process of collecting data, to their content, and maintenance. In some cases, GPS data will be removed or replaced by generic regional indication upon publishing (e.g. coordinates 48°51'29.6"N 2°17'40.2"E will be replaced by "France / Paris")
7. C&H	The Openstack access data (names and emails) and IP addresses are personal data and will be dealt with in accordance with GDPR regulations. This precludes data sharing.
8. 7BULLS	We do not plan to collect sensitive data. Only user accounts will contain private data (emails, usernames), but they are not intended to be shared.

9. AF	No sensitive, ethics wise, data is collected, hence there are no ethical issues associated with this data.
10. TRUST	No sensitive, ethics-wise, data is collected, hence there are no ethical issues associated with this data.
11. IDC	There are no ethical issues associated with IDC data.

2.10 Other issues

Additional issues associated with data management not covered above are addressed in this section. These may refer to relevant procedures required at different levels (e.g. country, sectoral, etc.). Information on this topic for each partner organisation is detailed in the table below.

Partner	
	For each partner, please provide in the cells below information on the following: - Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?
1. POLIMI	We will utilize Zenodo or our institutional research data repository when applicable.
2. BSC	No other procedures for data management are required beside those related to open access publications and institutional repositories (see section 2.1).
3. TUD	We will utilize the institutional research data repository when applicable.
4. UPV	Considering publications, we will follow the policy of our institutional library. With respect to research data, there is no institutional or national policy defined yet, although we will follow the recommendations of DIGITAL.CSIC.
5. GREG	No other procedures for data management are required
6. BECK	No other procedures for data management are required
7. C&H	In general, all the processes to guarantee data security management in C&H comply with ISO 27001 standard.
8. 7BULLS	No other procedures for data management are required.
9. AF	Not relevant
10. TRUST	Not relevant
11. IDC	Not relevant

References

1. European Commission (2016) *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*. (Version 3.0; 26 July 2016) [Online]. Retrieved at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Last accessed: 26th April 2021).

2. European Commission (n.d.) *Data Management - H2020 Online Manual* [Online]. Retrieved at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (Last accessed: 26th April 2021).