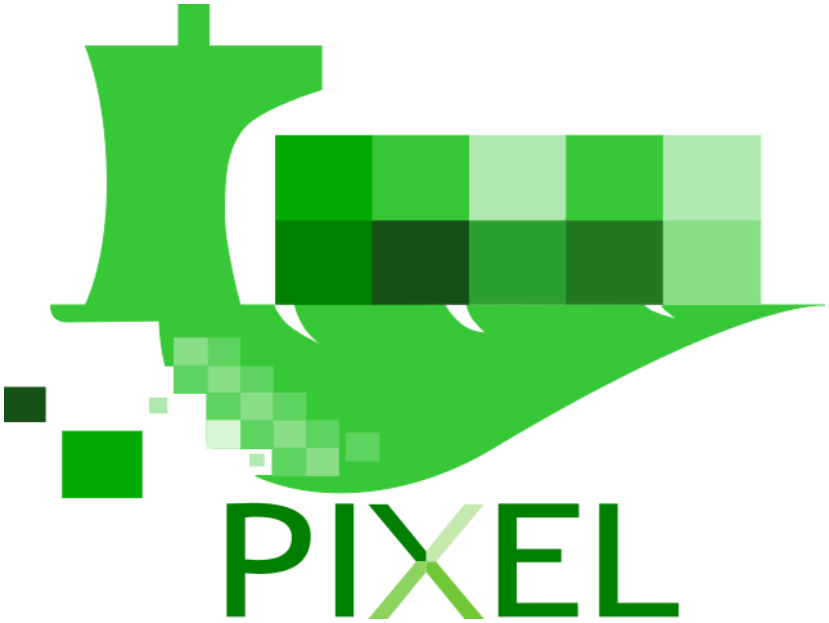
This project has received funding from the European's Union Horizon 2020 research innovation program under grant agreement No. 769355



D2.3 – Data – Management Plan v2

Deliverable No.	D2.3	**Due Date**	31-OCT-2019
Type	Report	**Dissemination Level**	*Public (PU)*
Version	1.0	**Status**	Final
Description	This deliverable includes information required in the Guidelines for Data Management in Horizon 2020		
Work Package	WP2		

Authors

Name	Partner	e-mail
Carlos E. Palau (PC)	UPV	cpalau@dcom.upv.es
Ignacio Lacalle	UPV	iglaub@upv.es

History

Date	Version	Change
5-SEP-2019	0.1	Template from D2.2 customized
15-OCT-2019	0.2	Introduction of detailed information of already known data sets
20-OCT-2019	0.3	Introduction of information of potential new datasets
31-OCT-2018	1.0	Version to be submitted to EC

Key Data

Keywords	Data sets Data protection, regulation, storage, processing, GDPR
Lead Editor	Carlos Palau, P01 UPV

Abstract

This deliverable has been created in the context of the Work Package 2 (*Work plan, coordination and document management*) of the H2020-funded project PIXEL (Grant No. 769355).

In this document the Consortium updated the DMP delivered in M6.

Deliverable D2.3 has advanced on the datasets identification and characterisation. This has been evidenced trough this document by including technical details over the previously detected ones and listing with a short description the new potential datasets to be generated in the project.

This DMP update and its future versions comply with Horizon 2020 FAIR Data Management Plan (DMP) Template.

Disclaimer

Table of contents

List of acronyms 4
1. About this document 5
1.1. Deliverable context 5
1.2. Methodology 5
2. Data Summary 7
2.1. Practical details of already known data sets 7
2.2. New potential data sets to be generated in PIXEL 11

List of acronyms

Acronym	Explanation
CA	Consortium Agreement
DMP	Data Management Plan
DPO	Data Protection Officer
EC	European Commission
EM	Ethics Mentor
GA	Grant Agreement
PC	Project Coordinator
PO	Project Officer
POPD	Protection Of Personal Data
ORDP	Open Research Data Pilot

List of tables

Table 1. Deliverable context 5
Table 2. Information of dataset: Project deliverables 8
Table 2. Information of dataset: Other internal documentation 8
Table 2. Information of dataset: Scientific publications 9
Table 2. Information of dataset: Other publications and outputs 9
Table 2. Information of dataset: Project deliverables 10

1. About this document

Following Data Management Guidelines for H2020 projects: Data Management Plan should be updated - if appropriate - during the project lifetime (in the form of deliverables).

Please create and submit a new version of the DMP whenever significant changes arise in your project such as:

- new data sets
- changes in consortium policies
- external factors.

This deliverable aims at updating D2.2 (PIXEL initial DMP) following the previous instructions and other indicated recommendations and commitments from the official current DMP of the project.

1.1. Deliverable context

Table 1. Deliverable context

Keywords	Lead Editor
Objectives	N/A
Exploitable results	Though not directly related with any Exploitable Result, updating and complying with the Data Management Plan will allow the project to develop a useful and regulatory compliant solution.
Work plan	This deliverable, even considered transversal, will be of application mostly in the preparation and execution of pilot every task in PIXEL in which data are involved.
Milestones	N/A
Deliverables	This deliverable is closely related with D1.5. While this document (D2.3) is rather generic for PIXEL data outcomes, D1.5 is focused primarily in the compliance of Data Management mechanisms with the current legislation, procedures and EC recommendations in relation to personal data.
Risks	Even not related directly to any of the Risks identified in the proposal, not following this Data Management Plan could imply certain IPR risks or further troubles related with the use and dissemination of data.

1.2. Methodology

In month M6, PIXEL Data Management Plan (DMP) was delivered. It outlined the measures that PIXEL project has put in place in order to accommodate for the requirements set for projects complying with FAIR data management and contributing to the Horizon 2020 pilot action on open access to research data.

The plan considered the protection of personal data and business confidential information. It identified the requirements for accessing existing datasets that form the basis of the work of the project. Pertaining to the data that the project will produce, the DMP initially identified the types of datasets that will be outcome of the project, namely: public deliverables, scientific publications, contributions to standards, software and applications:

- Project deliverables.
- Scientific publications.
- Contributions to standards.
- Software and Applications.

- Data (often traces) collected for analysis and evaluation.

But, according to DMP, these data may evolve during the project, e.g. anonymized data traces from the transport and logistics use case.

Additionally, also according to the DMP, the practical details on the management of the data sets will be provided later on through deliverables D2.3 (this one) and D2.4. During the project lifetime, additional information on the following aspects might be elaborated for all data sets on case by case basis before making consortium decision on handling of the particular data generated or collected.

Therefore, the scope of this deliverable has been two-fold:

- Add practical details of already known data sets to be generated by the project
- Identify new potential data sets to be generated by the project

And so it has been completed.

The original DMP was inherently compliant with the Horizon 2020 FAIR Data Management Plan (DMP) Template. It was stated that its future versions would as well be compliant with it, but some licenses have been taken for the writing of D2.3:

- Section 2, Data Summary has been kept. In the DMP this section served for summarising project objectives, architecture, main outcomes and main data concerns. For this deliverable (D2.3), we have made use of this section to introduce the main content needed: i) tables of detailed information on datasets and ii) list of new identified data sets. This structure has been put in the form of sub-sections of Section 2.
- The rest of sections of the FAIR DMP standard have been omitted since no change on the content has need to be clarified during these12 months.

2. Data Summary

In deliverable D2.2, these were the expected data sets to be collected and generated in PIXEL project is presented below:

- **Project deliverables.**
- **Scientific publications:** the scientific publications, mainly scientific papers, created by the consortium members, will contain technical results from the PIXEL project
- **Other publications and outputs:**

  Besides the scientific publications mentioned above, e.g. in journals or conference proceedings, it is expected that the project will generate further publications and other project outcomes, such as:

  - Promotion material (brochures, flyers, posters, etc.).
  - Press releases and further project announcements.
  - White papers created by the consortium on particular subjects.
  - Information regarding the open call.
  - Any further publication generated by the project.

- **Contribution to standards:** standardization constitutes an important dissemination activity in PIXEL project. It aims to contribute to the activities in major international standardization bodies, as defined in the PIXEL Description of Action (Part of the project Grant Agreement). For deployment of the new single metric to assess environmental impact (PIXEL's PEI) a standardization initiative within the targeted standardisation and regulation bodies is planned to be initiated. The activity will be performed by individual partners, group of partners, whole consortium or in the framework of different initiatives like CSA DocksTheFuture.

- **Software and applications:** the PIXEL project plans to develop and test several applications. In addition to the source code and binaries, documentation of the developed applications, their specifications, and other related material will be available in the project deliverables
- **Data collected for analysis and evaluation.**

In the next sub-section we are including the results of the analysis required of each of those data-sets for this time of the project.

2.1. Practical details of already known data sets

This section aims at covering the committed statement: *During the project lifetime, additional information on the following aspects will be elaborated for all data sets on case by case basis before making consortium decision on handling of the particular data generated or collected*:

- Nature and scale of the data in consideration.
- To whom it could be useful, targeted audience and its size and level of interest.
- Information on the existence of similar data and possible synergies.
- Possibility for integration and reuse of the provided data by external users and researchers.
- Any further related issue.

These explanations aim at covering the Annex 1 of DMP template by the European Commission.

The internal management of the information and the datasets is being handled using the Project Management tools deployed for the project, however because of the participation in ORDP, this procedure is ready to be accommodated to other different repositories.

With that aim, we have created a template table in which we can embed the associated information in a handy and useful format.

- **Project deliverables**

Table 2. Information of dataset: Project deliverables

Reference/name	Deliverables	**Data sharing method**		Public website (section deliverables)
Description	Public deliverables (according to GA) that will be generated during the execution of the project.	**To whom it could be useful**		Researchers, scientific community on the field
Type	Written documentation.	**Targeted audience**	**Size**	5-10 for official review by EC.
Related WPs and tasks	All WPs but WP1, which associated deliverables remain private.		**Description**	European Commission and technical reviewers
Format, standards	*DX.Y – [Name]*	**Possibility of integration and reuse of this data by external users and researchers**		Deliverables are not prone to integration but the information contained in them can be used for further researches (with the proper referencing and IPR respect).
Software	Generated after the writing of PIXEL partners. Software used can be text editors: MS Word, OnlyOffice or others.			
Estimated size	At this moment (M18), the total size of public deliverables (aggregated) is: 96Mb. It is foreseen about 200-250 Mb.	**Possible synergies with similar data**		N/A
Storage	In our secure repository OnlyOffice.	**Back-up**		A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included	Names of authors.	**Means for personal data protection**		Dealt with in CA.

- **Other internal documentation**

Table 3. Information of dataset: Other internal documentation

Reference/name	Other_internal_documentation	**Data sharing method**		Not to be shared publicly.
Description	Meetings minutes, agendas, internal documents of work, reporting docs…	**To whom it could be useful**		PIXEL partners.
Type	Written documentation.	**Targeted audience**	**Size**	70-80
Related WPs and tasks	All WPs		**Description**	People involved in the execution of the project.
Format, standards	*PPD_UC[No. use case]_Entity* and the rest of naming procedures detailed in deliverable D2.2	**Possibility of integration and reuse of this data by external users and researchers**		N/A
Software	Generated after the writing of PIXEL partners. Software used can be text editors: MS Word, OnlyOffice or others			
Estimated size	Current size of internal repository is: . Foreseen: .	**Possible synergies with similar data**		N/A
Storage	In our secure repository OnlyOffice.	**Back-up**		A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included	Yes, names and emails of PIXEL partners people.	**Means for personal data protection**		Dealt with in CA:

- **Scientific publications**

Table 4. Information of dataset: Scientific publications

Reference/name	Scientific_publications	**Data sharing method**		Editorials/magazines/conferences proceedings RiuNet (UPV) Website (space "publications") Researchgate
Description	Scientific articles published at Scientific Journals, books or Conferences coming directly from technical work in PIXEL and authored by PIXEL participants	**To whom it could be useful**		Scientific Community
Type	Written documents.	**Targeted audience**	**Size**	-
Related WPs and tasks	WP4, WP5, WP6, WP7 and WP9		**Description**	Followers in ResearchGate, Current active researchers in the field of transport and IoT
Format, standards	Documents to be stored and shared in pdf. In the case of images, the default format should be svg if possible or png or jpg otherwise. Formats are depending on the journal/conference that they aim at being published. Namely, Springer LNCS or others.	**Possibility of integration and reuse of this data by external users and researchers**		Publications can be integrated in new papers and scientific articles with the common tools of automated/classic referencing. Furthermore, the information contained in them can be leveraged for SotA studies for further researches (with the proper referencing and IPR respect).
Software	Generated after the writing of PIXEL partners. Software used can be text editors: MS Word, OnlyOffice or other			
Estimated size	Up to 6, 8, 10, 12 (or other) limit of pages set by the editorial. Average size ~4Mb per scientific article.	**Possible synergies with similar data**		Inner references in our scientific articles can be used for cascade literature research.
Storage	One copy of all articles is always stored in our secure repository. Own methods provided by the other data sharing means (check top right cell).	**Back-up**		Internal repository: A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup). Check policies by other means.
Personal data included	Names, emails and affiliations of authors and co-authors.	**Means for personal data protection**		Informed consents and similar procedures established by the editorial prior to publication.

- **Other publications and outputs:**

Table 5. Information of dataset: Other publications and outputs

Reference/name	Other_material	**Data sharing method**		Public website Social media channels Email to newsletter members
Description	Poster, leaflet, supporting material for communication, newsletter.	**To whom it could be useful**		Followers, casual readers, PIXEL partners, event promoters, audience of certain events/fairs
Type	Images, presentations, designs, full-size materials.	**Targeted audience**	**Size**	200/300 per year in conferences 8.000 in total through website 500 in total in social media
Related WPs and tasks	WP9		**Description**	Attendees of events, visitors of our website, followers in social networks

Format, standards	In the case of images, the default format should be svg if possible or png or jpg otherwise. The different files are expected to be below the 10 MB mark	**Possibility of integration and reuse of this data by external users and researchers**	This data is not prone to be integrated or reused by external users. It can be re-shared through social media by our followers. It can be used by event promoters to be included in publicity, promotion or event summaries
Software	These materials are produced by PIXEL partners using different software for edition: MS Visio, draw.io, GIMP, Photoshop and others.		
Estimated size	Average size of each material is ~6 Mb. Total size expected is 120-150 Mb.	**Possible synergies with similar data**	N/A
Storage	In our secure repository OnlyOffice.	**Back-up**	A mirror server for backup. Backup is done every day (delta changes) and each Sunday (full backup).
Personal data included	No	**Means for personal data protection**	N/A

- **Contribution to standards**

At this moment of the project, no contribution to standards have been initiated yet. Contacts with relevant entities and research of the current solutions and standardisation have been conducted but with no immediate action foreseen. If further actions happen, this information will be completed and detailed to make this activity compliant with the Data Management Plan.

- **Software and applications**

Table 6. Information of dataset: Project deliverables

Reference/name	Software_components	**Data sharing method**		Website (private access) GitHub Zenodo (through GitHub)
Description	Software modules developed in the project and that can be made public.	**To whom it could be useful**		EC, developers, open source community
Type	Software code with associated methodology	**Targeted audience**	**Size**	-
Related WPs and tasks	WP4, WP5, WP6 and WP7		**Description**	EC, developers, open source community
Format, standards	Containerisation practices will be followed when possible. License Apache 2.0 will apply.	**Possibility of integration and reuse of this data by external users and researchers**		All components have been created in a modular way including an interface for integration (API or others). All developments will be made public with associated markdown-type documentation (e.g. Readthedocs)
Software	The software modules are being generated after PIXEL partners' coding activities. The IDEs for development, frameworks and languages are diverse, such as Python, Java, Vue.js, IDE Eclipse, IDE VS Code… among others. The development platform is currently using 29 VCPUs, 60 GB RAM, 3 floating IPs, and 4 security groups.			
Estimated size	The size of the software is expected to be between 10-15 GB, mainly because they are provided as Docker instances and it increases the final size. Just considering the code in Github, without any Docker image, the size should be probably reduced up to a 60%	**Possible synergies with similar data**		Some synergies have been identified: 1) Federation and SSO: if the target port includes some delegation support (e.g. SAML), there is no need to even store user profiles in the PIXEL platform

			2) Storage: if the target port is also using Elasticsearch as database, both can be probably merged to reduce infrastructure costs
Storage	The Development platform is deployed in FIWARE Lab, whereas the pilot sites are supposed to provide their own infrastructure. This development platform is currently using 240 GB spread among the different services, being the Elasticsearch database (160 GB) the one with more demanding requirements (storage of historical sensor data. Models and predictive algorithms).	**Back-up**	The use of Docker instances allows us to just focus on the data for back-up purposes. The IH allows to create clusters redundancy for scaling and safety purposes. It is also possible to dump/export data from Elasticsearch or other databases (E.g. Mongo)
Personal data included	N/A	**Means for personal data protection**	The architecture includes a security module based on FIWARE security components (Wilma, Keyrock) to add OAuth2-based authentication and authorization security to PIXEL services and applications. Apart from user profiles, no personal data is stored in the system

2.2. New potential data sets to be generated in PIXEL

During the period M6-M18, several technical advances are being done in PIXEL. Altogether with technological developments in WP4 and WP6, diverse methodological and procedural outcomes are being generated in WP4 and WP5. Additionally, activities in WP3 and WP9 have provided several useful documents and reflections from external people. PIXEL partners have been working in parallel trying to identify potential new data sets that will be generated by PIXEL (or used during it and with a need of data set Data Management Plan inclusion). Below there is a list of preliminary potential data sets that would need to be further analysed later. Detailed information of them (as well as its final inclusion/exclusion on the DMP coverage) will be tackled in the next iteration of the DMP: deliverable D2.4. By that time, all data sets will have been finalised:

- **Online questionnaires for AB members**
- **Literature review on ICT and environmental fields for the technical development of the solution**
- **Data collected for analysis and evaluation:** Such a variety of data is necessary to produce tailored information which could be used directly by decision makers and stakeholders for their activities
    - Raw data from own sensors
    - Observational data from own web services
    - Observational data from external services
- **Results generated from model and PEI calculation**
- **Guidelines and recommendations related with the PEI adoption**
- **Historical data collection and processing for model training**
- **Data predicted generated after execution of predictive algorithms**
- **Pictures of PIXEL members presenting at events**:
- **Pilot deployments in the four ports participating in PIXEL**