



D1.8 Data Management Plan adhering to the H2020 Open Research Data Pilot v3



This project has received funding from the European Union's
Horizon 2020 Programme under Grant Agreement N°. 825014.

Workpackage	WP1 – Project Management
Editor(s)	KCL, ZABALA
Responsible partner	KCL
Contributors	KCL, ZABALA
Status-Version	v1
Due date	30/06/2021
Submission date	30/06/2021
EC Distribution	PU
Abstract	This document contains the third iteration of the Data Management Plan of DMS Accelerator. We outline the datasets that are being generated by the project and describe how these data will be managed in accordance with the FAIR principles advocated by the H2020 Open Research Data Pilot.

Document History

Version	Date	Comment
v0.1	17/06/21	First draft produced by KCL
v0.2	21/06/21	First draft revised by partners
v1	30/06/21	Final version of the document with partners inputs revised by KCL and Zabala

Contents

Document History	1
Contents	2
Executive Summary	3
Introduction	4
Summary of changes	4
Purpose of the data collection and generation	4
Project partners	5
Types of personal data generated/collected	6
Data description	6
Overview of Datasets	6
Portfolio of European SMEs approached by DMS	7
DMS applicant SMEs	8
Applicant SME ratings	9
DMS participant portfolio	10
Webinar recordings	10
Webinar engagement reports	11
Course engagement reports	12
Impact Survey Data	13
Survey of Startup Needs	13
Research Data	14
Data security	15
Ethical aspects	15

Executive Summary

The creation of a Data Management Plan is one of the obligations that DMS Accelerator is required to fulfil under the Horizon 2020 work programme. The European Commission encourages projects to make their data “as open as possible, as closed as necessary”, while meeting FAIR principles to ensure that data are Findable, Accessible, Interoperable and Reusable.

This document is the third version of the DMS Data Management Plan and includes information about:

- what data are collected, processed, and/or generated by the project
- which methodology & format is used
- how data are curated & preserved (including after the end of the project)

The previous two versions of this deliverable focused on confidential data that are used internally by DMS for the purposes of service provision and monitoring. During cohorts 2 and 3, the project began monitoring the wider impacts of the project and creating whitepapers for public dissemination. For this reason, this version of the Data Management Plan includes additional information about the reuse possibilities of DMS datasets and a new subsection regarding our management of Research Data.

1. Introduction

DMS Accelerator is a Horizon 2020 project that aims to support European data-centric start-ups and SMEs in overcoming barriers around data skills, entrepreneurial opportunities, legal issues, and standardisation. The project is expected to deliver a portfolio of 150 data-driven companies that gain demonstrable benefits from the free services provided by DMS. This includes impact in five different categories: (1) Fundraising; (2) Acceleration; (3) Standards & Legal; (4) Data skills and (5) Promotion.

The effective delivery of these services relies on various kinds of personal and non-personal data collected from startups who participate in the programme. The remainder of this section elaborates on the changes that have been made to our data management practices since the last version of the plan, the purposes of data collection, the types of data that are generated by the project and the stakeholders who control access to the data. Section 2 provides detailed descriptions of the datasets and the methods used to process, preserve and share them, followed by security considerations in Section 3 and Ethical Aspects in Section 4.

Summary of changes

The Data Management Plan presented here builds upon an earlier version of the plan, which was presented in DMS deliverable D1.7. The present document includes the previous information and makes the following additions:

- Information about new datasets created by the project in Section 2, specifically:
 - Impact Survey
 - Survey of Startup Needs
- Updates to the reuse possibilities, data sharing mechanisms and archiving of research datasets.
- A new subsection on research data, which describes how DMS datasets are being used for research and dissemination purposes.
- An elaboration of how our research activity is being accommodated through data sharing and privacy policies.

Purpose of the data collection and generation

DMS collects data about SMEs for the following purposes:

- **Contacting and engaging** with the enterprises who will benefit from the services, in the form of marketing communications and individual communications when support is requested.
- **Providing services** such as fundraising, acceleration, standardisation & legal advice, data skills training, and promotion.
- **Monitoring and analysing** these services and their outcomes, in alignment with the project's Key Performance Indicators.
- **Disseminating the results** of such services, primarily in the form of white papers and research articles.

Project partners

The data generated by DMS Accelerator will be used by the project partners, whose roles and relevance to the data management plan are specified in Table 1.

Table 1. DMS project participants

Organisation	Short name	Country	Role
Zabala Innovation Consulting S.A.	ZABALA	ES	Coordination. Set up of the mechanism of monitoring. Relations with additional initiatives from Big Data Value PPP (European Data Incubator (EDI), Big Data Value Ecosystem Project (BDVe)), FIWARE Foundation, Innovation Radar, Start-up Europe. Zabala handles the DMS email account and therefore holds the email details of participants.
Ogilvy worldwide S.A.	OG1	ES	Building of the contents for the promotion portfolio of initiatives. They provide training and may hold details of training participants for communications specific to the training sessions.
University of Southampton	SOTON	UK	Data science courses provider. They hold details of course participants.
King's College London	KCL	UK	Data science courses provider. They hold details of course participants.
IPtector Consulting	IPtector	DK	IPR training and GDPR. They provide training, and may hold details of training participants for communications specific to the training sessions.
GEIE ERCIM	ERCIM/W3C	FR	Standardisation. They will provide training, and may hold details of training participants for communications specific to the training sessions.
Bright Pixel	BRPX	PT	Company builder owned by a corporation. They provide training and mentoring and may hold details of participants for communications specific to these purposes.
Spherik Accelerator	SPHK	RO	Accelerator in Eastern Europe. They provide training and mentoring and may hold details of participants for communications specific to these purposes.
Spinlab	SPINLAB	DE	Accelerator in Germany. They provide training and mentoring and may hold details of participants for communications specific to these purposes.
The Next Web	TNW	NL	Incubator, Media publication, Data market itself. They provide promotion in the TNW conference, training, and mentoring, and may hold details of participants for communications specific to these purposes.

Types of personal data generated/collected

A majority of the information generated by DMS is non-personal data, as it relates to the needs and ambitions of startups and SMEs who enter the programme. Any personal information that is collected by the project is used for the purposes of service delivery. There are three types of personal data collected by DMS:

- **Contact information** i.e. names, email addresses, and phone numbers – in order to communicate with applicants, both prospective, successful and unsuccessful. The legal basis for this are (unless otherwise specified) legitimate interests – to identify potential data providers.
- **Recordings of DMS Accelerator webinars.** These webinars will contain voice and written contributions from the participants. It will be necessary for DMS Accelerator legitimate interests to keep these recordings, and release them as services to participant SMEs.
- **Course participation data.** Participants' interactions in the online courses provided by DMS Accelerator will be recorded and analysed for monitoring and course evaluation purposes. The course participants will be identified through their names and email contact details.

2. Data description

Overview of Datasets

The project collects information and generates datasets of various types and formats. Table 2 is a list of the datasets that have been generated during each of the three DMS cohorts (cohort 3 data is currently undergoing collection until M36).

Table 2. Overview of datasets collected and generated by DMS.

Dataset reference	Work package	Dataset description
DMS-2A	2	Portfolio of European SMEs approached by DMS
DMS-2B	2	DMS applicant SMEs
DMS-2D	2	Applicant SME ratings
DMS-2E	2	DMS participant portfolio
DMS-3A	3	Webinar recordings
DMS-4A	4	Webinar engagement reports
DMS-4B	4	Course engagement reports
DMS-4C	4	Impact survey data
DMS-4D	4	Survey of startup needs

Each of the datasets used in this project is described below. Each description contains the following information:

- Dataset reference: A unique reference to each of the datasets, as listed in Table 2.
- Relevant work package/s
- Type
 - Collected

D1.8 Data Management Plan adhering to the H2020 Open Research Data Pilot v3

- Generated
- Origin
 - Where does the data in the dataset come from?
 - From which sources it has been collected?
- Scale: the size of the dataset in Mb/Gb
- Description: A description of the dataset, including fields when tabular.
- Useful to: the reason why it is worth generating, keeping and sharing when applicable
- Methodology & format:
 - How the data will be input
 - Which tool will be used
 - The format in which the data will be kept
- Data sharing policy: stakeholders with whom the data will be shared
- Archiving and storage: where the data will be stored.
- Preservation time
- Additional preservation cost

Portfolio of European SMEs approached by DMS

Dataset reference	DMS-2A
Work package	WP2 SME engagement
Type	Collected
Origin	Project partners' networking connections
Scale (approx.)	Small (~50Mb)
Description	<p>All SMEs approached to apply for DMS services. The dataset contains the following fields:</p> <ul style="list-style-type: none"> ○ SME name ○ SME URL ○ Partner-owner (who owns the contact) ○ Contact name ○ Contact email ○ Notes (description, comments)
Useful to	The DMS consortium in approaching companies.
Methodology & format	For cohorts 1 & 2, the dataset was manually created in Google Spreadsheets and data were manually entered by participants. In cohort 3, Airtable was used to create and automate the data inputs.
Data sharing policy	Internal
Archiving and storage	These datasets are stored in the project Google Drive (cohorts 1 & 2) and Airtable (cohort 3) and will be destroyed at the end of the preservation time. This includes the destruction of any copies on online platforms and personal computers.
Preservation time	Five years beyond the end of the project.
Additional preservation cost	None.

DMS applicant SMEs

Dataset reference	DMS-2B
Work package	WP2 SME engagement
Type	Collected
Origin	Online form in DMS website (cohorts 1 & 2) & TNW website (cohort 3) completed by applicants
Scale (approx.)	Small (~100Mb)
Description	<p>List of all SMEs who applied for DMS services by filling a form. The dataset contains the following fields:</p> <ul style="list-style-type: none"> o Company name and email o Applicant name o Company financial details (yearly revenue, % of public funding) o Growth stage of the company o Relationship with the data market o Company category o Types of services applying for o Description of equality and/or social responsibility policy o Description of needs relevant to DMS services.
Useful to	The DMS consortium in selecting SMEs for providing them with services.
Integration and Reuse possibilities	Reuse possibilities could include systematic analyses of the needs expressed by European data-driven startups and SMEs. This could inform other publicly-funded innovation programmes in future.
Methodology & format	<p>Automatically collected from the online form linked to the DMS website, over three iterations, one per year.</p> <p>The online form lies in INDEX, a daughter company of TNW. INDEX is built with Datalicious, and one of their tools are online forms.</p> <p>The online form generates a CSV, only accessible to TNW. The dataset is then shared with the other DMS partners in Airtable for the purposes of selecting successful startups.</p>
Data sharing policy	<p>Internal.</p> <p>Controlled access to an anonymized version of research data may be considered in future through a data sharing agreement (e.g. with trusted colleagues who want to undertake research).</p>
Archiving and storage	This dataset will be stored in the servers of INDEX TNW, the partner in charge of WP2.
Preservation time	<p>Five years beyond the end of the project.</p> <p>After this, an anonymized version of the data may be kept to meet requests for controlled access to research data.</p>

Additional preservation cost	None
------------------------------	------

Applicant SME ratings

Dataset reference	DMS-2D
Work package	WP2 SME engagement
Type	Generated
Origin	TNW's DMS Applicant SMEs dataset (DMS-1B)
Scale (approx.)	Small (~100Mb)
Description	<p>All data contained in DMS-2B plus numerical and categorical ratings to each of the SMEs from four of the accelerator partners on the following categories:</p> <ul style="list-style-type: none"> o Motivation (0-10) o Services required (0-5) o Estimated annual revenue (1-3) o Public funding <40% (Yes/No) o Equality (0-1)
Useful to	The DMS consortium in selecting SMEs for service provision
Integration and Reuse possibilities	Reuse possibilities could include analysing application characteristics (e.g. computer-aided text analysis ¹) in relation to human evaluation scores. This could be of interest to organisational scientists.
Methodology & format	<p>For cohorts 1 & 2, a CSV from DMS-2B was loaded in Microsoft Excel, and copies were distributed by Zabala to each of the rating partners. Partners then rated and returned the data to Zabala, who merged the ratings.</p> <p>In cohort 3, the method of sharing data and gathering ratings from partners was automated by TNW using the Airtable platform.</p>
Data sharing policy	<p>Internal.</p> <p>Controlled access to an anonymized version of research data may be considered in future through a data sharing agreement (e.g. with trusted colleagues who want to undertake research).</p>
Archiving and storage	This dataset will be stored in Zabala servers and destroyed at the end of the preservation time. This includes the destruction of any copies on Excel and personal computers.
Preservation time	<p>Five years beyond the end of the project.</p> <p>After this, an anonymized version of the data may be kept to meet requests for controlled access to research data.</p>
Additional preservation cost	None

¹ <https://journals.sagepub.com/doi/abs/10.1177/1094428109335949>

DMS participant portfolio

Dataset reference	DMS-2E
Work package	WP2 SME engagement
Type	Collected
Origin	TNW's DMS Applicant SMEs dataset (DMS-1B)
Scale (approx.)	Small (~100Mb)
Description	Services to which each of the 50 selected SMEs have signed up to.
Useful to	The DMS consortium in allocating services to selected SMEs
Integration and Reuse possibilities	The data were reused in combination with DMS-4C to monitor the impact of DMS services on selected startups and to assess the project's alignment with startups' self-perceived needs.
Methodology & format	CSV from the survey collected in Typeform.
Data sharing policy	Internal Controlled access to an anonymized version of research data may be considered in future through a data sharing agreement (e.g. with trusted colleagues who want to undertake research).
Archiving and storage	This dataset will be stored in Zabala's Typeform account and destroyed at the end of the preservation time.
Preservation time	Five years beyond the end of the project. After this, an anonymized version of the data may be kept to meet requests for controlled access to research data.
Additional preservation cost	None

Webinar recordings

Dataset reference	DMS-3A
Work package	WP3 Services
Type	Generated
Origin	DMS account in LogMeIn
Scale (approx.)	Medium (~100Gb)

D1.8 Data Management Plan adhering to the H2020 Open Research Data Pilot v3

Description	Recordings of all video teleconference webinar sessions run by all consortium members, participated by selected SMEs members of staff, over the three iterations of the project.
Useful to	Participant SMEs as beneficiaries of the DMS Accelerator services.
Integration and Reuse possibilities	None
Methodology & format	Webinars are recorded in GotoWebinar and shared privately (only those with the link can see) in YouTube, for streaming only. The Youtube video link is shared on the DMS Accelerator website, only accessible through a password.
Data sharing policy	Private, accessible to participant SMEs through password.
Archiving and storage	These files will be stored in Zabala's LogMeIn account, accessible for streaming through GoToStage (a tool within LogMeIn) during the preservation time period.
Preservation time	Five years beyond the end of the project.
Additional preservation cost	None

Webinar engagement reports

Dataset reference	DMS-4A
Work package	WP4 Monitoring and analysis
Type	Collected
Origin	DMS account in LogMeIn
Scale (approx.)	Small (~100Mb)
Description	Data about the interactions of participants in all webinars. Relevant fields are: <ul style="list-style-type: none"> o Name o Email o Engagement measurements o Satisfaction survey
Useful to	The WP4 leader for monitoring and analysing the performance of the services, and producing deliverables.
Integration and Reuse possibilities	The data were reused in combination with DMS-4C to monitor the impact of different DMS service categories.
Methodology & format	Automatically generated in LogMeIn. Downloaded by WP4 leader for aggregate analysis. Name and email fields deleted immediately after download.
Data sharing policy	Internal for individual-level data.

	An aggregated and anonymized version of the engagement data (e.g. collective webinar attendance, satisfaction ratings) are shared openly to support research articles.
Archiving and storage	<p>The individual-level dataset will be stored in DMS LogMeIn account, and destroyed when the account is deleted.</p> <p>An aggregated and anonymized version of the data to be archived in an open repository (e.g. Zenodo).</p>
Preservation time	<p>For individual-level data, five years beyond the end of the project.</p> <p>Aggregated and anonymized data are archived permanently.</p>
Additional preservation cost	None

Course engagement reports

Dataset reference	DMS-4B
Work package	WP4 Monitoring and analysis
Type	Collected
Origin	SOTON account in Canvas
Scale (approx.)	Small (~100Mb)
Description	<p>Data about the interactions of participants in online courses. Relevant fields are:</p> <ul style="list-style-type: none"> o Name o Email o Engagement measurements
Useful to	The WP4 leader for monitoring and analysing the performance of the services, and producing deliverables.
Integration and Reuse possibilities	None
Methodology & format	Automatically generated in Canvas. Downloaded by WP4 leader for aggregate analysis. Name and email fields deleted immediately after download.
Data sharing policy	Internal
Archiving and storage	This dataset will be stored in Soton Canvas account, and destroyed when courses in the account are deleted.
Preservation time	Five years beyond the end of the project.
Additional preservation cost	None

Impact Survey Data

Dataset reference	DMS-4C
Work package	WP4 Monitoring and analysis
Type	Collected
Origin	Google form completed by startups
Scale (approx.)	Small (~100Mb)
Description	Data about the impact of DMS on startups. Relevant fields include: <ul style="list-style-type: none"> o Name, email & company name o Increases in sales capacity, revenue, new clients o Additional funding gained o New products, datasets, services o Jobs created, gender composition of teams, collaborations o Success stories
Useful to	The WP4 leader for monitoring and analysing the performance of the services, and producing deliverables. Other partners involved in service delivery (WP3) and dissemination of success stories (WP5).
Integration and Reuse possibilities	Integrated with the startups' application data recorded in DMS-2B to assess how the DMS offering aligned with the participants' initial objectives.
Methodology & format	Automatically collected from the Google form and available in CSV format. The surveys are executed over three iterations, where one initial survey and one follow-up survey is released to each cohort at yearly intervals.
Data sharing policy	Internal. Controlled access to anonymized research data may be considered in future through a data sharing agreement (e.g. with trusted colleagues who want to undertake research).
Archiving and storage	These datasets are stored in the project Google Drive and will be destroyed at the end of the preservation time.
Preservation time	Five years beyond the end of the project. After this, an anonymized version of the data may be kept to meet requests for controlled access to research data.
Additional preservation cost	None

Survey of Startup Needs

Dataset reference	DMS-4D
Work package	WP4 Monitoring and analysis

Type	Collected
Origin	Google form completed by startups
Scale (approx.)	Small (~100Mb)
Description	Data about the needs of DMS startups. Relevant fields include: <ul style="list-style-type: none"> o Company name o Rankings of service categories by importance o Suitability ratings for different service types o Written explanations about the specific support required by the company
Useful to	Partners involved in targeting and delivering services in WP3. The WP4 leader for monitoring and analysing the performance of DMS services, and producing deliverables.
Integration and Reuse possibilities	Integrated with the startups' application data recorded in DMS-2B to assess how the DMS offering aligned with the participants' objectives.
Methodology & format	Automatically collected from the Google form and available in CSV format. The survey was implemented in cohorts two and three.
Data sharing policy	Internal. Controlled access to anonymized research data may be considered in future through a data sharing agreement (e.g. with trusted colleagues who want to undertake research).
Archiving and storage	These datasets are stored in the project Google Drive and will be destroyed at the end of the preservation time.
Preservation time	Five years beyond the end of the project. After this, an anonymized version of the data may be kept to meet requests for controlled access to research data.
Additional preservation cost	None

Research Data

Some of the data collected by DMS has research value outside the project. Specifically, the data can provide insight into topics of public and scientific interest such as the current state of the European startup ecosystem, the efficacy of certain interventions and the particular needs of data-driven startups and SMEs. These topics are being addressed by DMS through whitepapers and academic research articles that reuse the following datasets:

- o DMS-4A Webinar engagement reports
- o DMS-2E Participant portfolio
- o DMS-2B Applicant SMEs
- o DMS-4C Impact Survey

- DMS-4D Survey of Startup needs

In order to meet the European Commission's FAIR principles in relation to research data, DMS aims to ensure as much as possible that its research data are Findable, Accessible, Interoperable and Reusable.

Our project cannot make its research data fully open due to the commercially sensitive nature of company information and the possibility of de-anonymization in small sample sizes. In our case, there are 50 publicly shared companies in each DMS cohort whose confidential contributions could be identified even from generally anonymous information such as the industry in which they operate. Moreover, our decision to make secondary use of administrative data for research emerged as the project progressed, meaning that the participants' consent for open archiving of their data was not obtained from the beginning of the project. For this ethical reason, and to protect commercially sensitive information, DMS offers controlled access to anonymised research data. This can be requested by emailing the authors of DMS papers (e.g. manguiano@zabala.es and maria.1.priestley@kcl.ac.uk), who will evaluate each request and provide access to trusted parties who want to use the data for justified and compatible purposes (e.g. colleagues who want to undertake a comparative research study or a review).

The interoperability of DMS research data is supported through the use of the CSV file format, which does not require proprietary software for analysis.

In addition to making the research data accessible, interoperable and reusable by trusted parties, we are making them findable through metadata shared in public repositories. For example, the DMS monitoring data are described in a [Zenodo repository](#) and linked to DMS publications. These monitoring templates are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0). Other metadata will be shared as DMS continues to publish its research findings.

3. Data security

All personal and commercially sensitive data will be safely stored in the password-protected accounts of the platforms where the data is held (e.g. GoToWebinar, Typeform, and Canvas) or on the dedicated partners' internal servers. All DMS data security measures are compliant with the GDPR regulation. The coordinator makes sure to regularly update the access list of personnel within the teams of the partners to the files and sensitive information contained in them.

4. Ethical aspects

As a GDPR compliant data management plan, personal data will be requested only when strictly needed, and only for the purposes stated when data is requested (or 'compatible' purposes for reuse, such as research). Name and corporate email will be the only fields requested from participants. When requesting personal data, the respondents are shown a clear statement of the purpose of collecting and keeping such information. In cohorts 2 and 3, DMS updated its [Privacy Policy](#) to inform applicants of the research purposes for which their data may be used.