

TOYLABS

ENABLING AN OPEN INNOVATION MODEL FOR EU TOY INDUSTRY SMES
 THROUGH CO-CREATION WITH FABLABS, SAFETY EXPERTS AND CUSTOMER
 COMMUNITIES
 732559

Data Management Handling Plan

Deliverable No.	7.2		
WP no	7	WP title	Exploitation and Business Innovation Planning
Activity No.	7.4	Activity Title	Project Data Management and Ethics Handling
Authors	SLG		
Status (final, draft, revised draft)	Final		
Dissemination level	Public		
File Name:	TOYLABS_SLG_WP7_IR_D7.2-v1.00		
Delivery date	13.07.2017		

ABSTRACT

The TOYLABS Deliverable D7.2 ("Data Management Handling Plan") presents the first version of the (open) data management plan (DMP) for the TOYLABS project as anticipated on Month 6 of the project. Such a data management plan investigates the appropriate methodologies, processes and open repositories for data management and dissemination to be adopted in the TOYLABS project.

The guiding principle of the TOYLABS project is openness as it aims at offering as much information as possible in a 'FAIR' manner (so as it is findable, accessible, interoperable and re-usable). Such information includes: the scientific publications prepared by the project consortium, the white papers published, the open source code generated, the public deliverables, the mock-up datasets used for supporting the development process etc. The list of research data expected to be released during the project consists of open methodologies (under creative commons licences), open source software components, and anonymous user statistics (from pilot operation). Such datasets are expected to be collected during the implementation and evaluation phases of the TOYLABS project and are therefore subject to change.

The publishing platforms, that can be accessed openly and are adopted by the TOYLABS project, are: the project website, Zenodo for long-term archiving (as proposed by EC), ResearchGate (for archiving and preservation of publications) and GitHub for open source code.

This deliverable is released as a living document that is expected to be updated whenever it is needed and to be respected by all partners.



VERSION HISTORY TABLE

Version no.	Date	Author	Changes
0.10	10/05/2017	SLG	Initial table of contents
0.20	19/05/2017	SLG	Initial deliverable draft
0.30	02/06/2017	SLG	Updated Deliverable draft and Updated ToC
0.40	09/06/2017	SLG	Initial TOYLABS publishing methodology (Section 2)
0.50	20/06/2017	SLG	Update on TOYLABS publishing methodology (Section 2)
0.60	22/06/2017	SLG	Draft of Section 3 and Section 4
0.70	23/06/2017	SLG	Revision of Publication Platforms (Section 3)
0.80	29/06/2017	SLG	Updated Publishable Results (Section 4)
0.90	03/07/2017	SLG	Final Draft for Review
0.95	12/07/2015	AIJU, NTUA	Peer Reviewd Version
1.00	13/07/2017	SLG	Final Version to be submitted to the EC



TABLE OF CONTENTS

1	Introduction	6
1.1	PURPOSE OF THE DOCUMENT	6
1.2	STRUCTURE OF THE DOCUMENT.....	7
2	Data Management Handling Methodology.....	8
2.1	BASIC RULES.....	8
2.2	DMP MANAGEMENT PROCESS.....	10
2.3	DMP RESPONSIBILITIES.....	13
3	TOYLABS Data Publishing Platforms	14
3.1	TOYLABS PROJECT WEBSITE	14
3.2	RESEARCHGATE.....	15
3.3	ZENODO	15
3.4	GITHUB.....	15
4	TOYLABS Publishable Results.....	16
4.1	TOYLABS PROJECT PUBLIC DELIVERABLES.....	16
4.2	SOFTWARE COMPONENTS	19
4.3	PUBLICATIONS.....	19
4.4	TOYLABS PROJECT DATASETS	20
5	Concluisions and Next Steps.....	21
	Annex I: Literature	22



LIST OF FIGURES

Figure 1: Open access strategy for publications and research data	9
---	---

LIST OF TABLES

Table 1: Initial Dataset Template.....	11
Table 2 List of TOYLABS Public Deliverables.....	16
Table 3: List of open source software components.....	19
Table 4: Project Research Artefacts.....	20

ABBREVIATIONS LIST

DoA	Description of Action
DMP	Data management plan
WP	Work Package



1 INTRODUCTION

Each project funded or co-funded by the EC Horizon 2020 program has to follow a set of rules and guidelines for open access to scientific peer reviewed publications and research data as defined in the EC Guidelines for Open Access [1]. A Data Management Plan is also required for all projects participating in the extended Open Research Data pilot.

In order to effectively undertake the TOYLABS research data management activities and actively take part in the Pilot of Open Research Data, the TOYLABS consortium shall follow a concrete methodology for data management handling to diffuse the project outcomes in broad audiences as defined in this deliverable.

1.1 Purpose of the Document

The present deliverable D7.2 "Data Management Handling Plan" is created in the context of the activities of Task T7.4 "Project Data Management and Ethics Handling" under WP7 "Exploitation and Business Innovation Planning". According to the TOYLABS Description of Action, T7.4 will provide the required open data management plan that will aim to communicate and spread knowledge on project results to all interested communities and stakeholders. An open data management plan will also generate wider interest towards improvements achieved by the project in order to facilitate and potentiate exploitation opportunities.

In brief, the goal of this deliverable is to investigate the appropriate methodologies and open repositories for data management and dissemination, and to offer through open access as much TOYLABS information as possible. Such information would be scientific publications issued by the project consortium, white papers published, open source code generated, mock-up datasets used for supporting the development process etc.

As the data management plan methodology is expected to be followed along the project duration, D7.2 is envisaged as a "living" document that is expected to be refined in further detail and updated whenever significant changes arise such as (but not limited to):

- new data are generated
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

Such updates will constitute part of the TOYLABS deliverables D7.4 "Exploitation Report, Updated Plan and First Release of Business Innovation Plan" (where the



definition of the datasets and components developed in the project will be provided) and D7.5 "Final Exploitation Report and Business Innovation Planning", which are due on M12 and M18, respectively.

1.2 Structure of the Document

This deliverable is structured in the following sections:

- Section 1 defines the purpose of the document and introduces its structure.
- In section 2, the methodological framework for handling the results and the data collected or generated during the project, is elaborated. The overall analysis is aligned (actually a summarised version) to the methodology proposed by EC.
- In section 3, the data publishing platforms that will be put into use in order to ensure that the data will be exploited and/or shared/made accessible for verification and re-use along with the data preservation and maintenance processes, are briefly presented.
- In section 4, the TOYLABS publishable results are referenced. For each result, we provide - in accordance to the data management guidelines [1] and [2] - a short description, the chosen way of open access, and a long-term storage solution.
- In section 5, brief conclusions are provided.
- In Annex A, the references for this deliverable are listed.



2 DATA MANAGEMENT HANDLING METHODOLOGY

The TOYLABS Data Management Plan (DMP) is instrumental for effectively collecting, managing and publishing data produced during the lifecycle of the TOYLABS project, leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. It takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.

As part of making the research data findable, accessible, interoperable and re-usable, a DMP should include information on [1]:

- the handling of research data during and after the end of the project,
- what 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).

TOYLABS will strive to make its research data 'FAIR', that is findable, accessible, interoperable and re-usable following the Horizon 2020 FAIR Data Management Plan (DMP) template [2].

2.1 BASIC RULES

Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

1. peer-reviewed scientific research articles (published in scholarly journals) or
2. research data (data underlying publications, curated data and/or raw data).

Open access [1] to scientific publications means free online access for any user. The 2 main routes to open access are:

- Self-archiving / 'green' open access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.



- Open access publishing / 'gold' open access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out by the consortium. Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge. The next figure presents the process flow towards defining the open access type in scientific publications and research data.

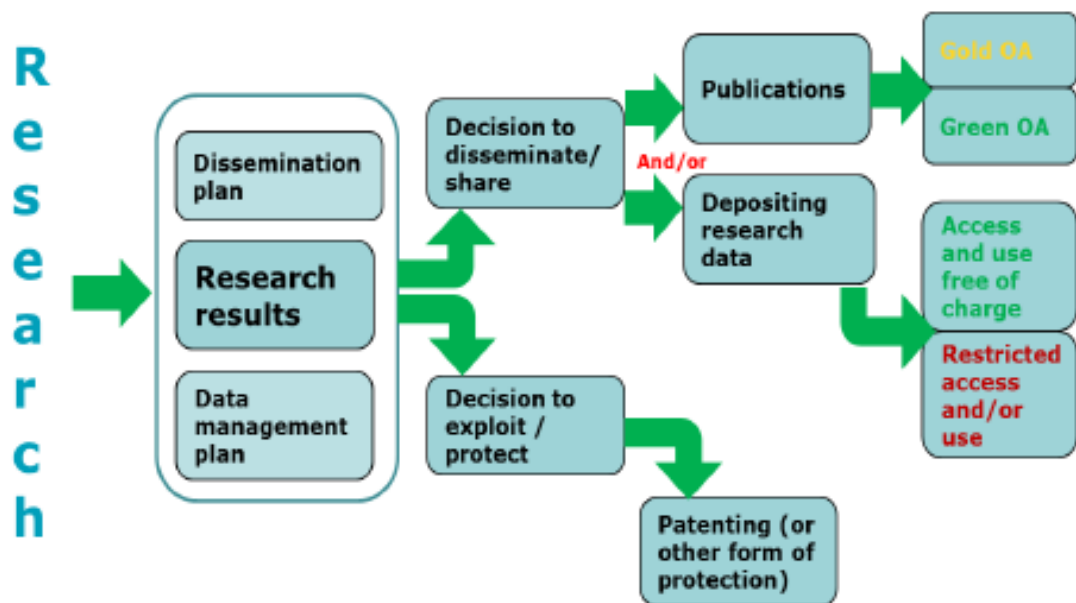


Figure 1: Open access strategy for publications and research data

The open access mandate comprises 2 steps:

1. depositing publications in repositories
2. providing open access to them



2.2 DMP MANAGEMENT PROCESS

Taking into account the preliminary analysis conducted in the first months of the project implementation, the TOYLABS Data Management Process is defined as a step-wise approach for each result generated or collected during the project lifecycle. The following questions must be answered to classify the different datasets:

1. Does a result provide significant value to others or is it a precondition to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. For example, pilot results that are very specific to the TOYLABS platform are usually of no scientific interest to anyone, nor does it add any significant contribution.

2. Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethics management plan of the project.

3. Does a result allow the identification of individuals even without the name?

This is also a step managed by the ethics management plan of the project as we have committed in the TOYLABS project to establish anonymisation techniques to conceal a single user's identity, e.g. abstraction, dummy users, or non-intersecting features. If this question is answered with yes, the result is classified as non-public.

4. Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and the project's ethics?

This is also a step managed by the ethics management plan of the project. If this question is answered with yes, the result is classified as non-public.

5. Does a result include business or trade secrets of one or more partners of the project?

If this question is answered with yes, the result is classified as non-public. Business or trade secrets needs to be removed in accordance to all partners' requirements before it can be published.

6. Does a result name technologies that are part of an ongoing, project-related patent application?



If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after the patent has been filed.

7. Does a result break security interests for any project partner?

If this question is answered with yes, the result is classified as non-public.

This is a simple structural approach to determine the different data types defined as part of the DMP.

Taking into account the Horizon 2020 FAIR Data Management Plan and the metadata used in open data platforms, the following table has been formulated and shall be used for completeness purposes in the TOYLABS project for the definition of the metadata associated to a dataset. It needs to be noted that the template is based on the Dublin Core and some of the metadata may not be applicable depending on the platform where it will be published.

Table 1: Initial Dataset Template

Metadata	Description
Reference ID	A unique identifier for the data set to be produced based on a common naming convention for the project (e.g. TOYLABS_Dataset_[number]).
Name	The name for the data set to be produced by the project.
Description	Description of the data that will be generated or collected, its nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existences (or not) of similar data and the possibilities for integration and reuse.
Creator	The entity responsible for the dataset to explain the origin of the data (in case it is collected).
Date_Published	The date when the dataset was originally published by the TOYLABS consortium (i.e. YYYY-MM-DD depending on the platform).



Date_Last Update	The date when the dataset was last updated (i.e. YYYY-MM-DD depending on the platform).
Type	The type and format of the dataset.
License	<p>The license associated to this dataset (e.g. exact licencing framework, embargo periods, commercial exploitation, etc.)</p> <p>If an embargo is sought to give time to publish or seek patents, it needs to be specified why and how long this will apply. It also needs to be clarified whether the re-use of some data is restricted, and how long it is intended that the data remains re-usable.</p>
Version	The version of the dataset (i.e. vn-m where n is the major version and m the minor version)
DOI	The Digital Object Identifier, if applicable.
Search Keywords	A list of keywords to facilitate search and optimise possibilities for re-use.
Standards	<p>Reference to existing suitable standards of the discipline related to this dataset. If these do not exist, an outline on how and what metadata will be created.</p> <p>Reference to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins, if applicable.</p>
Archiving and preservation	Description of the procedure that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end, volume, what the associated costs are and how these are planned to be covered.



2.3 DMP RESPONSIBILITIES

The TOYLABS consortium is committed to ensure that the DMP processes and policies are respected and applied throughout the project activities. A dedicated time slot will be reserved at each plenary meeting and, if needed, at selected consortium teleconferences. The EC and project reviewers will be informed about related work done and publications provided in the respective WP7 deliverables and the project management reports.

In general, data have to be created, managed and stored appropriately and in line with the applicable legislation. Appropriate metadata accompany the data / datasets description to facilitate access to particular instances of data.

Individual roles related to the data management activities are envisaged in the project consortium, including:

- Data Management Plan Leader (SLG) who is responsible for preparing and leading discussions related to the DMP at the project meetings and for maintaining the channels for publication of project outcomes.
- Scientific and Technical Project Manager who is responsible for identifying data collected or produced by the project and technical project outcomes that are eventually appropriate for publication.
- Dissemination leader in order to help partners identify the appropriate Open Access strategy when preparing their academic and industry publications and decide whether / when they suitable for posting in the considered repositories in the DRM.

Each individual partner shall:

- Identify its own project results that are suitable for publication.
- Consult the concerned partner(s) before publishing data that can be associated with an exploitable result in an open access strategy domain.
- Validate and register the data / datasets and metadata for the data a partner has generated in each WP.
- Back up data (that are under the partner's responsibility and ownership) for sharing through open access repositories.
- Quality control such data with support by the Data Management Plan Leader.
- Manage the different versions of data when updated and ensure that the latest version is timely available in the case of publically available data.



3 TOYLABS DATA Publishing Platforms

TOYLABS will pay particular attention to the definition of the public data, as well as to the selection of the platform to archive and preserve its datasets. When choosing a repository, it is important to consider factors such as whether the repository [4]:

- Gives the submitted dataset a persistent and unique identifier. This is essential for sustainable citations – both for data and publications – and to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants.
- Provides a landing page for each dataset, with metadata that helps others find it, tell what it is, relate it to publications, and cite it. This makes your research more visible and stimulates reuse of the data.
- Helps to track how the data has been used by providing access and download statistics.
- Responds to community needs and is preferably certified as a 'trustworthy data repository', with an explicit ambition to keep the data available in the long term.
- Matches particular data needs (e.g. formats accepted; access, back-up and recovery, and sustainability of the service). Most of this information should be contained within the data repository's policy pages.
- Offers clear terms and conditions that meet legal requirements (e.g. for data protection) and allow reuse without unnecessary licensing conditions.
- Provides guidance on how to cite the data that has been deposited.

The following list presents the platforms selected in the TOYLABS DMP to present the datasets/publications during the project and describes their concepts for publishing, storage, and backup. All Data Publishing Platforms adopted by TOYLABS are managed by SLG, who periodically updates the material.

3.1 TOYLABS PROJECT WEBSITE

The TOYLABS consortium designed and deployed an intuitive project website (<http://www.TOYLABS.eu/>) in the first months of the project. Such a website describes the mission and the general approach of the project and its development status. A blog informs about news on a regular basis. A dedicated section for downloads is used to publish reports and white papers. All documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title and the type of the document. All information on the TOYLABS website can be accessed without creating an account,



though a private section is also available (linked to Alfresco). The webpage is backed up once per month.

The TOYLABS project website will be available during the project runtime, and will still be available for at least two years after the official project end.

3.2 RESEARCHGATE

The TOYLABS ResearchGate account (<https://www.researchgate.net/project/TOYLABS-EU-Project>) has been created to promote the dissemination of scientific publications of the project. Open Access (or self-archived) documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title, a short description and the type of the document.

3.3 ZENODO

Zenodo (<http://zenodo.org>) is an open data repository service to archive, and make available research outputs in all scientific disciplines in alignment with the open data requirements of Horizon 2020. All public results generated or collected during the TOYLABS project will be uploaded to Zenodo for long-term storage and open access.

3.4 GITHUB

GitHub (<https://github.com/>) is a leading software development platform for distributed source code development, management, and revision control, as well as for tracking issues and retrieving documentation. All open source components that are developed during the TOYLABS project will be uploaded to an open access GitHub repository, linked also with the Zenodo account of the project. The GitHub TOYLABS account will be available in due time, together with the release of the open source code of the TOYLABS platform and components.



4 TOYLABS Publishable Results

In this section, a list of all existing or foreseeable publishable results for dissemination is presented, separated into public deliverables, software components, publications and open research data. For each result and in accordance to the FAIR data management guideline [5] we provide a description, name the standards used for storage and metadata (to make data findable & interoperable), and define which open access platform is chosen. Data Security aspects are also defined in this document, while the detailed ethics management policy of TOYLABS project is defined in D7.3. In brief, the TOYLABS partners will comply with the ethical principles as set out in Article 34 of the Grant Agreement, which asserts that all project activities must be carried out in compliance with:

- (a) Ethical principles (including the highest standards of research integrity - as set out, for instance, in the European Code of Conduct for Research Integrity - and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).
- (b) Applicable international, EU and national law.

4.1 TOYLABS PROJECT PUBLIC DELIVERABLES

The TOYLABS Project public deliverables are considered as part of the data management plan and will be published under appropriate Creative Commons licences (that will be defined in due time). The following table presents the list of public deliverables of the TOYLABS project in detail.

Table 2 List of TOYLABS Public Deliverables

Deliverable	Deliverable name	Leader	Type	Delivery date
D1.1	Exploring Progress and Innovation in TOYLABS Tackled Domains and TOYLABS Concept Definition- v1	AIJU	R	M3
D1.2	Stakeholders' Requirements Identification and Validation Framework Determination- v1	NTUA	R	M6



Deliverable	Deliverable name	Leader	Type	Delivery date
D1.3	Updated TOYLABS Concept Definition, Stakeholders' Requirements and Validation Framework-final version	NTUA	R	M16
D2.1	TOYLABS Open Innovation and Co-creation Integrated Methodology-v1	NTUA	R	M6
D2.2	TOYLABS Open Innovation and Co-creation Integrated Methodology-final version	NTUA	R	M13
D3.1	Technology Requirements and TOYLABS Components Analysis-v1	SLG	R	M6
D3.2	TOYLABS Added-Value Components Design and Development-v1	SLG	DEM	M7
D3.3	TOYLABS Added-Value Components Design and Development-v2	SLG	OTH	M10
D3.4	TOYLABS Added-Value Components Design and Development-v3	SLG	OTH	M14
D3.5	TOYLABS Added-Value Components Design and Development-final version	SLG	OTH	M18
D4.1	Collaboration Platform Design and Development- v1	SLG	DEM	M8



Deliverable	Deliverable name	Leader	Type	Delivery date
D4.2	Integrated Platform Design and Development-v1	SLG	OTHER	M11
D4.3	Integrated Platform Design and Development-v2 and Platform Validation-v1	SLG	OTHER	M15
D4.4	TOYLABS Integrated Platform Design, Development and Validation- final version	SLG	OTHER	M18
D5.2	Documentation on Pilot Scenarios Execution-v1	FABLAB LECCE	R	M13
D5.3	Updated Pilot Scenarios Execution and Pilots Evaluation-final version	FABLAB LECCE	R	M18
D6.1	Dissemination, Communication and Community Engagement Plan-v1	AIJU	R	M2
D6.2	First Dissemination, Communication and Community Engagement Report and Updated Plan	AIJU	R	M12
D6.3	Final Dissemination, Communication, Community Engagement Report	AIJU	R	M18
D7.2	Data Management Handling Plan	SLG	ORDP	M12



4.2 SOFTWARE COMPONENTS

In TOYLABS, the software to be developed and deployed is open source and will be distributed under appropriate licenses that will be defined by the partners in due time. The list of exploitable assets in terms of software components includes:

Table 3: List of open source software components

ID	Software Title	Responsible Partner
S_1	TOYLABS Collaboration Platform	SLG
S_2	Market Analytics & Trends Analysis Engine	NTUA
S_3	TOYLABS Augmented Reality Component	SLG
S_4	Partner Matching and Selection Component	SLG

Software published in github will obtain a URL via which it can then be accessed and exploited. It can also be made discoverable via respective web-based search engines, like Google. If software is also published in Zenodo, a proper DOI will be assigned to it.

4.3 PUBLICATIONS

Management of the TOYLABS dissemination and publication process will be performed in WP6 taking into account the open access strategies as defined in the TOYLABS DMP.

In general, TOYLABS will pursue open access publishing of its research outcomes in journals and conferences, and will at least ensure self-archiving for all its scientific publications that will be made available to Zenodo and ResearchGate as well as in the project website.

The following table attempts to enumerate specific research artefacts (from which publications can be produced) and the platforms for their publication.



Table 4: Project Research Artefacts

No	Artefact	Leader	Publication Means
R_1	TOYLABS Open Innovation Model and Co-Creation Methodology	NTUA	Zenodo, open-access journals, project web site (blogs), ResearchGate
R_2	Toy Sector Market Analytics and Trends Analysis Approach	NTUA	Zenodo, open-access journals, project web site (blogs), ResearchGate
R_3	Augmented Reality for End User Feedback Collection	SLG	Zenodo, open-access journals, project web site (blogs), ResearchGate
R_4	Partner Matching and Selection Methodology	NTUA	Zenodo, open-access journals, project web site (blogs), ResearchGate

4.4 TOYLABS PROJECT DATASETS

The private datasets gathered from the different pilots, namely the Mechanical puzzle toys Pilot and Dolls & Accessories Pilot, will be anonymised and published to the degree that they abide to the principles laid out in the present DMP, e.g. taking into account whether they will be of interest beyond the project consortium.



5 CONCLUSIONS AND NEXT STEPS

Deliverable D7.2 “Data Management Plan” is released in M6 in its initial version and describes the steps for data management which are to be followed during the implementation of the TOYLABS project. The current document provides preliminary information on the data types used and generated by the project consortium partners including focus on the means of sharing data captured by TOYLABS.

As this report is generated in the early stage of the project execution, it is considered as a living document which will be further complemented and enriched in subsequent WP7 deliverables.



ANNEX I: LITERATURE

- [1]. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf
- [2]. Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [3]. How to create a DMP Plan, <https://www.openaire.eu/opendatapilot-dmp>
- [4]. How to select a repository?, <https://www.openaire.eu/opendatapilot-repository>
- [5]. EC (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

