Data Management Plan

Deliverable D8.10

CIRCULATION
PU

VERSION
1.0

DATE
30-11-2020

AUTHORS
Tor Dokken

LEAD PARTNER
SINTEF

CONTRIBUTING PARTNERS
BOC

QUALITY CONTROLLERS
Robert Woitsch

Consisting of

SINTEF	SINTEF AS
TTTECH-IND	TTTECH INDUSTRIAL AUTOMATION AG
Jotne	JOTNE EPM TECHNOLOGY AS
FBA	FUNDINGBOX ACCELERATOR SP ZOO
TNO	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO
BOC	BOC ASSET MANAGEMENT GMBH
UNIBO	ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA
CLOUDBROKER	CLOUDBROKER GMBH
IR	ASSOCIATION IMAGES & RESEAUX
PSNC	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK
SPS	SPACE STRUCTURES GMBH
CORDIS	CORDIS AUTOMATION B.V
Unit040	UNIT040 ONTWERP BV
Author-e	AUTHOR-E BV
Additive	ADDITIVE INDUSTRIES BV
CT INGENIEROS	CT INGENIEROS AERONAUTICOS DE AUTOMOCION E INDUSTRIALES SL
AETNA GROUP	AETNA GROUP SPA
Graphenstone	INDUSTRIA ESPANOLA PARA EL DESARROLLO E INVESTIGACION 2100

Document History

Version [1]	Issue Date	Stage	Content and Changes
1.0	30-11-2020	Complete	

[1] Integers correspond to submitted versions

LIST OF DEFINITIONS, CHANGE2TWINS AND ABBREVIATIONS

TABLE 1: LIST OF ABBREVIATIONS

Term	Explanation
BibTeX	A reference management software for formatting lists of references.
CC licence	Creative Commons licenses are tools to grant copyright permissions to creative work.
CC-BY	This CC-license lets others distribute, remix, tweak, and build upon your work, even commercially, if they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.
CC-BY-NC	This CC-license lets others remix, tweak, and build upon your work non-commercially, and although their new work must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.
CC-BY-SA	This CC-license lets others remix, tweak, and build upon your work even for commercial purposes, if they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use.
CSL	Citation Style Language An open XML-based standard to format citations and bibliographies.
DMP	Data Management Plan
DoA	Description of the Action
DOI	Digital Object Identifier
FAIR data	Findable, Accessible, Interoperable, Re-usable data
GDPR	General Data Protection Regulation. Regulation (EU) 2016/679.
JSON	JavaScript Object Notation An open-standard file format.
MARCXML	An XML schema based on the common MARC21 standards
OAI-PHM	The Open Archives Initiative Protocol for Metadata Harvesting.
Research data	Refers to information, facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings, and images.
REST API	REST is an architectural style that defines a set of constraints to be used for creating web services. API means Application Programming Interface
SSL/TLS	Secure Sockets Layer / Transport Layer Security These are protocols offering secure communication on the internet.
Zenodo	Zenodo is a catch-all research data repository that enables researchers, scientists, EU projects and institutions to share research results, make research results citable, and search and reuse open research results from other projects. Zenodo is harvested by the OpenAIRE portal and hosted by the CERN cloud infrastructure.

EXECUTIVE SUMMARY

The purpose of the Data Management Plan (DMP) is to contribute to good data handling through describing what research data the project expects to generate and which parts of the data that can be shared with the public. Furthermore, it gives instructions on naming conventions, metadata structure, storing of the research data and how to make public data available.

During the 48 active months of the project, a project-internal TEAMS/SharePoint infrastructure is provided by the Coordinator SINTEF, which will be used as the online working and collaboration platform for data relevant issues. It is only accessible to registered project participants (beneficiaries) and the access rights depend on their roles within the project. All datasets will be uploaded to this site to ensure that data handling is in accordance with national and European rules on data protection and privacy. Metadata are added to datasets, and instructions on how to create relevant dataset are provided in this document.

Change2Twin will use the open research data repository *Zenodo* to follow the Horizon 2020 Open Access Mandate. This mandate applies to the underlying research data of publications, but beneficiaries can also voluntarily make other datasets open. In Change2Twin, all public deliverables and publications including related research data will be uploaded to the *European Commission Funded Research (OpenAIRE)* Community in Zenodo. Other datasets not directly linked to deliverables and publications that has the dissemination level "*Public*" will also be uploaded to Zenodo. Uploads will be done upon approval of the deliverables by the European Commission, upon publication or acceptance of scientific publications, or, for underlying datasets, at the end of the project at the latest.

Each dataset will be given a persistent identifier (Digital Object Identifier, DOI), supplied with relevant metadata, and linked to the project name and grant agreement number. Publications and underlying research data will be linked to a Creative Commons license that will regulate reuse of the Change2Twin research data. Data security arrangements are defined in this document and the monitoring of the those is part of the management task.

Ethical aspects related to data collection, generation and sharing have been considered and nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, as well as rules and norms regarding ethics in conducting research.

The DMP is a living document and will be updated periodically for each review throughout the project to reflect the actual research data generated. The final version of the DMP concluding all changes will be made available at the end of the project focusing on instructions to reuse the data provided by Change2Twin. Day-to-day data management and monitoring will be continuously updated to reflect actual data generation.

The monitoring of this list is performed by the Project Coordinator.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 PURPOSE OF THE DOCUMENT

This Data Management Plan (DMP) provides an overview of the handling the datasets the project is expected to generate. Relevant meta data like types, formats and how to access or process the data are considered following the so-called “FAIR” data management principles. FAIR stands for “Findable”, “Accessible”, “Interoperable” and “Re-usable” of datasets.

The purpose of the DMP is to specify the guidelines of handling dataset during the project's lifetime and ensure a monitorable plan to enable a re-usage of the data after the project end.

1.2 INTENDED READERSHIP

As the DMP is a project management instrument, this document is mainly concerned with the planning, specification and monitoring of the generation, quality approval and provision of data sets that are considered as sustainable results of the project. Beside the project management, the DMP addresses:

Internally in the project:

- Project participants who are responsible for, or in any way involved with, data collection and data handling can use this document for instructions on how to handle, store and process data.
- All project participants can use this document to get an overview of data collected in the project and how this is processed, stored, and made accessible.

External audience:

- The **Data Summary** and **FAIR Data Management** sections can be used by all relevant stakeholders who are interested in Change2Twin related activities and research topics to get an overview of the data collected in the project, how to access this data, and, if applicable, how to re-use this data in their own activities.

1.3 STRUCTURE OF THIS DOCUMENT

This document is structured as follows:

- Section 1 is an introduction chapter describing the main purpose, structure and intended readership of the DMP, as well as relationships to other project deliverables.
- Section 2 provides an overview of possible research data in the Change2Twin project, with details on types, formats, origin and metadata provisions.
- Section 3 introduces the current list of datasets
- Section 4 describes how the Change2Twin project will comply with the principles of FAIR data management and H2020 Open Access Mandate
- Section 5 describes the execution of the DMP considering the resources allocate, data security and ethical aspects.

This DMP is based on DMPOnline[2] provided templates addressing the guidelines for FAIR data management in Horizon 2020[3].

[2] https://dmponline.dcc.ac.uk/about_us
[3] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2 DATA MANAGEMENT INTRODUCTION

The initial set of the of data set categories, are presented in Table 1, on page 9. This list is evolving while Change2Twin progresses, a continues update of this list is provided for each review period.

2.1 PURPOSE OF THE DATA COLLECTION AND GENERATION

The goal of the Change2Twin Project is to ensure that 100% of manufacturing companies in Europe have access to 100% of technologies needed to deploy a digital twin.

The overall motivation for data collection in Change2Twin is to:

- Demonstrate, test and enable the assessment of digital twin technologies for interested stakeholders
- Facilitate evaluations, validations and learning as well as gather feedback to improve solutions.

Only data that is needed to perform project activities will be collected, and as far as possible personal data will not be collected unless this is necessary.

2.2 DATA TYPES, FORMATS AND SIZE

Types of data

The Change2Twin project will collect, generate and reuse various types of data, such as:

- Manually collected data from interviews, surveys, and questionnaires
- Automatically collected data from [e.g. mobile applications/software installed in…etc.]
- Contact information [e.g. project internal and external stakeholders/participants in pilots/workshops/evaluation activities etc.]
- Existing data from [e.g. literature reviews, databases, statistics, etc.]
- Pilot case and experiment specific data

Data will be organised in datasets according to type and content of the data.

Some of the data may be classified as *personal data*, such as names, e-mail addresses, IP addresses, phone numbers, etc. This data must be irreversibly anonymised before being made public. If such data cannot be irreversibly anonymised, it will remain confidential and the corresponding partner is responsible to respect the privacy as well as follow the procedures for confidential data specified in the Consortium Agreement. A corresponding template is provided in form of appendix A, which is considered as an input for bilateral agreements of the data providing and data receiving person or entity.

Only anonymised data or analysis of the aggregated data, containing no details that can be linked to individual participants, will be made public.

Data formats

The Change2Twin project aims to use widely accepted data formats, but due to the innovative character of the project, is may also work with proprietary formats or quasi-standards used within niches.

Publications, Deliverables and Documentations:

- Documents/Reports/Publications: e.g. .PDF, docx
- Audio, Video files or pictures: e.g. .avi, mp4, wav, jpg, png

Transformation, Data:

- Spreadsheets and Databases: e.g. xslx, cvs,

- Models or Ontologies: e.g. ttl, RDF(S), OWL, BPMN
- Workflows and Deployment Bundles: e.g. BPMN, Docker Images,
- Questionnaires and Assessment Matrix: current unknown format

Digital Twin Data:

- ISO 10303 (STEP) files: e.g. stp, step
- Timestamps: e.g. JSON, XML,
- Key Performance Indicators: current unknown format

Aforementioned data formats are considered as samples, as one data set can consist of several formats when combining either different sub-dataset in different – e.g. workshop results are packaged as video, picture and documentation files – or if sub-data are on purpose provided in different formats – e.g. an ontology in different tool formats to raise re-usability.

Size of data

Although the size of the data will vary the appropriate size of the data will be approved during the quality approval process before transferring the data form the project internal site to the public repository in order to avoid unnecessary challenges regarding the capacity or handling of the data.

2.3 ORIGIN OF DATA

The Change2Twin project will collect data at pilot partners and open call partner premises. Dependent on the type of data, there will be various methods and origins of data collection involved at each site.

Publications, Deliverables and Documentations:

- Survey responses
- Market survey

Transformation, Data:

- Interviews with groups and individual participants in the pilots/demonstrations/etc. at each site
- Feedback from participants at stakeholder workshops
- Literature study/review and open research data (re-use of existing data)

Digital Twin Data:

- Automated data collection through software installed on a selected device.

Although the aforementioned list is only an initial frame of possible data sources, it draws the awareness to the fact that data also have a lifecycle and the corresponding rights of the data owner or data producers need to be clarified and respected.

2.4 INITIAL METADATA PROVISION

Partners providing public data are responsible for uploading them onto the project-internal space. Each dataset will be catalogued in a list providing an overview of all datasets in the project, as well as uploaded to a dedicated data folder. Datasets use version and access control as long as they are not approved and publicly available. These metadata will be provided for each data set:

- Unique Identifier: e.g. Change2Twin-{WP*X-#* / D*X-# / Publication Citation*}
- File name: e.g. Change2Twin-*Name*
- Date of entry: e.g. YYY-MM-DD
- Version of the dataset: e.g. major changes V1, V2, V3 vs. minor changes V1.1, V1.2, V1.3
- Data type and reference to corresponding application – including version – to use the data

- Description: Executive summary and reference to additional documentation
- WP number: Indicating where most of the work had been performed for the data set
- Responsible person: Contact person in form of name, affiliation and email
- Dissemination level: Intended dissemination level as input for the approval

2.5 OPEN RESEARCH EUROPE

In the second half of November 2020, during the final steps of preparation of this deliverable Change2Twin was made aware of Open Research Europe - The European Commission open access publishing platform. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/other/comm/open-research-europe_horizon-h2020_en.pdf

This will provide all Horizon 2020 and Horizon Europe beneficiaries and their researchers with an easy, high quality peer-reviewed venue to publish their results in open access, at no cost to them, and in full compliance with our open access policies. The platform will use a model of immediate publication of submissions followed by transparent, invited, and open peer review with inclusion of all supporting data. The formal launch of the platform will take place in early 2021.

The launch of Open Research Europe will make it necessary to assess the reference model for uploading datasets depicted in Figure 1, as Change2Twin already realises an open data strategy, but in case the new framework requires adaptation it will be reflected accordingly.

2.6 ZENODO

Zenodo is a "catch-all" open research data repository which gathers research data across all disciplinary fields. It is for non-military purposes only, and the repository is hosted and managed by CERN. All data deposited to Zenodo is stored securely in the CERN Data Centre's cloud infrastructure[4] (see section 5.2).

Change2Twin will use the open research data repository *Zenodo* to comply with the H2020 Open Access Mandate[5]. All scientific publications, including public deliverables and public parts of underlying datasets will be uploaded to the *European Commission Funded Research (OpenAIRE)* Community[6] in Zenodo. In addition, the project will upload other datasets (not directly linked to publications and deliverables) with dissemination level "*Public*" and make them openly accessible via Zenodo.

[4] https://zenodo.org/
[5] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[6] https://zenodo.org/communities/ecfunded/?page=1&size=20

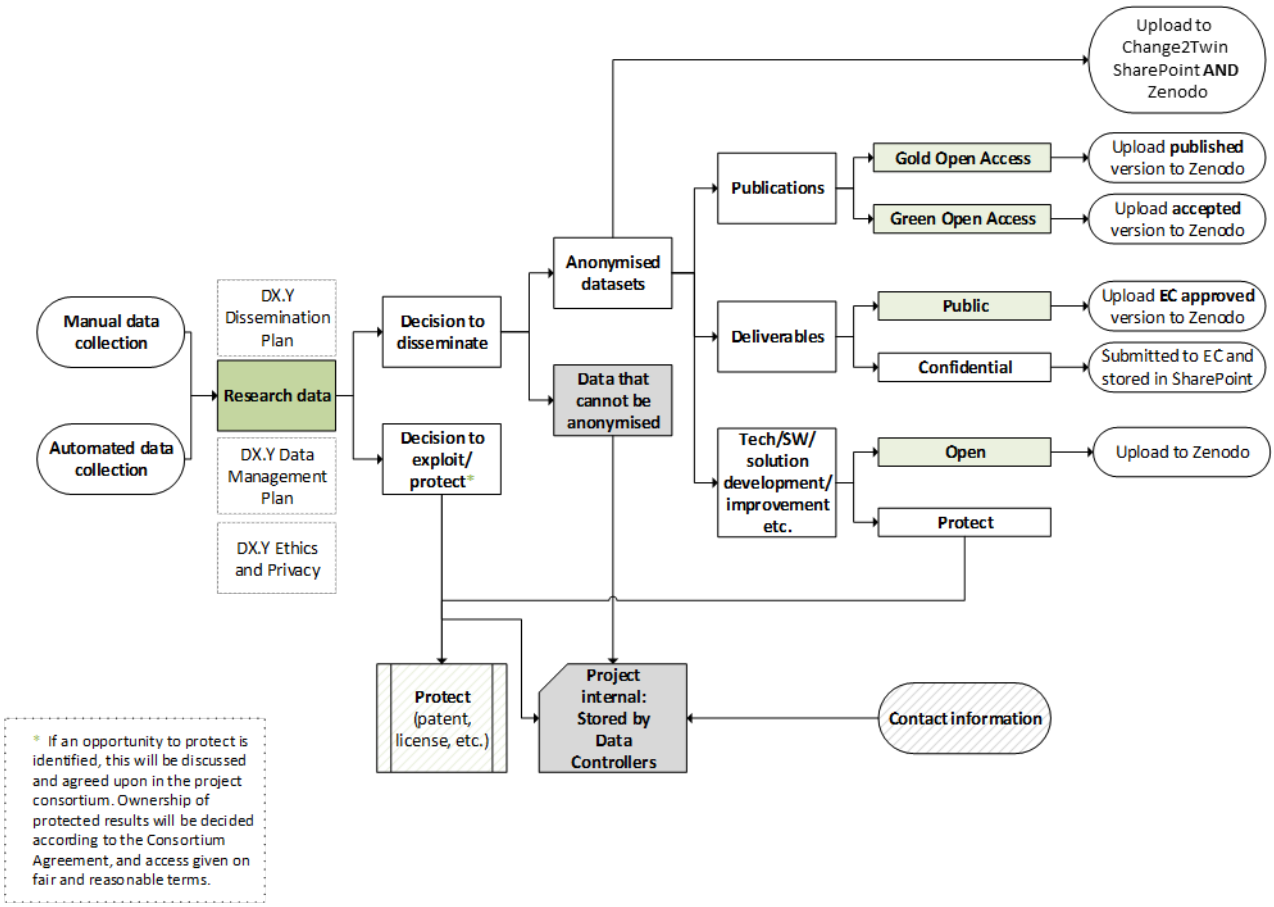

FIGURE 1. REFERENCE PROCESS FOR UPLOADING DATASETS

3 CHANGE2TWIN DATA SETS

Table 3 below lists the initial dataset types, which will evolve during the project duration.

TABLE 1: DATASET TYPES IN CHANGE2TWIN

WP	Name of dataset	Description/Purpose	Format	Origin
2	Digital Transformation Marketplaces	We expect data sets describing marketplaces such as but not limited to: • Offerings Templates • Deployment Configuration • Offerings Bundles • Offering Annotation	Proprietary Docker-Image TOSCA, CAMEL TTL, RDF(S)	Marketplace recherche, Research and experience gained during demonstration
3	Assessment relevant datasets	We expect data set for assessments such as but not limited to: • Questionnaires • Benchmarks • KPI Models	proprietary	Literature research, Research and experience gained in workshops
1, 3, 5	Digital Twin dataset	We expect data sets for digital twin concerning but not limited to: • Production data • Product data, Design and simulation data • Additive manufacturing • Production KPIs •	BPMN, ISO 10303 STEP, ply STL, AMF Proprietary	Pilot and experiment research.
2, 3, 5, 6, 7	Surveys and Success Stories	We expect data sets for success stories, surveys, documentation, and publications.	Office Format	Use Case Partners, DIH

4 FAIR – PRINCIPLES IN DATA MANAGEMENT

Change2Twin follows the so-called **"FAIR" data management[7] principles**, which is an acronym for **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable data. In the following each section explains how Change2Twin implements those aspects. Table 3 provides a current overview of the generated datasets in the Change2Twin project and their accessibility.

4.1 FINDABLE - DATA

4.1.1 Community in Zenodo

Change2Twin will use the Zenodo repository as the main tool to store data. Change2Twin will link all uploads to the *European Commision Funded Research (OpenAIRE)* community. Those publicly available links can then be used in publications, workshops, and conferences, webpages, emails and social-media activities. Zenodo provides version control and assigns DOIs to all uploaded elements.

4.1.2 Metadata in Zenodo

All uploads will be enriched with proposed Zenodo metadata as well as additional meta data the Grant Number and Project Acronym. Metadata associated with each published data set in Zenodo will by default be as follows:

- Digital Object Identifiers (DOI): The DOI number is created by Zenodo
- Version numbers: The version approved in the internal data repository
- Bibliographic information: Author, Affiliation and email
- Keywords: Agreed keywords that are defined during the approval process
- Abstract/description: Approved executive summary
- Associated project and community: Acronym "Change2Twin"
- Associated publications and reports: Referenced DOI in form of publications and deliverables
- Grant information: Grant Number
- Access and licensing info: Explanation on the data format, the corresponding tool that is capable to read the dataset as well as the licences and conditions under which the data can be used.
- Language: Definition of the natural language in case it is needed to interpret the dataset.

4.1.3 Versioning and Digital Object Identifiers (DOI)

Zenodo provides DOI versioning for datasets. Editing and updates of datasets result in new Zenodo identifiers but the DOI can be the same. This allows to either cite the DOI as the stable identifier of several sub data sets, hence the data sets can be handles as so-called "living documents" that evolve over the time but still can be cited, as well as cite a particular sub-dataset which is not evolving and hence ensure a stability[8].

A particular data set can be versioned like:

- v1.0 (specific version): DOI number / zenodo.XXX
- v1.1 (specific version): DOI number / zenodo.YYY
- Concept (all versions): DOI number / zenodo.ZZZ

[7] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[8] http://help.zenodo.org/ (DOI versioning)

The first two DOIs for versions v1.0 and v.1.1 represent the specific versions of the data set. The last DOI represents all the versions in one package.

4.1.4 Approach to search keywords

Each data set is describe using publicly aware keywords to support that dataset can be found from third parties outside of the project. Hence, the dataset specific keywords describe the context in such a way that a person who is not familiar with the project, would use it for a search query. The process of finding appropriate keywords is performed during the approval of the dataset. The keywords and the corresponding explanation – the taxonomy – will be published in the updated versions of this Data Management Plan.

4.1.5 Naming conventions

Data will be named using the following naming conventions:

Descriptive text_H2020_Change2Twin_Deliverable-/WorkpackageNumber_UniqueDataNumber

*Descriptive text_H2020_ Change2Twin _PublicationNumber_UniqueDataNumbe*r

Explanation of the naming convention:

- "Descriptive text" refers to a *short* description of the content of the dataset (see example)
- "Change2Twin " refers to the project's short name or acronym; *Change2Twin*
- "DeliverableNumber" refers to the deliverable number as described in the DoA
- "PublicationNumber" refers to the number the publication has in the project internal directory of all publications from the project
- "UniqueDataNr" is a unique number of the dataset.

Example of dataset name:

APP user acceptance_ H2020_Change2Twin _D1.1_003

Weather data OSL Jan 2019_H2020_Change2Twin_07_017

4.2 ACCESSIBLE - DATA

Data are only useful if they can be easily accessed. There are technical issues as well as legal or licence issues that influence the accessibility of data.

With respect to accessibility hampered by legal and licences issues, we ensure that personal or sensitive data due to privacy concerns and/or commercial or security reasons are protected. The analysis of the data is performed during the approval process to ensure that only public data are uploaded on Zenodo. Furthermore, only freely accessible datasets are provided. In case datasets are offered under commercial conditions, those are directed to data marketplaces under the exploitation path, but not via the open research data path. Data sets with dissemination level "confidential" are not shared independent of privacy/security/ethical or commercial concerns.

With respect to technical accessibility, the aforementioned metadata do not only explain for which purpose the data can be used – which is more addressed by the finding principles – but how the data can be technically accessed by stating the data format, the corresponding tool that can access the data format and if necessary additional information like versions or special configurations that are necessary to enable the tool to access the dataset. Although we foster the use of standards like we expect STEP (ISO 10303) files, we assume that references to alternative solution for accessing the data will be necessary.

The OAI-PHM protocol enables the harvesting of the metadata of the dataset. Metadata is also retrievable through the public REST API. Much of the data will be accessible using web browsers.

4.3 INTEROPERABLE DATE

The interoperability of data depends first on the interoperability of the meta data and the format of the dataset.

With respect to the meta data, Zenodo uses a JSON schema as the internal representation and enables the export in other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo whenever it is appropriate, including external vocabularies like for license (Open Definition), funders (FundRef) or grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

With respect to the datasets, we aim to use widely accepted data formats and standards where possible, but expect that due to the innovative character of the project and the need to apply also quasi-standards or proprietary formats for niches that other data formats will be used. It is expected that also such datasets are based on transformable data formats – e.g. CSV, XML, JSON – that enable a transformation into other formats.

During the data set approval, the interoperability will be also assessed and where possible the corresponding measures will be taken; however, it is the clear goal that Change2Twin creates datasets as one form of results and hence will not ban the publication of a particular data set due to a low interoperability. Hence, we prioritise the publication of qualitative data.

4.4 REUSABLE DATA

The Change2Twin project will enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) all public data sets, and regulate this by using Creative Commons Licences.

4.4.1 Recommended Creative Commons (CC) licences

Change2Twin will use Creative Commons licences (CC), which grant copyright permissions to creative work. As a default, we consider the CC-BY-SA license for research data. This license lets others remix, tweak, and build upon additional work even for commercial purposes, as long as they credit the author of the original data set and license their new creations under the identical terms. This license is often compared to “copyleft”, free and open source software licenses. With this licence all data will carry the same license, which also allows commercial use. This does not preclude the use of less restrictive licenses as CC-BY or more restrictive licenses as CC-BY-NC, which does not allow commercial usage.

Application of licences will be assessed on a case-by-basis in close collaboration with the Coordinator, Innovation Manager and partners concerned to ensure compliance with the Grant Agreement and the Consortium Agreement.

4.4.2 Availability and longevity of the Change2Twin research data sets

Public (anonymous) data

For data published in scientific journals, the underlying data will be made available no later than by journal publication. The data will be linked to the publication. Data associated with public deliverables will be shared once the deliverable has been approved and accepted by the EC. For other public datasets not directly linked to a scientific publication or deliverable, such datasets will be made available upon assessment by the responsible partner that it is ready for publishing, and in the final month of the project at the latest.

Open data can be reused in accordance with the Creative Commons licences. Data classified as confidential will as default is not reusable due to privacy/security concerns.

The public data will remain reusable **via Zenodo for at least 20 years**. This is currently the lifetime stated by the host laboratory CERN. In the event that Zenodo has to close their operations, they have provided a guarantee that they will migrate all content (including metadata) to other suitable repositories.

Confidential (non-anonymous) data

It is foreseen that confidential data will not be provided by Change2Twin.

Classification of research outputs

It is foreseen that classified data according to H2020 Guidance for the classification of information in research projects"[9] will not be provided by Change Twin.

[9] H2020 Programme. *Guidance for the classification of information in research projects*. Version 2.1. 26 October 2016. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/secur/h2020-hi-guide-classif_en.pdf

5 EXECUTION OF DATA MANAGMEENT PLAN

5.1 ALLOCATION OF RESOURCES

The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget. The internal repository is provided as part of the project management infrastructure by the coordinator. The external and long-term preservation of the public data is ensured through Zenodo.

Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

The overall responsibility for data management lies with the Project Coordinator, who can and will involve members from the project management board to assist whenever applicable.

5.2 DATA SECURITY

In the absence of confidential or classified data the security aspects are limited to the correct and safe storage of the datasets. For this purpose, we rely on the following security statements from Zenodo:

- 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: 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 stated to be 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.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- 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.

5.3 ETHICAL ASPECTS

Change2Twin will fully comply with the regulations set out in Regulation (EU) 2016/679, the General Data Protection Regulation (GDPR)[10]. In addition, Change2Twin comply with the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless the country in which research is carried out.

Contact information

The name, affiliation and email address is provided by the author of the data set who then agrees to make those data public.

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

Pictures and videos for communication purposes

Change2Twin will collect pictures and videos for use in communication activities (website, newsletter, social media). Such data will only be collected with prior consent from the people involved and only used for as long as consent is given.

Pictures and video can contain personal data if an individual is the focus of the image or video.

Examples of pictures and videos containing personal data are: 1) pictures/video of individuals stored together with personal details (e.g. identity cards); 2) pictures/video of individuals posted on the project website along with biographical details; 3) individual images published in a newsletter.

Examples of pictures and video that is unlikely to contain personal data are: 1) pictures/video where people are incidentally included in an image or are not the focus (e.g. at a big conference/workshop); 2) images of people who are no longer alive (the GDPR only applies to living people, see section 2.2).

When collecting pictures and video Change2Twin will follow established guidance and best practice on collecting and processing such data to ensure that we adhere to the legal requirements (e.g. guidance established by the University of Reading, UK[11]). Under no circumstances will pictures containing personal information be publicly shared without the subject's explicit consent.

[11] https://www.reading.ac.uk/internal/imps/DataProtection/DataProtectionRequirements/imps-d-p-photographic.aspx

6 CONCLUSIONS

As coordinator of the Change2Twin project, SINTEF will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner.

The DMP is a living document that will expand as the project evolves and new information on data collection, generation and handling arise. Continues update and reporting is provided in conjunction with the review of the reporting periods.

Day to day data management will happen as described in this document, and through continuous collaboration between the partners.

A revised and extended version of this DMP that are updated for the review of the reporting periods will be finalised at the end of the project to document the final status of data management in Change2Twin.

7 REFERENCES

European Commission. *H2020 Programme. Guidance for the classification of information in research projects*. Version 2.1. 26 October 2016

European Commission. *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0, 26 July 2016

European Commission. *Horizon 2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. Version 3.2, 21 March 2017

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC

- APPENDIX A: TEMPLATE FOR DATA PROCESSING AGREEMENT

DATA PROCESSING AGREEMENT

Between

[ENTER DATA PROCESSOR], Company Reg. No. [xxx xxx xxx] ("**the Data Processor**")

and

[ENTER CLIENT/PARTNER], Company Reg. No. [xxx xxx xxx] ("**the Data Controller**")

being an agreement regarding the processing of personal data ("**the Agreement**") to be performed by the Data Processor on behalf of the Data Controller as a consequence of [ENTER BACKGROUND, e.g. agreement entered into with the Data Processor as the Supplier and the Data Controller as the Client dated [date] ("**the Principal Agreement**")].

1. The purpose of the agreement

The Data Processor shall process personal data on behalf of the Data Controller based on the background indicated above.

The purpose of the data processing, the duration and nature of the processing, the type of personal data to be processed and the categories of registered individuals are specified in attachments to this Agreement.

The Agreement shall ensure that personal data are processed in accordance with prevailing statutory requirements for processing personal data, including EU Directive 95/46/EF of 24 October 1995 on the protection of individuals relating to the processing of personal data and on the free movement of such data, which has been implemented in Norway through Act No. 31 of 14 April 2000 relating to the processing of personal data (the Personal Data Act) and associated statutory regulations, as well as the requirements pursuant to decree of the European Parliament and Council relating to the protection of individuals with regard to the processing of personal data and on the free movement of such data, superseding on 27 April 2016 Directive 95/46/EF (the Data Protection Directive), and Norwegian law and associated statutory regulations adopted pursuant to the Data Protection Directive and replacing the Norwegian Personal Data Act. Both the current and the subsequent Personal Data Act are referred to below as "the Personal Data Act".

The Data Processor shall process the personal data as described in the Agreement, and in other ways as may be agreed in writing between the Data Processor and the Data Controller.

Terms and definitions used in the Agreement shall be construed in the same way as in the Personal Data Act.

2. The rights and obligations of the Data Controller. The obligations of the Data Processor

The Data Controller shall ensure that the relevant personal data can be processed. Specifically, Data Controllers shall ensure that adequate legal authority exists and that the agreements entered into with the registered individuals and any consents given are commensurate with and facilitate the processing of personal data as specified in Attachment 1.

The Data Processor confirms that it will implement suitable technical and organisational measures to ensure that all processing pursuant to this Agreement satisfies the requirements of the Personal Data Act with respect to the protection of the rights of the registered individual, as well as complying with all the requirements of Articles 32 of the Data Protection Directive. See also Section 4 for additional obligations. The Data Controller shall at all times maintain full legal ownership of the personal data.

The Data Processor shall only process personal data on the basis of written instructions received from the Data Controller. The Data Processor shall at all times be able to provide documentation of such instructions. The Data Processor shall not process personal data to which it obtains access in any other manner than is necessary to carry out the assignments it receives from the Data Controller.

The Data Processor shall assist the Data Controller in responding to requests submitted by a registered individual wishing to exercise his or her rights pursuant to Chapter III of the Data Protection Directive, taking into account the nature of the processing, also assisting as far as possible by way of suitable technical and organisational measures. The Data Processor shall also assist the Data Controller by ensuring compliance with the requirements connected with personal data security and the assessment of the consequences for personal protection and prior consultation in Articles 32 of the Data Protection Directive , taking into account the nature of the processing and the information available to the Data Processor. If approved standards of conduct exist, pursuant to Articles 40 of the Data Protection Directive or approved certification pursuant to Article 42, with which the Data Processor has undertaken to comply or according to which it has undertaken to be certified, the Data Processor is obliged to comply with said standards of conduct or certification requirements.

The Data Processor shall maintain a log of processing activities it carries out on behalf of the Data Controller, which shall include at least the information specified by Article of the Data Protection Directive. The Data Controller may at any time demand to be provided with a copy of such log.

The Data Processor shall make available to the Data Controller any information necessary to demonstrate that the obligations specified in this Section 2 are fulfilled, as well as facilitating and contributing to audits, including inspections, performed by the Data Controller or any other inspector authorised by the Data Controller. This also applies to providing access to security documentation. The Data Controller itself has direct responsibility for liaison with the relevant supervisory authorities.

The Data Processor has an obligation of secrecy with regard to personal data to which it obtains access as a consequence of the Agreement and its processing of personal data, and shall ensure that persons authorised to process the personal data have undertaken to do so confidentially or are subject to appropriate statutory professional confidentiality. This provision applies also after the expiry of the Agreement.

The Data Processor shall not divulge data or information that it processes on behalf of the Data Controller to third parties without explicit instructions from the Data Controller. The Data Processor shall forward any enquiries received in this respect to the Data Controller without undue delay.

If the Data Processor is of the opinion that an instruction from the Data Controller is in conflict with the Data Protection Directive, the Personal Data Act or any other regulation regarding the processing of personal data, the Data Processor shall immediately inform the Data Controller of this. The Data Processor undertakes to discharge its obligations pursuant to the Agreement irrespective of its opinion.

3. **The use of sub-vendors**

When processing personal data, the Data Processor shall only use sub-vendors (data processing sub-vendors) which have been approved in writing by the Data Controller and which have been confirmed as implementing suitable technical and organisational measures to ensure that processing pursuant to this Agreement complies with the requirements of the Personal Data Act and the need to protect the rights of registered individuals.

Approved data processing sub-vendors at the time of entry into the Agreement are specified in an attachment to the Agreement.

The Data Controller grants the Data Processor general permission to use data processing sub-vendors to process personal data pursuant to the Agreement. If the Data Processor plans to use other data processing

sub-vendors or substitute other data processing sub-vendors, the Data Processor shall inform the Data Controller of such plans and give the Data Controller an opportunity to oppose such changes.

Any data processing sub-vendor shall be made familiar with the obligations of the Data Processor pursuant to this Agreement and with the regulations governing the processing of the Data Controller's personal data, and shall be subject to the same obligations with regard to the protection of personal data as are stipulated in the Agreement. The data processing sub-vendor shall furnish adequate guarantees that technical and organisational measures will be adopted to ensure that its processing complies with statutory requirements. If a data processing sub-vendor fails to satisfy its obligations with regard to the protection of personal data and the requirements of the Agreement, the Data Processor shall assume full responsibility vis-à-vis the Data Controller for the sub-vendor's failure to satisfy those obligations.

4. **Security and non-conformances**

The Data Processor shall satisfy the requirements with regard to security measures as specified by the Personal Data Act and associated statutory regulations. The Data Processor shall be able to document its procedures and other initiatives for satisfying these requirements. The documentation shall be made available to the Data Controller on request.

Security audits shall be carried out regularly at times agreed by the parties to the Agreement. An audit may embrace a review of procedures, random inspections, more comprehensive local inspections and other appropriate verification measures. Agreement shall be reached with regard to the Data Controller's obligation to cover the cost of any use of personnel and resources necessary in connection with the performance of such audits.

In the event of breach of security or personal protection stipulations, the Data Processor shall notify the Data Controller without undue delay. Such notification shall include at least the following:

1. A description of the nature of the breach of personal data security, including, wherever possible, the categories and approximate number of registered individuals affected, and the categories and approximate number of personal data records affected,

2. the name and contact details of the personal protection advisor or any other contact site at which additional information may be obtained,

3. a description of the probable consequences of the breach of personal data security,

4. a description of the measures taken or proposed to handle the breach, including where relevant measures for reducing any adverse effects resulting from the breach.

If all the information cannot be provided in the first instance, it shall be provided successively as soon as it becomes available.

The Data Controller is responsible for submitting notification to the supervisory authority and the Data Processor shall not submit such notification nor contact the supervisory authority unless instructed to do so by the Data Controller.

5. **Transfer of data to foreign countries**

Personal data shall only be transmitted to third countries outside the EU or EEA according to instructions from the Data Controller. The Data Processor shall therefore not transmit, or allow parties in third countries in any way to obtain access to, personal data without explicit prior approval and instructions to this effect from the Data Controller. Consent and instructions must specify the countries to which the information may be transmitted. Even with consent and instructions, transfer to third countries shall only take place on condition that the requirements regarding security and the protection of the rights of the registered individuals pursuant to the Personal Data Act and other rules are satisfied.

6. **The duration of the Agreement, termination orders, obligations in the event of expiry or cancellation**

The Agreement is valid as long as the Data Processor processes or has access to personal data on behalf of the Data Controller pursuant to the Principal Agreement.

In the event of breach of this Agreement, the Personal Data Act or other relevant rules, the Data Controller is entitled to order the Data Processor to cease processing of the information with immediate effect.

On completion of the services connected with processing, the Data Processor shall, as instructed by the Data Controller, delete or return any personal data to the Data Controller and delete all existing copies unless required by law to continue to store the personal data. This also applies to any back-up copies, where it is enough to overwrite according to established routines for back-up creation.

The Data Controller shall receive written confirmation from the Data Processor that all personal data have been returned or deleted according to the instructions of the Data Controller, and that the Data Processor has not retained copies, printouts or personal data in any other form.

1. Other obligations and rights

Other obligations and rights ensue from the Principal Agreement between the Data Processor and the Data Controller regarding the services that necessitate the processing of personal data, and from this Agreement. The same contact representatives will serve in connection with this Agreement as for the Principal Agreement.

This Agreement shall not expand the Data Controller's right to impose sanctions, including the Data Processor's liability for damages, beyond the rights pursuant to the Principal Agreement.

In the event of conveyance of the Principal Agreement to other parties, this Agreement shall be conveyed correspondingly.

_____. _____ 2017

The Data Processor	**The Data Controller**
_____	_____
NAME	NAME

Two original copies of this Agreement have been prepared, of which each party has received one.

Attachments

The purpose of processing

[The purpose and intention of the processing shall be entered here.]

The duration of the processing

[Enter the length of time the processing shall take. If the processing is according to an agreement between the parties, the duration shall be expressed as "The processing shall last for as long as the Data Processor provides services to the Data Controller pursuant to the Principal Agreement.]

The nature of the processing

[Specify here what the processing consists of, for example, "Storage of personal data" or whatever circumstances necessitate the processing pursuant to the Principal Agreement.]

Types of personal data to be processed

The following types of personal data shall be processed pursuant to the Agreement:

[Enter the type of personal data, e.g. personal name, e-mail address, telephone number, etc. It is not necessary to give details of the information to be processed, only *the type* of information.]

Categories of registered individuals

[Specify categories of registered individuals, such as employees, clients, etc.]

Data processing sub-vendors at the time of entry into the Agreement

[Specify data processing sub-vendors and countries in which the data will be processed]

•