D9.6 Final Data Management Plan

H2020-DS-08-2017: SMOOTH

Project No. 786741
Start date of project: 01-05-2018
Duration: 33 months

Revision: 01

Date: 31-01-2021

Document Information

Document Name: Deliverable 9.6 Final Data Management plan

WP: 9

Revision: 01

Revision Date: 31.01.2021

Author: Rosa Araujo

Dissemination Level

Project co-funded by the EC within the H2020 Programme		
PU	Public	x
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)**	

Approvals

	Name	Entity	Date	Visa
Author	Rosa Araujo	Eurecat	31/01/2021	✓

Document history

Revision	Date	Modification
Version 1	31/01/2021	V1

SMOOTH project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 786741 (Art. 29.4)

Executive summary	This deliverable is the final version of the data management plan that was originally submitted in Oct-18 (D9.1), and updated April-19 (D9.5). It describes life cycle for the data collected, processed and generated during the SMOOTH project. The purpose is to detail how the data is being created, stored and backed-up, who owns it and who is responsible for the different data and which data will be preserved and shared according to the participation of the project in the Open Research Data Pilot.

Index

1.- Introduction

This document describes the final version Data Management Plan, and it constitutes and actualization of the Deliverable D9.1 Data Management Plan (DMP) outlined for the SMOOTH project submitted in month 6 (October 2018), and revision of the Data Management Plan submitted in April 2019. As the SMOOTH project, funded by the Horizon 2020, opted-in for the Open Research Data Pilot is required to develop several versions of the Data Management Plan (DMP), to specify what data will be kept, how, where, and the measures taken for the controlled access to data.

The Consortium has followed the guidelines described in **OpenAire** platform and the document "***Guidelines on Data Management in Horizon 2020***". A DMP describes the data management life cycle for all datasets to be collected, processed, or generated by a research project. It covers:

- The handling of research data during and after the project
- What data will be collected, processed, or generated
- What methodology will be applied
- Whether data will be shared or made open access and how
- How data will be curated and preserved

Data management refers to the everyday handling and workflow of research data during the active phase of a project as well as the practices that support long-term preservation, access, and use after the project has been completed. These activities can include planning, documenting data, formatting data, storing data, anonymizing data, and controlling access to data.

In summary, the management of research data in the SMOOTH project is based on the following rules:

- Provide a maximum level of security for sensitive data and personal data, including the exchange of personal and/or sensitive data between selected partners
- Use well-known, established repositories for publishing and archiving non-sensitive research data
- Encourage data providers to make non-sensitive data available using Creative Commons licences, e.g. CC-BY
- Raise awareness among researchers, companies, and public stakeholders for the importance of making non-sensitive research data available to the public

Being a project that ensures compliance with the GDPR and also, within the call of Digital security, we have made a significant effort to verify that all data management within the project and between partners have been done with maximum guarantees, and proof of this, two additional measures have been taken, which are:

1- A **Data Protection Impact Assessment (DPIA)**: *Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.* Even if the SMOOTH legal experts in the project (KUL, DVI, AEPD till its termination date) concluded that no DPIA was needed, because the SMOOTH processing activities are not supposed to be considered

as high risk, for showing good practice DPIAs have been conducted for each phase of the project.

2- A **Penetration test**: Which consists of attacking a computer system with the intention of finding the security weaknesses and everything that could have access to it, its functionality, and data. An assessment of the integrity of the platform to analyse potential vulnerabilities has been performed through security tests, using the **Common Vulnerability Scoring System (CVSS)**. The potential vulnerabilities have been classified and reported in terms of real impact and probability. A number of recommendations have been indicated to prevent the vulnerabilities identified, which we have acted upon and have been mitigated accordingly. In view of the final audit performed, the PEN test has concluded that the SMOOTH platform has high level of security. More detail about the Penetration test is given in D6.3 *Final version of SMOOTH cloud platform*.

1.- Data, Materials, Resources Collection Information

SMOOTH is an innovation action that delivers ICT-based solutions assisting Micro-enterprises (MEnts) to adopt and be complaint with the GDPR by designing and implementing an easy-to –use and affordable cloud-based platform service. A major part of the project outcomes are potentially susceptible of being protected for exploitation, therefore the data management plan identifies clearly which data will be kept confidential and which will be made openly available.

The purpose of this section is to provide a full description of the data that has been generated and stored during this project. Datasets of various nature have been collected, processed, and generated within the SMOOTH project.

Data collected	Format	Size	Partner	Where
Project management: timetables, deliverable plans, information sheets, minutes, reports	Textual (.docx) Tabular (.csv, xlsx)	<1 GB	EUT	EUT data repository, Redmine, Microsoft SharePoint
Anonymous surveys from interviews	Textual (.docx) Tabular (.csv, xlsx)	<1GB	EUT	EUT data repository
Official deliverables	Textual (.docx, pdf)	<2GB		EUT data repository, Redmine, Microsoft SharePoint, Project website
Journal articles, conference papers	Textual (.docx, pdf)	<1GB		Partners repository and SMOOTH Zenodo Community

Data in relation to the dissemination and communication articles	Bibliography of publications (.docx, .html, .xml) Publications & data (.pdf, .pdf/a, .docx, .xls, .csv) Text documents (.docx, .xls, .pdf) & photos (.jpg, .eps, .tiff) Website (.html, .css, .jpg) Usage statistics from website and social	<1GB		EUT data repository
Code and data used only for project analysis	Calculation tables, graphs (xls, jpg, tiff,)	<2MB	Technical partners	Partners data repository
Designs, images, and videos	* JPEG (.jpg, .jpeg) * TIFF (.tif, .tiff) * PNG (.png) *EPS (.eps) * MPEG-2 (.mpg, .mpeg, …) * MPEG-4 H264 (.mp4) * Lossless AVI (.avi) * QuickTime (.mov)	<3MB		Partners data repository

Table 1. SMOOTH project_Types of datasets

For what regards the data related to the implementation of the platform, the schedule envisaged during its development has set the framework that have conditioned the data to be processed during the different phases of the project:

March 2019	First pilot phase: algorithms training
July 2019	First initial ad hoc feedback to companies
August 2019 - July 2020	Beta testing of the platform
August 2020	Launch of the platform
August 2020 - October 2020	Validation of the platform
October 2020 - January 2021	Market validation

Table 2. Phases of technical data collection

1.1.- Description of the data

Apart from the data generated in the context of the project management (deliverables, minutes, agendas, reports, financial statements,…) and in the context of the dissemination & communication activities (publications, newsletters, analytics from the project website,…), for what relates the data linked to the technical implementation of the project, as mentioned in the previous versions of the DMP, the SMOOTH platform is composed of the following technological modules:

a. SMOODATA:
For the automated analysis of the Enterprises' databases. This module identifies the categories of the data that an Enterprise has stored. Based on this data, it can be assessed if the Enterprise:

i is only storing the personal data items in accordance with its policies or is storing personal data (by mistake) that it is not allowed to process.

ii is applying the data minimisation principle (i.e., it is only storing the personal information required to run its business) and only that. To this end, the module uses as input the questionnaire provided by the Enterprise upon registration in the SMOOTH Platform.

iii processes "Sensitive Personal Data" in the data repository.

The above analysis will be used by the Platform to report which data the Enterprise stores and whether the Enterprise:

1) has a legal basis for their personal data processing,
2) applies the data minimisation principle and,
3) stores sensitive personal data.

b. SMOOTEXT
It is for the analysis of the legal texts that are used by Enterprises to comply with Privacy Legislation, such as Privacy Policies or Cookie Policies and will have to be able to provide feedback about:

i the presence of all required mentions in the policy,

ii the readability, highlighting ambiguous or overly-complicated phrases,

and advise on required ameliorations.

The development of SMOOTEXT requires the development of algorithms and tools to process those textual documents, extracting key information and presenting it to the user for interpretation.

c. SMONLINE
SMONLINE analyses websites and mobile apps and will be composed of the following modules:

i the SMONLINE website

ii the SMONLINE-advertising

iii the SMONLINE-MobileApp

The data processing during the project has been adapted according to the different development phases that are composing the advanced technological modules.

The results of the automated compliance tests that these technologies have developed have been

supplemented with contextual information of the micro-enterprise under analysis. During the registration process in the SMOOTH Platform, the micro-enterprise is asked to fill in a questionnaire with contextual information on its processing activities. This information is to be used, together with the automated tests results, to generate a compliance report for the micro-enterprise, providing specific feedback on aspects of compliance/non-compliance, and in case of non-compliance guidance and recommendations on how to remedy the identified problem(s).

In Deliverable **D9.2 SMOOTH ELSA first year report** (Ethical, legal, and societal aspects), **D9.3 SMOOTH ELSA second year report**, and **D9.4 SMOOTH ELSA final report**, ***Section 2 SMOOTH data management and compliance with GDPR***, the data processing strategy under the GDPR is detailed in depth.

1.2.- Phases of collected and/or created data

For the SMOODATA module three phases were planned:

- **Phase 1: Data transfer / collection**

  In this phase the Data files were transferred from the recruited MEnts (as a data controller) to NEC and EURECAT (who were also data controllers in this phase). The MEnts were required to provide the Data to NEC and EURECAT by uploading the data files to a cloud-based platform in accordance with the instructions of EURECAT. The platform was provided by EURECAT, who intervened also as a processor at this stage, offering a Microsoft Office 365 private site to which Eurecat has subscription. SMOOTH was allocated in a private environment and EURECAT administered the rights of access, users, and roles.

  For the regulation of the relationship between the companies testing the platform and the SMOOTH Consortium, a Participation Agreement has been established, and it has been effective during all phases of the project (template Attached in Annex I).

  Once the Participation Agreement was signed by both sides, the contact person of the MEnt received a link to the **Microsoft** site where each MEnt had its own independent folder for the delivery of information. Users could upload documents directly to **OneDrive.** OneDrive is a file hosting service and synchronization service for hosting files in the cloud. OneDrive offers a simple way to store, sync and share various types of files.
  The access to OneDrive space was done directly from a web browser.
  The site where the information was shared, counted with advanced protection from viruses and cybercriminals, and kept the information private and secure.

  The folder that was enabled for each MEnt, was composed of three sub-folders that served to categorize each of the pieces of information to be analysed.

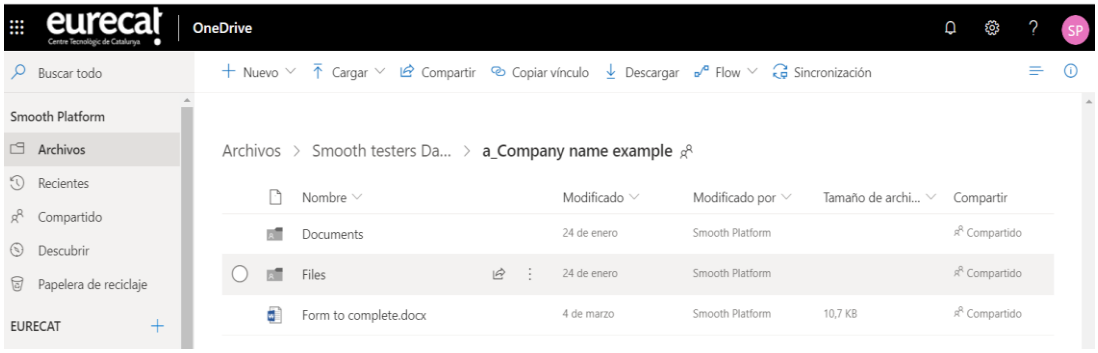

Figure 1. SMOOTH SharePoint

- In folder **Documents**, MEnts were asked to upload the legal documents needed for the SMOOTEXT analysis (Privacy policies, cookie policies). The formats of documents provided were mainly .doc, .pdf, .xls or csv.
- The folder **Files** was intended to collect the databases that were needed for the SMOODATA analysis, such as customer, providers, and employees' data files.
- The file **Form to complete** was a template in docx. format to be filled-in by the MEnt's representative and in which a summary of the basic data of the company was collected, as well as their **APK** (file format used by Android operating system for distribution and installation of mobile apps and middleware) and **URL** (Uniform Resource Locator, address that shows where a particular page can be found on the World Wide Web). APK and URL were to be analysed under SMOOWEB tools.

Once the documents and files were uploaded to the OneDrive site by the MEnts recruited, Eurecat shared the links to the MEnts' folders with the partners that needed the data collected for the analysis to be done by each of the technological modules, being:

- ⇨ Folders Documents: with technical members participating in WP3 (NAVER, EUT)
- ⇨ Folders Files: with technical partners participating in WP4 (NEC, EUT)
- ⇨ Form to complete files: with technical partners participating in WP5 (NEC, UC3M, IMDEA, LSTECH)

➢ **Phase 2: Development and training of the algorithms**

In this phase the Research Partners developed different models for personal data identification, based on the analysis of the content uploaded by the MEnts. New algorithms making use of the latest technology in machine learning were developed for allowing the analysis of multi-modal data (text, structured data, etc.).
The Research Partners were data controllers in this phase.
Once the technological modules have concluded its analysis, the data files have been deleted.

➢ **Phase 3: Interactive testing and assessment pilots**

In this phase SMOOTH project has evaluated the performance and usability of the Platform and its modules through two interactive piloting activities. The first pilot trialled an interim version of the Platform under a **controlled environment** and assisted by EURECAT. The link to the platform was not public and the credentials for accessing the site were provided by EURECAT.
The second pilot has revised an almost final version of the Platform with minimal support from Eurecat. In this stage, the link to the platform was public and the authentication and login was made directly by the users when browsing the link to the platform in the internet.
Following the GDPR minimisation principle, the information required to Sign up the system has been reduced to the necessary fields to run the platform and generate the report. The fields to be completed are:

- Email: because when the report is generated is automatically sent to the user's email
- Username: To be able to Log in the platform for accessing if a previous session has not been completed or to download the compliance report directly from the system
- Password: a strong security policy has been set and the system displays some requirements when establishing the password, being mandatory that the password selected contains at least four types of characters, such as lower case letters (a…,z), upper case letters (A,..,Z), numbers (0,..,9), special characters (¨;~%^)
- Company: For referring the compliance report to a given use case
- Country:
- Language: For the Data base module to select the library of the language for which it will need to analyse the potential files that the user may upload.

In order to ensure the integrity of the site, a **PENETRATION test** was carried out beforehand, which helped us to revise and assess measures and tools to make the environment more protected and consequently, more secure.

The participants were required to provide some feedback through several questionnaires and interviews, as detailed in the relevant Deliverables for the Platform validation D7.*2 Report of first evaluation of SMOOTH platform*, and D7.3 *Report of second evaluation of SMOOTH platform*.

For the trials during the data processing operations during the pilots, EURECAT has interacted with representatives of focus group of micro-enterprises to gather their feedback regarding the usability of the interfaces through questionnaires and monitoring of their physiological indicators as cardiac activity, skin conductance, respiratory activity, etc. A specific informed consent form has been designed to explain which data was to be collected and for which reasons, and it has been attached in Annex II of this Deliverable.

For the second pilot in order to ensure that the user had all the information regarding the type of data to be processed, our purpose and detail of who would

have access to the files and the security measures put in place, the process of log in the platform included a step in which the user is forced to open and scroll down the Participation Agreement. Until the user reaches the end of the document, the accept button is not enabled. Once this button has been activated, the user can continue to access the questionnaire on the platform.

Phase 4: Market assessment pilot

Once the platform was at a more mature stage and having incorporated the improvements derived from the feedback received from users in the previous phases, the final validation stage for the market assessment, was launched.

The platform in this phase has functioned as if it was a product launched on the market, with a public link and with total autonomy on the part of the users.

To gather the necessary feedback for the market assessment, a link was added from the same platform to a LIME survey. Within the instructions provided to the users, they were asked to fill in the survey with questions about the price they were willing to pay, the frequency of use of the platform, subscription modalities, etc. For the survey the user was anonymous, and the results and statistics where compiled in Eurecat's servers.

Once they had finished interacting with SMOOTH platform and after having read the Compliance report, they were asked to spend an extra minute to fill in the survey.
All the data files and document uploaded by a user, are automatically deleted once the compliance report is generated and sent through email to the address provided by the user when login in.

2.- Responsibilities of the partners

Each partner has been responsible for the data generated in their own premises and has assigned a responsible person from their institution to this task. All the Consortium has been obliged to follow the outlined data management policy in the best way possible.

Eurecat, as project coordinator, has been responsible for ensuring the implementation of the DMP and specifically for all aspects related with data storage and back up as data processor.

Over the life of the project, the Research Partners have:

- used the Data in the Project to develop the Platform,
- provided a technical solution to allow the recruited MEnts to upload the Data in a secure way, compliant with the Privacy Laws,
- processed the Data in strict compliance with the Privacy Laws,
- provided support to the participants.

Each Research Partner has received the entire Data set and has processed the Data separately from the other Research Partner, in its own name and on its own behalf as data controller. NEC has only been a data controller in Phases 1 and 2. EURECAT has remained data controller throughout both phases.

In relation to all personal data that has been processed by the Research Partners, the Research Partners always have complied in all material respects with their respective obligations under the Privacy Legislation.

The Research Partners have not used or otherwise process the Data for any other purpose than the purposes described in the project objectives.

The Research Partners have implemented and maintained all appropriate technical and organisational security measures to:

- Protect Data against unauthorised or unlawful processing, accidental loss, accidental or unlawful destruction, damage, alteration or unauthorised disclosure or access and against all other unlawful forms of processing and
- Ensure a level of security appropriate to the nature of the Personal Data, the possible risks, the state/ level of science and the cost of the implementation of such measures. In particular, the Research Partners have implemented all the security measures in accordance with the Privacy Legislation. Each relevant employee of the Research Partners having access to the relevant personal data have been advised of the contents of the technical and organizational security measures and legal obligations related to the Data that each one has access to.
- Assisted the participants by appropriate technical and organizational measures, for the fulfilment of the MEnts's obligation to respond to requests for exercising the data subjects' rights when and to the extent these relate to the Data processed by the Research Partners.
- Had it been the case, notified personal data breaches to the authorities according GDPR Art 33 without undue delay and within the legal delays, on becoming aware of any personal data breach, and taken all other measures as required by the Privacy Legislation.
- Had it been the case, notified personal data breaches to the data subjects when required (GDPR Art. 34).
- Performed data protection impact assessments when required.
- Not stored the Data longer than required and to, in any case, delete the Data within 6 months after the completion of the Project.

- Documented all steps undertaken to comply with the Privacy Legislation, to be able to demonstrate compliance with the Privacy Legislation.

2.1.- Intellectual Property

The Research Partners are the (co-)owner(s) and/or licensee(s) of the intellectual property rights (including, without limitation, any patent rights, model and design rights, topography rights, trademark rights, and/or any applications for such rights, copyrights, neighbouring rights, portrait rights, database rights, trade names and know how, as well as any similar rights hereafter "Intellectual Property Rights") in the Project, the outcome thereof, the Platform, any developments made during the project, documentation delivered and work carried out.

Software

The Platform functions through and in combination with the use of software necessary for its operation and good functioning. The Platform, or specific parts of it, may be installed on one or more internal and/or external processing units or computer systems of the recruited MEnts or may be made accessible via web applications.

The Platform uses intelligent machine learning / deep learning software that is continuously trained by analysing data to improve its analytical and recognition skills. Such training is done by the use of the Data of the recruited MEnts, and of own data sets.

3.- Data storage and back up during the research

With the purpose of safeguarding the appropriate preservation of the data, the Consortium has backed-up the data bases generated during the lifespan of the project. The data has been stored in databases installed on the partners' servers or in internal secured clouds. These databases are only accessible locally in order to prevent connections from outside.

Storage and maintenance of SMOOTH data has been handled according to the data category, privacy level, need to be shared among the consortium and its size.

SMOOTH has used Redmine as open-source project management web application for management files. All the Consortium members that have access to the Redmine have previously signed a Non-Disclosure Agreement with Eurecat:

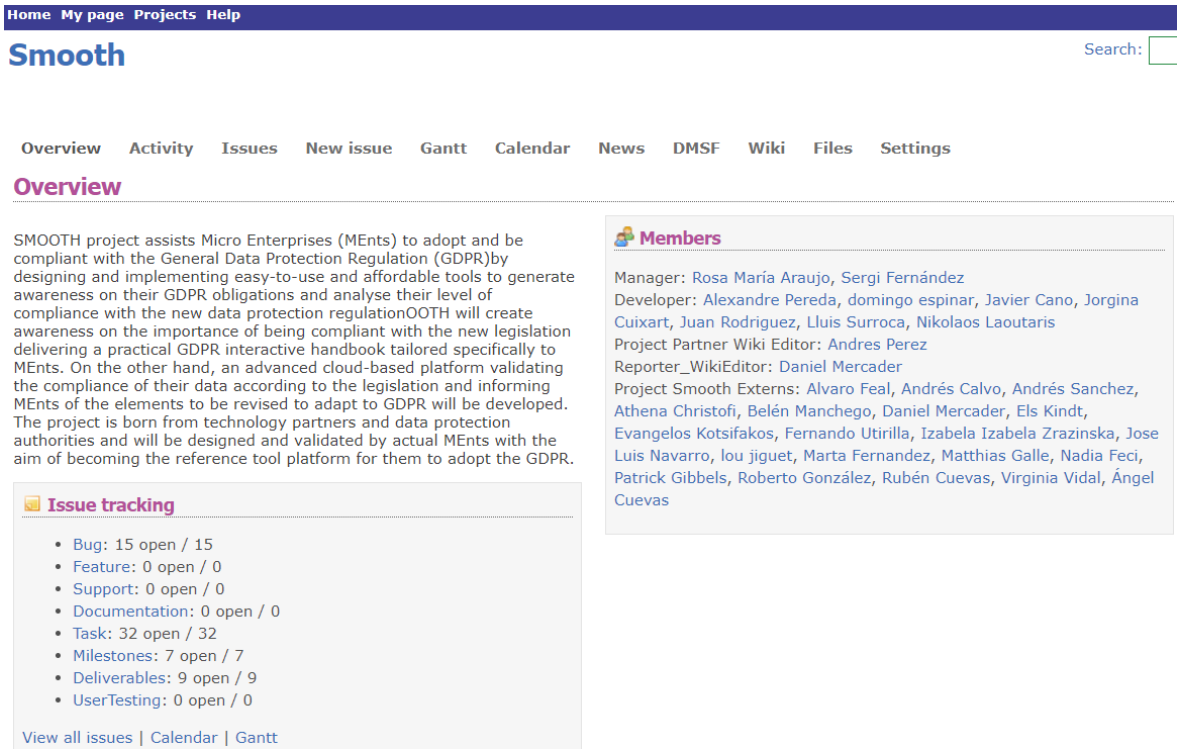

Figure 2. Redmine screenshot_SMOOTH project landing page

For the SMOOTH project different sections have been designed for compiling the information so that all the partners have access to the relevant information.

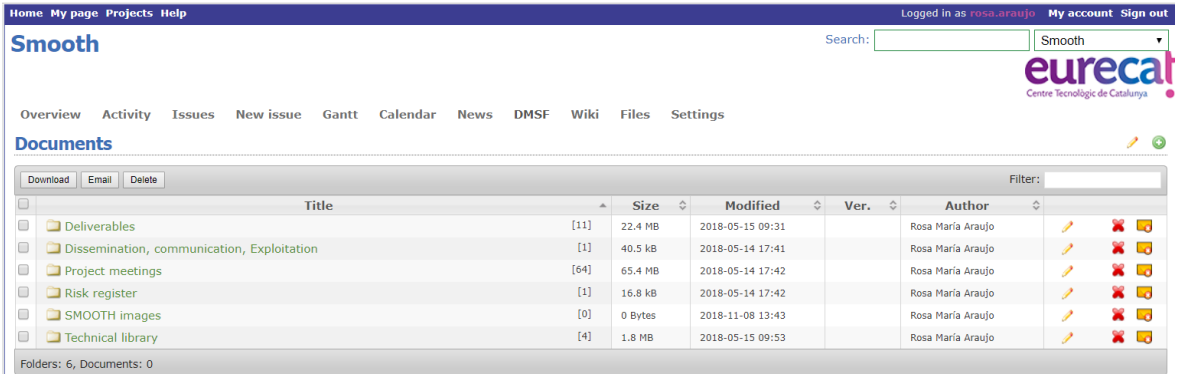

Figure 3. Redmine screenshot_project managent folders

The section **DMSF** has been intended to the collection of the relevant documents generated during the project:

Deliverables: The definitive version of each Deliverable has been uploaded to its individual folder.

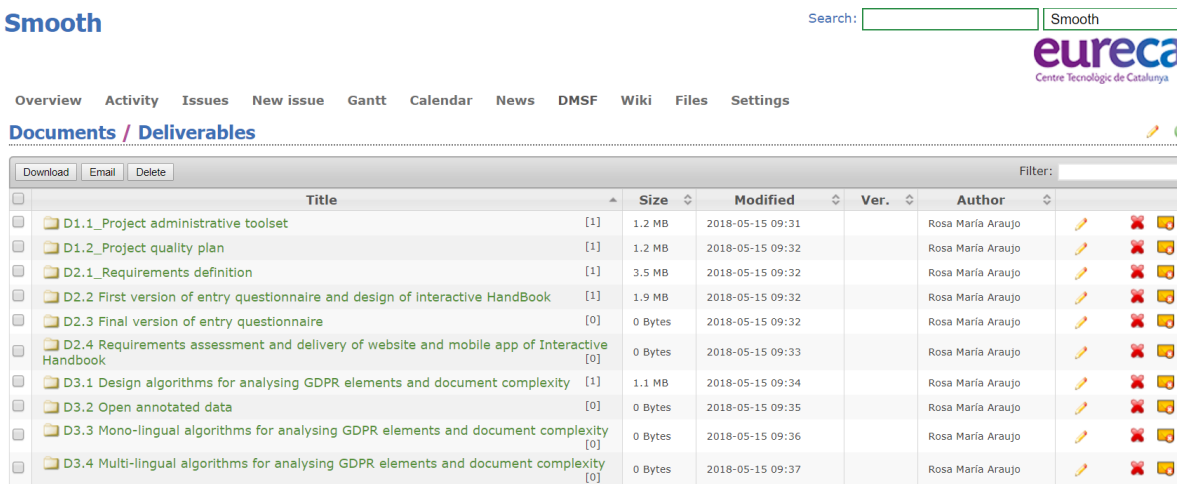

Figure 4. Redmine screenshot_Deliverables folder

Dissemination, communication and Exploitation: Relevant documents related to these items (brochures, dissemination materials, SMOOTH images) have been stored so that all the partners have access to them.

Project meetings: Every meeting has its own folder, inside each of the meetings folder, the minutes, list of attendants, agenda, the presentations used and practical information have been stored.

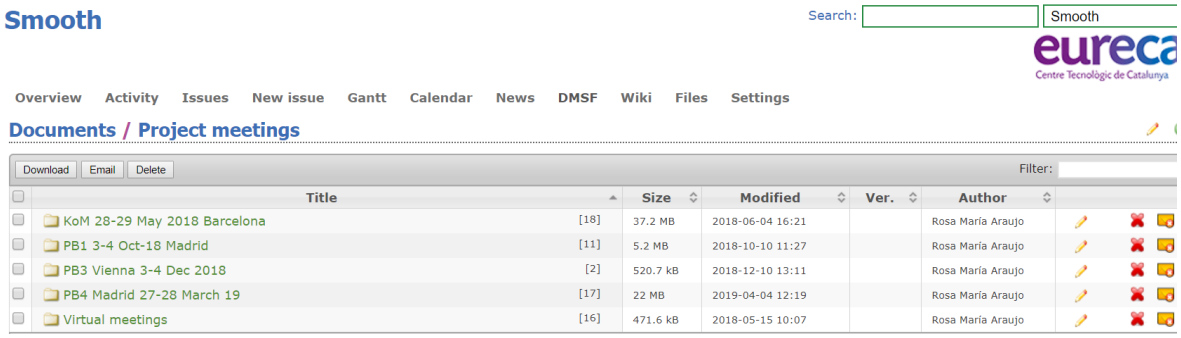

Figure 5. Redmine screenshot _management folders

SMOOTH images: To collect the images that are being designed for SMOOTH visual identity

Technical library: It is aimed at collecting the information, graphics, files, etc., that have served as source for each of the work packages and tasks.

Source code has been archived locally at partners' servers because of its privacy level. Local version control and software development platforms at partners' servers has been used. All software data has been being backed up in daily basis.

The electronic data generated during the research activities has been stored at partners' workstations and servers. Locally, research partners have secure servers on which all information is stored. The server drives are backed up periodically. The servers have a local firewall that only allows secure web connections to the Internet and verified IP addresses for developments or updates. Local log files have recorded every access to the servers.

The details about the warehouse model implemented and data information stored by the SMOOTH platform, can be found in D6.3 Final version of SMOOTH cloud platform, section 2.7.

Maintenance of datasets stored in partners' servers has been carried out according to the partners' backup policy.

For the data managed by Eurecat as data processor, the server is in Eurecat premises, a dedicated facility with controlled access and personal ID cards for authorised staff. The server has dedicated bandwidth and back up power system with the objective of guaranteeing availability.

4.- Copyright and Intellectual Property Rights (IPR) issues

The Platform is exclusive Partners' Intellectual Property, including any experience, improvements, and evolutions in the functioning of the Platform through the deep learning process, regardless of which and whose data was used to train the Platform. This does not affect the MEnts' Intellectual Property Rights on the material that such participant examines with the Platform, nor on the outcome of such examination or research. The Research Partners are entitled to analyse the Data, to mix the Data up with other data and to maintain such Data in anonymized form.

The project has delivered the following results on which the generated data sets depend:

	Key results	How is it going to be used?	Application areas
1	Module 1: GDPR key text document analysis	• Practical use, enabling the analysis of text documents related to privacy protection • In future research and technological solutions on multi-lingual text mining and machine learning analysis of document complexity	Data protection, privacy and transparency; as well as in other data science applications (e.g. other legal applications, insurances, medical, technical building code,…)
2	Module 2: GDPR personal data repository analysis	• Practical use, enabling the analysis of personal data repositories. • In future research and technological solutions on methods for data ingestion and analysis of database storage practices	
3	Module 3: GDPR auditing for websites and apps	• Practical use, enabling the analysis of the use of personal data in websites and mobile apps. • In future research and technological solutions on online advertising, data leakage…	Data protection, privacy and transparency; and more generally on online advertising, online crime, data security, …

	Key results	How is it going to be used?	Application areas
4	GDPR online interactive handbook for MEnt with kit of resources	• Practical use, as a communication and dissemination tool to reach a large number of MEnts and raise their awareness about GDPR • Freely accessible materials for GDPR compliance: templates, information and advice	Communication, awareness, education Data protection, privacy and transparency, Ments and citizens
5	Platform for GDPR compliance analysis	• Practical use, as integrated solution for EU MEnts to adopt the GDPR. • To be used as reference on how a software solution can support the adoption of new legislation by MEnts	Data protection, privacy and transparency; regulation, communication, awareness, education
6	CEN Workshop Agreement (CWA) Guidelines for Traditional Micro-SMEs' GDPR compliance	• Freely accessible materials for GDPR compliance: templates, information and advice for the direct use by MEnts (dissemination tools)	Data protection, privacy and transparency, MEnts of all sectors

Table 3. Project assets table

Research data and intellectual property generated within SMOOTH will be owned by the respective beneficiaries (solely or jointly when several beneficiaries have jointly carried out the work and if their respective share cannot be ascertained), according to article 26 of the Grant Agreement and section 8.2 of the Consortium Agreement, respectively.

The ownership and IPR of these assets belong to the partner that has generated them. For the platform exploitation the resulting agreements will be compliant with the corresponding legislation (i.e. GDPR, Copyright, Freedom for Information Act, etc.). More detail about the IPR claimed can be found in D8.5 Final dissemination, communication, exploitation, and standardization report.

5.- Sharing the data and OPENDATA

Following the general principle for Open access in H2020, and for facilitating an effective access of the public materials generated during the 33 months of the project, the SMOOTH partners have already delivered results to the identified SMOOTH target audience such as end users, groups or organisations, governance bodies and Scientific Community in concrete contributions as:

- ⇨ Web-site with easy access to retrieve public material generated within the project activities
- ⇨ Presence with technical papers, demonstrations or talks at relevant international conferences, workshops, ICT initiative, technical events and cooperation with European stakeholders
- ⇨ Cooperation with other projects in related areas
- ⇨ Production of leading-edge research material suitable for publication in international Journals and conferences.

⇨ Published research material publicly available following the "green" open access strategy.

A sample of the research results that SMOOTH Consortium has make accessible to contribute to better and more efficient science and to innovations in the public and private sectors are displayed in the following table:

Scientific publication (name of the journal/book)	Publisher	D.O.I. (*)	Title	Partner
The World Wide Web Conference	ACM Press	10.1145/3308558.3313601	An Intelligent system for Real-Time Filtering of Invalid Ad Traffic	IMDEA/UC3M
Communications of the ACM	ACM	10.1145/3426361	Does Facebook Use Sensitive Data for Advertising Purposes? Wordwide Analysis and GDPR impact	UC3M
		10.5281/zenodo.3461603	50 Ways to Leak Your Data: An Exploration of Apps' Circumvention of the Android Permissions System	IMDEA
Proceedings of the Internet Measurement Conference	ACM	10.1145/3355369.3355583	Tales from the Porn - A Comprehensive Privacy Analysis of the Web Porn Ecosystem	IMDEA
Proceedings of the 2019 Conference on Empirical Method in Natural Language Processing	Association for Computational Linguistics	10.18653/v1/d19-1222	To Annotate or Not? Predicting Performance Drop under Domain Shift	NAVER
Proceedings of the 15th International Conference on Emerging Networking Experiments and Technologies	ACM	10.1145/3359989.3365428	Detecting targeted ads via distributed counting	IMDEA
HAMLETS workshop, NeurIPS 2020	NeurIPS		Robust active learning strategies for model variability	EUT

IEEE Access	IEEE	10.1109/ACCESS.2020.3047343	Establishing Trust in Online Advertising with Signed Transactions	UC3M/IMDEA
IEEE Symposium on Security and Privacy	IEEE	10.1109/sp40000.2020.00013	An Analysis of Pre-installed Android Software	IMDEA
Internet Measurement Conference	ACM	10.1145/3419394.3423662	Understanding Incentivized Mobile App Installs on Google Play Store	IMDEA
Proceedings of the VLDB Endowment	ACM	10.14778/3415478.3415511	X2R2 A tool for explainable and explorative reidentification risk analysis	EUT
Journal of Advertising	Taylor and Francis Online	10.1080/00913367.2020.1749914	Tracking Fraudulent and Low-Quality impressions	UC3M
Electronics	MDPI	10.3390/electronics9081332	Malvertising in Facebook: Analysis, Quantification and Solution	UC3M
Electronics	MDPI	10.3390/electronics9111822	Digital Marketing Attribution: Understanding the User Path	UC3M

Table 4. SMOOTH publications

As one of the objectives of the SMOOTH project is to develop a solution into a commercial product, the Consortium may decide that potentially publishable data will not be available for open access until the end of the project, once the exploitation paths have been defined.

All public deliverables will be available at the SMOOTH website once they have received the formal approval from the European Commission. The Deliverables will be kept at least for three years after the project completion at the project website.

Selected datasets, databases, standalone documents, and software may be made public or open for exploitation at the end of the project. Metadata will be linked to the materials so that they can inform any future user about their usefulness and purpose of the dataset.

In order to make the non-confidential results openly available, the Consortium has used ZENODO. Zenodo servers are in Switzerland and that it still follows the GDPR requirements. The scientific publications have been shared on the website repository, where SMOOTH has a dedicated Community https://zenodo.org/communities/h2020-ds-sc7-2017-smooth-ga786741/

This public repository offers the following features:

- Safe — SMOOTH research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- Trusted — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.

- Citeable — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- No waiting time — Uploads are made available online as soon as we hit publish, and our DOI is registered within seconds.
- Open or closed — Share e.g. anonymized trial data with only specific professionals via Zenodo restricted access mode.
- Versioning — Easily update SMOOTH dataset with versioning features.
- GitHub integration — Easily preserve your GitHub repository in Zenodo.
- Usage statistics — All uploads display standards compliant usage statistics

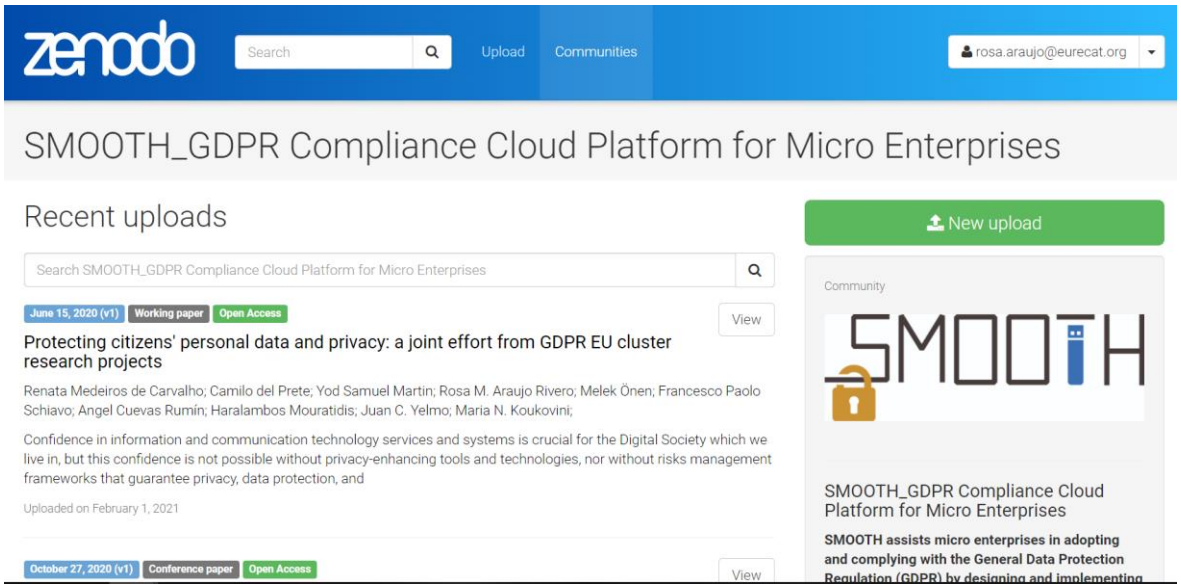

Figure 6.

A total of 25 publications about topics in connection with GDPR, which include anonymisation, data privacy, data utility, online tracking, secure digital identities, data minimisation, digital society, etc have been uploaded to the public site in Zenodo, being:

- 24 scientific publications in pdf format (the details of the publications can be also found in the Deliverables for dissemination, communication, exploitation (D8.3, D8.4, D8.5), and
- one datafile in tsv format, consisting of crawled privacy policies from European Privacy policies containing personal data

6.- Conclusion

With the current DMP, the SMOOTH project reports on the organisational measures which have been applied regarding the management of data during the SMOOTH project. Beside using established and renowned repositories which have been already successfully provided by partners, the report lists in detail the generated data formats that have been processed, stored, and shared among partners, stakeholders, and other targeted groups during and after the SMOOTH project. This report further includes data characteristics, security plans, as well as outlines of workflows regarding the data processing, physical security protection and privacy protocols which have been implemented.