European Commission | Horizon 2020
European Union funding
for Research & Innovation

Data Management Plan

D6.3

Grant Agreement number	730082
Call identifier	EO-3-2016 Evolution of Copernicus services (RIA)
Project Acronym	**E²mC**
Project title	**E**volution of **Em**ergency **C**opernicus services
Types of action	Research and innovation actions (RIA)
Project Starting date	01/11/2016
Project Duration	30 months
Project Coordinator	e-GEOS S.p.A.
Deliverable reference number and full name	D6.3 - Data Management Plan
Delivery Date	31/08/2019
Issue/Revision	2.0
Document produced by	e-GEOS
Document validated by	e-GEOS
Document type *	R
Dissemination Level	PU

* Type of the deliverable codes:
R Document, report
DEM Demonstrator, pilot, prototype
DEC Websites, patent fillings, videos, etc.
OTHER;
ETHICS Ethics requirement

* * Dissemination level codes:
PU = Public,
PP = Restricted to other programme participants (including the Commission Services).
RE = Restricted to a group specified by the consortium (including the Commission Services).
CO = Confidential, only for members of the consortium (including the Commission Services).

Dissemination list:

This deliverable is a public document

Applicable documents

Ref. / Document Title	Issue	Date
730082_Annex 1 - Description of the action (part A)	n/a	13/10/2016
730082_Annex 1 - Description Of Action (part B)	1.0	13/10/2016
730082_Data Management Plan	1.0	14/04/2017

Document Change Record

Issue	Date	Page / paragraph affected
1.0	14/04/2017	First issue
2.0	31/08/2019	Second Issue as requested after FR

Table of contents

Acronyms & Definitions	
EC	European Commission
EU	European Union
REA	Research Executive Agency
WP	Work Package
WPC	Work Package Committee
ORD	Open Research Data
DMP	Data Management Plan

1 INTRODUCTION

The Data Management Plan (DMP) is a deliverable required for all projects participating in the ORD Pilot (Open Research Data Pilot). In the previous work programmes, the ORD pilot included only some areas of Horizon 2020, now, under the revised version of the 2017 work programme, the Open Research Data pilot has been extended to cover all the thematic areas of Horizon 2020. It aim to encourage the data management following the principle "as open as possible, as closed as necessary".

"*The ORD pilot aims to improve and maximize access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions*" (from "Guidelines on FAIR Data Management in Horizon 2020").

The structure of the document follows the "Guidelines on FAIR Data Management in Horizon 2020", provided by the Directorate-General for Research & Innovation of the European Commission

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

1.1 Scope of the document

The final goal of this document is to show how the research data of the project are **findable, accessible, interoperable and reusable** (FAIR).

This deliverable describes the data management life cycle for the data to be collected, processed and/or generated by the project. It includes information on:

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

It reports also some considerations on the ethical aspects and data security and also an assessment on costs related to data management.

The DMP is intended to be a living document and it will be updated over the course of the project whenever significant changes arise. In any case, it will be updated as a minimum in time with the periodic reviews of the project.

2 OVERVIEW ON PROJECT DATA

2.1 Data types, collection and generation

Data types, collection and generation	
▪ What types and formats of data will the project generate/collect?	The E2mC will use data from crowdsourcing and social media (e.g. images, videos, text messages). The project will generate, from that data, mainly maps in support to the Emergency responders, which include a layer relevant to the social elaborated information.
▪ Purpose of the data collection/generation and its relation to the objectives of the project	The main goal of the project is to produce better, more qualitative and more timely maps using information either to be collected from social media or solicited from the crowd information, enriching the already available information from the satellite sources.
▪ Expected size of the data	The size of the data will be in the range of few Terabytes
▪ Origin of the data	Crowdsourcing and social media

2.2 Existing data

Existing data	
▪ Re-use of existing data	Existing datasets from social media and crowdsourcing sources related to past disaster events will be re-used as well as products delivered by the Copernicus Emergency management Service. This re-use will enable the testing of the system on a set of "cold cases", giving hints and suggestions on the set-up of the procedures and their reliavility.
▪ Way to re-use, to what end	Data will be re-used through re-processing of social media and crowdsourced data in combination with satellite images and visual interpretation in order to demonstrate the feasibility of

Existing data	
	deriving useful and geolocated information on the disaster location and extent form these unconventional sources.

2.3 Data usability

Data usability	
▪ To whom might it be useful?	These data might be useful to the Service Provider of the Copernicus Emergency Management Service in order to provide an enhanced service to the European Commission and, ultimately, to the Civil Protection of the European Member States and to the Humanitarian Aid actors. In fact, the outcomes of the project and the released data and maps, can suggest also other applications of the delivered system, by selecting other keywords and enabling the location and the mapping of other events of interest, also exploiting the multilingual capability of the system.

3 "FAIR" DATA APPROACH

3.1 Data "Findability"

Data "Findability"	
▪ 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)?	E2mC deployed a master PostgresSQL/PostGIS database where to store the data produced by the project (indexes over social media data streams, data contributed form crowdsourcing, satellite images …). All the data are saved to this database through a REST interface called DataAPI. The API is developed by using the Swagger/Open API standard and the documentation is available here.
▪ What naming conventions do you follow?	The only naming convention adopted is the one related to the creation of each specific event workspace: dYYYY-MM-DD_NameOfTheEvent The values coming from the social networks do not follow any naming convention, nevertheless the data findability for each kind of resource is guaranteed through the usage of well defined tags.
▪ Will search keywords be provided that optimize possibilities for re-use?	Yes, through search interfaces using local cache/db and existing service API available by social
▪ Do you provide clear version numbers?	Yes
▪ 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	We will use INSPIRE compliant metadata for the generated geospatial datasets and services

3.2 Data "Accessibility"

Data "Accessibility"	
▪ 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 (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)	Maps and derived information will be made openly available. As for data from social media and crowdsourcing, it has to be decided which data can be shared, indeed there could be ethical and legal issues related to this data that need to be further discussed with the ethical expert.
▪ How will the data be made accessible (e.g. by deposition in a repository)?	A web-based access will be allowed for working on the project master database and the project Social&Crowd Platform.
▪ What methods or software tools are needed to access the data?	A web browser to access the project Social&Crowd Platform and different FOSS SW package to access the data generated (images, PDF documents, database)
▪ Is documentation about the software needed to access the data included?	Yes, it will be part of the metadata
▪ Is it possible to include the relevant software (e.g. in open source code)?	Yes, the SW of the Social&Crowd Platform will be released as an open source software for Commission use only
▪ Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.	Yes, are available through the project Social&Crowd Platform here with the following usr/pwd: demouser/demouser The project code is stored on the e-GEOS Gitlab repository here. The credentials to access must be requested at the following email address: mariano.biscardi@e-geos.it
▪ Have you explored appropriate	It is not necessary to take any particular

Data "Accessibility"	
arrangements with the identified repository?	arrangement with Gitlab.
▪ If there are restrictions on use, how will access be provided?	The administrator access to Gitlab will be grated to Commission and Code maintainers only
▪ Is there a need for a data access committee?	No
▪ Are there well described conditions for access (i.e. a machine readable license)?	No
▪ How will the identity of the person accessing the data be ascertained?	The access to the master database and to the Social&Crowd platform will be strictly controlled and credential will be generated/preconfigured through a service admin facility. Nonetheless we have no means to ensure that the provided credentials are not used by a different person

3.3 Data "Interoperability"

Data "Interoperability"	
▪ 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)?	Yes, intermediate and final data/products will be stored in the project storage with unique resources identifiers. In some cases (e.g. Social Data), there will be the need to respect the Terms & Conditions of the specific Social Network used.
▪ What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	The project will use INSPIRE compliant metadata associated, where possible, to vocabularies such as GEMET
▪ Will you be using standard vocabularies for all data types present in your data set, to allow inter-	As far as possible standard vocabularies will be used, exploiting also progress done by other FP7/H2020 projects in the field of

Data "Interoperability"	
disciplinary interoperability?	standardization and 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?	Yes

3.4 Data "Re-usability"

Data "Re-usability"	
▪ How will the data be licensed to permit the widest re-use possible?	Data and products generated within the Copernicus frame (CEMS) will be publicly available according to the Copernicus terms and conditions https://www.copernicus.eu/en/how/how-access-data …. Data and products generated within exploitation frame other than Copernicus, will be available under the license Creative Commons Attribution + ShareAlike (CC BY-SA) https://en.wikipedia.org/wiki/Creative_Commons_license Social data are accessible under the terms and conditions of services (Twitter, Youtube and Flickr)
▪ 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.	Data will be made available for re-use at the end of the project. During the course of the project, some datasets may be released after validation. There could be some limitations due to big amount of data.
▪ Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	Yes

Data "Re-usability"	
▪ How long is it intended that the data remains re-usable?	The data will remain re-usable at least for one year after project completion
▪ Are data quality assurance processes described?	Data Quality assessment will be part of final maps preparation during demonstration phases. The quality assessment of outputs is ultimately a project goal.

4 ETHICAL ASPECTS AND DATA SECURITY

4.1 Ethical aspects

Ethical aspects	
▪ 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).	As stated in the DoA-part B section 5.1.1 *"The E₂mC Project involves humans, but only for registering information relevant to an emergency crisis, <u>not requiring sensitive information about identity, health, etc.,</u> and as volunteer contributor to a crowdsourcing platform."* So, the ethical/legal issues in place are: 1. Management of personal data (i.e. registering information) 2. Information related to the images and videos collected (via crowdsourcing application) during a disaster as persons or vehicles can potentially be detected Detailed information on privacy/confidentiality and the procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction along with the confirmation that they comply with national and EU legislation are provided in D7.1 and D7.2.
▪ Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?	The Digital Informed Consent form is filled when registering and creating an account on the platform, to agree to the Terms and Conditions and to the Privacy Policy. Details are provided in D7.1. No other use of personal data is performed by the platform service.

4.2 Data security

Data security	
▪ What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?	The master database will be stored in a controlled cloud environment to ensure high data security and reliability
▪ Is the data safely stored in certified repositories for long term preservation and curation?	No. See also 3.4 Data "Re-usability"

5 ALLOCATION OF RESOURCES

5.1 Costs related to data management

Costs related to data management	
▪ What are the costs for making data FAIR in your project?	Most of the FAIR Data Principles are already taken into account. Further evaluation will be made on the basis of the final data ensemble (code repository, demonstration phases) in terms of code, output data and source data.
▪ 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)	Data will be stored in a commercial cloud facility. Cost for storage and management are included in the project budget to cover the period of usage within the project lifecycle.
▪ 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)?	The amount of data stored for each event is in the order of tens of MB. Given that, no rolling archive policies have been put in place. Nevertheless, the system database has been designed in a way that each activation is not related to another one. This separation allows the administrator to backup and put offline old data in any moment, if needed.

5.2 Data management responsible

The nominated data management responsible is: Mariano Alfonso Biscardi (Software Developer and Architect at e-Geos).

6 CONCLUSIONS

This DMP issue contains all information available at this point in time, further information will be included in the next issues of the document.

The E2mC project rely on the collection and use of a variety of crowdsourced data relevant to an emergency crisis. Such data (e.g. images, video, text messages etc. not requiring sensitive information about identity, health, etc.), will be collected through social media, crowdsourcing means and through volunteer contributors. They will be further analysed, processed (e.g. geolocated) and arranged for exploitation in the project master database and in the Social&Crowd Platform developed by the project consortium.

Further to this, any ethical and legal issues related to the usage and sharing of the collected data shall be carefully discussed with the ethical expert as to put in place proper informed consent forms, disclaimer, protection and security measures etc. in line with personal data protection policies and applicable IPR, if any.

Operators of the service (e.g., Project partners, commission personnel, Copernicus Emergency services providers…) will access the project master database and the Social&Crowd Platform on a controlled basis (username/pwd mechanism).

All the data will be stored in an external Cloud facility for the duration of the project.

At the finalisation of the Service Design phase of the project, details on the identification of the repositories or archive as well as data management responsibilities and roles will be discussed and allocated to partners.