eTryOn - Virtual try-ons of garments enabling novel human fashion interactions

Project Title:	**eTryOn - Virtual try-ons of garments enabling novel human fashion interactions**
Contract No:	951908 - eTryOn
Instrument:	Innovation Action
Thematic Priority:	H2020 ICT-55-2020
Start of project:	1 October 2020
Duration:	24 months

Deliverable No: D8.6

Data Management Plan, updated

Due date of deliverable:	31 March 2021
Actual submission date:	31 March 2021
Version:	Final
Main Authors:	Andreas Stasinakis (CERTH)

Project funded by the European Community under the H2020 Programme for Research and Innovation.

Deliverable title	Data Management Plan (updated)
Deliverable number	D8.6
Deliverable version	Final
Contractual date of delivery	31 March 2021
Actual date of delivery	31 March 2021
Deliverable filename	eTryOn_D8.6_final.docx
Type of deliverable	Report
Dissemination level	PU
Number of pages	41
Work package	WP8
Task(s)	T8.3
Partner responsible	CERTH
Author(s)	Andreas Stasinakis (CERTH), Tasos Papazoglou Chalikias (CERTH), Machi Boikou (CERTH), Jim Downing (Metail), Thomas De Wilde (QC)
Editor	Spiros Nikolopoulos (CERTH)
Reviewer(s)	Michaela Jauk (Odlo)

Abstract	This deliverable is an updated version of the guidelines for the data collection, storage and ownership for all eTryOn activities that was initially presented on D8.2. This updated version incorporates the new guidelines that have arisen along with the applications development, updating all relevant datasets and also enlarging the list with additional datasets required for the needs of the development phase and the pilots. In addition to defining and ensuring that all of the collected data are intended to be processed, relevant and limited solely to the purposes of eTryOn, it also foresees the essential security measures that will be implemented to prevent unauthorized access to personal data or the equipment used for their processing.

Keywords	Data management

Copyright

Deliverable history

Version	Date	Reason	Revised by
alpha	10/03/2022	Table of Contents	Machi Boikou (CERTH)
Beta	25/03/2021	Input on Datasets	All partners
Review	29/03/2021	Review comments	Michaela Jauk (Odlo)
Final	31/03/2021	Final version	Spiros Nikolopoulos (CERTH)

List of abbreviations and Acronyms

Abbreviation	Meaning
AR	Augmented Reality
DMP	Data Management Plan
DoA	Description of Action
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation
H2020	Horizon 2020
MR	Mixed Reality
N/A	Not Applicable
VR	Virtual Reality
WP	Work Package
SMEs	Small and medium-sized enterprises
TBD	To be determined

Table of Contents

1. Executive summary 7
2. Introduction 8
3. Data Management Methodology 9
   3.1 Data Summary 9
   3.2 Making data findable 11
   3.3 Making data openly accessible 13
   3.4 Making data interoperable 15
   3.5 Increase data re-use (through clarifying licenses) 16
   3.6 Allocation of resources 18
   3.7 Data security 18
   3.8 Ethical aspects 19
   3.9 Other issues 21
4. Data management plan for eTryOn datasets 22
   4.1 Datasets for the development of technologies 24
   4.2 Datasets for piloting activities 35
5. Conclusions 41

1. Executive summary

This deliverable is an updated version of the guidelines for the data collection, storage and ownership for all eTryOn activities that was initially presented on D8.2, where the basic framework and an initial listing of data management activities for the project was recorded.

This updated version incorporates the new guidelines that have arisen along with the applications development and the progress of the project's implementation, updating all relevant datasets in section 4 "Data Management Plan for eTryOn Datasets" and also enlarging the list with additional datasets required for the needs of the development phase and the pilots.

In addition to defining and ensuring that all of the collected data are intended to be processed, relevant and limited solely to the purposes of eTryOn, it also foresees the essential security measures that will be implemented to prevent unauthorized access to personal data or the equipment used for their processing.

A project which aims at (3D) digitizing several types of objects (people, clothes) presents several data management challenges besides the collection of data. Therefore, the data management lifecycle must be described precisely. The DMP is based on the H2020 Online Manual for *Data Management Plan*, which describes the data management life cycle for the data to be collected, processed and/or generated by the project. The methodology proposed by the European Commission Guidelines in line with the EU General Data Protection Regulation (GDPR) has been adopted for the deliverable compilation. FAIR Data Management is highly promoted by the European Commission and since eTryOn is a data intensive project, relevant attention has been given to this task.

Following these guidelines for making research data findable, accessible, interoperable and re-usable (FAIR), this deliverable includes information on:

- The handling of research data (during and after the end of the project)
- The types and formats of data collected (processed and/or generated)
- Methodologies and standards applied
- Data accessibility and restrictions
- How data will be curated and preserved (during and after the end of the project)

2. Introduction

The fashion industry has been one of the industries that have been slow in incorporating new technological advancements (such as VR, AR, MR) in their business operations, in order to enhance both the creative process of garment design and the consumer interaction with fashion items. Indeed, the available interaction pathways have changed minimally over the past decades and mostly in the direction of encapsulating the recent social media frenzy (e.g. Instagram). With eTryOn, the use of interactive technologies will become mainstream in the fashion industry, focusing on three distinct fashion experiences that target both fashion designers and consumers: i) **creative experience**: while the creative process of garment design has changed over the past years from 2D sketches to using 3D design software, the visualization capabilities are still rather limited to just fitting the garments on grayscale predefined still avatars without considering the response of the garments during movements; ii) **social experience**: while the social experience of fashion has changed with the wide adoption of platforms like Instagram, it is still limited to just uploading images of people wearing physical clothes; and iii) **shopping experience:** the online shopping experience is essentially the same whether people buy clothes or electronics (i.e. they look at a few images of the items and their specs).

It goes without saying that a huge part of the project depends on different types of data (questionnaires for user requirements, data for developing and testing algorithms etc.), which makes the Data Management Plan (DMP) more than important. The project Data Management Plan aims at defining the management strategy of all data within the eTryOn framework, and describes all activities and procedures to ensure that all data is FAIR, following the template and approach recommended by the EC. While the project is in favour of making key data assets produced within the project openly available and accessible, data sharing may also be restricted in several cases, taking 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"*[1]

In order to gather all the related information for each of the datasets used (or will be used) during the lifetime of the project, a template (see Section 4) has been sent to all the partners. The template will continue to be used during the course of the project for any new datasets that are found to be necessary and with any substantial changes the DMP document will also be updated.

This document consists of two main parts, Section 3 and Section 4. In Section 3, we present the general data management methodology of eTryOn-according to H2020 Guidelines and FAIR data- along with the FAIR data financing plan, data security measures and ethical aspects. In Section 4, we present and explain the eTryOn dataset template, and then, using this template, we document the eTryOn datasets – up to the composition date of this deliverable. Finally, section 5 concludes the deliverable.

[1] Guidelines on FAIR Data Management in Horizon 2020

3. Data Management Methodology

The methodological approach that has been used to compile this deliverable follows the "Template for Horizon 2020 Data Management Plan (DMP)", version 1.0, released on 13.10.2016 by the European Commission. The eTryOn DMP presented in this deliverable addresses the following aspects of eTryOn data:

- Data summary
- FAIR data
  - Making data findable, including provisions for metadata
  - Making data openly accessible
  - Making data interoperable
  - Increase data re-use
- Allocation of resources
- Data security
- Ethical aspects
- Other issues

In the following subsections, we briefly present the kind of questions associated with each of these aspects. For each question we also provide a summary of the general strategy adopted by the project consortium for handling different dataset categories. Detailed answers for each dataset are provided in Section 4.

Updating Methodology and future versions of the eTryOn DMP

Generally, as no modifications are expected on the DMP methodology during the project's lifetime, this updated version mainly includes include updates in Section 4 of this deliverable. In this document, each dataset owner has added new or updated existing dataset entries for which he/she is responsible, in close cooperation with the WP8 Leader. The document is stored in the file repository (Google Drive) of the project and will follow the Data Security measures described in section 3.7.

3.1 Data Summary

The Data Summary addresses the following issues:

- Outline the purpose of the collected/generated data and its relation to the objectives of the eTryOn project;
- Outline the types and formats of data already collected/generated and/or foreseen for collection/generation at this stage of the project;
- Outline the reusability of existing data;
- Outline the origin of the data;
- Outline the expected size of the data;
- Outline the data utility.

This field describes the data that will be generated or collected, including references to their origin (in cases where data is collected), nature, scale, to whom it could be useful, and whether it underpins a scientific publication. With regard to the individual questions, our generic DMP approach is summarized below (detailed answers for each dataset are given in Section 4).

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The main goal of eTryOn is to design new interactive applications that will be tailored to the end users' needs (i.e. fashion designers and consumers). In this direction, our plan is to gather continuous feedback from the user base, starting from the beginning of the project (through collecting the user requirements) to its end (through iterative testing as well as through the envisaged pilots demonstrating the developed technology and the system update based on the users' evaluation). For this vision to unfold, the following **types of dataset** are expected to be used, collected or generated:

- **Requirements analysis data** (in the form of questionnaires, interviews, focus groups, etc.) will be collected by use case partners from users, to identify user needs, use case scenarios and desired software functionalities. The objective of collecting such data is to orient the design and development of the eTryOn applications, tools and assets towards the needs of actual users, and support the eTryOn use cases altogether. For instance, such data has already been used in WP6 for D6.1 "The user requirements".
- **Evaluation data**, such as user activity and survey data, will be collected from the end users with the aim to assess the impact and effectiveness of the proposed set of apps/tools.
- **Technical data** (existing or generated) will be collected by technical partners in order to develop and test the eTryOn applications and tools. A variety of data will be necessary, including images, text, 3D objects etc.
- **Data related to eTryOn dissemination and communication activities**, to allow better organization of events and offer better services to attendees. Video content and photos from participants will also be used for creating dissemination content.
- **Contact data of eTryOn consortium members** (e.g. name, email, organization, etc.) used for project management activities. Selected video-conference calls may be recorded, so there is also audio-visual content of the partners involved in this category

Details on these datasets can be found in section 4

What types and formats of data will the project generate/collect?

The project will use different types of data (video, images, text, communication data, system log data, etc.), both personal and non-personal, from a variety of sources (web, partners, testers, etc.) and will probably generate datasets in the process of creating the eTryOn tools and assets.

Will you re-use any existing data and how?

There will be reuse of existing datasets, for instance DeepFashion dataset for object detection in WP3, in tasks related to developing and benchmarking algorithms. Other datasets will be reused via machine learning models trained on them, e.g. pose detection for use case 2, which uses an open source model trained on the COCOMO dataset.

What is the origin of the data?
The data comes from various origins, including, but not limited to, the following: ● Individual researchers that openly share their data in open repositories such as GitHub, Zenodo or via their webpages; ● Research and academic organizations that openly share data in open or institutional repositories; ● Use case partners that share data with the technical partners of the consortium to help them train and test their algorithms and software; ● Web pages; ● Participants, end users or project partners, after filling out a consent form; ● Questionnaires and surveys filled in by end users (user requirements and evaluation questionnaires); ● Interviews and focus groups conducted with end users; ● Audio-visual content recorded for the project needs; ● Use of eTryOn software tools by the users (automatically collected data analytics). ● Dataset generated in order to build data driven software components as part of the project development.

What is the expected size of the data?
Dataset sizes are discussed in section 4.

To whom might it be useful ('data utility')?
The datasets listed in this DMP are necessary to project partners for identifying user and technical requirements for the use cases, designing, developing and testing the eTryOn methodologies, algorithms and tools, and assessing the effectiveness of these tools in real-life trials involving end users. It is also crucial for increasing the project outreach and achieving high dissemination impact. Technical and evaluation data may also be useful to researchers with a focus on the development of similar fashion tools.

The following subsections (3.2-3.5) about making data FAIR refer to the datasets that are produced by the project and not those that already exist and are being used by the project.

3.2 Making data findable

This point addresses the following issues:

- Are the data produced in the project discoverable and identifiable?
- What naming conventions are followed?
- Will search keywords be provided that optimize possibilities for re-use?
- Are clear version numbers provided?
- What metadata will be created?

In general, the data collected and generated by the project will be identifiable and discoverable. With regard to the individual questions, our DMP approach is summarized below (again detailed answers for each dataset are given in section 4).

Are the data produced and/or used in the project discoverable and identifiable?

Datasets that will be made publicly available will be uploaded to open repositories like Zenodo etc., thus making it both easily discoverable and identifiable externally. With regard to datasets that will only be used internally in the project, either because of confidentiality reasons, IPR constraints, importance to commercial exploitation or due to limited value to external parties, they will only be discoverable and identifiable by consortium partners or selected institutional users involved in the processing of this data. Consequently, these datasets are not subject to the FAIR data principles.

What naming conventions are followed?

A specific naming convention is suggested to identify the various eTryOn datasets:

eTryOn_<WPno>_<serial number of dataset>_<data type>_<dataset title>

- The <WPno> reveals the WP in the context of which this data is collected or generated and processed
- The <serial number of dataset> is assigned manually in the order of presentation in this deliverable
- The <data type>field is determined according to the categorization presented below:

Acronym	Description
RTD	Supporting research and technical development
PILOT	Resulting from pilot activities

- Finally, the <dataset title> is a descriptive dataset title

Will search keywords be provided that optimize possibilities for re-use?

Keywords will be provided in the cases where this is applicable.

Are clear version numbers provided?

For datasets that will be made publicly available in open repositories, versioning will be supported by appropriate naming conventions.

What metadata will be created?

For datasets that will be shared via open repositories, the metadata standards used by these repositories will be used.

Metadata for data uploaded at the project Google Drive repository is also supported.

In general, Data discoverability will be further enhanced by associating search keywords along with the data, as well as promoting the datasets through the project's

communication activities (e.g. blog posts, tweets, etc.). As a part of metadata provision, keywording must comply with the following principles:

- Who, what, when, where and why: these questions must be covered.
- Consistency among the different keyword tags needs to be ensured.
- Keywording must be relevant, understandable and clear.

3.3 Making data openly accessible

- Which data produced and/or used in the project will be made openly available as the default?
- How will the data be made accessible (e.g. by deposition in a repository)?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Where will the data and associated metadata, documentation and code be deposited?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are there well-described conditions for access (i.e. a machine-readable license)?
- How will the identity of the person accessing the data be ascertained?

With regard to the individual questions about data accessibility, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

Which data produced and/or used in the project will be made openly available as the default?

Some of the datasets to be used in this project (as described in Section 3.6) is open data already, made openly available by third parties (for example DeepFashion). Since this data is already open, as a general policy, eTryOn will not re-share it. Sharing such data will only be pursued in cases where the data license allows it and when eTryOn researchers estimate that re-sharing of the data (in some new form) provides additional benefit to third parties.

In addition to open data, there are also privately owned datasets. These are owned by the organizations involved in eTryOn, as well as some technical and academic partners, and have been collected and created over a period of years or in the context of other projects or internal processes, independently from eTryOn. Such data may be provided to the project for research purposes, but will not be shared openly. However, effort will be made to make this (or part of this) data openly available in cooperation with the data owners, wherever this is possible.

Data that will be collected by the project in the form of questionnaires or forms addressed to end users (for user requirements analysis and/or for evaluation) will not be made openly accessible, since they may contain personal or confidential information. Wherever possible and in case there is added value from their sharing, such data will be anonymized before being shared (mainly with regard to the evaluation data).

In any case, the aforementioned data (whether public, private, or personal) will be used exclusively for achieving the project objectives. Where appropriate, the analysis results

will be made open as part of public project deliverables and publications available in open repositories.

How will the data be made accessible (e.g. by deposition in a repository)?
Open data will be deposited in open repositories like Zenodo and GitHub. The datasets will also be shared through the eTryOn website. Datasets destined to be used internally by project partners will be stored either on the project's file repository on Google Drive and/or in the servers of project partners.

What methods or software tools are needed to access the data?
Different methods and software tools will be required to access the data depending on the dataset. More details are provided in Section 4 (e.g. web-browser, API).

Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Where this is applicable, the relevant software and its documentation will be included.

Where will the data and associated metadata, documentation and code be deposited? Have you explored appropriate arrangements with the identified repository?
Open data will be deposited in open repositories, like Zenodo and GitHub. These adopt standard and simple procedures to allow data sharing by researchers. No need for appropriate arrangements is foreseen.

If there are restrictions on use, how will access be provided?
If such cases are identified, access could be provided either through use of consent and anonymization, or by regulating and restricting access to specific users.

Is there a need for a data access committee?
No such need has emerged yet.

Are there well-described conditions for access (i.e. a machine-readable license)?
Such licenses will be used for the data we plan to make openly available.

How will the identity of the person accessing the data be ascertained?
This will be dealt with on a case-by-case basis. For the open datasets, no identification of the person accessing the data will take place. For the data that will be used only

internally by project partners (which is stored on the project file repository or partners' servers), access control procedures are in place that define access rights and provide secure access with username/password credentials.

3.4 Making data interoperable

This point specifies what data and metadata vocabularies, standards or methodologies are followed in order to facilitate interoperability. It also addresses whether a standard vocabulary is used for all data types within the dataset, in order to allow interoperability. The specific issues covered are the following:

- Are the data produced in the project interoperable?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Are the data produced in the project interoperable?

Effort will be made to achieve interoperability on most of the data produced in eTryOn. More information will be provided as the project unfolds.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

In order to ensure interoperability and maximum re-use of eTryOn data, project partners will try to collect existing and new data in standardized formats, following well-known data representation models and metadata vocabularies.

Standard and simple data vocabularies will be adopted for different types of datasets (image data, text data, user analytics, etc.). Additionally, we will consult the OpenAIRE Guidelines for Data Archives[2]. As the project progresses and data is identified and collected, further information on making data interoperable will be outlined in subsequent versions of the DMP.

Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?

Whenever possible, standardized vocabularies will be used to encourage the wide exchange of information and sharing of data.

[2] OpenAIRE Guidelines for Data Archives: https://guidelines.openaire.eu/en/latest/

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
This will be examined on a case by case basis.

Further actions on making data interoperable will be outlined in subsequent versions of the DMP, as the project progresses. These actions refer to revisiting data and metadata vocabularies, imposing additional standards or methodologies and optimizing interoperability overall.

3.5 Increase data re-use (through clarifying licenses)

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

With regard to the individual questions about increasing data re-use, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

How will the data be licensed to permit the widest re-use possible?
This will be examined on a case by case basis depending on the dataset. Our general approach can be summarized as follows: 4. In case of data coming from external open sources or in cases where the data comes with a license on its own, the data will be shared under the same license. 5. For others, a CC-BY 4.0 (Creative Commons Attribution 4.0 International License) license will be selected, which allows open sharing but also allows keeping some control over the data (e.g. requires attribution). Most common CC –BY 4.0 licensing types are the following: ▪ **Creative commons Attribution-Share Alike 4.0** (CC BY-SA 4.0): any third party can freely copy, distribute, display and modify the datasets for any purpose. Remix, transform, or built upon data, must be distributed under the same license as the original. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution 4.0 International** (CC BY 4.0): any third party can freely copy, distribute, display and modify the datasets for any purpose. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made. ▪ **Creative Commons Attribution-NoDerivatives 4.0 International** (CC BY-ND 4.0): any third party can freely copy, distribute, display and modify the

datasets for any purpose. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.

- **Creative Commons Attribution-NonCommercial 4.0 International** (CC BY-NC 4.0): third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.
- **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International** (CC BY-NC-ND 4.0): third parties can copy, distribute, display and modify the datasets for any purpose other than commercial unless they get a permission by project partners first. Remix, transform, or built upon data, however, must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.

Licensing will be discussed in later stages of the project with all involved parties. Alternative license schemes may also be adopted at the discretion of the dataset owner.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

This will be examined on a per case basis. In general, effort will be made for the data to be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

This will be examined on a case-by-case basis (see section 4). The openly shared datasets will be reusable after the end of the project on Zenodo and any additional available platforms/outlets (GitHub, Gitlab, eTryOn website, partners' websites).

How long is it intended that the data remains re-usable?

The openly shared datasets will in general be perpetually reusable.

Are data quality assurance processes described?

Automatic *data cleaning* techniques will be employed to improve data quality. Data cleaning consists of identifying incomplete, incorrect, inaccurate, or inconsistent parts of the data and then replacing, modifying or deleting such data. This is necessary for improving data quality and producing a clean, uniform, and consistent dataset for integration; the quality of the data reflects directly upon the quality and accuracy of the analysis results.

For datasets including questionnaire data, a manual quality control will be performed by partners to ensure data quality.

3.6 Allocation of resources

Since the beginning of the design of the project, data management was taken into consideration and every partner has been allocated effort for this purpose. This is embedded into the tasks dealing with data management activities, either collecting, processing, or creating datasets. Hence, all related costs for data management are already covered by the project and no additional resources will be needed. Costs for publications are covered by the project budget. Other costs for making the data FAIR will be covered by the individual partners that will share the data. Zenodo, which is free of charge, will be used to make available papers and datasets (Green Open Access model) under the eTryOn community.

Regarding the Data management role, although not specifically mentioned in the DoA, a data manager role has been established in the project to ensure that data processing actions within eTryOn are in line with the law. CERTH has been appointed as the beneficiary responsible for data management and has cooperated with technical and pilot partners to draft a detailed data management plan that clearly identifies how each dataset used or created by the project will be handled. CERTH will be responsible for closely monitoring the execution of the data management plan and ensuring that project partners handle project datasets appropriately.

3.7 Data security

This addresses secure storage and transfer of sensitive data as well as data recovery, including the following questions:

- Is the data safely stored in certified repositories for long-term preservation and curation?
- What provisions are in place for data security?

All software tools and data storage mechanisms used within eTryOn are designed to safeguard collected data against unauthorized use and to comply with all national and EU regulations. Engineering best practices and state-of-the-art data security measures along with GDPR legislation will all be incorporated following their respective guidelines and principles.

As explained, eTryOn datasets will either be openly shared (by uploading them in open repositories) or shared internally among specific partners (stored in the project file repository or partners' servers). Below, we examine the data security strategy for these options.

Open repositories

Datasets to be openly shared will be deposited in repositories such as Zenodo that have in place strong mechanisms and protocols for data recovery and long-term data preservation.

eTryOn file repository

In order to be able to share files within the consortium, a Google Drive repository has been set up. Inside the repository, a Work Packages folder with one subfolder for each WP have been created so as to share internally deliverables and relevant documents for each WP. In addition, a data repository folder has been set up to share data such as avatars, garments, etc. Furthermore, folders for sharing data with respect to the project meetings, participation in events, communication kit, literature, templates and meetings have been created. Finally, in the Google Drive folder there is information about project administrative issues such as contract documents, contact details, the mailing list etc.

Google started in the cloud and runs on the cloud, so it's no surprise that they fully understand the security implications of powering the business in the cloud. Because Google and its enterprise services run on the same infrastructure, any organization will benefit from the protections they have built and use every day. Their robust global infrastructure, along with dedicated security professionals and our drive to innovate, enables Google to stay ahead of the curve and offer a highly secure, reliable, and compliant environment.

In order for Google Drive to be protected from unauthorized access, alteration, disclosure, or destruction of information, Google privacy policy[3] includes, but not limited to, the following:

- Encryption to keep data private while in transit
- A range of security features, like Safe Browsing, Security Checkup, and 2 Step Verification to help you protect your account
- Reviewing their information collection, storage, and processing practices, including physical security measures, to prevent unauthorized access to our systems
- Restricting access to personal information to Google employees, contractors, and agents who need that information in order to process it. Anyone with this access is subject to strict contractual confidentiality obligations and may be disciplined or terminated if they fail to meet these obligations

All partners have access to this repository and store content that needs to be shared among the consortium and it is expected that this infrastructure will be used for most project datasets.

Google Drive is in full compliance with GDPR in order to strengthen personal data protection in Europe. In order for that to be achieved, Google supports the GDPR compliance[4] efforts by:

- Committing in their contracts to comply with the GDPR in relation to their processing of customer personal data in all Google Cloud Platform and Google Workspace services
- Offering additional security features that may help the customers to better protect the personal data that is most sensitive
- Giving the documentation and resources to assist any customer in their privacy assessment of google services
- Continuing to evolve google capabilities as the regulatory landscape changes

Partners' servers

eTryOn partners have significant experience in data handling and protection both in the context of their institutional operation as well as in the context of their participation in other H2020 projects. As a result, the beneficiaries already have in place operational policies regarding potential ethics issues as well as privacy and security guidelines for data protection, adhering to national and EU regulations. Ultimately, each partner is responsible for the data protection and security mechanisms in their own servers.

3.8 Ethical aspects

This section covers any ethical or legal issues that can have an impact on data sharing, including references to ethics deliverables and ethics chapter in the DoA. Specifically, it addresses the following issues:

[3] https://policies.google.com/privacy?hl=en-US

[4] https://cloud.google.com/security/gdpr

- Are there any ethical or legal issues that can have an impact on data sharing?
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

When a dataset cannot be shared, the reasons for this will be outlined (e.g. ethical restrictions, rules governing privacy and personal data protection, intellectual property, and commercial sensitivity).

With regard to the individual questions, our generic DMP approach is summarized below (again detailed answers for each dataset are given in section 4):

Are there any ethical or legal issues that can have an impact on data sharing?
Addressing legal and ethics challenges is an important part of the eTryOn work plan. As already indicated in section 5 "*Ethics and Security*" of the DoA, special attention has been paid to these issues since the very beginning of the project. eTryOn will pay particular attention to any ethical issues that will arise and will address them in a professional way following established EU regulations and corresponding national laws about user privacy, confidentiality and consent. On that direction we have foreseen the eTryOn's Ethics Board in the organizational structure (see deliverable D8.1) of our project, which will be the responsible committee to depict and face all the rising issues that refer to ethics. In detail, the adopted ethical practices are described below. Moreover, all the necessary information regarding the human participation during the project and all the practices that eTryOn will follow in order to assure the data privacy, protection and confidentiality of the participants and their data generated throughout the project implementation and the pilots can be found in the deliverable D8.4 "Ethical, legal and privacy requirements and guidelines for implementation". The objective of this deliverable was to review all the related legislative framework and present guidelines on how to manage all the project generated data assuring that all ethical, privacy and legal directives are followed by the project. All personal data that will become available during the project will be kept secure and unreachable by unauthorized entities. The data will be handled with appropriate confidentiality and technical security, as required by law in the individual countries and EU laws and recommendations. A general policy on ethical conduct will be adopted by the eTryOn Consortium. Prior to the start of relevant eTryOn activities approval form responsible ethics committees will be requested in line with current regulations and guidelines and will explicitly address specifics related to the conduct of analysing personal data, including procedures of (electronic or written) informed consent, remote data collection, user's feedback and privacy and confidentiality and cybersecurity in the data chain of the data and possible data sharing. The eTryOn activities will comply with all applicable national, EU and international legislation, regulations and conventions around the Research on Humans (such as GDPR, Data Protection Directive 95/46/EC of European Parliament and of the Council, OHCHR, etc.).

Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?
As stated in D8.4 “Ethical, legal and privacy requirements and guidelines for implementation” and its updated version D8.5, in order to ensure safety, the available data will be automatically anonymized or at least become pseudonymous and will not be transmitted to third parties. Therefore, at least in this phase of the project, there is no need for informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data. In case that, during the lifetime of the project, storing and sharing personal data is needed, the participants will be fully informed through the informed consents (as provided analytically in D8.5).

3.9 Other issues

Other issues refer to other national/ funder/ sectoral/ departmental procedures for data management that are used.

In general, all the research organisations and SMEs that participate in the project have in place their own data privacy and security policies, which are compliant with EU regulations and especially the GDPR.

4. Data management plan for eTryOn datasets

This section includes information on all datasets that can be foreseen as necessary, at the time of writing this deliverable. Every table provides information on the dataset, along with explanations on whether and how this dataset will be FAIR and secure, as far as the datasets produced by the project are concerned. In the first table, we present the structure of the table along with explanations for every field contained. This table template also includes the partner responsible for data collecting and maintaining the dataset, along with an indication on whether these data will be based on existing datasets.

The datasets are categorised under two sections depending on their purpose, including a) supporting research and technical development, and b) resulting from pilot activities. According to this classification we also codify their naming under RTD and PILOT.

NAME	**eTryOn_<WPno>_<Serial number of dataset>_<data type>_<dataset title>**
Data summary	**Responsible partner:** Partner responsible for producing and/or using the specific dataset **Purpose:** Short description of data (include a sample of the features of the dataset if possible). Also, what is the purpose of data collection/generation (and its relation to project objectives) in the context of eTryOn? **Type/format**: What is the type/format of the data generated/collected? **Re-use of existing data:** Are existing datasets reused and how? **Data origin:** What is the origin/source of the data? **Expected size:** What is the expected data/dataset size (if known)? **Data utility:** To whom will this data be useful and how? (inside the project, for instance WP and/or partners, and also to third parties, if applicable)
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: Are the data produced in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? **Search keywords:** Will search keywords be provided that optimize possibilities for re-use? **Versioning:** Will clear version numbers be provided? **Metadata creation:** Specify standards for metadata creation (if any). If there are no standards in your discipline, describe what type of metadata will be created and how.
FAIR Data: **Accessibility**	**Data openly accessible:** Will data produced in the project be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. **How it will be accessible:** How will the data be made accessible (e.g. by deposition in an open repository)?

	Methods/software tools to access data: What methods or software tools are needed to access the data? Also, is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? **Repository:** Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. **Restrictions on access:** If there are restrictions on use, how will access be provided?
FAIR Data: **Interoperability**	**Interoperability:** Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organizations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)? **Data and metadata vocabularies**: Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability **Use of standard vocabularies:** Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? **Mappings to commonly used vocabularies**: In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
FAIR Data: **Reusability**	**License:** Specify how the data will be licensed to permit the widest reuse possible. E.g. Open Data License (Creative Commons CC0 License, Creative Common Attribution License-CC-BY v4.0, etc.). **Availability for re-use:** When will data be made available for re-use. If applicable, specify why and for what period a data embargo is needed **Usable by third parties after end of project**: Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why. **Re-use timeframe:** Specify the length of time for which the data will remain re-usable **Data quality assurance process:** Describe data quality assurance processes
Allocation of resources	**Costs for making the data FAIR:** Covered by the individual partners that will share the data **Costs for long-term preservation:** TBD

Security	**Security measures:** Security measures implemented for data protection (incl. controlled access, user authentication, firewalls, VPNs, encryption, back-ups, etc.)
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Are there any ethical or legal issues that can have an impact on data sharing? **Is informed consent for data sharing and long term preservation given:** Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?
Other issues	Refer to other national/funder/sectorial/departmental procedures for data management that you may be using (if any)

4.1 Datasets for the development of technologies

NAME	**eTryOn_WP2_001_RTD_3D Garments**
Data summary	**Responsible partner:** ODLO **Purpose:** 3D garments created by ODLO will be used for all 3 applications in eTryOn. All different end-users (online shoppers, influencers and designers) are interacting with the 3D virtual created garments in slightly different ways. These virtual developed garments are from the SS23 collection and some special outfits in case of DressMeUp (Heritage, World Champion, Scott Sram replica outfit), the same time when the pilot phase will officially run and the real counterpart of that garment is available in the online shop. 3D developed garments are build out of actual real size patterns, digitized fabrics, which contain on one hand the visual appearance of the material through several captured texture maps and on the other hand the physical properties, which can be tested with specific machines and define how this fabric should behave in the 3D software when you interact with it. Additionally, it can have some trims as 3D objects like zippers, cord stoppers, etc. **Type/format**: Browzwear files and exported fbx, gltf and obj files **Re-use of existing data:** The data is original **Data origin:** Patterns, fabrics, trims, color information, labels from ODLO,Browzwear parametric avatars (In order to create the 3D garment) **Expected size:** Depends on the content. One BW file can be between 9 and 190 MB, depending on how detailed fabrics have been scanned, on the number of colour ways in the file, on complexity and number of trims (hard objects) used - in average most of the final files are around 50MB. **Data utility:** Especially for WP2 avatar- garment interaction, the 3D garments are the base of all 3 applications

FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The 3D garments are stored on a project`s google drive folder, in which only the partners have access. In the piloting phase the end-users will interact with the 3D garments through the 3 different apps. The source BW files will not be discoverable or accessible outside the consortium. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** The 3D garments will be only accessible for people outside of the project through the 3 different apps and the way how they can interact with the application. The accessibility is limited within the application. For people in the project group they can access them on Google Drive **How it will be accessible:** Accessible through the 3 applications and only usable within the application. And for project partners on Google drive. **Methods/software tools to access data**: Web-browser for project partners, web application and mobile application for the end users in pilot phase. **Repository:** Stored on Google Drive -Data Repository – Garments Storage for the 3 applications: VR Designer: The files are imported and stored inside the VR Application. Dress Me Up: The files are stored in Google Firebase cloud storage. Magic Mirror: The files are stored in Google Firebase cloud storage. **Restrictions on access:** Shared among partners with access to project`s Google Drive. End-users can access only if they register to the applications.
FAIR Data: **Interoperability**	**Interoperability:** N/A. End users are not expected to consume the data in the applications directly, only as a part of the developed User Experiences. **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: No
FAIR Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project**: No, the 3D garments are ODLO`s property (IP) and as such should not be used by third parties (except for consortium members, during the demonstrator phase) after the project has ended. **Re-use timeframe:** N/A

	Data quality assurance process: 3D garments are developed by ODLO and they pass the internal development process and this secures the quality level of the data
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The 3D garments are stored on a Google Drive folder of the project. This Google Drive is restricted only to registered users while registration is possible only by invitation. Access requires username/ password authentication. 3D garments with which end-users can interact through the eTryOn applications: end-user can only interact with them when he has created an account for the application with username/password authentication. In VR Designer there is no authentication process, with the files being embedded directly in the app.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** 3D garments are ODLOs property (IP) and for this reason it is not allowed to share them with competitors and parties outside the consortium. **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	N/A

NAME	**eTryOn_WP3_002_RTD_MLTrainingSets**
Data summary	**Responsible partner:** MLZ, CERTH **Purpose:** Sets of fashion products images, text and annotations used to train Machine Learning models for the D3.1 of WP3 (fashion attributes extractor) plus mappings between datasets for the standardisation of values across them. Datasets of user ratings on fashion products from the historical data collected by MLZ in time for D3.2 (fashion trend detector). Datasets for user ratings and reviews on fashion products from public sources and D3.4 (fashion recommendations). **Type/format**: images (JPEG), text (TXT), annotations (JSON, TXT, XML), spreadsheets (Google Sheets), ratings (JSON/AWS S3 dumps). **Re-use of existing data:** The datasets used for training are partly public data and partly own-generated using the MLZ assets and creating annotations across the consortium stakeholders in the WP. **Data origin:** MLZ-owned images and texts for products, public fashion annotations datasets. **Expected size:** A few hundreds of GBs in total. **Data utility:** The datasets will be used in WP3 to train Machine Learning algorithms (for trend detection and garment recommendations) the results of which will be then useful for the rest

	of the technical partners for other activities related to the three eTryOn applications
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The public datasets used are already discoverable (therefore, they will not be re-shared). The data annotated and built by consortium stakeholders will not be discoverable. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** The data annotated by members of the consortium will not be made accessible outside the eTryOn project as it is built on proprietary data owned by MLZ. **How it will be accessible:** Mapping sheets are stored in the consortium's Google Drive folder, images/text/annotations datasets built on proprietary data are shared between the WP responsible partners using Slack and/or email channels due to their relatively large size. Bigger sets for D3.2 and D3.4 (ratings and historical time series) will be shared across the partners via private links to AWS S3. **Methods/software tools to access data**: Web-browser for the data in the Google drive (just by members of the consortium), private Slack channels/email for bigger datasets. **Repository:** The Google Drive hosts the small datasets used in the deliverables as well as general documentation (on Google docs, updated constantly) on all sources of data. **Restrictions on access:** To be shared across eTryOn consortium stakeholders only.
FAIR Data: **Interoperability**	**Interoperability:** The data formats comply to interoperable formats for machine-readability. **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** The public datasets used in WP3 are licensed for non-commercial use only, the proprietary datasets built by partners will not be licenced for the course of the project, as they are not for public use; they might be licenced at a following stage as proprietary assets for commercial exploitation if the partners decide so. **Availability for re-use:** The public datasets are de facto available for (non-commercial) reuse. The private datasets might be made available for reuse under commercial agreement at a later stage. **Usable by third parties after end of project**: Private datasets may be made available for use by third parties at the end of the project under terms to be discussed. **Re-use timeframe:** Will be decided at a later stage.

	Data quality assurance process: Data is cleaned and analysed/edited accordingly by partners of the project to assure the best quality (e.g. broken images are removed, annotations are created under a protocol).
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** Will be discussed and decided at a later stage
Security	**Security measures:** The data stored on the Google Drive folder of the project is restricted only to registered users while registration is possible only by invitation; access requires username/password authentication and it fully complies with the European and international framework and the GDPR. Private datasets shared across partners are passed through private slack channels when size allows, bigger datasets will be shared using private links to AWS S3 repositories.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The project's legal terms cover for data sharing initiatives amongst partners of the consortium so there are no further legal aspects to be covered in that the private data is not shared outside the consortium. **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	N/A

NAME	**eTryOn_WP1_003_RTD_QCAvatarSet**
Data summary	**Responsible partner:** Metail, ODLO, QuantaCorp **Purpose:** Collect any and all scans taken by consortium members using QuantaCorp technology. **Type/format**: images (PNG), 3D model (OBJ), skeleton (BVH) **Re-use of existing data:** No existing data is being re-used in this data set. All scans in the set are new and taken by consortium members. **Data origin:** Scans taken by consortium members. **Expected size:** ~25 **Data utility:** The data in this set is used as input to Metail's processing pipeline with the purpose of producing a standardized scanatar for WP2 and any work package that may depend on WP2.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The data is only discoverable by consortium members with access to the project's Google Drive folder. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

FAIR Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. The data set is very sensitive of nature (images) and any person scanned is informed his/her scan will be used exclusively for research purposes by consortium members for the duration of the eTryOn project. **How it will be accessible:** The data is only accessible to consortium members with access to the project's shared Google Drive. **Methods/software tools to access data**: Desktop clients for Google Drive or a web browser. **Repository:** The data set is stored in the 'Data Repository/Avatars' folder on the project's Google Drive. **Restrictions on access:** Access is restricted to consortium members only.
FAIR Data: **Interoperability**	**Interoperability:** The data uses standardized formats like Portable Network Graphics (PNG), Bounding volume hierarchy (BVH) and Wavefront .obj file (OBJ). **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** N/A **Availability for re-use:** This data set is not intended to be re-used outside of the eTryOn project. **Usable by third parties after end of project**: N/A **Re-use timeframe:** This data set will be used for the entire duration of the project. **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It is fully complying with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The data contains sensitive personal information of a person, but will only be shared with consortium members. The project defines clear legal terms for sharing data, and has put an ethics board in place to monitor data usage and to make sure guidelines are being followed. **Is informed consent for data sharing and long term preservation given:** N/A (participants have been asked by mail for their voluntary participation. Additionally, being ODLO fit models, they have signed a contract where they have agreed to collect their measurements and take pictures). It is worth noting that this data will only be used

	for the development of the eTryOn application during the project duration. The data that come from scans of the end users are included in the PLT datasets and contain detailed informed consents.
Other issues	N/A

NAME	**eTryOn_WP2_004_RTD_ObiParamEstimation**
Data summary	**Responsible partner:** Metail, CERTH **Purpose:** 3D sequences of the results of physics-based-simulation on avatars undergoing a number of animated movements, for Browzwear and ObiCloth physics simulations. Used to predict ObiCloth simulation parameters from Browzwear material parameters. **Type/format**: Animated body models: FBX (~2MB / motion), internal temp Reference garment animation: ABC (~150MB), internal temp A sequence of rendering images for internal loss computation: PNG (100KB / sample), internal temp Obi param images: PNG (100KB / sub-mesh), output Optimised obi scales: JSON (~250 Byte), output **Re-use of existing data:** Doppleganger body mody models: FBX, input Skinned garment: FBX, input **Data origin:** Will be created by Metail during the project, with commercial and project-created software. CERTH will support by creating the ObiCloth simulations. **Expected size:** During the process of a simple garment (e.g. t-shirt consisting of 5 sub-meshes) for one body size model, it will create around 170MB of data in multiple formats. The size may increase depending on various factors including a) the complexity of a 3D garment b) the number of sub meshes c) the length of a testing body motion. **Data utility:** Only of specific utility within the project, and for ongoing Metail R&D. Re-use by others for other purposes is not anticipated.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: The dataset will only be made available inside the project. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** We're not expecting this data to have longer term value, but if it does it will fall under the exploitation plan for the project. If this does not have a clear plan for exploitation as a

	closed asset it will be made open. It's possible that we'll publish the technique of parameter prediction, in which case this dataset will be submitted as supporting information and made available through the journal and / or OA preprint service. **How it will be accessible:** Subset of data might be shared to support publication. **Methods/software tools to access data**: GDrive **Repository:** Metail internal GDrive **Restrictions on access:** FAIR access to consortium partners as per CA.
FAIR Data: **Interoperability**	**Interoperability:** We will use the standard formats such as FBX where they are supported by the software used and to the extent that they're supported. We note that format standardisation and interoperability is generally poor in this area. **Data and metadata vocabularies**: None applicable **Use of standard vocabularies:** n/a **Mappings to commonly used vocabularies**: n/a
FAIR Data: **Reusability**	**License:** None available by default **Availability for re-use:** TBD depending on exploitation plan. **Usable by third parties after end of project**: TBD depending on exploitation plan. **Re-use timeframe:** TBD depending on exploitation plan. **Data quality assurance process:** TBD depending on exploitation plan.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** During the project the data will be held on AWS S3 and backed up to Google Storage Cloud or Google Drive. Both systems will have authentication to limit access to consortium partners.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** None known. **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	None.

NAME	**eTryOn_WP5_005_RTD_UserMedia**
Data summary	**Responsible partner:** CERTH

	Purpose: A photo of the user that is taken through the Dress Me Up application, in order to use it in conjunction with a garment and create a synthesized garment fitted photo. **Type/format**: .jpg **Re-use of existing data:** N/A **Data origin:** User submitted through the Dress Me Up app. **Expected size:** N/A **Data utility:** This dataset is useful for the service that will be creating the synthesized photo. The User Media dataset will be periodically purged.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: No. **Search keywords:** No. **Versioning:** No. **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** No. **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** Google Firebase Storage **Restrictions on access:** N/A
FAIR Data: **Interoperability**	**Interoperability:** No. **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** N/A **Availability for re-use:** No. **Usable by third parties after end of project**:No. **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** During the project the data will be held on Google Storage Cloud. The datasets will be periodically purged from the cloud.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The dataset depicts users, therefore it is classified as sensitive data. **Is informed consent for data sharing and long term preservation given:** Yes.
Other issues	N/A

NAME	eTryOn_WP4_006_RTD_SynthesizedPhotos
Data summary	**Responsible partner:** Metail **Purpose:** A synthesized 2D image that is created from Metail's composition service. It's the output file from the core action initiated by the user in the Dress Me Up application, where a user photo and a garment file are provided to create an image with a virtual fit in it. **Type/format**: .jpg **Re-use of existing data:** N/A **Data origin:** It is generated from a service. **Expected size:** N/A **Data utility:** This data is useful to the user of the application.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: No. **Search keywords:** No. **Versioning:** No. **Metadata creation:** No.
FAIR Data: **Accessibility**	**Data openly accessible:** No, because each created image is tied to a specific user and depicts them. **How it will be accessible:** The data will not be accessible. **Methods/software tools to access data**: N/A **Repository:** Google Firebase Storage **Restrictions on access:** Access is provided to each user after logging in the platform.
FAIR Data: **Interoperability**	**Interoperability:** No. **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** N/A **Availability for re-use:** No. **Usable by third parties after end of project**: No. **Re-use timeframe:** Until the user deletes it. **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** During the project the data will be held on Google Storage Cloud. Each user is authenticated through the app in order to access their files.

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long term preservation given:** N/A
Other issues	N/A

NAME	**eTryOn_WP5_007_RTD_UserScreencap**
Data summary	**Responsible partner:** CERTH **Purpose:** A screenshot that depicts the user wearing a garment from inside the Fit View of Magic Mirror application. **Type/format**: .jpg **Re-use of existing data:** N/A **Data origin:** User created through the Magic Mirror app. **Expected size:** N/A **Data utility:** This dataset is useful for the user, to be able to save fitted garments that they like.
FAIR Data: **Findability, including provisions for metadata**	**Is data discoverable**: No. **Search keywords:** No. **Versioning:** No. **Metadata creation:** N/A
FAIR Data: **Accessibility**	**Data openly accessible:** No. **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** Google Firebase Storage **Restrictions on access:** N/A
FAIR Data: **Interoperability**	**Interoperability:** No. **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
FAIR Data: **Reusability**	**License:** N/A **Availability for re-use:** No. **Usable by third parties after end of project**:No. **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A

Security	**Security measures:** During the project the data will be held on Google Storage Cloud. Each dataset entry can be deleted by users, requiring access rights.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The dataset depicts users, therefore it is classified as sensitive data. **Is informed consent for data sharing and long term preservation given:** Yes.
Other issues	N/A

4.2 Datasets for piloting activities

NAME	**eTryOn_WP6_005_PILOT_UserRequirmentsQuestionnaries**
Data summary	**Responsible partner:** ODLO **Purpose:** Two structured questionnaires (one for designers and one for fashion lovers/customers) has been developed by ODLO for the collection of user requirements as a part of the deliverable D6.1 "The user requirements". It will help us define a list of requirements, including both general requirements concerning the eTryOn technical functionalities and features but also requirements focusing on the specific use cases (e.g. VR experiences, AR experiences). The questionnaires include questions regarding the features that users think are important for this kind of apps, their past experience with similar applications, demographic data (age, gender) etc. The findings of the questionnaires have been summarized in D6.1. **Type/format**: Google form questionnaires. For analysing: export into Excel documents containing questions and user responses **Re-use of existing data:** The data is original **Data origin:** Questionnaires filled by ODLO designers and partners' customers and employees and the general public (shared through social media) **Expected size:** A few MB in total **Data utility:** This data will be used to structure the user requirements in WP6. The insights of which will be then shared to technical partners for developing the three applications of eTryOn. Therefore, these questionnaires address all technical partners.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The questionnaires are stored on a project's google drive folder, in which only the partners have access. So the raw data is not discoverable for third parties outside the project. The aggregated knowledge from this data was made accessible to the public through D6.1. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. After aggregation and processing, analysis results based on this data was shared with the consortium (only) and the findings of the questionnaires were part of the D6.1, but no sensitive information was gathered since the questionnaires were anonymous and only demographic non-identifying questions were included (age and gender). **How it will be accessible:** Stored in google drive and it is **only** internally accessible by project partners. Therefore the data is not accessible to third parties outside the project. **Methods/software tools to access data**: Web-browser (only by project partners) **Repository:** The raw questionnaires are stored in Google Drive. D6.1 is available on the project website. **Restrictions on access:** Shared among project partners with access to project's Google Drive.
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** The data will not be licensed since it will only be used internally **Availability for re-use:** This data is not expected to be re-used. It will be used once from partners to specify the user requirements for the three applications. **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** Raw data is cleaned and pre-processed as described in section 3.5. Moreover, given the small sample of the questionnaire a manual control was performed by ODLO to ensure data quality.
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The data is stored on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data are anonymous and do not contain any personal information, besides non-identifying demographic information (gender and age). Only aggregated forms of the data are made publicly available through D6.1.

	Is informed consent for data sharing and long term preservation given: N/A (Raw data will not be shared. Regarding the use of the data, the participants of the questionnaires were informed on it by adding the data usage information at the start of the questionnaires)
Other issues	N/A

NAME	**eTryOn_WP6_006_PILOT_EVALUATION-DATA**
Data summary	**Responsible partner:** ODLO **Purpose:** Structured questionnaires will be developed by WP6 partners for the evaluation of the developed tools in the context of the various use cases during the pilot trials. The questionnaires will include questions that cover issues such as usefulness, usability, visualisation and interaction, learnability, encountered problems and future expectations, etc. as well as user demographics. This dataset includes questionnaires filled by the end users to assess the tools developed. Data collected through the questionnaires is used exclusively for analysis and statistical purposes. **Type/format**: Word/Excel documents containing questions and user responses **Re-use of existing data:** No **Data origin:** Questionnaires filled by end-users in the context of pilot evaluation **Expected size:** A few KBs per questionnaire. A few MB in total **Data utility:** This data will be used in the context of WP6 to evaluate eTryOn technologies. The evaluation results of the three pilot cases will be used by the technical partners to improve eTryon's tools as part of further development and commercial exploitation activities. The data will be also used in D6.3 "Pilot evaluation report" in M24.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The questionnaires and the answers will be stored on a project's google drive folder, in which only the partners have access. Access to questionnaires will be given to participants through Google Forms. The answer data is not discoverable for third parties outside the project. Access to the statistical analysis of the results will be given to the public through D6.3, which will be a public deliverable. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** Raw data is considered internal working material. Hence, interviews, questionnaires, assessments, etc. will be considered confidential and will only be accessible by the consortium. After aggregation and processing, analysis results based on this data will be shared with the consortium and included in relevant deliverables. In case of a report or paper submitted for

	publication, all research findings will be integrated into the report or paper. Datasets will not be added to the publication. **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** N/A. **Restrictions on access:** N/A
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A **Costs for long-term preservation:** N/A
Security	**Security measures:** The datasets will be stored in the premises of the respective pilot leaders and (if needed) on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** These datasets may contain personal information of end-users, but they will not be shared. **Is informed consent for data sharing and long term preservation given:** As stated in D8.4, Informed Consent Forms (electronic or written) will be shared to all participants (designers, influencers and customers) for the participation in the pilots. These forms will describe all the rights of the participants and will make clear that the treatment of the data is confidential, complies with GDPR and is carried out exclusively for analysis and statistical purposes. If there is a need for data sharing and long term preservation, the participants will be informed accordingly in the Informed Consent Forms.
Other issues	N/A

NAME	**eTryOn_WP6_007_PILOT_EVALUATION-LOG-FILES**
Data summary	**Responsible partner:** CERTH, Metail, QC, MLZ, ODLO

	Purpose: This dataset will include log files that will be automatically created during pilot evaluation sessions. These files will include, but not limited to, context information, timestamps and general event information along with associated analytics which will be created during the test session by users of eTryOn technologies. This content will be used for evaluation purposes and demonstration of the results of the project. **Type/format**: Log files regarding usage of eTryOn applications. Google form questionnaires. For analysing: export into Word/Excel documents containing questions and user responses. **Re-use of existing data:** No **Data origin:** Log files created by the usage of eTryOn from end-users in the context of pilot evaluation **Expected size:** Depends on the number of participants. **Data utility:** This data will be used in the context of WP6 to evaluate eTryOn technologies, understand how useful the tools can be to the users and demonstrate the capabilities of the project technologies.
Fair Data: **Findability, including provisions for metadata**	**Is data discoverable**: The data will not be discoverable for third parties outside the project. **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Fair Data: **Accessibility**	**Data openly accessible:** The data will not be openly accessible. If there is a need for publishing some of the analytics (for instance as a part of a deliverable) all the sensible information will either be removed or pseudoanonymized (see deliverable D8.4). **How it will be accessible:** N/A **Methods/software tools to access data**: N/A **Repository:** N/A. **Restrictions on access:** N/A
Fair Data: **Interoperability**	**Interoperability:** N/A **Data and metadata vocabularies**: N/A **Use of standard vocabularies:** N/A **Mappings to commonly used vocabularies**: N/A
Fair Data: **Reusability**	**License:** N/A **Availability for re-use:** N/A **Usable by third parties after end of project**: N/A **Re-use timeframe:** N/A **Data quality assurance process:** N/A
Allocation of resources	**Costs for making the data FAIR:** N/A

	Costs for long-term preservation: N/A
Security	**Security measures:** The datasets will be stored in the premises of the respective pilot leaders and (if needed) on a Google Drive folder of the project. The google drive is restricted only to registered users while registration is possible only by invitation. Access requires username/password authentication. It fully complies with the European and international framework and the GDPR (see section 3.7).
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** These datasets may contain personal information of end-users, but they will automatically anonymized or at least become pseudonymous if a part of them needs to be shared. **Is informed consent for data sharing and long term preservation given:** As stated in D8.4, Informed Consent Form (electronic or written) will be shared to all participants (designers, influencers and customers) for the participation in the pilots. These forms will describe all the rights of the participants and will make clear that the treatment of the data is confidential, complies with GDPR and is carried out exclusively for analysis and statistical purposes. If there is a need for data sharing and long-term preservation, the participants will be informed accordingly in the Informed Consent Forms.
Other issues	N/A

5. Conclusions

This deliverable updates the datasets managed by the eTryOn consortium as initially provided in D8.2. In the present document a more detailed list of the different datasets is provided, including both pre-existing ones and newly-created within the duration of the project. For its conclusion, a template was shared to all partners in order to fill in the information related to the datasets aligned with the “Template for Horizon 2020 Data Management Plan (DMP)”.

We kept the initial cluster of the datasets in the following two categories depending on their purposes:

- Supporting research and technical development (RTD)
- Resulting from pilot activities (PILOT)

We enlarged the above-mentioned clusters by providing additional datasets and updating the existing ones with more detailed information incorporating the new guidelines that have arisen along with the applications development and the progress of the project’s implementation.

The consortium guarantees that it will take all the suitable measurements to make the data FAIR. In any case, eTryOn datasets will either be openly shared (by uploading them in open repositories) or shared internally among specific partners (stored on the project Google Drive). Datasets to be openly shared, will be deposited in certified repositories like Zenodo that have in place strong mechanisms and protocols for data security and long-term data preservation. Similar mechanisms exist in both the Google Drive repository and the partners’ servers to ensure data protection.