ALTERNATIVE Deliverable D8.2

Data Management Plan

Editor, Organisation:	EURECAT
Deliverable nature:	Report
Dissemination level: (Confidentiality)	Consortium (CO)
Contractual delivery date:	31/03/2022
Actual delivery date:	23/03/2022
Suggested readers:	Project partners
Version:	1.0
Total number of pages:	38
Keywords:	Data Management, FAIR, datasets, metadata, accessibility, licences, repository, GDPR, security,

Abstract

This deliverable represents the Data Management Plan for the ALTERNATIVE project. This document includes the prospects for the data that will be collected and created during the project, to ensure its correct implementation and addresses the project participation in the Pilot on Open Research Data for H2020 applicants. It includes the application of the FAIR principles and tackles aspects such as data security, allocation of resources and ethical aspects for data management.

[End of abstract]

Disclaimer

In case of Public (PU):

All ALTERNATIVE consortium parties have agreed to full publication of this document.

In case of Restricted to Programme (PP):

All ALTERNATIVE consortium parties have agreed to make this document available on request to other framework programme participants.

In case of Restricted to Group (RE):

All ALTERNATIVE consortium parties have agreed to full publication of this document. However, this document is written for being used by <organisation / other project / company etc.> as <a contribution to standardisation / material for consideration in product development etc.>.

In case of Consortium confidential (CO):

The information contained in this document is the proprietary confidential information of one or more ALTERNATIVE consortium parties and may not be disclosed except in accordance with the consortium agreement.

The research leading to these results has received funding from the European Union Horizon 2020 Programme under grant agreement number 101037090 – ALTERNATIVE.

Imprint

Full project title	Environmental Toxicology of chemical mixtures through an innovative platform based on aged cardiac tissue model
Short project title	ALTERNATIVE
Number and title of work-package	WP8. Impact Creation
Number and title of tasks	T8.2 – Data, Knowledge and IPR management
Document title	Data Management Plan
Editor: Name, organisation	Laura Armayones, EURECAT
Work-package leader: Name, organisation	Milon Gupta, EURESCOM
Internal reviewer(s): Name(s), organisation(s)	Susanna Sartori, POLITO Federico Vozzi, CNR
Estimation of PM spent on the Deliverable	1.5 PM

Executive summary

This deliverable DMP describes the methodology for data management planned to be employed in the framework of the ALTERNATIVE project. The described methodology aims to safeguard the sound management of the data collected and generated during the project's activities across their entire lifecycle, while also making them FAIR. Moreover, the DMP identifies the anticipated activities required for making data FAIR, outlines the provisions about their security and revolving around their collection/generation.

The DMP is considered to be a living document in the framework of ALTERNATIVE and will be updated as needed throughout the course of the project taking into account its latest developments and available results. The DMP will be reviewed annually, and any updates will be included in the periodic project management reports. Ad hoc updates may also be realised when deemed necessary, to deliver an accurate, up-to-date, and comprehensive DMP before the completion of the project. These updates will also be appended on the periodic project management reports. This deliverable is delivered in M6 of the project and is updated based on the latest information available up to the month of delivery.

List of authors

Company	Author	Contribution
EUT	Laura Armayones	Author, Editor
CNR	Federico Vozzi	Review
CNR	Nunzia Linzalone	Contributor
EUR	Anastasius Gavras	Final review and submission
EUR	Milon Gupta	Contributor
IRFMN	Alessandra Roncaglioni	Contributor
MUI	Martin Paparella	Contributor
POLITO	Gianluca Ciardelli	Final review
POLITO	Susanna Sartori	Review

Table of Contents

List of figures and tables

List of figures:

List of tables:

Abbreviations

ALTERNATIVE	environmentAL Toxicology of chEmical mixtuRes through aN innovATIVE platform based on aged cardiac tissue model (the project)
AOP	Adverse Outcome Pathways
COL	Collected
CRE	Created
DMP	Data Management Plan
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GDPR	General Data Protection Regulation
MIE	Modular Initiating Event
ML	Machine learning
PBPK	Physiologically-Based Pharmacokinetic (model)
QSAR	Quantitative Structure-Activity Relationship
TBD	To Be Discussed
TK/TD	Toxicokinetic/Toxicodynamic (TK/TD)
WP	Work Package

1 Introduction

In H2020, applicants will, by default, participate in the Pilot on Open Research Data, which aims to improve and maximise access and re-use of research data generated by actions. However, participation in the Pilot is flexible in the sense that it does not mean that all research data needs to be open.

The **Data Management Plan (DMP)** addresses the relevant aspects of making data FAIR – findable, accessible, interoperable and re-usable, including what data the project will generate, whether and how it will be accessible for verification and re-use, and how it will be cured and preserved. This deliverable presents the first version of the Data Management Plan (DMP) for the ALTERNATIVE project. Thus, the purpose of this Data Management Plan (DMP) is to organise and acknowledge the data that the project partners will use to carry out the implementation of the Alternative project.

As established by the Grant Agreement, the project's Data Management Plan (DMP) will detail the type of data generated by the ALTERNATIVE project, whether and how it will be exploited or made accessible for verification and re-use, and how it will be stored and preserved. In addition, the DMP aims to ensure that ALTERNATIVE activities are compliant with the H2020 Open Access policy and the recommendations of the Open Research Data pilot, such as the FAIR policy.

The application of the DMP by all ALTERNATIVE partners will be essential to ensure the project greatest impact. The responsible for task T8.2 – Data, Knowledge and IPR management will strongly interact with all the WPs to collect all the information from partners and ensure that the information included in the Data Management is clear and transparent.

This document is public, and it will serve as official guidance to Project Partners on data management.

The DMP is structured in 7 chapters:

- Chapter 1 provides introductory information about the DMP, the context in which it has been elaborated as well as about its objectives and structure.
- Chapter 2 presents a summary of the data to be collected or/and generated during ALTERNATIVE activities, including the purpose of its collection/generation and its types and formats. Additionally, it outlines its origin, expected volume and the stakeholders that may find it useful.
- Chapter 3 describes the methodology applied in the framework of ALTERNATIVE to safeguard the effective management of data across their entire lifecycle, making it FAIR.
- Chapter 4 estimates the resources required for making the project's data FAIR, while also identifying data management responsibilities.
- Chapter 5 outlines the data security strategy applied within the context of ALTERNATIVE along with the respective secure storage solutions employed.
- Chapter 6 addresses ethical aspects such as the management of personal data
- Chapter 7 concludes on the next steps foreseen in the framework of the project with respect to its data management plan.

Finally, this document has been produced following the EC guidelines and templates for project participating in the Open Research Data Pilot:[1]

- *H2020 Annotated Model Grant Agreement - Open access to research data*
- *Guidelines to rules on Open Access to Scientific Publications & Open Access to Research Data in Horizon 2020*
- *Guidelines on FAIR Data Management in Horizon 2020*
- *Template for the Data Management Plan*

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

2 Data Management Plan - Updates

The project's first development and future updates mainly rely on the collecting Data Management tables filled out for each project dataset by project partners responsible for producing data. The DMP is a living document. Every time there may be a modification in the Data, their impact, or their accessibility, the document will be modified accordingly, for constant update and further development all along with the project lifespan. A dedicated session will be set up during periodic General Assembly meetings annually at least.

Different versions will be identified by a version number and a date. The version number will be composed of two digits separated by a period: the digit before the period represents in ascending orders the official versions submitted to the European Commission as deliverable; digits after the period means the periodic internal revisions of such official versions.

For the time being, the official submitted versions will be stored in the Alternative Only Office repository. An editable word copy of the latest version will also be stored to facilitate revision and update of the already identified datasets and policies. The DMP datasets, on XLSX files, will be stored in the ALTERNATIVE repository to allow their modification by the WP leaders and consultation by the other Partners.

Should a new dataset be identified along with the project implementation, partners can easily contact the responsible EUT contact on data management, who will then be in charge of updating the document and informing of the changes to the Consortium.

The Project consortium will decide in the General Assembly further modification of the repository, the nature of the data to be stored and their accessibility, also bearing in mind the need to agree on data formats and procedures to integrate datasets in the ALTERNATIVE's platform.

3 Data Summary

3.1 General Overview of data according to WP

In this section, to give an overview of the purpose of the data generated and its relation to the WP objectives, individual tables for each WP have been created and were originally collected during the DM workshop (S*ee Data Management and FAIR principles workshops*). These tables also include the relationship between WPs creating data and WPs using the data as also, the foreseen data utility.

Table 1: WP1 objective, data utility and related WP

WP1 Ethics requirements
Objective: The objective is to ensure compliance with the 'ethics requirements' set out in this work package.
Data purpose: WP1 is focused on two main aspects: 1) the analysis of the ethical aspects of the proposed research concerning the use of cells applied in the development of ALTERNATIVE cardiac tissue replica, in particular hIPSC and endothelial cells; 2) the health and safety procedures conforming to relevant local/national/European guidelines/legislation, adopted by the whole ALTERNATIVE consortium, regarding occupational safety and health and waste handling. Thus, there is not data to be created or collected in this WP.
Involved partners: POLITO, CNR
Related WPs: WP4, WP5
Data utility: Researchers, Biologists, Chemists

Table 2: WP2 objective, data utility and related WP

WP2 Regulatory Uptake
Objective: Scientifically summarize in a concise and traceable way the regulatory needs for the development and validation of an ALTERNATIVE platform, structure knowledge on cardiotoxicity within the AOP framework, draft guidance for its potential regulatory use and prepare the necessary documentation for the envisaged follow up, i.e., the international regulatory validation process. To this aim the data will be collected based on descriptive data collected from literature, scientific reports, regulatory documents, discussions during workshops and outcome of interviews.
Data purpose: WP2 descriptive data will be collected from literature, scientific reports, regulatory documents, discussions during workshops and outcome of interviews. The purpose of the data will be mapping existing regulatory requirements for cardiotoxicity, identifying regulatory needs and the construction of AOPs and AOP networks for cardiotoxicity.

Involved partners: SC, MUI, POLITO, IRFMN, CNR
Related WPs: WP5 and WP6 will also use the data generated in WP2.
Data utility: The data and information will serve regulators and scientists for discussing, mutual understanding and making decisions on the regulatory use and/or further validation needs of non-animal-methods for the purpose of assessing the potential cardiotoxicity of chemicals.

Table 3: WP3 objective, data utility and related WP

WP3 Epidemiology and toxicology
Objective: Systematic review approaches are promoted in regulatory hazard and risk assessment of chemicals in the EU. This process assesses the strength of evidence associating environmental chemical mixture with a health outcome. The process aims to expand of methods used in clinical sciences, which typically focus on evidence from human studies, as it incorporates both systematic review and integrated assessment (SYRINA) of different lines of evidence. Population epidemiology findings will be integrated with toxicological effect findings. This approach will achieve an integrated assessment of different strand of evidence, those from population epidemiology and toxicological experiments.
Data purpose: Descriptive and statistical data extracted from primary studies in order to build evidence.
Involved partners: CNR, UU, SC, MUI
Related WPs: Any directly and WPs 2,4,6 indirectly will use summary figures to highlight prioritised cardio-toxic chemicals determining exposures at population level.
Data utility: Highly supported associations of toxic chemicals with direct myocardial damage will be represented in statistical significance terms, i.e., using measures of Odds Ratios or Relative Risks with respective confidence intervals. Data will be useful to all actors involved in developing promotion, prevention and awareness strategies, e.g.: Policymakers, Healthcare providers, Industry and Media, Civil societies, Research, Surveillance, and Academia.

Table 4: WP4 objective, data utility and related WP

WP4 In vitro aged cardiac system
Objective/s: 1. to design a 3D scaffold replicating in vitro aged myocardium properties.

2. to develop a microfluidic system including new pH and O_2 sensors, able to connect one to several cell culture chambers, with a bench-size dimension, developed at ELV (in collaboration with IVT) for bench tests and validated at CNR for cell culture. 3. to develop a 3D bioengineered cardiac tissue model replicating in vitro its functional features. 4. to develop and test the designed sensorized bioreactor, with a standard multi-well plate slot dimension, developed at ELV and EUT (in collaboration with IVT) for bench tests and validated at CNR for cell culture. 5. to test cell behaviour and chemicals in vitro system. 6. to identify biochemical and omics markers for cardiotoxicity
Data utility: - Creation of data concerning experimental results describing the physico-chemical properties of polymers: DSC, TGA, rheology, mechanical testing, SEC, ATR-FTIR, etc. to ensure the material selection for the design of the 3D models of cardiac control and aged tissue. - Creation of Firmware repository through code with version control for the electronics to function. - Creation of perfusion data concerning flow sensor measurements to assess the performance of the microfluidic system for the bioreactor perfusion - Collection of literature, patent search and direct contact with potential end-users to carry out a microfluidic system market study to determine the market needs for the products under development. - Created prototype version data to be shared with partners in the R&D phase of the technology. - Creation of cell culture samples for omics analysis and market quantification in the experimental results category to allow the analysis of chemical mixtures effects.
Involved partners: POLITO, ELV, IVT, EUT, CNR
Related WPs: WP6 and WP5 will use the part of the data created in WP4. Specifically, the prototype version and the cell culture datasets.
Data utility: Researchers, Toxicologists, Biologists, Chemists, Bioinformatics, Bioengineers, Industry, chemical and pharmaceutical companies.

Table 5: WP5 objective, data utility and related WP

WP5 Omics
Objective/s: 1. Harmonization of sample collection and optimization of extraction protocols to ensure reproducibility and quality data. 2. Performance of metabolomics, proteomics and transcriptomics analysis. 3. Assessment of metabolic, proteomic and transcriptional signatures of cardiac tissue. 4. Performance of proteomics analysis by OrbiTrap mass spectrometry technology.

5. Establishment of biomarkers of toxicity due to the chemical exposure. 6. Characterization of chemical derivatives from metabolization of toxic compounds.
Data purpose: - Created database integrated by derived metabolized compounds from polluants through analysis in LC/GC-MS for the creation of computational models. - Created database of QSAR for cardiotox to allow the prediction of cardiotoxic effects of chemicals obtained through analysis in LC/GC-MS. - Creation of metabolomics database for the discovery of biomarkers and the creation of computational models. - Creation of proteomics database based on proteins obtained through the analysis in nLC-OrbiTrap-MS for the discovery of biomarkers and the creation of computational models. - Created transcriptomics database formed by gene expression levels through analysis in Illumina platform for the discovery of biomarkers and the creation of computational models - The metabolomics, proteomics and transcriptomics databases will be used as input data for the in-silico ML model
Involved partners: EUT, CNR, POLITO
Related WPs: WP6 and WP7 will use the data generated by WP5.
Data utility: ESCHA, European Medicines Agencies (and similar agencies outside the Europe, like FDA), chemical and pharmaceutical companies, Analyst, Biologist, Chemists, Toxicologist, Researchers, Bioinformatics.

Table 6: WP6 objective, data utility and related WP

WP6 Computational assessment of cardiotoxicity
Objective/s: 1. Develop a TK/TD model that explains the effect of a chemical compound in the cardiac tissues. 2. To develop machine learning methods relevant to assess the toxicity of mixtures towards the heart using toxicological data and knowledge of the Adverse Outcome Pathways (AOP) of cardiac toxicity. 3. Enabling the deployment of ML model into a cloud data platform. 4. Integration of all data sources and output ML data in the ALTERNATIVE platform. 5. Creation of APIs which allow external systems to exchange data and utilise the ML in silico model developed on the ALTERNATIVE platform.
Data purpose: - Collected MIE experimental data from PubChem, CheMBL and other sources on the integration with MIE and associated to cardiotoxicity for the creation of computational models.

- Collected literature on physiological data (Anatomic and physiological characteristics of target populations) in order to parametrize the PBPK models - Created mathematical models (code) on literature, calculated and fitted to experimental data to predict the concentrations at the biophase that drive effects (PD) - Created TK/TD models (mathematical model) to model the action of the toxins in the cardiac tissue (code). - Created data (code) to build the ALTERNATIVE could platform in order to store and process data generated from the other WPs. - Created toxicity data generated as output of the in-silico ML model to be used by ML algorithms.
Involved partners: IRFMN, EUT, UU, CST, POLITO, CNR, SC
Related WPs: WP6 will use data from WP2, WP4 and WP5
Data utility:ESCHA, European Medicines Agencies (and similar agencies outside the Europe, like FDA), chemical and pharmaceutical companies.

Table 7: WP7 objective, data utility and related WP

WP7 Proof of concept and recommendations
Objective/s: 1. Propose a candidate procedure for cardiotoxicity and co-exposure. 2. Integrate the different technical elements composing the ALTERNATIVE platform. 3. Provide a proof of concept for the ALTERNATIVE system. 4. Provide a series of recommendations for a new risk assessment and regulatory approaches for chemical mixtures.
Data purpose: validation of toxic compounds/mixture and the respective biomarkers
Involved partners: IRFMN, SC, IVT, MUI, CNR, POLITO, EUT, ELV, UU
Related WPs: WP7 will use data from WP5**.**
Data utility: ESCHA, European Medicines Agencies (and similar agencies outside the Europe, like FDA), chemical and pharmaceutical companies.

Table 8: WP8 objective, data utility and related WP

WP8 Impact Creation
Objective: ALTERNATIVE's overall goal regarding impact creation (Obj.-6) is to create a sustainable impact towards better assessing and containing harmful co-exposures to industrial

chemicals and pharmaceuticals through the effective dissemination and exploitation of the project results. This will be achieved through the following sub-goals/tasks: 1. Effectively create awareness and disseminate ALTERNATIVE results to target audiences, particularly relevant research communities, regulators, and the chemical and pharmaceutical industry. (T8.1) 2. Effectively manage the knowledge and Intellectual Property Rights (IPR) created in the project, to prepare their widespread commercial and non-commercial use by consortium partners and third parties. (T8.2) 3. Develop suitable exploitation routes for ALTERNATIVE results. The project will identify and explore promising exploitation opportunities for the project results and develop plans for pursuing these opportunities individually and jointly.
Data purpose: - Collected data regarding the participants of workshops, newsletter subscribers and Alternative tracker repository with the aim to fulfil WP8 tasks such as the registration and sharing of the newsletter, monitoring the impact of dissemination activities and results and the information regarding workshops.
Involved partners: EUR, EUT, POLITO, IVT, IRFMN, SC, CST, CNR, UU, MUI, ELV
Related WPs: Dissemination-related information used from all WPs
Data utility: Policymakers and regulators on EC level and EU member state level, National health services, Pharma companies, chemical companies, Environmental and health related NGOs and citizen initiatives in Europe, Scientific community.

ALTERNATIVE makes use of the Only Office provided by Eurescom. Access rights to that system are based on credentials (username and password). Without other agreements made, the data collector/creator/owner stays responsible for his own data.

In case other data may be useful for others outside the consortium, the general principle of ALTERNATIVE is to make data as openly accessible as possible. However, this principle and the degree to which data are openly shared will be discussed, as there may be cases with good reasons to keep data confidential.

Regarding re-using existing data, this will be determined by the “Background” brought by the partners. “Background data” is data and/or data bases already existing before the project, that are brought to the project. Partners must give each other access to the background that is needed to implement their own tasks and exploit their own results. However, there is no obligation to provide access if there are restrictions or limits (legal or otherwise) and the partner has informed the other partners about it. The Background is specified in the Consortium Agreement (CA). Partners must give access to the background that is needed to implement their own tasks and exploit their own results.

3.2 ALTERNATIVE Datasets

In the following table, the preliminary project datasets are presented in listed on a WP1 to WP8 order. For each dataset, the origin is specified, the types and formats of the data and whether the data is being re-used, collected from other sources, named (col.). In the table or created, identified by (cre.). To differentiate datasets, a title has been given to each to easily identify them.

The expected size of the datasets is still “To be Discussed” (Tbd) at this stage of the project (M6) because the information is starting to be collected and there is still no certainty of its extent. However, during the following DMP submissions, this information will be updated, as it is also essential to foresee the platform capacity of storing data.

Finally, datasets’ locations are also considered; however, as the project progresses and some of this is shared, FAIR principles will be considered to ensure long-term preservation and the use of Open Access Repositories such as Zenodo. ALTERNATIVE’S partners are at M6 normally storing the data on its own entity storage systems or in the only office platform, when being shared. Thus, the starting raw data is not shared with others, but it will be the own partner who, will then share the information in compatible formats after reviewing and processing the data.

Table 9: Summary of ALTERNATIVE dataset

No	Origin	Origin-1	Title	Category	Resp.	WP cre.	WP use	Format	Size	Location
1	Col.	Literature, Discussions during PARERE workshop, Interviews with regulators	**Regulatory needs for cardiotoxicity**	Interviews + Public information	SC/MUI	WP2	WP2, 4, 5, 6, 7	PDF and Word	Tbd	Raw data: SC server (not shared)
2	Col.	Literature, AOP wiki	**AOP**	Literature data	SC/MUI	WP2	WP2, 4, 5, 6, 7	PDF and Word	Tbd	Raw data: SC server (not shared)
3	Col.	Literature, PubMed, EMBASE and Scopus databeses	**Reviews of published evidence in Epidemiologic and Toxicologic literature**	Literature data	CNR	WP3	WP2, 3, 4, 5, 6, 7	Raw data: .pdf Elaborated data: XLSX	Tbd	Raw data: Rayyan server (not shared) Extracted data: Google Drive
4	Col	Literature	**Phisico-chemical characterization on polymers and native tissue**	Literature data	POLITO	WP4		PDF and Word	Tbd	OneDrive (not shared)
5	Cre.		**Polymer physico-chemical characterization**	Experimental results	POLITO	WP4		XLSX, PDF, PPT, ZIP	Tbd	OneDrive
6	Cre.		**FW Repository (simulation)**	Firmware	EUT	WP4		Code	Tbd	OneDrive
7	Cre.		**Perfusion data**	Experimental results	ELV	WP4		CSV, XLSX	< 1 MB	Raw data: ELV server (not shared) Extracted data: On8eDrive

No	Origin	Origin-1	Title	Category	Resp.	WP cre.	WP use	Format	Size	Location
8	Col.	Literature, survey	**Microfluidic system market study**	Interviews + litterature	ELV	WP4		PDF	Tbd.	Raw data: ELV ser1ver (not shared) Extracted data: OneDrive
9	Cre.		**Prototype version**	Hardware (prototype design) and protocols to use the systems	IVTECH / EUT/ELV	WP4	WP6	PDF, ZIP, XLSXL, PPT	Tbd	Tbd
10	Cre.		**Cell Culture**	Experimental results	CNR	WP4	WP5	Raw data: *.D Extracted data: XLSX	1 Mb	Dropbox
11	Cre.		**Metabolized polluants database**	Experimental results	EUT	WP5	WP6, WP7	Raw data: *.D Extracted data: XLSX	Raw data: 100 Mb/file Extracted data: 1 Mb	Raw data: EUT server (not shared) Extracted data: OneDrive
12	Cre.		**QSAR models for cardiotox**	Software (expanded)	IRFMN	WP5, WP6	WP7	Code	Tbd	VEGA software, downloadable from the internet
13	Col	Literature	**Search of known biomarkers for model validation**	Literature data	EUT	WP5	WP6, WP7	PDF and Word	1 Mb	OneDrive
14	Cre.		**Metabolomics analysis**	Experimental results	EUT	WP5	WP6, WP7	Raw data: *.D Extracted data: XLSX	Raw data: 100 Mb/file Extracted data: 1 Mb	Raw data: EUT server (not shared) Extracted data: OneDrive

No	Origin	Origin-1	Title	Category	Resp.	WP cre.	WP use	Format	Size	Location
15	Cre.		**Proteomics analysis**	Experimental results	EUT	WP5	WP6, WP7	Raw data: *.raw / *.msf Extracted data: XLSX	Raw data: 100 Mb/file Extracted data: 1 Mb	Raw data: EUT server (not shared) Extracted data: OneDrive
16	Cre.		**Transcriptomics analysis**	Experimental results	EUT	WP5	WP6, WP7	Raw data: *.fastq Extracted data: XLSX	Raw data: 100 Mb/file Extracted data: 1 Mb	Raw data: EUT server (not shared) Extracted data: OneDrive
17	Cre.		**QSAR models for MIE/KE at the basis of cardiotoxic effects**	Source code	IRFMN	WP5, WP6	WP6	Code	Tbd	VEGA software, downloadable from the internet
18	Col.	Data	**Metabolomics database, Proteomics database, Transcriptomics database**	Database	CST	WP5	WP6	Tbd	Tbd	European public cloud provider (tbd which one)
19	Col.	PubChem, ChEMBL, other sources	**MIE/KE experimental Data**	Experimental results	IRFMN	WP6		XLSX, CSV, sdf	Tbd	Seafiles (IRFMN server not shared) Supplementary to papers, zenodo
20	Col.	Literature	**Physiological data**	Database	UU	WP6		XLSX	Tbd	OneDrive

No	Origin	Origin-1	Title	Category	Resp.	WP cre.	WP use	Format	Size	Location
21	Cre.	Literature, calculated, fitted to experimental data	**PBPK**	Code	UU	WP6		R	Tbd	OneDrive
22	Cre.	Created	**TK/TD models**	Software	EUT	WP6.2		Code, generic input output data file (XLSX, CSV, ...), other text files.	Tbd	Tbd
23	Col.	Software	**Apache Hadoop**	Software	CST	WP6.6		Code	Tbd.	European public cloud provider (tbd which one)
24	Cre.	Data	**Toxicity data**	Database	CST	WP6.6		Plain text format (CSV, tsv, Json or other)	>100 GB	European public cloud provider (tbd which one)
25	Col.	Web form	**Participants of Workshop X**	Personal (registration) data	EUR	WP8 - T8.1		CSV, XLSX	< 1 MB	Eurescom Web server in Germany, OnlyOffice project repository hosted by Eurescom in Germany

No	Origin	Origin-1	Title	Category	Resp.	WP cre.	WP use	Format	Size	Location
26	Col.	Web form	**Newsletter subscribers**	Personal (registration) data	EUR	WP8 - T8.1		CSV, XLSX	< 1 MB	Eurescom Web server in Germany
27	Col.	Public information on the web, entries by Consortium partners	**ALTERNATIVE Tracker repository**	Information about events and publications	EUR	WP8 - T8.1		CSV, XLSX	< 50 MB	Eurescom Web server in Germany

4 FAIR DATA

4.1 Making data findable, including provisions for metadata

Data will be made findable by its storage and organization in a structured and defined way.

Metadata provision and standards for metadata creation

Metadata is defined as “data about data” or “information about information”. It is usually structured textual information that describes something about a digital resource’s creation, content, or context – be it a single file, part of a single file, or a collection of many files. Metadata is the tool that helps people to discover, manage, describe, preserve, and build relationships with and between digital resources. Three distinct types of metadata have been identified and are presented below:

- **Descriptive metadata,** used to identify and describe collections and related information resources. Descriptive metadata at the local level helps with searching and retrieving. In an online environment, descriptive metadata helps to discover resources. Most of the time include information such as the title, author, date, description, identifier, etc.
- **Administrative metadata** is used to facilitate the management of information resources. It is helpful for both short-term and long-term management and processing of data. This information that will not usually be relevant to the public but will be essential for staff to manage collections internally. Such metadata may be location information, acquisition information, etc.
- **Structural metadata** enables navigation and presentation of electronic resources. It documents how the components of an item are organized. Examples of structural metadata could be the way in which pages are ordered to form chapters of a book, a photograph that is included in a manuscript or a scrapbook or the JPEG and TIF files that were created from the original photograph negative, linked together.

Taking this information into consideration, ALTERNATIVE will address the creation of metadata and the compliance to standards by depositing part of its data to Zenodo.

In particular Zenodo creates metadata to accompany the datasets that are uploaded to its repository, reaching a wider audience of interested stakeholders. This metadata can be exported in several standard formats, including open and machine-readable ones (such as MARCXML, Dublin Core, and DataCite Metadata Schema), following the guidelines of OpenAIRE.

For the data which will not be made available in the short-term, as the data is kept in the platform, there is another open and machine-readable metadata standard called Dublin Core Metadata Standard. This standard under the International Standard ISO 15836 is compatible with Zenodo, so it is very useful if the data is made open at a later stage and it is simple to understand. The tool Dublin Core Generator is and will be highly recommended to the Consortium to generate metadata by providing the input required and obtaining an output with metadata that can be downloadable in XML, HTML or XHTML format.

Persistent identifiers (PIDs)

For data published to a wider audience, persistent identifiers/references will be used. An example of a PID is a digital object identifier (DOI) number from the International DOI

Foundation https://www.doi.org/assigned to a publication. In the ALTERNATIVE project as claimed in the Grant Agreement "we will deposit our research data in an accessible format in the Zenodo repository (https://zenodo.org/)". Zenodo registers Digital Object Identifiers (DOIs) for all submitted data, meaning that the data preserved will be accessible for years and the DOIs will function as perpetual link to the resources. Hence, they are findable by 3rd parties interested in the project results.

In addition to Zenodo, datasets will also be uploaded to the ALTERNATIVE's platform and to the webpages in some cases; however, long-term preservation will not be ensured in these last two.

Keywords

Suitable search keywords are used for each data set to optimize that potential user of the data can find the data set. In general, keywords are a subset of metadata and include words and phrases used to name data. In the context of ALTERNATIVE, keywords are used to add valuable information to the data collected/generated as well as to facilitate the description and interpretation of its content and value.

The keywords used will match the content of the data and the standards of the research field and as usual, will address the following principles:

- The "who", the "what", the "when", the "where", and the "why" should be covered.
- Consistency among the different keyword tags needs to be ensured.
- Relevant, understandable and clear keywording ought to be sought

The keywords will generally comprise terms like Environmental toxicology at the population and ecosystems level, Safety Pharmacology, Toxicology, Risk assessment, Cardiac 3D tissue models; proteomics, metabolomics, transcriptomics; in silico models; chemical mixtures toxicity, regulatory uptake, etc.

The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

Naming conventions

According to the UK Data Archive[2], a best practice is to create short yet meaningful names for data files, that facilitate classification. The naming convention should avoid the utilisation of spaces, dots and special characters (such as & or !).In contrast the use of underscores is endorsed, to separate elements in the data file name and make them understandable. At the same time, versioning should be a part of a naming convention to identify the changes and edits in a file clearly.

A naming convention recommended to the Consortium is:

ALTERNATIVE_[Name of the study]_[Number of dataset]_[Issue Data]_[Version number]

- ALTERNATIVE: The name of the project
- Name of the study: A short version of the name of the activity for which the dataset is created
- Number of the dataset: An indication of the number assigned to the dataset
- Issue date: The date on which the latest version of the dataset was modified

[2] Organising — UK Data Service

- (YYYY.MM.DD.).
- Version number: versioning number of the dataset

With the above in mind, an example to showcase the naming structure applied in the context of ALTERNATIVE is provided below. This example does not correspond to an actual dataset.

- ALTERNATIVE_NeedsAndRequirements_Dataset1_2021.11.31_v1 – The first dataset generated within the survey framework was conducted to identify the needs and requirements of the diverse stakeholders. This is the first version of the dataset that was last modified on the 31st of November 2021 (31/11/2021).

Versioning.

Versioning of information revises datasets uniquely identifiable and can be used to determine whether and how data changed over time and define specifically which version the creators/editors are working with. Different versions will be identified by a version number and a date. The versioning number will be composed of two digits separated by a period: the digit before the period represents in ascending orders the official versions submitted to the European Commission as deliverable; digits after the period represents the periodic internal revisions of such official versions.

4.2 Making data openly accessible

The degree to which data are openly shared is case-specific, determined by the involved partners owning the data, which is defined by the rules set in the Consortium Agreement.

The partners will ensure that adequate steps towards IPR protection are taken before exploitation, dissemination and communication, preventing unapproved public disclosure of data, results, tools, products and services. The Data Management Task responsible will continuously work out the plan with the relevant partner(s).

ALTERNATIVE will ensure a) the obligation to protect results b) the confidentiality obligations, c) the security obligations and d) the obligations to protect personal data, as stated in the Grant Agreement, all of which apply. Reasons for data not being shared may include commercial exploitation, protection of IPR, securing of future data production, project agreements, privacy policy or other legislation.

Apart from the above-mentioned restrictions, the general principle of ALTERNATIVE is to make data openly accessible when this is both possible and useful, preferably after publication, when they will be deposited in an open-access research data repository to facilitate further use of data.

As claimed in the Grant Agreement “to ensure impact and cost-efficiency, we will use Gold Open Access publishing in renowned journals and Green Open Access via self-archiving. By default, we will make published scientific articles available at no cost via Green Open Access, using the Zenodo repository, ensuring the long-term accessibility of publications. ALTERNATIVE has reserved budget for a limited number of Gold Open Access publications in selected renowned peer-reviewed journals. ALTERNATIVE pursues openness by making all its public results freely available via its website.”

At this stage of the project, the accessibility of data is still to be discussed, thus the IP claims and foreseen exploitation routes are being collected next to partners for D8.3 deliverable, to

identify potential patents and other relevant aspects related to IPR agreements, and the accessibility of data will be dependent on this topic.

The only data that is stage is known that it won't be openly accessible will be the datasets from WP8-T8.1. They are under GDPR rules no exploitation of data permitted beyond the direct purpose of monitoring, for example, event participants, newsletter subscribers and other communication and dissemination matters.

If there are some restrictions concerning some datasets, the access to be provided is still under discussion, as IPR concerns need to be collected first.

Finally, as stated in Table 9, the location of the data and the metadata, which is raw data mainly, will be stored in the individual partner's servers. The extracted data will be upload it to shared spaces within the Consortium such as Dropbox, Only Office, or OneDrive.

Also, in the not so long-term future, part of the datasets will also be kept in the Web Page and the Platform of the ALTERNATIVE projects, more information regarding these storage locations will be uploaded in the next DMP version.

4.3 Making data interoperable

Data will be in the form of its common standard in this field. This makes the data sets interoperable allowing data exchange and re-use between researchers, institutions, organisations, and countries.

Variables and value names will be constructed following general data processing conventions. A list of value names and used vocabulary will be provided as per need in the Metadata schema, which will be created following Zenodo and Dublin Core standards.

The formats of the extracted data from raw datasets will use the most common data formats are: .docx, .pdf, .mp4, .mp3, .jpg, .csv, .xlsx collected through experiments, lab tests, surveys, etc. All data must be stored properly -data for consortium partners will be accessible on Only Office (the main project repository) and/or other shared repositories.

4.4 Increase data re-use (through clarifying licences)

The project will produce several re-usable assets, including knowledge and datasets. In general, data re-use is encouraged by assuring that easy findable and accessible and that data is supported by documentation and long-term preservation.

According to the ALTERNATIVE Grant Agreement:

- Projects documents: After being finalised, they will be publicly available and released under a **Creative Commons license** to allow maximum reuse of the content
- Presentations: All presentations by project participants about the project will be made publicly available on the project website, third-party website, and social media channels. They will all be licensed via a *Creative Commons* license to maximise the reuse of the knowledge

Among the typologies of CC licences, the following can be found:

- Attribution 4.0 International

- Attribution-Share Alike 4.0
- Attribution-No Derivatives 4.0 International
- Attribution-Non-Commercial 4.0 International
- Attribution-Non-Commercial-Share Alike 4.0 International
- Attribution-Non-Commercial-No Derivatives 4.0 International

Partners voted for the best option and the typology agreed was **CC Attribution 4.0 International**, which is considered the most open, used by default in the Zenodo repository, in which the project partners will deposit their Open Access publications.

Licensing

For data published on open-access repositories the data will be free of charge for any user. Still, specific licenses may apply to this data. For the data with restricted access, specific contracts can be generated between the owner and the third party, to safeguard the owner and to agree on general rules for the data use.

Embargo

In some cases, an embargo for opening the data may be needed to give time to publish or seek patents, keeping in mind that the aim is to make data available as soon as possible. The embargo may be needed to protect company business, product development and its IPR, patent applications, or repeated use of the same data by the partners for additional output. Again, the need for embargo and its length is case specific, determined by the involved partners owning the data.

Data Quality Assurance

In general, partners will be responsible for the data they generate, applying a review process by the team leader who has collected/generated the data before making them open.

5 Allocation of resources

5.1 Covering the cost of open data

The work to process, store and document the data is part of the partner's budget in each WP and to the general budget of the project.

Some public open access data repositories are free of charge to use. The Green Open Access included in the Grant Agreement, allows to publish of scientific articles available at no cost, using the Zenodo repository and ensuring long-term preservation. Related to the platform and website, their costs are covered for conservation of 2 years of a lifetime after the project. Nonetheless, all the relevant data generated will also be preserved through the Zenodo repository.

Finally, ALTERNATIVE has a reserved budget for a limited number of Gold Open Access publications in selected renowned peer-reviewed journals.

5.2 Responsible person

Individual partners will be responsible for the collection, transfer, storage and documentation of data from their research activities, within the framework of the DMP. However, the person in charge of the data management and the coordination of its collection will be Laura Armayones, task leader from **T8.2 Data, Knowledge, and Intellectual Property Rights management**, who is also the author of this deliverable. She will validate the document annually with partners, add the modifications needed towards the end of the project and track any changes that may occur next to the coordinators and WP leaders.

6 Data security

During the project, datasets will be available only to those project partners or project consortium members who have been accredited by and their data usage has been approved by the coordinator or authorized project consortium member.

Project partners will be responsible for curating, preserving, disseminating, and deleting the datasets in their possession in an appropriate manner. The retention time for curated datasets will be the same as for other project results at the project consortium partners.

Data collected or acquired within the project will be stored in a secure IT environment behind a firewall. Access will need authentication.

ALTERNATIVE makes use of Only Office, provided by EURESCOM, which fulfils the highest data security standards. Access to the data is managed by credentials (username and password) including occasional two-factor authentication (multi-factor authentication (MFA)). Other raw data is stored on servers inside the EU from individual partners, who count on security for their servers.

The consortium, within its competencies and available infrastructure, will assure secure storage, delivery and access of personal information, and will manage the rights of the users. In this way, there is a complete guarantee that the accessed, delivered, stored and transmitted content will be managed by the right persons, with well-defined rights, at the right time. State-of-the-art firewalls, network security, encryption and authentication will be used to protect collected data (specific details will be developed in the course of the project). Firewalls prevent the connection to open network ports, and the exchange of data will be through consortium known ports, protected via IP filtering and password. Where possible (depending on the facilities of each partner) the data will be stored in a locked server, and all identification data will be stored separately. Intrusion Detection systems will monitor anomalies in network traffic and activate restraint policy if needed.

For long -term storage and publication (including beyond the project's lifespan) certified repositories as Zenodo that ensure data security are preferred.

Regarding personal data collected by WP8, it will be stored in private servers and the project Only Office repository during the project's life, but it will be destroyed after.

In Table 10 the datasets backup frequency, duration in years, privacy level and whether it is not destroyed at the end of the project is included:

For the destruction of the datasets at the end of the project, it has been considered that:

- No (1): Last version can be kept for exploitation purposes under the license agreement, and a royalty-free version will be provided to end-user partners for their own use
- No (2): Dongle-protected versions can be kept for their own use
- No (3): Data can be kept only for joint exploitation
- No (Tbd): it will be updated in the next DMP version

Table 10: Data security from the ALTERNATIVE datasets

No	Title	File Location	Back-up frequency	Destroyed at the end?	Duration (preservation)	Privacy level
1	Regulatory needs for cardiotoxicity	Raw data: SC server (not shared)	Daily	No (3)	Tbd	Consortium
2	AOP	Raw data: SC server (not shared)	Daily	No (3)	Tbd	Consortium
3	Literature databases	Raw data: Rayyan server (not shared) Extracted data: Dropbox	Once	No (3)	Tbd	Consortium
4	Phisico-chemical characterization on polymers and native tissue	Onedrive	Tbd	No (3)	Tbd	Consortium
5	Polymer physico-chemical characterization	OneDrive	Tbd	No(3)	Tbd	Confidential
6	FW Repository (simulation)	OneDrive	Version control	No (Tbd)	Tbd	Private Eurecat
7	Perfusion data	Raw data: ELV server (not shared) Extracted data: OneDrive	Monthly	No (Tbd)	Tbd	Consortium
8	Microfluidic system market study	Raw data: ELV server (not shared) Extracted data: OneDrive	Daily	No (Tbd)	Tbd	Consortium, confidental or stricter

No	Title	File Location	Back-up frequency	Destroyed at the end?	Duration (preservation)	Privacy level
9	Prototype version	Tbd	Monthly. Tbd	No (Tbd)	Tbd	Consortium
10	Cell Culture	Dropbox	Daily	No (3)	Tbd	Consortium
11	Metabolized polluant analysis	Raw data: EUT server (not shared) Extracted data: OneDrive	Daily	No (3)	Tbd	Consortium
12	QSAR for cardiotox	VEGA software, downloadable from the internet	Tbd	No (Tbd)	Tbd	Public
13	Search of known biomarkers for model validation	OneDrive	Once	No (Tbd)	Tbd	Public
14	Metabolomics analysis	Raw data: EUT server (not shared) Extracted data: OneDrive	Daily	No (3)	Tbd	Consortium
15	Proteomics analysis	Raw data: EUT server (not shared) Extracted data: OneDrive	Daily	No (3)	Tbd	Consortium
16	Transcriptomics analysis	Raw data: EUT server (not shared) Extracted data: OneDrive	Daily	No (3)	Tbd	Consortium

No	Title	File Location	Back-up frequency	Destroyed at the end?	Duration (preservation)	Privacy level
17	QSAR models for MIE	VEGA software, downloadable from the internet	Tbd	No (Tbd)	Tbd	Public
18	Metabolomics database, Proteomics database, Transcriptomics database	European public cloud provider (tbd which one)	Tbd	Tbd	Tbd	Consortium
19	MIE experimental Data	Seafiles (IRFMN server not shared)	Daily	No (Tbd)	Tbd	Public (after publication)
20	Physiological data	OneDrive	Daily	No (Tbd)	Tbd	Public
21	PBPK	OneDrive	Daily	No (Tbd)	Tbd	Public code
22	TK/TD models	don't know	Tbd	No (Tbd)	Tbd	Consortium
23	Apache Hadoop	European public cloud provider (tbd which one)	Tbd	No (Tbd)	Tbd	Consortium
24	Toxicity data	European public cloud provider (tbd which one)	Tbd	Tbd	Tbd	Consortium
25	Participants of Workshop X	Eurescom Web server in Germany, OnlyOffice project repository hosted by Eurescom in Germany	Daily	Yes	0	Consortium

No	Title	File Location	Back-up frequency	Destroyed at the end?	Duration (preservation)	Privacy level
26	Newsletter subscribers	Eurescom Web server in Germany	Daily	Yes	0	Consortium
27	ALTERNATIVE Tracker repository	Eurescom Web server in Germany	Daily	No (2)	10 years	Consortium

7 Ethical aspects

When processing personal data ALTERNATIVE ensures full compliance with article 39 of the GA and EU Regulation 2016/679 on the protection of personal data through the implementation of data checks and data sanitization procedures. The ALTERNATIVE consortium partners confirm compliance with the EU General Data Protection Regulation (GDPR), including relevant systems and privacy practices and deploy privacy-by-design and privacy-by-default. The ALTERNATIVE consortium partners declare that technical and organisational measures compliant with GDPR will be implemented.

Human cells used in the project will be obtained from accredited commercial providers like ATCC, which adhered to the highest ethical standards for obtaining cell lines and procurement of human biospecimens. No personal data will be involved and as such the European General Data Protection Regulation (EU GDPR) 2016/679 does not apply. An exception are personal data that may be collected as part of the dissemination and communication activities (see below under PERSONAL DATA).

Eurescom, as leader of WP8 “Impact creation” and of Task T8.1 “Dissemination and communication” will supervise the collection, storage, protection, retention and destruction of personal data used during the project’s dissemination and communication activities. Informed consent will be used to collect data from project partners and relevant stakeholders (EU officials, Member States delegates, etc.) and target audiences via social media (Twitter, LinkedIn, YouTube, etc.). A template for the informed consent to be shared with users is included in the annex.

The ALTERNATIVE project will not collect and/or process sensitive personal data (i.e., ethnicity, political opinions, religion, health, finances, etc.), as obtaining such types of data is outside the project scope. In the specific case of social media analysis, the ALTERNATIVE partners will not gather private personal information, including personal sensitive data, or use secretive methods to gain access to social media data. The ALTERNATIVE social media accounts will be public.

7.1 Internal procedure for legal and ethical data compliance

The type of data that will be generated by the ALTERNATIVE project are detailed in this Data Management Plan, also whether and how these data will be exploited or made accessible for verification and re-use, and how it will be storage and preserved.

Eurescom will share the Data Management Plan with all partners to guide them on how to accomplish the legal and ethical requirements regarding data collection, storage, protection, retention and destruction, and their responsibilities regarding research data quality, sharing and security.

8 References

[1] https://alternative-platform.eu/ ALTERNATIVE website 2021

[2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm Open Access and Data Management Resources by the European Commission

Annex A Annexes

A.1 Data Management and FAIR principles workshops

During the first 6 months of the ALTERNATIVE project, two sessions have been organized to ensure that all partners engaged in the Data Management activities. During the kick-off meeting, which took place the 19th and 20th of October of 2021, a **data management workshop** was organised. Thanks to this workshop, a first preliminary table with foreseen datasets within the project was created. During this session, which lasted half an hour all consortium partners filled out a table considering aspects such as the use of sensitive data, the privacy level and the category of their data.

From these results, a more extensive table available in the Only office repository was explained and shared with the partners during a subsequent meeting organised the 4th of November of 2021 by the WP8 members.

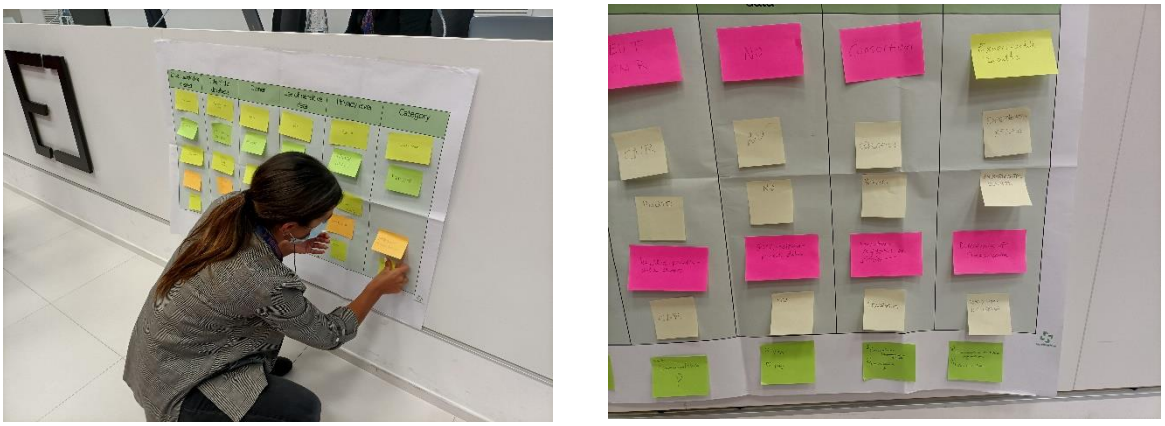

Figure 1: Data Management Workshop organised the 20th of October of 2021 during the Kick-off Meeting

Following these awareness activities, a **FAIR Principles theoretical session** was organized during the 2nd Plenary Meeting, which took place the 31st of January and the 1st of February of 2022.

During the session, some relevant concepts as 'metadata' were explained, as also good practices to develop by researchers and a repository of practical tools (FAIR principles self-assessment questionnaires) and resources (EC explanatory document) was created and shared in the Only Office platform, internal for project partners.

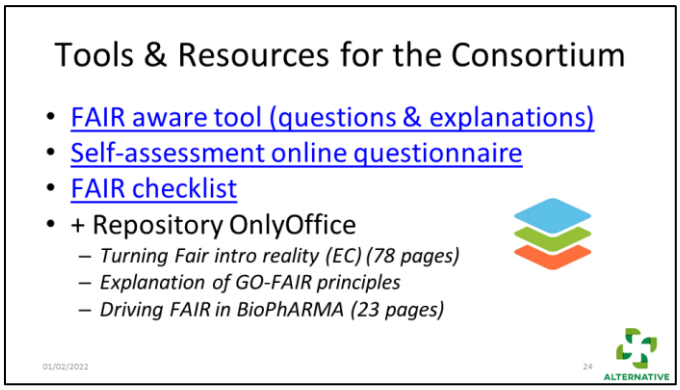

Figure 2: FAIR Principles session organised the 1st of February of 2022 during the Plenary Meeting

A.2 Template of the informed consent form

The following informed consent form/information sheet included in the Grant Agreement will be used by the project to collect and store personal information from the project partners and stakeholders (name, email, function within own organization).

In conformance with GDPR, we would like to kindly ask for your consent to: – keep your contact information (full name, email address) in the project's internal and external communication platforms (SharePoint, ERIC Forum website), – and to use them for the mailing groups presented to you during the Kick-off meeting, which will be used to keep you informed about the projects' activities and for a better interaction among different work packages. For that, please reply to this email with: "I at this moment give my consent to store and use my contact information as mentioned. If you have questions or objections to this, please let us know by replying to this email

[end of document]

