LOCARD

DELIVERABLE

D1.6 Data Management Plan

Project Acronym:	LOCARD	
Project title:	Lawful evidence collecting and continuity platform development	
Grant Agreement No.	832735	
Website:	http://locard.eu/	
Contact:	info@locard.eu	
Version:	1.0	
Date:	30 April 2020	
Responsible Partner:	ARC	
Contributing Partners:	ARC, APWG, VIC	
Reviewers:	All consortium	
Dissemination Level:	Public	X
	Confidential – only consortium members and European Commission Services	

This project has received financial support from the European Union Horizon 2020 Programme under grant agreement no. 832735.

Revision History

Revision	Date	Author	Organization	Description
0.1	05/07/2019	Fran Casino, Constantinos Patsakis, Tom Dasaklis	ARC	Initial draft
0.9	22/04/2020	Fran Casino, Constantinos Patsakis,	ARC	Final version for review
1.0	30/04/2020	Fran Casino, Constantinos Patsakis	ARC	Final version

Table of Contents

List of Acronyms

Abbreviation / acronym	Description
D	Deliverable
DMP	Data Management Plan
DMS	Data Management Strategy
DoA	Description of the Action
IPR	Intellectual Property Rights
M	Month
T	Task
UC	Use Case
WP	Work Package
EUCI	EU classified information
REA	Executive agency for research

1 Executive Summary

This deliverable constitutes the second version of the Data Management Action Plan (DMP) of the LOCARD project and provides a general outline of the project policy for data management. The overall purpose of this DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project, describing what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. Moreover, we devote a specific section for EU restricted deliverables and access granted through strict management policies, as requested by the EU Commission.

The Data Management Strategy (DMS) to be followed by LOCARD is based on the identification and classification of data generated and collected, identification of standards and metadata to be used, exploitation and availability of data as well as the required ethical, legal compliance and the responsibilities in the implementation of the DMP.

The initial set of the project datasets are described as follows:

- DATASET 1: *Information about the consortium*
- DATASET 2: *Project files*
- DATASET 3: *Research activities*
- DATASET 4*: Development data, implementations, and codes*
- DATASET 5: *Pilots and testing activities*

A tailored-down DMS for each dataset will be described in this document. In addition, this document reflects the current state of consortium agreements regarding data management and is consistent with those referring to the exploitation and protection of results. It is a living document that is expected to mature during the project lifetime and will be updated accordingly in M24, and M36.

The next version of the LOCARD DMP will put a strong emphasis on the complete definition of procedures to be implemented by the project to efficiently manage its research data in terms of storage and backup (backup provision, recovery procedure), selection and preservation (which data will be retained/shared/ preserved, length of time data to be preserved and preservation preparation time).

2 Introduction

This deliverable constitutes the second version of the DMP of the LOCARD project and provides a general outline of the project policy for data management, including the following aspects:

- What types of data will the project collect/generate?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? Reasons why the data cannot be made available in some cases.
- How will this data be curated and preserved?
- How data restriction levels are managed?

This DMP outlines how research data will be handled during the project and after it is completed. The overall purpose of this DMP is to support the data management life cycle for all data that will be collected, processed or generated by the project. It will contribute to:

- Ensure project research data and records are accurate, complete, authentic, interoperable and reliable,
- Save time and resources in the long run,
- Enhance data security and thereby minimize the risk of data loss,
- Ensure research integrity and reproducibility by others,
- Prevent duplication of effort by enabling others to use the project's data,
- The described policy herein reflects the current state of consortium agreements regarding data management and is consistent with those referring to exploitation and protection of results. It is a living document that is expected to mature during the project lifetime and will be updated accordingly.

3 Data Management Strategy

The general strategy for data management, according to the Guidelines on Data Management in Horizon 2020 [1] will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archived, the preservation of the information as well as the ethical, legal compliance and the responsibilities in the implementation of the DMP.

The LOCARD DMP will cover all the data life cycle following the H2020 guidelines regarding Open Research Data [2].

To formulate an effective DMP, which helps to keep track of the varieties of data generated by the project, it is useful to categorize such data according to their form, source and type.

3.1 Data to be collected/ generated

The data to be collected/extracted/generated in LOCARD can be grouped in the following categories:

- *DB1 Information about the consortium*: Data about the consortium, such as personal information, emails and etc will be handled and stored in private and secure storage (i.e. Owncloud). Access will be restricted to the members of the consortium. Note that EAB members are considered members of the consortium, yet with limited access.
- *DB2 Project files:* In this regard, all data gathered from meetings, workshops, and any type of internal communication will be protected according to each required level of confidentially (i.e. and stored using the local cloud repository as in DB1). Fruitful and general outcomes of the project will be disseminated without restriction if no sensitive data are disclosed. In the case of confidential discussions or outcomes (e.g. EU Restricted deliverables), only specific partners will have access and files will be stored encrypted ineach corresponding organization premises (in the case of sensitive data). For example, in the case of meetings between LEAs or these which require confidentiality, any result will be disclosed, and only authorized consortium members will be able to participate. More details about the EU restricted data management is given in a separate document, the EUCI handling policy.
- *DB3 Research activities:* The LOCARD research activities and their corresponding outcomes (deliverables, publications) are different (see WP4). State-of-the-art methods and public resources will be used to carry such activities. In the case of research that involves data gathering from social networks or other platforms, such data will be stored and protected locally by the corresponding organization. In this regard, any sensitive material (i.e. including personal data, as later defined in Section 3.2) will leave the premises of the organizations. In the case that specific materials need to be shared, each partner will provide their own restrictions and policies (according to national and EU laws), which should be agreed by the interested members. In each case, proper anonymization and/or encryption mechanisms will be applied to guarantee that any personal or sensitive data is disclosed. EU restricted data will be managed accordingly. These procedures will comply with the LOCARD ethics reports and deliverables, especially in regard to sensitive data storage and mass surveillance policies reported in D8.4, D8.9 and D8.12.

- *DB4 Development data, implementations and codes:* Development and codes, as well as implementations derived from the project, will be performed in a private repository (GitLab). Periodic versioning backups will be made for each module of the LOCARD project, which will be stored in Owncloud.
- *DB5 Pilot and testing activities*: In the case of pilots, the information of the users involved, the pieces of synthetic evidence and the processed data will be stored and managed locally by each organization. At the beginning we consider the use of synthetic data; nevertheless, each organization (LEAs) may use their repositories for such pilots. Therefore, in the case of an end user using their own repositories (or a copy of specific evidences for testing purposes), their local policies and data management restrictions will apply. In any case, data will be deleted when the corresponding validation finishes. In addition, the outcomes of such pilots will be reported anonymized if some sensitive data needs to be shared. Other types of users and participants may be considered during the project lifetime, according to new system requirements, whose data will be stored and managed accordingly.

A global overview of the different data level present in the project databases is depicted in Figure 1. In this regard, the corresponding clarifications and descriptions are presented in the next sections.

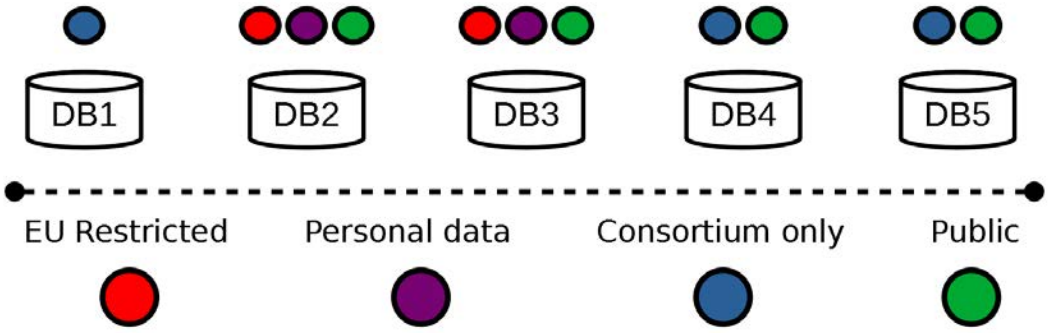

Figure 1 – Correlation between LOCARD database contents and data restriction levels.

3.2 Data types

The type of data to be collected and processed can be either:

- No personal data: such information which is not affected by Data Protection legislation
- Personal data: Data containing individual's sensitive information, according to the EU definitions[1]
- EU restricted data

The Council's decision on the security rules for protecting EU classified information lays down the basic principles and minimum standards of security for protecting EUCI. These principles and standards apply to the Council and its General Secretariat, and they also need to be respected by member states when they handle EUCI.

1https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

EU classified information is **categorised** in four levels, which are defined by the severity of the impact of disclosure:

- TRÈS SECRET UE/EU TOP SECRET: the unauthorised disclosure of this information could cause exceptionally grave prejudice to the essential interests of the EU or one or more of the member states.
- SECRET UE/EU SECRET: the unauthorised disclosure of this information could seriously harm the essential interests of the EU or one or more of the member states.
- CONFIDENTIEL UE/EU CONFIDENTIAL: the unauthorised disclosure of this information could harm the essential interests of the EU or one or more of the member states.
- RESTREINT UE/EU RESTRICTED: the unauthorised disclosure of this information could be disadvantageous to the interests of the EU or one or more of the member states.

The Council decision on security rules covers several ways to protect this information, including personnel security, physical security, management of the information, information assurance, industrial security, or the ways EUCI is shared within the EU institutions and with third states and international organisations. In LOCARD, we will have to manage EU RESTRICTED deliverables and information.

In LOCARD, a set of policies and access management to EU restricted data are described in a specific document, namely the *LOCARD EUCI handling policy.* All the details about how access is managed and granted, what are each partner's responsibilities and what data are restricted and to whom are described in such document, as well as the procedures for protection and dissemination of EU restricted data.

Personal data: data which relate to an individual who can be identified

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

In the second case since LOCARD follows the European directives, all personal data will be anonymized according to the guidelines derived from activities focussed on ensuring the ethical and privacy issues and compliance with legislation. Such activities are carried out as part of WP2 and WP8.

When trials/pilots are carried out (WP6), each volunteering participant, including members of LOCARD consortium, will be informed before each data collection exercise and test session on the ethics directives, principals and implications and they will be invited to sign a consent form.

3.3 Standards and metadata to be used

The metadata for the different identified datasets will be generated either automatically by the system or through manual content annotation. A metadata framework will be used to identify the data types, owners and allowable use. The framework will be complemented with a controlled access mechanism and in the case of data transmission with efficient encoding and encryption mechanisms.

More detailed descriptions of the metadata will become available as part of the work in various project WPs – i.e., requirements for social, ethical, legal and privacy issues (WP2, WP8), definition of case scenarios, and reference architecture (WP3), platform implementation (WP5) and testing (WP6).

3.4 Exportation, Exploitation, availability of data and re-use

In general, LOCARD, following the requirements of disseminating the results and developments of the project, is willing to make open access as much material as possible, and open access to data will be the default option for the consortium. Especially with relation to dissemination tasks in WP7 and with WP4 (Research contributions, except for these considered confidential). Therefore, deliverables, publications and reports will be made public whenever possible. Nevertheless, as the project will engage individuals and, respectively, process personal data, open access to the deliverables and project data might evoke potential issues (all ethical and legal aspects will be thoroughly examined by the Project Ethical Manager, Data Controller and Security Officer). Therefore, the consortium will carefully examine to what extent data can be publicly shared. In case of a positive decision, careful anonymization will take place to remove any personal data present in such datasets.

In regard to data exportation, LOCARD will not transfer data to third countries. Personal data will be processed in the territory of the European Union. As the social media analysis and deviant behaviour detection tools will crawl the web, partners assume that personal data of non-EU citizens or EU citizens outside the EU will be processed as well. However, the protection of their personal data will conform the EU data protection rules. Partners ensure that the implemented technical and organizational measures facilitate the adequate protection of their rights and freedoms.

3.5 Archiving and preservation

LOCARD will use state-of-the-art technologies for secure storage, delivery and access of personal information, as well as managing the rights of the users. In this way, there is a complete guarantee that the accessed, delivered, stored and transmitted content will be managed by the right persons, with well-defined rights, at the right time.

During the execution of the project, each partner will provide detailed information on privacy/confidentiality and the procedures that will be implemented for data collection, storage, access, sharing policies (especially when third party countries are concerned), protection, retention and destruction. The consortium will confirm that the project complies with national and EU legislation. For all data, the names of the participants can be replaced with ID codes to maintain anonymity. The identity of any participants will be fully masked in any printed materials, project reports or dissemination. The personal data, if any, will be treated confidentially and carefully (taking proper technical means of information protection, for example, storing general and personal data separately, using encryption for personal data and identities, deleting personal data when it becomes unnecessary). Some examples include public-key encryption and symmetric encryption with session keys negotiation over HTTPS. Considering that some transmitted data may be regarded as highly sensitive, the highest security standards would be used (i.e. asymmetric cryptography with at least 1024-bit keys). Where necessary (e.g. sensitive and evidence information collected by LEAs) the data will be stored in a locked server, and all identification data will be stored separately. Moreover, no one outside of the research team will have access to these or be able to listen to transcripts or watch video footage/view photographs. The access to the database will be restricted to authorised personnel only. Moreover, an access log will be maintained as so to ensure the proper use of the accessed data.

Other research data will be stored and backed up regularly through existing back-up mechanisms in place at Owncloud. Qualitative data will be backed up and secured by the coordinator on a regular basis and metadata will include clear labelling of versions and dates.

The software Zed! 4.0[2] will be used for the secure exchange of EU RESTRICTED information between partners when necessary. Licenses will be acquired by the coordinator and shared with the partners that request them. More information about EU restricted documents can be found in the LOCARD handling policy document.

All data containing private information will be destroyed upon completion of the respective study/research task. In any case, all personal data will be destroyed automatically at the end of the project and only anonymous or non-identifiable data will be retained after the completion of the final report.

3.6 Testing and Validation data

The development of other LOCARD components will occur with the processing of dummy or synthetic data, therefore the GDPR will not be applicable. However, partners keep in mind during their activities that after the conclusion of the project personal data will be processed through these components, therefore SELP requirements remain relevant.
However, partners are able to describe the synthetic data that will be used for the pilots, according to each user case. For this a list of the synthetic data that is planned to be used in the pilots can be categorised in the list below:

- Multimedia files (Snapshot, Screenshot, Video, Screen records),
- Text messages (E-mail communications),
- Streaming data (URL, illegal content, linked web pages, etc.),
- Contact details (username, email, phone numbers, etc.),
- Timestamp and hashes of the data.

All the data will be collected only for the scope of the project. Though the data will be synthetic, the main purpose is to validate the LOCARD technologies and LOCARD platform modules implemented are fit to be incorporated into a complete solution and are compatible with different use case scenarios, as it is said in Work Package 6.

Regarding the volume, it is hard to define at this stage of the project and will probably scale with the number of end users. Most likely, it is to be in the scale of millions.

Concerning the finalized system, to be deployed after the conclusion of the project, there will be a possibility of manual entry of evidence in very specific cases (when a case is transferred to other end user and the evidence exchange between them can only be done in a specific way). Partners also consider that some data may come from the connector engine, so that future users can query external systems of databases from other entities (e.g. Europol). It must be emphasised that LOCARD will only provide the means through a connection engine, but it will not connect directly to any external repository. This will require further steps and ad-hoc configurations with the local authority and their connector systems, managed by each corresponding end user or LEA. LOCARD will help with that but till that extent. A total connection is out of scope of the testing and validation.
After the conclusion of the project, when gathering or seizing evidence by LOCARD it must be done lawfully or it will be inadmissible in a court. The responsibility to define the legal ground will be on the user of the platform.

2 https://www.consilium.europa.eu/en/general-secretariat/corporate-policies/classified-information/information-assurance/eu-restricted/offline-file-encryptor/zed-v4-0/

3.7 EU restricted data

As a requirement in LOCARD, some of the data generated by the consortium is classified as EU restricted. A secure and standardized procedure to send such data will be followed by Project Security Officer (PSO) to ensure that documents are protected and sent to the commission with the required security guarantees.

As stated in previous sections, to further extend this information and to disseminate the proper procedures in the case of managing EU restricted data, the SAB has developed a handling policy as a normative reference for all partners. It includes rules and procedures that all involved partners shall have to follow in order to ensure proper handling of EUCI. Individuals that need access to EU classified information must be briefed/trained by their respective organisations as to their responsibility for the security of the information that they manage.

3.8 Personal data

Processing of personal data in LOCARD will occur primarily through the social media analysis and deviant behaviour detection tools (as well as through the engagement of stakeholders in e.g. workshops). As the primary purpose of the project, and respectively the processing of personal data, is to carry out research, the GDPR and its national implementations are applicable. It must be noted however that if the platform will be used by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, the 2016/680/EU Law Enforcement Directive and its national implementations will become applicable. As regards the processing operations carried out by the social media analysis and deviant behaviour detection tools, ARC, as the responsible partner for the development of these tools, ensures that the respective data protection laws will be adhered to. To ensure the accuracy of data respective to the purpose of processing (to extract/identify deviant behaviour via machine learning and extract the corresponding information), specific social networks and platforms will be identified. The range of categories are expected to be the following, but this is subject to change (prior to the first application partners are unable to define the exhaustive list):

- Username,
- IP address,
- Comments,
- Shared audio-visual content,
- Data which facilitates the identification of interacting friends.

At this stage (M12), it is difficult to make a strict and specific list because when data (or after the project potential evidence) is collected we can find diverse types of files, yet the ones listed above are the classical ones.

Data collected via web-crawling will be stored locally at the premises of ARC and will be automatically encrypted. Currently only two members of ARC have (physical) access to the database. No one from the LOCARD consortium has requested further access so far. In case the data processing operation achieves its purpose, it means that ARC will be in possession of criminally relevant information concerning a potentially criminal activity. Therefore, an incidental findings policy will be drafted and implemented before the first use of the tools, which will lay down the rules of communication with the relevant authorities (i.e. with Greek authorities and indirectly with Europol).

To facilitate the minimization of processed data, manual deletion mechanisms will be implemented. The utility of data will be continuously reviewed, and unnecessary data will be deleted.

The legal basis of the data processing done by the social media analysis and deviant behaviour detection tools is article 6 (1) e) of the GDPR: “processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller”. Although recital (33) GDPR

emphasises that "data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research", it also identifies an exception: "data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose." Respectively, notification of data subjects about the existence of such processing operation would undermine the research purpose as individuals would behave differently (i.e. hiding their potentially criminally relevant behaviour intentionally). Such notification and reaction would result in inaccurate sets of data, making the development of the LOCARD social media analysis and deviant behaviour detection tools impossible. Therefore, both art 6 (1) a) and art 6 (1) f) are excluded. The latter cannot be used as the legitimate interest test requires inter alia proximity with the data subjects. For the reasons explained above this would also undermine the purpose of the processing and respectively the success of the research activity.
The legal basis of processing sensitive data is article 9 (2) g) of the GDPR: "processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject." The justification thereof aligns with the one presented above. According to European case law, necessity and the public interest imply a 'pressing social need'.[3] In LOCARD, processing of personal data will be carried out by a Greek university (ARC), which is empowered by the Greek state to perform research.

Accordingly, LOCARD implements appropriate safeguards to protect the right and freedoms of data subjects; adequate technical and security measures entrenching the principle of data minimisation and using pseudonymised data as default; and achieves compliance with recognised ethical safeguards. Since no other personal data will be used during the project lifetime (note that testing and validation will be performed with synthetic data), the corresponding technical and organizational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants is submitted in deliverable 8.4.

In regard to other external data inputs, during the project no personal data from third parties (e.g. criminal database) will be processed. Concerning the finalized system, to be deployed after the conclusion of the project, there will be a possibility of manual entry of evidence in very specific cases (when a case is transferred to other end user and the evidence exchange between them can only be done in a specific way). Partners also consider that some data may come from the connector engine, so that future users can query external systems of databases from other entities (e.g. Europol). It must be emphasised that LOCARD will only provide the means through a connection engine, but it will not connect directly to any external repository. This will require further steps and ad-hoc configurations with the local authority and their connector systems, managed by each corresponding end user or LEA. LOCARD will help with that but till that extent. A total connection is out of scope of the testing and validation.

3.9 Ethics and legal compliance

All details about ethics and legal compliance in terms of Current EU legislative initiatives, anonymisation procedures, consent needed, restrictions on 3rd parties, procedures for handling sensitive data and data owners will be included in the corresponding deliverables of WP2 and WP8. Procedures and clear protocols for collection and management of research data (gathering, processing and dissemination) will be defined and implemented in with the support of the consortium, the External Advisory Board members as well as the

3 Handyside vs the UK App no. 5493/72 (ECHR, 7 December 1976); Leander v. Sweden App no. 9248/81 (ECHR, 26.03.1987); see also EDPS, Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit, 11 April 2017

Ethics manager. Meanwhile, information about the expected in future updates of this document (in M12, 24 and 36).

Additionally, the Grant Agreement and the LOCARD Consortium Agreement are to be referred to for further details on the ownership, management of intellectual property and access.

3.10 Responsibility and resources

ARC, as project coordinator and leader of WP1, VICTOR as a technical coordinator and VUB as leader of Social Ethical Legal and Privacy WP2, have a particular responsibility in creating and updating the Data Management Plan. Moreover, the documents elaborated in WP8 are also useful for this task.

Each LOCARD partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

4 Project Datasets

The current version of the DMP addresses the following aspects on a dataset basis and presents the current status of reflection within the consortium concerning the set of data managed by the project.

- Dataset reference, name and reference: Identifier for the data set to be gathered/processed/generated.
- Dataset description: Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
- Standards and metadata: it includes the reference to existing suitable standards of the discipline. If these do not exist, an outline of how and what metadata will be created.
- Data sharing: A detailed description of how data will be shared, including access procedures, outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be wide open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating, in particular, the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g., ethical, rules of personal data, intellectual property, and commercial, privacy-related, security-related).
- Archiving and preservation: Description of the procedures that will be put in place for the long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

Currently, LOCARD plans to manage 5 different datasets. Next sections describe those datasets following a structure in accordance with the Guide of Horizon 2020 for the Data Management Plan [1].

4.1 Dataset 1: Information about the consortium

1	**Dataset reference**
	Information about the consortium
2	**Dataset Description**
	This dataset includes information about the consortium (e-mails, phones, project resources and etc)
3	**Standards and Metadata**
	Excels, SQLs and text files containing data about the consortium.
4	**Data Sharing**
	Data is only for the consortium
5	**Archiving and preservation (including storage and backup)**
	The database will be stored in private Owncloud.

4.2 Dataset 2: Project files

1	**Dataset reference**
	Project files
2	**Dataset Description**
	This dataset includes all meetings reports, research activities from their creation till their dissemination, the information and data collected in workshops and other communication activities.
3	**Standards and Metadata**
	Multimedia files, text documents.
4	**Data Sharing**
	Confidential data will not be shared and if it has to, it will be protected accordingly (e.g. encrypted). The rest of the data can be publicly disseminated.
5	**Archiving and preservation (including storage and backup)**
	Sensitive data will be stored locally by each organization under their responsibility and strict protection policies. Non-sensitive data can be stored in Owncloud and disseminated through LOCARD communication channels

4.3 Dataset 3: Research activities

1	**Dataset reference**
	Research activities
2	**Dataset Description**
	This dataset stores the outcomes of the research activities, in terms of deliverables, publications as well as multimedia data collected during such activities.
3	**Standards and Metadata**
	Metadata, Multimedia, Documents, Sensitive/personal Data
4	**Data Sharing**
	Non-sensitive data will be shared between the consortium and research outcomes will be made publicly available when necessary. If during the research activities some organization collects sensitive/personal data from the internet (e.g., social network crawlers and other data gathering procedures) such data will only be shared if it is strictly mandatory and only prior anonymization of sensitive data and encrypted. Moreover, access to such data will be controlled and logged (see Annex) and data will be deleted immediately after the investigation/research is concluded, discarding irrelevant information when possible.
5	**Archiving and preservation (including storage and backup)**
	Each organization will store research data locally and protected accordingly (e.g. in a computer with no access to the internet and encrypted). In the rest of cases, data will be stored in Owncloud.

4.4 Dataset 4: Development data, implementations and codes

1	**Dataset reference**
	Development data, implementations and codes
2	**Dataset Description**
	This dataset includes all the implementations, codes and development outcomes regarding to the architecture of LOCARD
3	**Standards and Metadata**
	Documents, programs, ontologies, architectures, code
4	**Data Sharing**
	Data will be shared in GitLab between required consortium members.

5	**Archiving and preservation (including storage and backup)**
	The code of the project will be stored in GitLab and periodic backups, as well as associated documentation, will be stored in Owncloud

4.5 Dataset 5: Pilot and testing activities

1	**Dataset reference**
	Pilot and testing activities
2	**Dataset Description**
	This dataset stores data related to the pilot and testing activities of the project. It includes the executables, datasets (synthetic and real) and other user and participants information. It also includes the information made available in the blockchain during the validation phase.
3	**Standards and Metadata**
	The outcomes include documents and reports, code and execution logs, architectural details and forensic data as well as other kinds of metadata
4	**Data Sharing**
	Pilot and testing outcomes will be shared between members if they do not disclose sensitive data. If organizations use real data, all experiments will be performed in their premises and the outcomes of such tests will be shared prior sanitization of data.
5	**Archiving and preservation (including storage and backup)**
	Synthetic and non-confidential data will be stored in Owncloud. Data about real case scenarios and participant users as well as other sensitive data will be stored and managed locally by each participant according to strict protection measures, such as an isolated storage/computer resource as well as encryption mechanisms.

5 Conclusion

This deliverable provides a second iteration on the description of the data that LOCARD will manage during its lifetime together with challenges and constraints that need to be considered to ensure project's research data and records will be accurate, complete, interoperable and reliable; to enhance data security and minimize the data loss risks; to prevent duplication of efforts allowing others to use some of the data managed by the project.

The next version of the LOCARD DPM will provide a more robust description of data management policies and a log of actions performed to sensitive data collected (if any), which will follow the guidelines described in WP8. Moreover, we will ensure that all generated datasets do not infringe either partner IPR rules or regulations related to personal data protection. A clear and complete mechanism for systematic anonymization of personal data will be defined and is planned to be in place before the first stage of pilots will start at M20.

6 References

[1] Guidelines on Data Management in Horizon 2020, V2.0, 30 October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020, Version 2.0, 30, October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

7 Annex

7.1 Access control template

Date: DD/MM/YYYY

Location: Street, Number, Postal code, Country

Full Name: Name Surname

ID/Passport: 0123456789 X

Purpose of the access: Examples: Research activity on case id=x. Forensic analysis of dataset id=y.

Encrypted Data or Datasets Retrieved/Sent (if any): Dataset, hash or the original dataset, hash of the encrypted version, procedure (e.g. ZED 4.0, other standardised methodology or protected channel), destination details.