Deliverable 7.3

DATA MANAGEMENT PLAN – UPDATE 1

CHORIZO PROJECT

D7.3

Data management plan – update 1

Full title project	Changing practices and Habits through Open, Responsible, and social Innovation towards ZerO food waste
Call	HORIZON-CL6-2021-FARM2FORK-01
Topic	HORIZON-CL6-2021-FARM2FORK-01-13
Type of action	HORIZON-RIA
Project coordinator	EV ILVO
Project URL	https://chorizoproject.eu/
Due date of delivery	31st March 2024
Submission date	30th March 2024
Dissemination level	PU - Public **Document type** DMP – Data Management Plan
Work package	WP7
Author(s)	Erika De Geest Fleur Marchand
Document version	1.0

Grant agreement	101060014	**Duration**	36 months
Start Date	1st October 2022	**End date**	30th September 2025

Contributors

Name	Organisation
Erika De Geest	EVILVO
Isabeau Coopmans	EVILVO
Fleur Marchand	EVILVO
Capwell Echo Forbang	EVILVO
Dakerlia Claeys	EVILVO
Bart Ampe	EVILVO

Revision history

Version	Date	Reviewer	Modifications
0.1	7-03-2024	Capwell Echo Forbang	Minor corrections
0.2	22-03-2024	Gabriela Alcat	Minor corrections
0.3	26-03-2024	Concha Ávila	Doc and corrections review- minor changes
0.4	30-03-2024	Erika De Geest	Corrections based on the reviewing feedback + finalizing for upload

**Funded by
the European Union**

*This project has received funding from the Horizon 2020 research and innovation programme
under Grant Agreement No 101060014*

Disclaimer

The content of this document reflects only the author's view. Neither the European Commission nor REA are responsible for any use that may be made of the information it contains.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the CHORIZO consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the CHORIZO Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the CHORIZO Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

Table of content

List of figures

List of tables

Glossary of terms and acronyms used

Acronym/Term	Description
CC0	Creative Commons no rights reserved
CC BY	Creative Commons Attribution License
CERN	Conseil Européen pour la Recherche Nucléaire
CS	Case Study
CSV	Comma-separated values
D	Deliverable
DDMMYY	Day Month Year in 2 numbers e.g. 100323
DM	Data Management
DMP	Data Management Plan
DMT	Data Management Team
Docx	Microsoft Word text document
DOI	Digital Object Identifier
DPO	Data Protection Officer
DPS	Data Protection Steward
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
FLW	Food Loss & Waste
FW	Food Waste
GDPR	General Data Protection Regulation
GIT	Global Information Tracker
HE	Horizon Europe
HUMAT	A socio-cognitive agent architecture in the SMARTEES H2020 project
IP1, 2, 3, …	Impact Pathway 1
ISO	International Standards Organisation
JRC	Joined Research Center
M	Month
MD5	Message Digest Algorithm 5
MOA	Motivation, Opportunity, Ability
MP3, MP4	Motion Pictures Expert Group Audio Layer 3, 4
MS	Microsoft

Horizon Europe

Acronym/Term	Description
O1	Objective 1
OEI	Other Ethics Issues
ORD	Office of Research and Development
PID	Personal Identifiable Data
R1, 2, 3, …	Result 1
R&I	Research & Innovation
RAR	Roshal Archive format
SSL/TLS	Secure Sockets Layer/Transport Layer Security
TXT	Text
USB	Universal Serial Bus
WP	Work Package
ZIP	Compressed file

Executive summary

This deliverable is the first update of the Data Management Plan (DMP) for the CHORIZO project.

In this report, the ethics requirements, the data management policies and data protection policies are be defined in accordance with the current EU data protection regulations (GDPR), the European Code of Conduct for Research Integrity and national regulations in the CHORIZO partners' countries. The ethics requirements ensure compliance with the above-mentioned regulations, identifying potential ethical issues and ensure that the research is conducted at the highest level of integrity, quality and transparency. The data management plan describes the data life cycle for the data (to be) collected, processed and generated in CHORIZO, in alignment with the Data Protocol in T1.1: handling of research data during and after the end of the project; data (to be) collected, processed and/or generated; methodology & standards (to be) applied; whether data is shared/made open access; how data will be curated & preserved (including after the end of the project). The plan evolves and is updated during the lifetime of the project. So a wide variety of data is to be created, collected and/or re-used.

The DMP contributes to meeting laws and regulations obligations regarding data management. It follows the template provided by HE on data plans and data management, describing what research data the project expects to generate and how the principles of FAIR data management should be met.

Furthermore, it gives instructions on naming conventions, metadata structure, storing of the research data and how to make public data available.

Finally, it outlines the security and ethical consideration that must be met.

This DMP is the second version of a total of 3 versions planned and will consequently be updated and revised. The final version is planned at M36. These are all public versions and will be uploaded on the Participants Portal. They are maintained in the CHORIZO collaborative workspace and are accessible for all consortium partners. All partners are required to review each new version of the DMP and to comply to the processes and requirements set in them.

1   INTRODUCTION

1.1   Main aims of CHORIZO

CHORIZO aims to improve the understanding of how social norms influence behaviour related to FLW generation. To significantly accelerate progress towards zero food waste, CHORIZO aims to use this knowledge to increase the effectiveness of decision-making and engagement of food chain actors, in changing social norms towards zero food waste.

To achieve its aim, the project provides evidence on the role of existing social norms in actors' FLW behaviours through translating results from previous FLW actions into evidence and generate new evidence on social norms & FLW behaviours.

Second, CHORIZO embeds these research results into innovation products that can foster change of FLW-related social norms. These products include more effective (sector-based) guidance, communication & science education packages and capacity building actions.
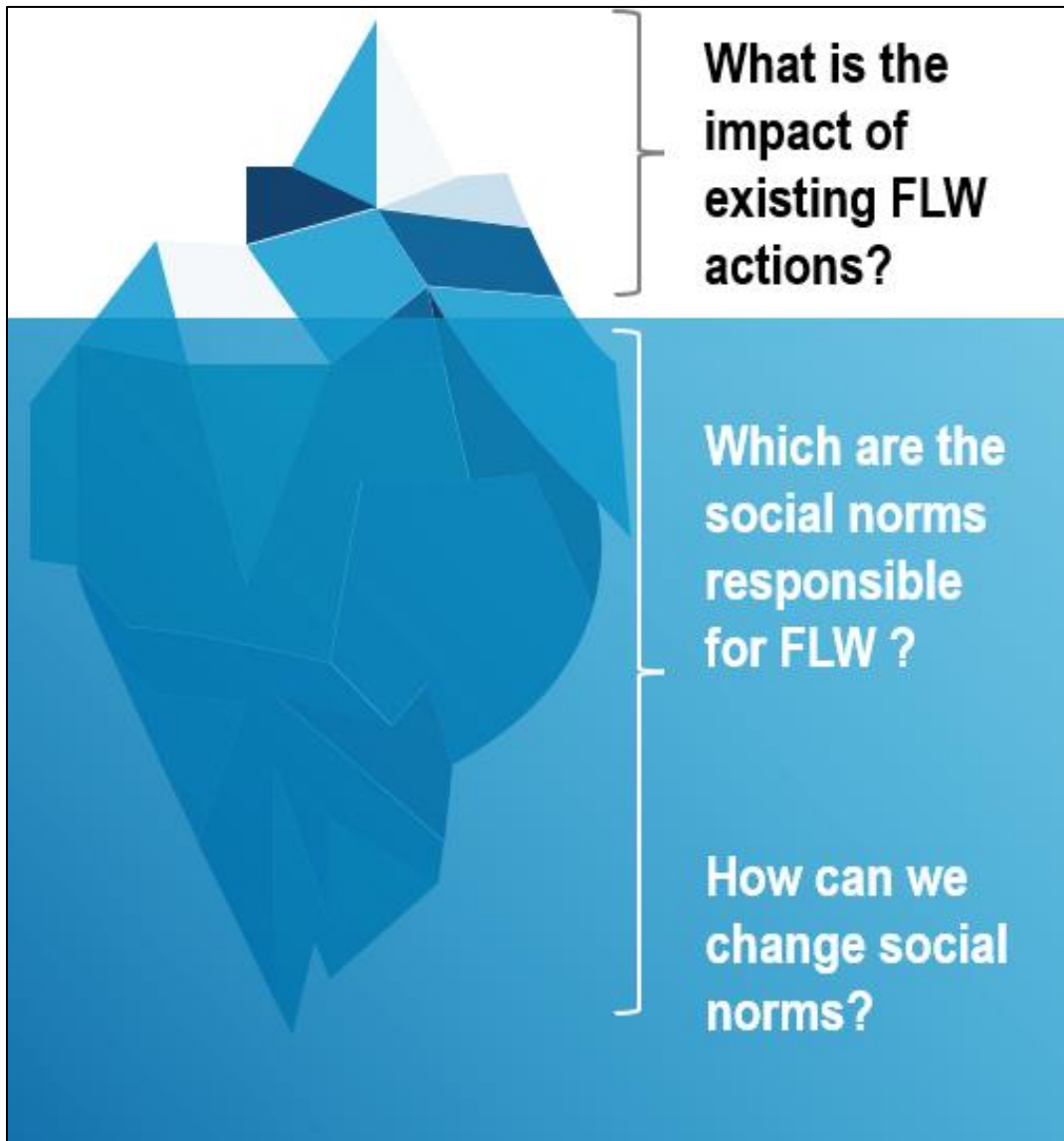

Figure 1 CHORIZO at a glance

CHORIZO employs real-life Case Studies (CSs) to serve its three interlinked purposes:

1. to provide information and data on the context and impact of previous FLW prevention/reduction actions undertaken by the Case Study members, thus enriching the evidence-based analysis on previous FLW actions;
2. to generate new evidence on the interaction between social norms, behaviour and food waste, to feed into the project's FW models and innovation products;
3. to validate the communication & science education packages of the project.

Moreover, on the technical side, CHORIZO utilises advanced modelling techniques to produce solutions that integrate behavioural and economic theories and integrate gender and intersectional analysis to interpret social norm and behavioural data and to effectively engineer innovation processes and outputs.
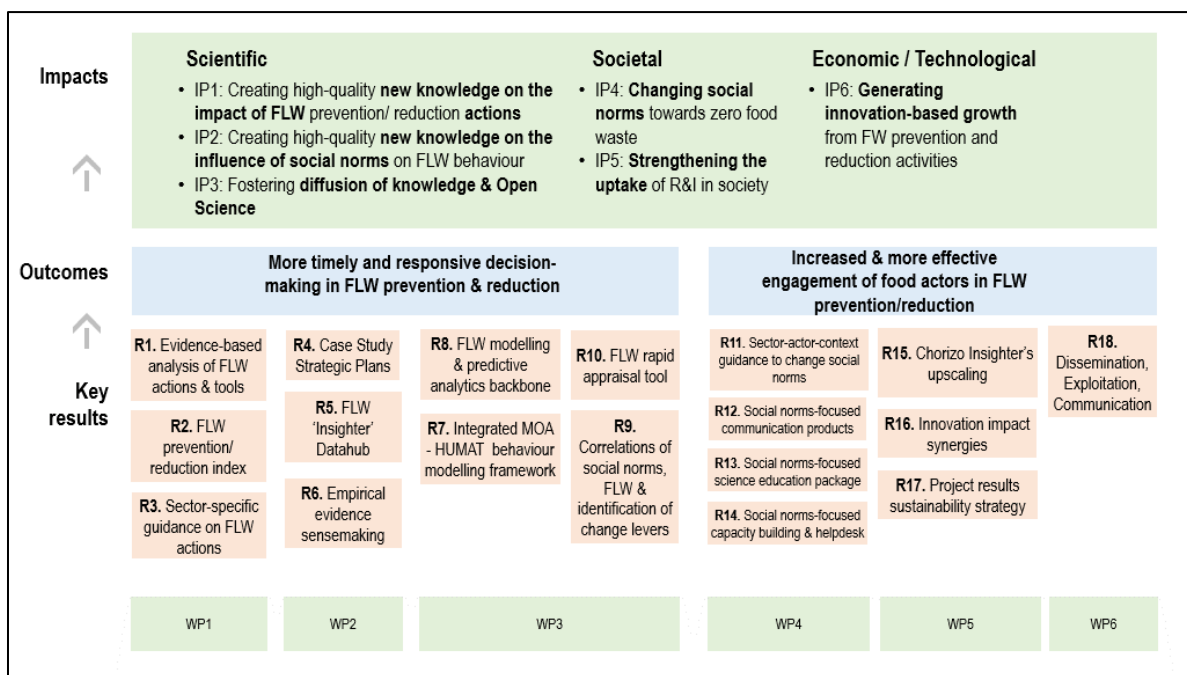
Figure 2 CHORIZO's results, outcomes and impacts

To reach its goals CHORIZO is divided into 8 WPs (see Figure 2) with different goals, tasks and deliverables. This document is originating from WP7, Project Management. **This document supports these objectives by outlying a Data Management Plan that describes the principles and procedures compliant with EU rules and regulations, which the project should employ to manage data generated and/or collected (e.g. re-used) throughout the project duration.**

1.2 From CHORIZO outputs to impact

CHORIZO delivers 3 kinds of outputs aiming for different kinds of impact.

First, new evidence is generated on different aspects leading to IP1, 2 and 3 (Figure 2). The project provides new knowledge on the appropriateness and impact of specific FLW prevention/reduction actions and on how social norms influence FLW behaviour. CHORIZO fosters the diffusion of this new knowledge through Open Science.

CHORIZO has developed sector specific guidance to change social norms, including policy. Communication and education packages together with capacity building activities will contribute in changing social norms towards zero FLW at the consumer, corporate and policy level (IP4).

A second set of outputs are tools that support the innovation within FLW prevention and reduction activities (IP 5 and 6). For example, a FLW prevention/reduction index can guide food actors in prioritising alternative actions based on results previously achieved. The FLW rapid appraisal tool being a simplified and interactive version of the modelling backbone, provides actors with knowledge on the key social norms that drive FLW in their contextual environment. Furthermore, there will be emphasis on the exploitation of the CHORIZO 'Insighter' datahub involving the rules-based data sharing of empirical evidence between data providers and consumers, and the provision of services to corporate actors pursuing FLW prevention initiatives.

1.3   Data management in CHORIZO

As foreseen in the first DMP, EVILVO took up the role of the **Data Manager (DM)** is for this project. Furthermore, each project partner has appointed a **Data Protection Steward (DPS)** in their respective organization (see Table 1).

The **Data Management Team (DMT)** is the team that consists of the DPSs at the level of the CHORIZO project.

The DPS is a participant/partner of the CHORIZO project, that collects, stores and/or analyses the data for CHORIZO purposes, on daily basis. This person needs to ensure that all data is handled according the agreements and procedures in the DMP. He/she interacts with the Data Protection Officer (DPO) within their organization, as well as with the DMT of the CHORIZO project.

The primary role of the DPS is to ensure that his/her organisation processes the personal data of the research participants in compliance with the applicable data protection rules.

Another function of the DPS is, if a security incident (e.g. data breach, loss/theft of laptop or USB stick, ...) occurs, then this must be compulsorily reported without delay to the DMT.

The DMT determines, in mutual consultation, whether reporting is necessary to the relevant supervisory body at that time and, if necessary, to those involved and this, within 72 hours after discovery of the incident.

This **data management plan (DMP)** describes the data and procedures to collect, process and store the data, thus handling of research data during and after the end of the project. All in alignment with the deliverable D1.1 Data protocol. Figure 3 shows the flow of data handling in CHORIZO and the issues addressed in this DMP:

- methodology and standards to be applied, naming conventions, metadata etc.
- pseudonymisation/anonymisation guidelines
- storage during lifetime of the project, security measures
- main output, how data will be curated and preserved (including after the end of the project)
- whether data will be shared/made open access, licensed and stored

The plan evolves and be updated during the lifetime of the project.

As described in the first DMP, the DM organized a meeting on Wednesday 24th May 2023 to update all DPSs about the procedures. The DMT also had a meeting on Tuesday 16th January 2024 to update and discuss changes in the procedures. The meeting minutes are accessible for all partners on the collaborative Chorizo Teams channel.

1.4   Consortium members in CHORIZO

Since the beginning of the project on 1st October 2022, there have been changes in the consortium composition. Nordic Choice Hotels changed to Strawberry and CTIC CITA has been taken over by CNTA. All administrative procedures are still taking place to formalise this. Table 2 shows an overview of the active members of the consortium on 31st March 2024.

① **Naming conventions**

② **Pseudonimisation/ anonymisation**

③ **Storage during project lifetime for analyses**

④ **Main Output data**

⑤ **Open Access, License & Storage**

CHORIZO collection methods	CHORIZO DATA	Naming conventions	Pseudonimisation/ anonymisation	Storage during project lifetime for analyses	Main Output data	Open Access, License & Storage

See D1.1 DATA PROTOCOL and D2.1 CS Strategic Plans

RESEARCH DATA
e.g.: Interview transcripts
Excel spreadsheet containing survey responses

Section 2.1, 2.3 and D1.1. DATA PROTOCOL

See Annex 1 for guidelines

Security
See guidelines in section 4.2
- Local storage
- Encrypted on Teams
- Key password
- Only accessible for researchers that need it!

1. Metadata

See D1.1. DATA PROTOCOL

2. Deliverables

See Table 3

3. Tools
- FLW index
- Insighter datahub
- Behavioural change model
- FLW rapid appraisal tool

Open Access for anonymised data/results (See section 2.2)

No embargo
CC BY 4.0C (See section 2.4)

- on website CHORIZO
- on Teams
- on Zenodo
- Insigther datahub

CHORIZO activities

Meetings
Reporting
Communication
…

OTHER DATA
e.g.:
Meeting Minutes
Agenda
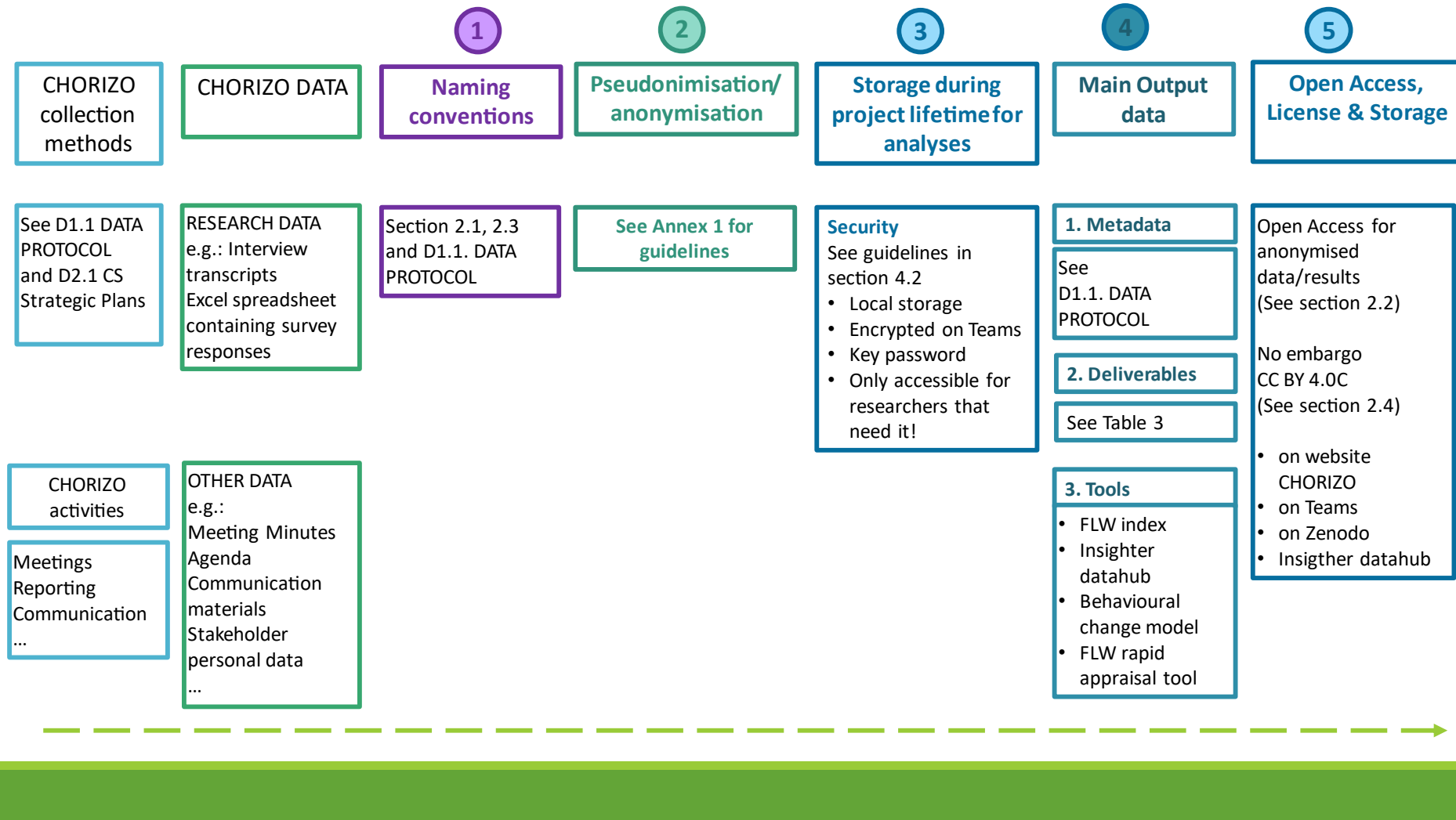Communication materials
Stakeholder personal data
…

Figure 3 Flow of data handling in CHORIZO

List of designated DPSs

Participant No	Participant organisation name	Short name	Country	Name	E-mail
1 (Coordinator)	Eigen Vermogen Van Het Instituut Voor Landbouw – En Visserijonderzoek	EV-ILVO	BE	Erika De Geest	erika.degeest@ilvo.vlaanderen.be
2	VLTN GCV	VLTN	BE	Charis Latinopoulos	charis.latinopoulos@vltn.be
4	Collaborating Centre for Sustainable Consumption and Production	CSCP	DE	Rosalyn Old	Rosalyn.old@cscp.org
5	Norwegian Research Centre AS	NORCE	NO	Patrycja Antosz	paan@norceresearch.no
6	University of Bologna	UNIBO	IT	Caterina Rettore	caterina.rettore2@unibo.it
				Filippo Pini	filippo.pini5@unibo.it
7	ICLEI European Secretariat Gmbh	ICLEI	DE	Simon Gresset	simon.gresset@iclei.org
8	Hungarian Food Bank Association	HFBA	HU	CSEH, Balázs	cseh.balazs@elelmiszerbank.hu
9	Asociacion Para La Investigacion Desarrollo e Innovacion Del Sector Agroalimentario	CTIC CITA	ES	Ixone Alonso Miquélez	ialonso@cnta.es
10	ITC - Innovation Technology Cluster Murska Sobota	ITC	SI	Miran Bunderla	miran.bunderla@itc-cluster.com
11	Pomurje Chamber of Commerce and Industry	PCCI	SI	Robert Grah	robert.grah@gzs.si
12	Spanish Federation of Food and Beverage	FIAB	ES	Concha Ávila	c.avila@fiab.es
13	University of Copenhagen	UCPH	DK	Bent Egberg Mikkelsen	bemi@ign.ku.dk
14	Nordic Choice Hotels	NCH	NO	Sigve Eliassen	Sigve.eliassen@strawberry.no
15	INLECOM Innovation	INLEIN	GR	Sissi Koronaiou	Sissi.Koronaiou@inlecomsystems.com

Table 1 List of CHORIZO's designated Data Protection Stewards

CONSORTIUM MEMBERS

Participant No	Participant organisation name	Short name	Country	Contacts	E-mail
1 (Coordinator)	Eigen Vergmogen Van Het Instituut Voor Landbouw – En Visserijonderzoek	EV-ILVO	BE	Fleur Marchand	fleur.marchand@ilvo.vlaanderen.be
				Isabeau Coopmans	Isabeau.Coopmans@ilvo.vlaanderen.be
				Jürgen Vangeyte	Jurgen.Vangeyte@ilvo.vlaanderen.be
				Erwin Wauters	erwin.wauters@ilvo.vlaanderen.be
				Bart Van Droogenbroeck	Bart.vandroogenbroeck@ilvo.vlaanderen.be
				Anna Twarogowska	Anna.twarogowska@ilvo.vlaanderen.be
				Capwell Forbang Echo	capwellforbang.echo@ilvo.vlaanderen.be
				Karima Abunada	karima.abunada@ilvo.vlaanderen.be
				Erika De Geest	erika.degeest@ilvo.vlaanderen.be
2	VLTN GCV	VLTN	BE	Aristos Halatsis	aristos.halatsis@vltn.be
				Stathis Zavvos	stathis.zavvos@vltn.be
				Charis Latinopoulos	charis.latinopoulos@vltn.be
				Veerle Leemen	veerle.leemen@vltn.be
				Paulina Szwed	paulina.szwed@vltn.be
				Ramanuja Rao Kotaprolu	ramanuja.rao@vltn.be
				Chantal den Broeder	chantal.denbroeder@vltn.be
4	Collaborating Centre for Sustainable Consumption and Production	CSCP	DE	Nora Brüggemann	nora.brueggemann@cscp.org
				Rosa Strube	rosa.strube@cscp.org
				Lea Leimann	lea.leimann@cscp.org
				Rosalyn Old	rosalyn.old@cscp.org
5	Norwegian Research Centre AS	NORCE	NO	Patrycja Antosz	paan@norceresearch.no
				Ivan Puga-Gonzalez	ivpu@norceresearch.no
				Markus Grendstad Rosseau	maro @norceresearch.no
				LeRon Shults	lesh@norceresearch.no
				Marius Berger	mber@norceresearch.no
6	University of Bologna	UNIBO	IT	Matteo Vittuari	matteo.vittuari@unibo.it
				Matteo Masotti	matteo.masotti8@unibo.it
				Caterina Rettore	caterina.rettore2@unibo.it
				Filippo Pini	filippo.pini5@unibo.it
				Eugenia Maria Modon	eugeniamaria.modoni@unibo.it
7		ICLEI	DE	Dominique Sandy	dominique.sandy@iclei.org

				CONSORTIUM MEMBERS	
Participant No	**Participant organisation name**	**Short name**	**Countr y**	**Contacts**	**E-mail**
	ICLEI European Secretariat Gmbh			Angèle Tasse	angele.tasse@iclei.org
				Simon Gresset	simon.gresset@iclei.org
				Nestor Gisasola Maiztegi	nestor.gisasola-maiztegi@iclei.org
8	Hungarian Food Bank Association	HFBA	HU	CSEH, Balázs	cseh.balazs@elelmiszerbank.hu
				Éva Bérces	berces.eva@edelmiszerbank.hu
				Katalin Katus	katus.laszlone@elelmiszerbank.hu
9	Asociacion Para La Investigacion Desarrollo e Innovacion Del Sector Agroalimentario	CNTA (CTIC CITA)	ES	Daniel De La Puente	ddelapuente@cnta.es
				Rodrigo Fernández Martín	rfernandez@cnta.es
				Clara Olave Sobrón	colave@cnta.es
				Irene González Pérez	igonzalezp@cnta.es
				Ixone Alonso Miquélez	ialonso@cnta.es
10	ITC - Innovation Technology Cluster Murska Sobota	ITC	SI	Daniel Copot	daniel.copot@itc-cluster.com
				Sasa Straus	sasa.straus@itc-cluster.com
				Tamara Kozic	tamara.kozic@itc-cluster.com
				Miran Bunderla	miran.bunderla@itc-cluster.com
11	Pomurje Chamber of Commerce and Industry	PCCI	SI	Robert Grah	robert.grah@gzs.si
				Roman Wolf	Roman.Wolf@gzs.si
12	Spanish Federation of Food and Beverage	FIAB	ES	Gabriela Alcat	g.alcat@fiab.es
				Concha Ávila	c.avila@fiab.es
13	University of Copenhagen	UCPH	DK	Bent Egberg Mikkelsen	bemi@ign.ku.dk
				Subash Rana	sr@ign.ku.dk
				Mukti Ram Chapagain	mrc@ign.ku.dk
14	Strawberry	STRAWBERR Y	NO	Harald Bjugstad-Holm	harald.bjugstad-holm@strawberry.no
				Sigve Eliassen	Sigve.eliassen@strawberry.no
15	INLECOM Innovation	INLEIN	GR	Ioanna Fergadiotou	ioanna.fergadiotou@inlecomsystems.com
				Makis Kouloumbis	gerasimos.kouloumbis@inlecomsystems.com
				Sissi Koronaiou	Sissi.Koronaiou@inlecomsystems.com
				Marissa Nousi	marisa.nousi@inlecomsystems.com

Table 2 Consortium members

2 FAIR DATA

In this section you can find the basic agreements and guiding principles for CHORIZO to support the handling of the data in a FAIR way. Several amongst them are still not decided and are topics of the DMT meetings.

2.1 Making data FINDABLE, including provisions for metadata

To make the data findable for project partners on the SharePoint, including the provisions for the metadata, the data are assigned a globally unique and eternally persistent identifier. Files names therefore are allocated consistently and are provided with a descriptive name so when organizing files it is obvious where to find specific data and what the files contain.

To order the documents on SharePoint, we store files and working documents using the WP structure.

For files and working documents, the following naming convention are used:

For any document:

- Project name_Document name_Event date DDMMYYYY

   o Example: CHORIZO_General Assembly Agenda_02022023

For (final) version of deliverables:

- Project name_Deliverable number + Deliverable Title_draft or final_version number

   o Examples: CHORIZO_D7.2 Data Management Plan_draft_0.2

      CHORIZO_D8.1 Ethics_final_1.2

Following remarks regarding the name of electronic records (files) should be followed:

- use _ instead of space
- preferably not exceed 255 characters (to ensure it is readable at 32bit and above operation systems)
- if the document is modified include a version number and the date of last modification as a denominator required for identification of file content

Clear Versioning

This refers to saving new copies of files, so previous versions can be referred to. Versioning is necessary to allow changes to be tracked and is useful to attribute changes when working on joint documents. When creating new versions of files, it is important to record what changes are being made to the files and to give the new files a unique name. A unique version number should be assigned to each version of a document depending on whether the changes are significant (major) or not (minor) and a new number should be allocated accordingly.

Strict versioning of the project files and documents is followed:

- The author of the document ensures the current version number is identified on the appropriate place on the first page of the document

- The first draft of a document is version 0.1, while subsequent drafts will be an increase of "0.1" in the version number (e.g., 0.2, 0.3, 0.4, 0.9, 0.10, etc.)
- The first final version of a document is version 1.0, while subsequent final documents will have an increase of "1.0" in the version number (1.0, 2.0, etc.)
- The second final version of a document after reviewing by a partner will be Version 1.1, 1.2, etc.

METADATA

CHORIZO datasets have followed the provenance metadata defined in the Data Protocol (D1.1):

- dataset PID
- dataset description
- dataset date of deposit
- dataset author(s)
- dataset venue
- dataset embargo
- HE funding statement
- project name
- project acronym
- project number
- licensing terms
- PID for the authors
- PID for the authors' organizations
- PID for the grant
- PID for related publications and other outputs

Moreover, the data protocol D1.1 defines additional datatype-specific metadata, enumerating all obligatory metadata to be collected by responsible partners.

Alongside the data from the project, relevant metadata will be deposited in repositories identified in D1.1. If the repository offers such an option, keywords for metadata will be available and metadata will be indexed and will be available for harvesting.

2.2   Making data ACCESSIBLE

Open data

All **data aggregated on the website and the Insighter datahub** are openly available. These include:

- Data regarding previous and on-going actions, case studies datasets; the latter are anonymised data and previously agreed with by the signature of an informed consent by the data owner.
- Data regarding the behaviour change model and resulting tools (FLW Index, FLW rapid appraisal tool when relevant, previously agreed with by the signature of an informed consent by the data owner).
- Data regarding events and activities, including venue, date, type of demonstrations, involved actors, etc.
- Deliverables, reports, scientific papers, educational and communication packages, capacity building materials, communication and dissemination materials.

No special software solutions are needed for accessing the databases. The required tools are: a web browser and MS Office or Linux for accessing the project documents, the Insighter Datahub and FLW rapid appraisal tool.

Datasets are anonymised where possible; when only pseudonymisation is possible, then the data set will not be open due to GDPR. See Annex 1 for guidelines.

Further information on informed consent procedures can be found in D8.2 OEI – Requirement No.2 (Ethics).

Closed data

All **sensitive data of stakeholders** in the possession of the consortium partners (e.g. email lists, contact lists) and used for communication and dissemination purposes are stored with password-based user authentication on the servers of the partners' institution. Furthermore, sensitive data (not only related to the identity of the contacts, but also related to the environment in which the data were collected: date of collection, format, hour, location, etc.) **must not** be made openly accessible.

Any information related to a natural person or 'Data Subject' that can be used to directly or indirectly identify the person, will be anonymised before storage on a server.

With respect to the business related data, any information related to identity (e.g. company ownership details) and property (e.g. land, building) are considered as sensitive data and have to be removed when anonymising data.

Selection of a Repository for Open access of data

CHORIZO will use the Zenodo repository as a main tool for making research data accessible in accordance with the HE Open Access mandate. If partners prefer another repository, they need to inform the DM. The selection of this repository has been discussed at the kick-off meeting; the identified repository is known to the partners from previous projects.

Zenodo is a portal based on the well-established git version control system and the Digital Object Identifier system [11]. The portal's aims are inspired by the FAIR principles, thus it represents a very suitable and natural choice in this context. It is operated by CERN. Zenodo is set up to facilitate the finding, accessing, re-using and interoperating of data sets, which are the basic principles that ORD (Office of Research and Development) projects must comply with. The repository services offered by Zenodo are free of charge and enable peers to share and preserve research data and other research outputs in any size and format: datasets, images, presentations, publications, and software. The digital data and associated meta-data are preserved through well-established practices such as mirroring and periodic backups. Each uploaded dataset is assigned a unique DOI rendering each submission uniquely identifiable and thus traceable and referenceable.

For data that will be publicly available through the Zenodo platform, Zenodo's metadata standards can be used. As all versions uploaded onto Zenodo are kept there, we will share the data on this platform after the European Commission Review (M21) in case a deliverable needs to be revised.

The following list briefly describes the Zenodo security settings:

- **Versions**: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- **Replicas and file preservation**: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.

- **Retention period**: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN has defined a lifetime for the repository of the next 20 years minimum.
- **Functional preservation**: Zenodo makes no promises of usability and understandability of deposited objects over time.
- **Fixity and authenticity**: All data files are stored along with an MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans**: In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject based repositories.

The dataset PID is assigned by the data repository.

2.3   Making data INTEROPERABLE

In CHORIZO, each data file/set is identified by a unique name, as proposed in this section and also set out in the D1.1 Data Protocol. The naming convention adheres to the following rules:

- A data file/dataset name consists of several parts (qualifiers) connected by periods,
- Each qualifier must begin with an alphabetic character (A-Z) or a special character ($,#,@),
- Each qualifier may contain alphabetic characters (A-Z), digits (0-9), a hyphen (-), or the special characters ($, #, @),
- Each qualifier should be as short and meaningful as possible,
- Each data file/set name should be as short and meaningful as possible,
- Data file/set name should include at least two quantifiers,
- The first qualifier refers to the data collection method/technique and starts with a capital letter (Table 3):

Data collection method/technique	Qualifier
Focus group interview	FGI
In-depth interview	IDI
Workshop	Wkshp
Webinar	Web
Online survey	Survey
Agent-based model simulation	ABM-sim
Qualitative desk research	Qual
Quantitative desk research	Quant

Table 3 First qualifier that refer to the data collection method

- If multiple files are produced with the use of the same method/technique in the same case study/task (e.g. multiple in-depth interviews), the first qualifier is appended with a unique number (e.g. IDI01, ABM-sim01),
- The second qualifier refers to the case study or task that produced the file/set and starts with a capital letter,
- Qualifiers denoting case studies begin with the CS and number (e.g. CS1, CS2) and include a hyphen followed by a two-letter ISO 3166-1 alpha-2 country code signifying where the data originates (e.g. Belgium = BE, Norway = NO). Examples of the second qualifiers denoting case studies are: CS1-BE, CS3-SI,
- Qualifiers denoting project tasks report the task number (e.g. T4.2),
- The third and further qualifier(s) can be used if needed and specify the content of the data file/set.

Proposed dataset names for all data types elicited in CHORIZO can be found in Table 1 in D1.1 Data protocol. The CHORIZO project data ontology is described and mapped in D1.1 Data protocol. Wherever possible, the data will include qualified references[1] to other data.

Figure 1 in D1.1 Data protocol presents the data collection methods/techniques as defined by the data ontology. Grouping the data collection methods and techniques enables us to standardize provenance metadata tailored to the collection approach.

CHORIZO has several dedicated tasks to guarantee the data quality: D1.1 Data protocol and D2.1 Case Study Strategic plan.

2.4 Increase data RE-USE

Licensing

Research data will be licensed under the latest version of **CC BY** (attribution required) or **CC0** (public domain), or equivalent. As a standard, we opt for: CC BY 4.0. However, both Share-Alike and Non-Commercial Share-Alike licenses can be considered for the parts of datasets for which the decision of making that part public has been made by the Consortium.

- **Attribution**: requires users of the dataset to give appropriate credit, provide a link to the license, and indicate if changes were made.
- **NonCommercial**: prohibits the use of the dataset for commercial purposes by others.
- **ShareAlike**: requires the others to use the same license as the original on all derivative works based on the original data.

Also the Metadata will be made openly available and licenced under a public domain dedication a Creative Commons Public Domain Dedication CC0 whenever the data repository allows for such an option.

Independent of the level of protection of the empirical data collected in the project, associated provenance metadata in CHORIZO must be open under CC0 or equivalent, to the extent legitimate interests or constraints are safeguarded. To facilitate the process of identifying and collecting relevant metadata, tools tailored to method types were prepared for project partners (D1.1).

By the end of the project, all simulation code present in the agent-based models from work package 3 will be open sourced with an Apache licence 2.0. This license strikes the right balance between being open source while allowing for commercial users to provide extensions or visualizations for them without having to open source their additions.

Re-use pathways

All open datasets, deliverables and scientific publications will be uploaded in Zenodo. Public deliverables are also made available on the official CHORIZO website http://chorizoproject.eu. They will also be linked to the OpenAIRE-community for maximum findability, after the first reporting period. The uploaded files will be tagged with metadata according to the Zenodo standards, as well as with the tag "CHORIZO".

---

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

An embargo period may be applied if the data (or parts of data) are used in published articles in "Green" open access journals. The recommended maximum embargo period length by European Commission is 6 months.

All provenance metadata types (general, project-specific, and dataset-specific as outlined in D1.1 Data protocol) will be available.

For now, but to discuss further, the data will be preserved for 5 years on the respective servers of the contributing research institutions. Data on the SharePoint platform will be preserved on the servers of EVILVO and data on the CHORIZO website on the hosting platform.

3   DATA OVERVIEW

3.1   Purpose of the data generation

CHORIZO aims to improve the understanding of how social norms influence behaviour and FLW generation and use this knowledge to improve the effectiveness of decision-making and engagement of food chain actors, towards zero food waste. To achieve its aim, the project will pursue six detailed objectives.

Objective 1:

To undertake a comprehensive evidence-based analysis of previous/ ongoing FLW prevention/ reduction actions and tools, including a cost/ benefit analysis and an impact assessment (economic, environmental, societal).

Corresponding WPs: WP1

Objective 2:

To develop an FLW Datahub (Chorizo FLW Insighter), which will incorporate: (i) the results of the evidence-based analysis of previous/ongoing FLW actions (O1); (ii) new empirical case study evidence on social norms, consumer behaviour, economic actor behaviour and charity (e.g. food banks) behaviour in relation to FLW. The FLW Datahub will provide more timely and easily accessible evidence to inform FLW decision making, accompanied by a stakeholder engagement and governance model to ensure the Datahub's evolution and sustainability beyond the project duration.

Corresponding WPs: WP2

Objective 3:

To develop a modelling & predictive analytics backbone based on data from the CSs, and use it to discover and explain the correlations between social norms, business practices, consumer behaviour and food waste, thus also providing insight into the reasons why previous activities did not bring benefits in line with the Green Deal's goal of 50 % reduced food waste by 2030 and how such benefits can be achieved.

Corresponding WPs: WP3

Objective 4:

To foster change in social norms and behaviours, by embedding the new evidence and understanding gained on FLW-related social norms and behaviours, into the design and delivery of: (i) more engaging and effective communication and school science education programs to prevent/reduce FLW; (ii) more effective capacity building activities, enabling food chain actors to deliver the change needed and to continue to innovate towards zero FLW.

Corresponding WPs: WP4

Objective 5:

To effectively manage the project's innovation upscaling, by: (i) guiding a strategy for the exploitation of the project results by the project partners (individually and jointly) as well as incentivising adoption by EU Food Waste Associations; (ii) incentivising an upscaling strategy for the services layer

of the CHORIZO datahub including a business model for sustaining the Datahub's operation after the end of the project; (iii) exploring opportunities for combining results with sister projects to achieve impact synergies; (iv) implementing responsible innovation management practices that guide and manage innovation from the project to exploitable and sustainable outcomes.

Corresponding WPs: WP5

Objective 6:

To engage with national EU, and global initiatives on a dialogue towards changing social norms to achieve a near-zero FLW and to undertake appropriate dissemination, exploitation and communication actions to maximize the project's impact outside the consortium.

Corresponding WPs: WP6

Objective 7:

To ensure the effective administration of the project activities according to the rules and regulations of the EU, and according to sound project management and coordination practices.

To ensure that all project outcomes are of high quality, meet their objectives and are delivered according to the agreed time- and resource planning.

To ensure that all potential risks are identified at an early stage, and appropriate mitigation actions are taken.

To ensure an effective interface with the EC.

Corresponding WPs: WP7

3.2   Types and formats of data

Due to the interdisciplinary nature of the project, the defined objectives and the various data needs, there will be several data origins, different data types generated, collected and stored and thus different formats needed.

The gathered data will be analysed or processed resulting into a set of deliverables (Table 4) and used in communication and dissemination activities. All public deliverables are uploaded to the data hub and publicly available.

Deliverable (number)	Deliverable name	WP	Short name of lead	Type	Diss. level	Delivery date	Quality reviewer
D1.1	Data protocol	1	NORCE	R	SEN	M3	EVILVO
D1.2	Evidence-based analysis of FLW actions/tools	1	VLTN	R	PU	M10	CSCP
D1.3	FLW prevention/reduction index	1	VLTN	R	PU	M12	UNIBO
D1.4	Sector-specific guidance	1	EVILVO	R	PU	M15	CSCP
D2.1	Case studies' Strategic Plans	2	EVILVO	R	PU	M3	NORCE
D2.2	CHORIZO FLW 'Insighter'	2	EVILVO	OTH	PU	M12	VLTN
D2.3	Empirical evidence sensemaking	2	VLTN	R	PU	M18	NORCE
D3.1	Conceptual framework for behavioural change understanding	3	UNIBO	R	PU	M12	CSCP
D3.2	0FLW interventions	3	UNIBO	R	PU	M14	PCCI
D3.3	Case-independent changing social norms predictive model	3	NORCE	OTH	PU	M18	ITC
D3.4	0FLW impact scenarios	3	UNIBO	R	PU	M24	UCPH
D3.5	FLW rapid appraisal/visualizer tool	3	UNIBO	R	PU	M28	ITC
D4.1	Actor specific guidance	4	CSCP	R	PU	M28	UCPH
D4.2	Social norms-focused communication products	4	UCPH	OTH	PU	M35	CTIC-CITA
D4.3	Social norms-focused science education package	4	UCPH	OTH	PU	M35	UNIBO
D4.4	Capacity building and Help desk	4	CSCP	OTH	PU	M35	UCPH
D5.1	Innovation Management & Exploitation	5	INLEIN	R	SEN	M36	VLTN
D5.2	CHORIZO Datahub Services Layer Upscaling Strategy	5	INLEIN	R	SEN	M36	EVILVO
D5.3	Innovation Impact Synergies	5	INLEIN	R	PU	M36	ITC
D5.4	Sustainability Strategy: EU FW Associations	5	FIAB	R	PU	M36	INLEIN
D6.1	CHORIZO website	6	FIAB	DEC	PU	M3	CTIC-CITA
D6.2	Dissemination, Exploitation and Communication Plan	6	FIAB	R	SEN	M6	INLEIN
D6.3	Dissemination, Exploitation and Communication Plan – update 1	6	FIAB	R	SEN	M18	HFBA
D6.4	Dissemination, Exploitation and Communication Plan – update 2	6	FIAB	R	SEN	M36	NCH
D6.5	Practice abstracts – batch 1	6	FIAB	R	PU	M18	ICLEI
D6.6	Practice abstracts – batch 2	6	FIAB	R	PU	M36	HFBA
D6.7	City Interest Group's' reporting	6	ICLEI	R	SEN	M36	CSCP
D6.8	Policy brief	6	EVILVO	R	PU	M36	ICLEI
D7.1	Project Handbook	7	EVILVO	R	SEN	M3	ICLEI

Deliverable (number)	Deliverable name	WP	Short name of lead	Type	Diss. level	Delivery date	Quality reviewer
D7.2	Data Management Plan	7	EVILVO	DMP	PU	M6	NORCE
D7.3	Data Management Plan – update 1	7	EVILVO	DMP	PU	M18	FIAB
D7.4	Data Management Plan – update 2	7	EVILVO	DMP	PU	M36	PCCI
D7.5	Quality & Risk Assessment Report	7	EVILVO	R	SEN	M12	VLTN
D7.6	Quality & Risk Assessment Report – update 1	7	EVILVO	R	SEN	M24	NCH
D7.7	CHORIZO's impact assessment	7	INLEIN	R	PU	M36	UNIBO
D8.1	OEI - Requirement No. 1	8	EVILVO	ETHICS	SEN	M3	FIAB
D8.2	OEI - Requirement No. 2	8	EVILVO	ETHICS	SEN	M18	HFBA
D8.3	OEI - Requirement No. 3	8	EVILVO	ETHICS	SEN	M36	CTIC-CITA

Table 4 Overview of the deliverables according the CHORIZO Grant Agreement commitments

In addition to the deliverables listed above, general activities of the CHORIZO consortium generates:

- Press articles in partners' media or other media related to food and agriculture, internet posts through social media
- Photos from events, demonstration activities, meetings
- Videos on YouTube channel
- Tabular, qualitative and geospatial data
- An email list, used for output dissemination, which will be confidential
- PowerPoint presentations or posters from events/fairs/conferences
- Scientific publications

Social media channels:

- X: https://twitter.com/CHORIZOproject
- LinkedIn: https://www.linkedin.com/company/chorizoproject/about/?viewAsMember=true
- YouTube: https://www.youtube.com/channel/UCkWWnjjsIzWZyYCEg5wng4A
- Facebook: Federación Española de Industrias de Alimentación y Bebidas | Madrid | Facebook
  FIAB has all the social networks and publishes information about the project on its own Facebook page.

This list of data produced is reviewed and updated during the DMT meeting on 16th January 2024 to ensure all datasets are included in the data management plan.

Examples of data in CHORIZO include survey results, interview recordings and videos obtained with the involved parties' consent.

The format of the recordings is: mp3, mp4.

The format of the transcription is: text, image, report format with summarised findings (e.g. interactive activity flipcharts/photographs with permission), .txt, .docx.

The format of the dataset: CSV and/or Word document with basic results/descriptive analysis, .txt, .docx.

3.3 Origin/provenance of the data

Related to the **data origin**, D1.1 Data Protocol details the data collected within each WP.

Primary data on case studies has been collected by the consortium members. Both the D1.1 Data protocol and D2.1 Case study Strategic Plans elaborates per case study on the data that have been collected.

Secondary data are and will be collected from multiple sources, including official databases, surveys, reports, scientific studies and other European projects and for multiple levels: European, national, regional and local.

The data origin is and will be reported in each deliverable. The use of existing data or the generation of new data is and will be clearly identified in each deliverable produced. The ethical clearance of all data, both new and existing, is and will be addressed by each work package leader and will be clearly reported upon.

The gathered data is and will be analysed or processed resulting into a set of deliverables and used in communication and dissemination activities.

All CHORIZO datasets have a dataset PID, assigned by the data repository.

3.4   Data utility

Data is used by three main groups: the consortium partners, the scientific community and the end-users. This last group is a very diverse group: Food businesses, Food processing, Retail, Hospitality, HoReCa, Public meals, Take-away and ready-to-eat, Food surplus redistribution logistics actors, Food banks, Policy makers (EU, National and Regional governments and local authorities), Educational institutions, Citizens/consumers/households/children, Schools.

At different levels, the data users during the course of the project are primarily the 14 consortium partners, who have specified their expected data needs.

All information aimed at the scientific community (project reports, deliverables, publications) is centralised on the CHORIZO website http://chorizoproject.eu.

In addition, the results and insights for end-users are made publicly available on the website of CHORIZO. Given the fact that each consortium partner has a network, a non-determined number of institutions, enterprises and organisations interested in European food system issues, these may also be potential users of the data generated and collected by the project.

Furthermore, the CDE plan (D6.1) details how to reach these end-users.

At last, we will collaborate with the sister project ToNoWaste and JRC to discuss possible joint communications for similar end-users. There was a meeting on Wednesday 14th December 2022 and Wednesday 25th January 2023 between CHORIZO and ToNoWaste.

The collaboration plan:
- Main conclusion
  - both projects have very different focusses and timelines
  - This makes the 2 projects very complementary
  - However, it hinders certain types of collaboration
- Collaboration actions
  - keep each other informed about key deliverables and events we organise
  - based on this, we will get more knowledge about each other's results and hence be able to identify possibilities for collaboration
  - Join certain events organised by sister project (e.g., webinars, workshops, etc.)
  - Maybe at some point organise a joint dissemination (e.g. joint webinar)

All used presentations and meeting minutes can be found on Teams.

3.5   Re-use of data

Data re-use was present, particularly in the collection and analyses of previous/ongoing FLW prevention/reduction actions and tools (WP1), the storage of the results on results of the evidence-based analysis of previous/ongoing FLW action (WP2) and the use of HUMAT and MOA approaches in the modelling work (WP3). Previous experiences and research served as data sources and were collected through (i) the involved partners; (ii) a screening of existing projects and databases, (iii) collaboration with JRC. Should we re-use data of our sister project TONOWASTE, this will only be anonymised data.

The access to sensitive raw data collected from previous actions, interviews and similar primary data generated within the project, is limited to those members of the CHORIZO team that work directly with these data. This is ensured by storing the sensitive raw data in separate restricted access folders on the MS Teams/SharePoint-platform.

The non-confidential data generated within the CHORIZO project is available for re-use by the consortium partners after the project ends and will be stored on the project SharePoint hosted by EVILVO. The duration of kept data will be discussed in the upcoming Project Steering Committee. The publicly available data will be stored on the CHORIZO website (https://chorizoproject.eu) for at least 2 years after the project ends.

3.6 Implementation of the FAIR principles per WP

WP + PARTNER	Expected size of data	Provide information	Reference to other data	Re-use of data	Purpose of data
WP1 VLTN	FLW actions & tools (>300 initiatives, classified into an FLW actions taxonomy) linked to the achievement of specific goals & context, and including an impact assessment & sector guidance	The Evidence Search Plan/Methodology (Word) and the Standardized Reporting Template (Excel) for Task 1.2 provides information on how and what data was obtained during research and interviews, and the internet or journal sources for already publicly available datasets on food loss and food waste. The subsequent and remaining two tasks in WP1 rely solely on the data provided in T1.2	Yes	Collection and analyses of previous/ongoing FLW prevention/reduction actions and tools	Collection and re-use of existing data on previous/ongoing FLW prevention/reduction actions and tools
WP2 EVILVO	6 case studies, sample sizes are detailed within the D1.1 Data Protocol and D2.1 Case Study Strategic plans			The storage of results on results of evidence-based analysis of previous/ongoing FLW action	Collection of data on 6 case studies (further details in D1.1 Data Protocol and D2.1 Case study Strategic Plans)
WP3 UNIBO	Modelling and social simulation: two models are developed	Metadata, methodology report, codes, file listing libraries/packages used for running the simulation	Yes	The use of HUMAT and MOA approaches in the modelling work	Existing knowledge on HUMAT and MOA approaches, empirical data collection through social simulation
WP4 UCPH	Product and service for Implementation of the behaviour change	Data is stored in open registry which defiantly facilitate data re-use.	Yes		Analytics from WP1 (D1.2 & D1.4), WP2 (D2.3), and WP4 (D4.1) including data from CS4

WP + PARTNER	Expected size of data	Provide information	Reference to other data	Re-use of data	Purpose of data
WP5 INLEIN	Relatively in small size	Original data is also made available for reference N/A	No		Gather data relating to business and upscaling strategies for exploiting project results that are innovative. Business plans are also created in order to upscale the services layer of the CHORIZO datahub. Depending on the owner of the results, the data may be sensitive. As sensitive we consider any 'business data' that could expose 'business sensitive' information. In this case we ask the solution-owners, if they have any sensitive data that require special treatment. Additionally, personal information/data of external consultants (food chain actors, key sector representatives, external advisors, etc.) engaged in the innovation and upscaling process are collected and safeguarded (when required) according to the project's GDPR (data protection) procedures.

WP + PARTNER	Expected size of data	Provide information	Reference to other data	Re-use of data	Purpose of data
					Lastly, we ensure that our 'findings' are easily findable (as per FAIR principles).
WP6 FIAB	Relatively in small size	Templates: Excel files, Word files and PowerPoint files. Logos which are internally shared within the consortium. Dissemination Material (Roll-up, leaflet, poster, Practice Abstracts). Public information is stored on the website and in the newsletters.	Yes	N/A	The purpose is to provide templates to the partners in the consortium to professionalize the working methods e.g. templates for deliverables, presentations, … Some documents also have the purpose to collect info from the partners e.g. dissemination, exploitation and communication. Other materials for communication or dissemination.
WP7 EVILVO	Relatively in small size	Some Excel files, Word files, templates, which are internally shared within the consortium.	No	N/A	The documents are only used for internal project management e.g. contact list, meeting minutes, … and for communication with external stakeholders.

Table 5 Implementation of the FAIR principles per WP

4 COSTS, SECURITY AND ETHICS

4.1 Costs for making data and other research outputs FAIR

Costs for data storage is to be kept at a minimum using free services and tools needed in the project, like Zenodo and project MS Teams site. Some costs may still incur.

Costs related to open access to research data in Horizon 2020 are eligible for reimbursement under the conditions defined in the CHORIZO Grant Agreement, in particular Article 6 ("Eligible Cost") and Article 6.2. ("D.2 Costs of other goods and services"), but also other articles relevant for the cost category chosen.

Project beneficiaries are responsible for applying for reimbursement for costs related to making data accessible to others beyond the Consortium.

To be eligible, costs must be:

a) purchased specifically for the action and in accordance with Article 10.1.1 (best value for money)

or

b) contributed in-kind against payment and in accordance with Article 11.1 (rules for in-kind contributions for against payment).

The current foreseen costs for making data FAIR includes:

- Fees associated with the publication of scientific articles containing project's research data in "Gold" Open access journals. The cost sharing, in case of multiple authors, shall be decided among the authors on a case-by-case basis.
- Project website development and operation: FIAB budget. The requirement is to maintain the website until the completion of the project.
- Data archiving at Zenodo and on other online data base: free of charge.
- Copyright licensing with Creative Commons: free of charge.
- CHORIZO FLW "Insighter" and Datahub development and operation: ILVO Budget. Requirement is to maintain the hub until the completion of the project.

Other costs may incur during the project's lifetime and will be evaluated as eligible or not as specified in the Grant Agreement. Each partner is responsible for the data they produce. Any fee incurred for Open Access through scientific publication of the data is the responsibility of the data owner (authors) partner(s). The default long term preservation will be through Zenodo. Zenodo states that: "Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least." Since uploading to Zenodo is free of charge this is expected to keep overall retention costs small.

4.2 Data security

Raw data will not be shared because it is subject to privacy laws. Pseudonymised data be shared with all members of CHORIZO utilizing security measures, but will not be made public. However, in accordance with Horizon Europe policy (anonymised) data is made available when anonymisation is possible. Anonymisation refers to removing any identifier that can reveal the identity of the participants both from data and metadata. See Annex 1 for guidelines.

At the level of the CHORIZO project partners

The **obtained personal data is processed on protected and backed-up network disks of the partner organisations** themselves, in folders to which only the researchers involved in the CHORIZO project have access.

After conducting the interview, **personal data is, as soon as possible, anonymised** (e.g. anonymised written report where there is no longer any link to the participant)/**pseudonymised** (e.g. transcription of interview where directly identifiable elements have been replaced by a code).

When storing/processing data, **identification and analysis data is stored separately** with a common code. In this way, access to the identification data can be strictly limited and monitored. Only the person, who has the key to the code, can retrieve the person behind the code number. **The key is kept on protected and backed-up network drives of institutions** themselves in folders to which only the relevant researchers of the CHORIZO project have access.

The **analyses are conducted only on the basis of pseudonymised data**. The pseudonymised data are stored on protected and **backed-up network drives of the institutions themselves and in encrypted folders on CHORIZO MS TEAMS** (where it should be checked whether everyone needs access to pseudonymised data).

Outside the CHORIZO project

If necessary, data transfer to and from end-users (including transfer of sensitive data if allowed) is performed encrypted, either sent by encrypted ZIP or RAR files, or by download directly as web-based services from servers. In any case strong password is required for accessing transferred datasets and passwords must be sent separately from the dataset (preferably using also different channels of communication e.g. SMS, Viber, WhatsApp).

4.3   Ethics

Impact on data sharing

The proposed work in CHORIZO does fully comply with the regulations set out in Article 14 of the Grant Agreement, which states that all activities must be carried out in compliance with:

- Ethical principles (including the highest standards of research integrity, e.g. the principles of the European Charter for Researchers and The European Code of Conduct for Research Integrity), and
- Applicable international, EU, and national laws (in particular the GDPR).

To fully comply, CHORIZO appointed an external ethics advisor who does advise us on a regular basis throughout the project. The specific issues that he provides advice on are listed below.

o   Humans:

The project implements qualitative-quantitative behavioural studies with stakeholders by involving the stakeholders in interviews, surveys, case studies and experiments. He advised use on the procedures and criteria that were used to identify/recruit research participants, but also the specific target groups.

o   Personal Data:

The project collects various data to analyse behaviour of the stakeholders. The undertaken activities may involve profiling of people. Further, data to be collected may involve sensitive categories and even be used for such purposes, e.g. social norms and pressure.

o   Artificial Intelligence:

The CHORIZO 'Insighter' datahub involves the rules-based data sharing of empirical evidence between data providers and consumers, and the provision of services to corporate actors pursuing FLW prevention initiatives.

o   Non-EU:

The activities performed in Norway, as well as the exchange of data between EU and non-EU countries will be duly considered. The ethics advisor provides guidance on the sound and correct treatment of such data transfers.

Next, the ethics advisor has/will complete(d) two reports, defined as deliverables 8.2 and 8.3, where he (will) describe(s) the procedures that are/were implemented to solve any remaining ethical issues.

Informed consent for data sharing and long term preservation

Both online and live interviews and questionnaires are conducted in accordance with the EU Law (no. 97/2008, 104/2009, 68/2012 and 107/2012). The prior information was provided to the interviewees in accordance with the Article 15 and their consent (i.e. authorisation to collect, process, use data, preserve on a long term and share) was asked for: in writing in case of oral interviews, and by clicking an "I Agree" button at the bottom of the page in case of online questionnaires, which contains all the information included in the informed consent form and the information sheets. All of this is addressed in D8.2.

5   CO-AUTHORSHIP IN CHORIZO

In the project, we are moving from data collection to data analysis to writing scientific papers about it. We apply Open Access concerning sharing data but also concerning insights and results.

The agreed criteria for (co-)authorship

- A researcher will **claim** a legitimate **co-authorship** if he/she has made an integrating, overarching and substantial contribution to <u>**at least two**</u> <u>of the following four research activities</u>:

  1. participated substantially in the **conception and design** of the work

  2. participated substantially in the **collection of the data**

  3. participated substantially in the **analysis of the data**

  4. participated substantially in the **writing of the manuscript**

- Additionally, each (co-)author at least:

  1. reviewed and approved the final version of the manuscript for submission

  2. takes public responsibility for its content

  3. declared that neither this manuscript nor one with substantially similar content under our authorship has been published or is being considered for publication elsewhere, except as described in an attachment

The agreed approach for co-authorship

1. Before using any CHORIZO data in publications, **the initiative taker/lead author** always **contacts all persons/partners who were involved in at least one of the four research activities and provides them the opportunity to become a co-author**

   ◦ If the person who was involved in at least one of the four research activities is not available anymore, the respective beneficiary is contacted

2. All interested parties meet to agree on:

   ◦ the research goal, scope, approach, and the prospected end result

   ◦ the **process** of performing all the necessary research activities and writing the paper,

   ◦ the **allocation of tasks and responsibilities** among the partners

3. **Being offered the opportunity and accepting it** versus **taking up responsibility afterwards** are 2 different things

   ◦ If a partner does not fulfill the agreed responsibilities, the lead author can hold that person accountable for this

◦ the lead author has the right to remove an author name from the list of authors in case the designated responsibilities as agreed in step 2 are not met

Communication = key in order to prevent conflicts.

First author and author ranking

- The **first author** is always the person who has made the most important integrating contribution.

- All other authors are ranked **in accordance with the general rules** or customs of the **specific discipline**.

- Authors may be ranked in <u>order of decreasing importance of their integrating contribution, or a special position may be assigned to the 2nd author (direct supervisor) and last author (research leader),</u> unless the editorial board of a scientific journal has determined other rules for author ranking explicitly.

- Keep in mind the tasks and competences of the author in order to balance the importance of their "substantial individual contributions".

- The **author who submits a manuscript** for publication accepts the responsibility of having included as co-authors all persons appropriate and none inappropriate. The submitting author will sent each co-author a draft copy of the manuscript and obtains the co-author's assent to co-authorship of it.

- The **co-authors of a paper** will be all those persons who have made significant scientific contributions to the work reported and who share responsibility and accountability for the results. **Other contributions will be indicated in a footnote or an "acknowledgments" section.**

Annex 1

Info sheet anonymising or pseudonymising

Why anonymise or pseudonymise?

To protect personal data (privacy-sensitive information), you are required by the GDPR (General Data Protection Regulation) to implement various organisational and technical security measures.

According to GDPR, "personal data" means:
"any information relating to an identified or identifiable natural person ("the data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more elements characterising the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

"pseudonymisation" means according to GDPR:
"processing personal data in such a way that the personal data can no longer be linked to a specific data subject without the use of additional data, provided that such additional data are kept separately and technical and organisational measures are taken to ensure that the personal data are not linked to an identified or identifiable natural person;

Here, **anonymisation or pseudonymisation** is seen as an important security measure to reduce the risk of the data subjects (the individuals to whom the data relates).

For example: to avoid losing valuable information during an interview, interviews or focus group discussions, are often recorded. However, audio files are highly privacy-sensitive information and, in the context of data minimisation, should be made 'safe' as soon as possible by anonymising or pseudonymising. After all, you should not process more data than is strictly necessary for your intended purpose. The ILVO guideline for retention of audio/video recordings is 2 to 3 months unless otherwise specified by the client (and then this must be communicated in the informed consent). Data minimisation ensures that the privacy risk of the data subject (in this case, interviewees/participants of focus group discussions) is reduced: after all, what you do not own, you cannot lose or end up in the wrong hands (unauthorised persons, hackers, etc.).

What is anonymisation or pseudonymisation?

Anonymisation is not the same as pseudonymisation (see table 1).

Table 1: Anonymisation is not the same as pseudonymisation

Identified personal data	Pseudonymised personal data	Anonymous data
E.g. Audio recording	E.g. Pseudonymised transcript	E.g. Anonymised written report
Linkable without additional data to data subject	Linkable with additional data to data subject	Unlinkable to data subject
Data subject* is identifiable	Person involved is re-identifiable	Data subject is no longer (re-)identifiable
Personal data	Personal data	No personal data
GDPR applicable	GDPR applicable	GDPR not applicable

Anonymise →

Pseudonymisation →

Re-identify ←

* The individual person to whom the data relates

With **anonymisation**, you process your raw data until all options for (re)identification are irreversibly removed. There is no longer any link to the data subject. The data subject can no longer be (re)identified.

This is why both direct identifiers and indirect identifiers must be removed or replaced.

- Direct identifiers: data that allow you to directly identify the data subject. For example: name, address, telephone number, e-mail address, IP address, recognisable image of a data subject, voice of a data subject, etc.
- Indirect identifiers: based on a combination of data it is possible to (re)identify the data subject. For example: the combination of farm type, municipality business seat and age.

The GDPR legislation no longer applies.

For example: anonymised written report of focus group discussion.

With **pseudonymisation**, you process your raw data until you can no longer link the pseudonymised data (= analysis data) to the data subject without using the data subject's additional data.

You store the link between the identity of the data subject and the pseudonym (= identifying data) in a separate file: the key file. Access to the key file is highly restricted and only accessible for a limited number of people.

Analyses are done only on the basis of the pseudonymised data.

The GDPR legislation does still apply because the pseudonymised data is linkable to the data subject on the basis of the key file.

Example: transcription of interview where directly identifiable elements have been replaced by a pseudonym (=code). List of interviewees' names and codes used in the pseudonymised transcript are kept in highly restricted key file.

Storage of the data

1 Process the **obtained personal data on protected and backed-up network disks of the partner organisations** themselves, in folders to which only the researchers involved in the CHORIZO project have access.
2 After conducting the interview, **personal data should be anonymised as soon as possible** (e.g. anonymised written report where there is no longer any link to the participant)/**pseudonymised** (e.g. transcription of interview where directly identifiable elements have been replaced by a code)
3 When storing/processing data, **identification and analysis data should be stored separately** with a common code. In this way, access to the identification data can be strictly limited and monitored. Only the person, who has the key to the code, can retrieve the person behind the code number. **The key is kept on protected and backed-up network drives of institutions** themselves in folders to which only the relevant researchers of the CHORIZO project have access.
4 The **analyses are conducted only on the basis of pseudonymised data**. The pseudonymised data are stored on protected and **backed-up network drives of the institutions themselves and in encrypted folders on CHORIZO MS TEAMS** (where it should be checked whether everyone needs access to pseudonymised data).

When to anonymise or pseudonymise?

When anonymisation is not possible or not desirable in function of your intended purpose of your scientific research, pseudonymisation is brought forward from the GDPR to protect your personal data.

For example: in the context of follow-up research, you want to be able to ask questions of the interviewee later on.

Sometimes, in the context of research, it is important that you do not anonymise or pseudonymise the recording. For example: consider a practice video or a video testimonial to promote and disseminate best farming practices and innovations.

In each case, however, it is important that you inform the data subjects about this in the informed consent and that the data subjects have given their explicit consent to this.

How to anonymise or pseudonymise?

Dataset with only direct identifiers

Anonymise (See Table 2, Table 5 and Table 6)

- **Strip data of direct identifiers** (name, address, phone number, email address, etc). (See Table 5)
  - o Direct identifiers are removed (e.g. omit columns in Excel).

  or
  - o Direct identifiers are masked: completely or partially overwrite with (*/x)

  or
  - o A random code is given to each data subject. E.g. Farmer51, Farmer07, Farmer04, ...

  or

- **Aggregate data** (e.g. average, sum, ...). Never figures from less than 5 farms. (See Table 6)

- Raw dataset to be deleted.

Pseudonymisation (See Table 2, Table 3 and Table 4)

- Direct identifiers (name, address, phone number, email address, ...) from raw dataset are replaced by a **pseudonym** using:
  - o Sequential counter: a sequential number (and possibly a prefix) is used as a pseudonym. E.g. farmer01, farmer02, farmer03, ... (See Table 4)

  or
  - o Random Number Generator: a random number (and possibly a prefix) is used as a pseudonym.
- Create **separate key file** containing a link between analysis data and personal data based on pseudonym. (See Table 3)
- The raw dataset is preferably deleted (or has severely restricted access). The researcher preferably performs analyses on the pseudonymised data (= analysis data). The analysis data are restricted to authorised researchers who need the data to conduct the study for which the data were obtained. The key file has highly restricted access (= identifying data).
- At the end of the study - e.g. if pseudonymised data should not be kept as part of follow-up research - it is easy to anonymise the pseudonymised data. Change the pseudonym by a random code and permanently delete the raw dataset - if not already done - and the key file. (See Table 5).

Table 2: Raw data

Names farmers	Email address	Housing type Eastern bunnies	Average number of Easter eggs per Easter bunny per year
Noah Peeters	n.peeters@telehoo.com	Enriched cage	461
Arthur Janssens	ajanssens@yanet.com	Outdoor system	409
Louis Maes	L2Maes@ghoo.com	Outdoor system	380
Olivia Jacobs	Oli.jacobes@yamail.com	Organic system	355
Emma Willems	emmaW@gtele.com	Organic system	340
Louise Dubois	Ldubois@telemail.com	Enriched cage	392
…	…	…	

Table 3: Key file (= identification details)

Pseudonym	Names farmers	Email address
Farmer01	Noah Peeters	n.peeters@telehoo.com
Farmer02	Arthur Janssens	ajanssens@yanet.com
Farmer03	Louis Maes	L2Maes@ghoo.com
Farmer04	Olivia Jacobs	Oli.jacobes@yamail.com
Farmer05	Emma Willems	emmaW@gtele.com
Farmer06	Louise Dubois	Ldubois@telemail.com
…	…	…

Table 4: Pseudonymised data (=analysis data). Based on pseudonym and key file, the farmer can be re-identified

Pseudonym	Housing type Eastern bunnies	Average number of Easter eggs per Easter bunny per year
Farmer01	Enriched cage	461
Farmer02	Outdoor system	409
Farmer03	Outdoor system	380
Farmer04	Organic system	355
Farmer05	Organic system	340
Farmer06	Enriched cage	392
…	…	

Table 5: Anonymised data by removing all direct identifiers and giving a random code to all data subjects

Random code	Housing type Eastern bunnies	Average number of Easter eggs per Easter bunny per year
Farmer51	Enriched cage	461
Farmer07	Outdoor system	409
Farmer04	Outdoor system	380
Farmer67	Organic system	355
Farmer78	Organic system	340
Farmer94	Enriched cage	392
…	…	

Table 6: Anonymised data through aggregation

Housing type Eastern bunnies	Number of specialised Easter bunny companies	Average number of Easter eggs per Easter bunny per year
Enriched cage	148	452
Outdoor system	166	390
Organic system	74	332

Dataset with direct and indirect identifiers

1. Anonymise (See Table 7, Table 10, Table 11, Table 15, Table 16 and Table 17)

- **Removing identifiers** from data
  - Identifiers are **removed** (e.g. omit columns in Excel).

  or
  - Identifiers are **masked**: overwrite fully or partially with '*' or 'X'.

  or
  - A **random code** is given to each person involved. (E.g. Farmer51, Farmer07, Farmer04, ...)

Or

- **Generalising** data: here, the data become less precise, specific to the intended purpose of scientific research. If there is too much loss of information in function of intended purpose, then anonymising is not the appropriate security measure and you should pseudonymise the dataset. For example:
  - Replacing a year of birth with an age or age range
  - Replace a location (e.g. municipality business seat) with a less precise location (e.g. province, agricultural region, ... business seat)
  - Replacing a specific farm type (e.g. specialised greenhouse vegetable farms) with a less specific farm type (e.g. specialised horticultural farm)

Or

- Setting **upper and lower limits:** group values of data above or below certain limits to avoid that data subjects can be identified based on outliers. For example
  - Age range: '65 years and above' and the age range: 'less than 35 years'
  - Class of business size of '<25,000 euros' and business size of '>=500,000 euros'

Or

- **Data perturbation:** rounding, adding noise, replacing real values with simulation values or group averages rounding numeric values and dates. The degree of value change determines anonymisation. When data accuracy is essential, data perturbation should not be applied.

Or

- **Aggregate** data (average, sum, ...). Never figures from less than 5 companies.

The raw dataset are removed.

To verify that qualitative data has been properly anonymised, you can e.g. read the anonymous report from the data subject's point of view or have the anonymous report read by the data subject. If there are no more links to the data subject, the GDPR no longer applies.

2. Pseudonymisation

- **Direct identifiers** (name, address, phone number, email address, ...) from raw dataset **are replaced by a pseudonym**.

- Create a separate key file containing a link between analysis data and personal data based on the pseudonym.

- As a function of intended purpose of scientific research
    - **Separate indirect identifiers from analysis data** and add to the key file (See Table 7, Table 8 and Table 9)

    Or
    - **Do not separate indirect identifiers from analysis data** because they are of necessary importance for the analysis of the dataset and in consideration of the intended purpose: (See Table 12, Table 13 and Table 14)
        - **Generalise**: in this process, the data become less precise, specific. (Possibly not desired in the context of research.) For example:
            - ✓ Replacing a date of birth with an age or age range
            - ✓ Replacing an address of place of business with a less specific location: e.g. municipality or province or agricultural region, ... in which the place of business is located
            - ✓ Replacing a specific farm type (e.g. specialised greenhouse vegetable farms) with a less specific farm type (e.g. specialised horticultural farm)

        Or

        - **Setting upper and lower limits**: group values of data above or below certain limits to avoid that data subjects can be identified based on outliers. (May not be desirable in the context of research.) For example:
            - ✓ Age class: '65 years and above' and the age class: 'less than 35 years'
            - ✓ Class business size of '<25,000 euros' and business size of '>=500,000 euros'

        - **Data perturbation:** rounding, adding noise, replacing real values with simulation values or group averaging rounding numeric values and dates. When data accuracy is essential, perturbation should not be applied.

- When pseudonymising/anonymising transcripts, the above techniques can also be used. Substitutions can be indicated with square brackets e.g. [Farmer01] has a specialised Easter bunny farm in [Limburg] and says that keeping Easter bunnies in free range is better for animal welfare. [Farmer02] has a specialised Easter bunny farm in [West Flanders] and says that keeping Easter bunnies in enriched cage is less labour intensive.

- The raw dataset is preferably deleted (or has severely restricted access). The researcher preferably performs analyses on the pseudonymised data (= analysis data). The analysis data are restricted to authorised researchers who need the data to conduct the study for which the data were obtained. The key file has highly restricted access (= identifying data).

- At the end of the study - e.g. if pseudonymised data should not be kept in the context of follow-up research - you can consider - if possible/if desirable - anonymising the pseudonymised data. From the point of view of the data subject, consider whether the data subject is re-identifiable from the pseudonymised data.

- o If re-identifiable, the dataset is not sufficiently pseudonymised. GDPR legislation continues to apply. Data must be deleted at the end of the study, unless the data subject has given permission for the personal data to be retained for possible follow-up research.
- o If not re-identifiable, then the dataset is sufficiently pseudonymised. Change the pseudonym by a random code. Permanently delete the raw dataset - if not already done - and the key file. The GDPR legislation no longer applies to anonymous data. The anonymised data may be retained for follow-up research but can no longer be linked to specific individuals/firms. Longitudinal research at company level is therefore not possible with anonymised data.

Table 7: Raw data

Names farmers	Municipal business seat	Email address	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Noah Peeters	Peer	n.peeters@telehoo.com	Enriched cage	461
Arthur Janssens	Pelt	ajanssens@yanet.com	Outdoor systems	409
Louis Maes	Pittem	L2Maes@ghoo.com	Outdoor systems	380
Olivia Jacobs	Poperinge	Oli.jacobes@yamail.com	Organic systems	355
Emma Willems	Putte	emmaW@gtele.com	Organic systems	340
Louise Dubois	Puurs-Sint-Amands	Ldubois@telemail.com	Enriched cage	392
…	…	…	…	…

Table 8: key file (= identification data including location data)

Pseudonym	Names farmers	Municipal business seat	Email address
Farmer01	Noah Peeters	Peer	n.peeters@telehoo.com
Farmer02	Arthur Janssens	Pelt	ajanssens@yanet.com
Farmer03	Louis Maes	Pittem	L2Maes@ghoo.com
Farmer04	Olivia Jacobs	Poperinge	Oli.jacobes@yamail.com
Farmer05	Emma Willems	Putte	emmaW@gtele.com
Farmer06	Louise Dubois	Puurs-Sint-Amands	Ldubois@telemail.com
…	…	…	…

Table 9: Pseudonymised data (= analysis data). Based on pseudonym and key file, the farmer can be re-identified

Pseudonym	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Farmer01	Enriched cage	461
Farmer02	Outdoor systems	409
Farmer03	Outdoor systems	380
Farmer04	Organic systems	355
Farmer05	Organic systems	340
Farmer06	Enriched cage	392
…	…	

Table 10: Anonymised data by removing all direct and indirect identifiers and giving a random code instead of a sequential code to all data subjects

Random code	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Farmer51	Enriched cage	461
Farmer07	Outdoor systems	409
Farmer04	Outdoor systems	380
Farmer67	Organic systems	355
Farmer78	Organic systems	340
Farmer94	Enriched cage	392
…	…	

Table 11: Anonymised data through aggregation

Type housing Easter bunnies	Number of specialised Easter bunny companies	Average number of Easter eggs per Easter bunny per year
Enriched cage	148	452
Outdoor systems	166	390
Organic systems	74	332

Table 12: Raw data

Names farmers	Municipal business seat	Email address	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Noah Peeters	Peer	n.peeters@telehoo.com	Enriched cage	461
Arthur Janssens	Pelt	ajanssens@yanet.com	Outdoor systems	409
Louis Maes	Pittem	L2Maes@ghoo.com	Outdoor systems	380
Olivia Jacobs	Poperinge	Oli.jacobes@yamail.com	Organic systems	355
Emma Willems	Putte	emmaW@gtele.com	Organic systems	340
Louise Dubois	Puurs-Sint-Amands	Ldubois@telemail.com	Enriched cage	392
…	…	…	…	…

Table 13: Key file (= identification data including location data)

Pseudonym	Names farmers	Email address
Farmer01	Noah Peeters	n.peeters@telehoo.com
Farmer02	Arthur Janssens	ajanssens@yanet.com
Farmer03	Louis Maes	L2Maes@ghoo.com
Farmer04	Olivia Jacobs	Oli.jacobes@yamail.com
Farmer05	Emma Willems	emmaW@gtele.com
Farmer06	Louise Dubois	Ldubois@telemail.com
…	…	…

Table 14: Pseudonymised data (= analysis data). Based on pseudonym and key file and the use of a less specific location the farmer can not be re-identified, unless the key file is used.

Pseudonym	Province	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Farmer01	Limburg	Enriched cage	461
Farmer02	Limburg	Outdoor systems	409
Farmer03	West-Vlaanderen	Outdoor systems	380
Farmer04	West-Vlaanderen	Organic systems	355
Farmer05	Antwerpen	Organic systems	340
Farmer06	Antwerpen	Enriched cage	392
…	…	…	

Table 15: Anonymised data by removing all direct and indirect identifiers and giving a random code to all data subjects

Random code	Province	Type housing Easter bunnies	Average number of Easter eggs per Easter bunny per year
Farmer51	Limburg	Enriched cage	461
Farmer07	Limburg	Outdoor systems	409
Farmer04	West-Vlaanderen	Outdoor systems	380
Farmer67	West-Vlaanderen	Organic systems	355
Farmer78	Antwerpen	Organic systems	340
Farmer94	Antwerpen	Enriched cage	392
…	…	…	

Table 16: Anonymised data through aggregation (total number companies per province per type household is at least 5)

Type housing Easter bunnies	Number of specialised Easter bunny companies	Antwerpen	Limburg	Oost-Vlaanderen	West-Vlaanderen	Vlaams-Brabant
Enriched cage	148	21	35	12	48	32
Outdoor systems	166	15	21	32	25	73
Organic systems	74	11	26	9	14	14

Table 17: Anonymised data through aggregation: average number Easter eggs per province per type housing (knowing that the total of the number of compagnies per province per type housing is at least 5)

Type housing Easter bunnies	Antwerpen	Limburg	Oost-Vlaanderen	West-Vlaanderen	Vlaams-Brabant
Enriched cage	432	460	480	470	418
Outdoor systems	330	400	382	411	427
Organic systems	312	349	350	309	340

Sources:

AVG: persoonsgegevens pseudonimiseren | (onder)zoektips (ugent.be)

Anonymisation & pseudonymisation - Research Data Management (kuleuven.be)

Anonimisatie Vs. Pseudonimisatie (smalsresearch.be)

Horizon
Europe