D11.4 DATA MANAGEMENT PLAN V3

Revision: v.1.0

Work package	WP 11
Task	Task 11.3
Due date	31/12/2023
Submission date	21/12/2023
Deliverable lead	Martel
Version	1.0
Authors	Giacomo Inches (Martel) leading, contributions from all WP leaders
Reviewers	DCAL (Kearsy Cormier), Uwe Reichel (AUD)

| Abstract | This document identifies the best practices and specific standards for the generated data, assess their suitability for sharing/reuse in accordance with official EC guidelines. The Data Management Plan is prepared, detailing what data the project generated, whether and how it will be exploited or made accessible for verification and re-use and how it will be curated and preserved – adhering to the FAIR guiding principles as adopted by the EC for research data management.

This is version 3 (final) of the Data Management Plan. |
|---|---|
| Keywords | Data Management Plan, Data Protection, Open Data, Open Research Data Pilot |

WWW.PROJECT-EASIER.EU

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	21/11/2023	Partners contribution requested to update V2	Martel
V0.2	Until 11/12/2023	All partners contributed in a shared repository	All partners
V0.3	15/12/2023	Consolidation and version ready for internal review	Martel
V0.4	19/12/2023	Internal review finalisation	AUD, DCAL
V1.0	21/12/2023	Final check and submission	Martel

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the "Intelligent Automatic Sign Language Translation" (EASIER) project's consortium under EC grant agreement 101016982 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

© 2021 - 2023 EASIER Consortium

Project co-funded by the European Commission in the H2020 Programme			
Nature of the deliverable:		**R**	
Dissemination Level			
PU	Public, fully open, e.g. web		✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC		
CO	Confidential to EASIER project and Commission Services		

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

Funded by the Horizon 2020
Framework Programme of the European Union

OTHER: Software, technical diagram, etc.

EXECUTIVE SUMMARY

This deliverable is the third and last version (V3) of the EASIER Data Management Plan (DMP), defined at the end of the project, i.e., at M36. The DMP is a living document that identifies which data is produced during the EASIER project, who owns that data, how it is documented, how it is preserved, and with whom and under what form it was and will be shared. This is the final release of the DMP, which already had two previous versions, at M12 (V1) and at M24 (V2).

Most of the data produced during the project correspond to documentation, source-code and experimental results. Additional data containing personal information are anonymized to comply with the General Data Protection Regulation (GDPR) and the highest ethical standards.

The purpose of the DMP is to contribute to good data handling by indicating what research data the project is generating and describe which parts of the data can be shared with the public. Moreover, it gives instructions on naming conventions, metadata structure, storing of the research data and how to make public data available.

This plan describes:

- what kind of data are generated, collected, processed, and shared,

- which methodology and standards are applied during data collection and handling,

- the procedures for sharing and providing open access to the EASIER data and that used for the curation and preservation of the data,

- the procedures to comply with the GDPR to ensure the protection of the involved companies' data, information, and privacy rights.

As part of Horizon 2020, the EASIER project participates in the Pilot on Open Research Data (ORDP) a pilot action that aims to improve and maximise access to and reuse of research data. It aims to respect the Findable, Accessible, Interoperable and Reusable (FAIR) guidelines of the H2020 programme, which state that data will be made available as long as possible, but without negatively affecting the commercial advantage of the partners.

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

BSL	British Sign Language
CA	Consortium Agreement
DGS	German Sign Language
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
DSGS	Swiss German Sign Language
EMSL	European Meta Sign Language
FAIR	Findable, Accessible, Interoperable and Reusable guidelines
GDPR	General Data Protection Regulation
IP	Internet Protocol
LIS-CH	Italian Sign Language of Switzerland
LREC	Language Resources and Evaluation (International Conference on)
LSF-CH	French Sign Language of Switzerland
MT	Machine Translation
ORDP	Open Research Data Pilot
SL	Sign Language
TCP	Transmission Control Protocol
Vx	Version x of the document e.g. v1, v2, v3, …
WP	Work Package

1 INTRODUCTION

1.1 ORDP PARTICIPATION

This deliverable D11.3 is of the type Open Research Data Pilot (ORDP), which by definition of the European Commission, enables open access and reuse of research data generated by Horizon 2020 projects. The ORDP is constituted by two main pillars:

    a. the Data Management Plan (DMP) and

    b. an open access to research data.

A project that opts-in to ORDP has to adhere to the following conditions:

- Develop (and keep an up-to-date) DMP,
- Deposit the data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate this data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research.

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that are specified in the DMP.

1.2 DATA MANAGEMENT PLAN

This document constitutes the third and last version of the Data Management Plan required by the ORDP and the nature of the deliverable is to outline how research data were handled during the research project and after it is completed. It includes clear descriptions and rationale for the access regimes that are foreseen for all the collected data sets within the EASIER project.

The DMP was not a fixed document; it evolved and gained more precision and substance during the lifespan of the project. Therefore, it now describes the data management life cycle for all data sets that were collected, processed or generated by the project. It outlines how research data were handled both during the research project and after the project has been completed, describing what data were collected, processed or generated and following what methodology and standards, whether and how this data was and will be shared and/or made open, and how it was curated and will be preserved.
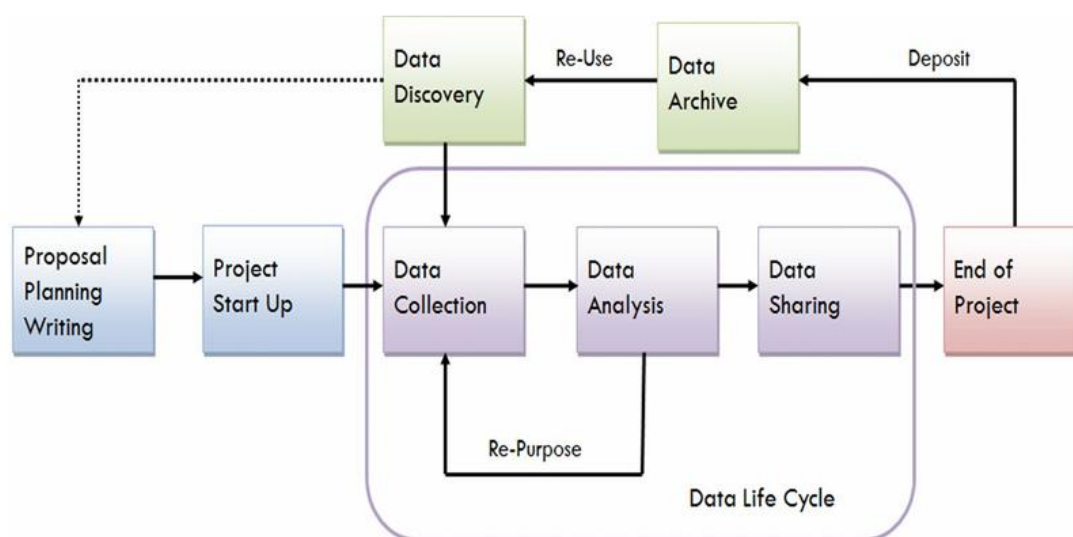
FIGURE 1: DATA MANAGEMENT LIFECYCLE

This first version of the DMP focused on the identification of the different kind of data that each Work Package (WP) would employ, while the second and this third versions of the DMP are complementing that list and enriching the properties each dataset currently specified, as well as their metadata and tools for processing and repository location. Data are therefore organized on a WP level and presented in an exhaustive way in Section 2. We also describe the properties of each collection towards the FAIR (Findable, Accessible, Interoperable, Reusable) principles[1].

1.2.1 Notable changes

With respect to the previous version of the deliverable (V2) submitted at M24, in this third and final version (V3) changes can be found in almost all WPs.

The following datasets were added:

- EASIER_WP1_02_PER_InterimEvaluation

- EASIER_WP1_03_PER_FinalEvaluation

- EASIER_WP4_09_MUL_DGS_translation_test_cases

The following datasets were removed:

- EASIER_WP9_01_PER_ContactList because the list of contacts was not structured into a consolidated dataset and only consists of a few email addresses/exchange.

- EASIER_WP4_08_MUL_training_data_quality_estimation_models because a duplication of the dataset already described in WP5: EASIER_WP5_03_MUL_qualityEstimationTrainingData.

---

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

While descriptions of the following datasets were updated:

- EASIER_WP3_01_SRC_EMSL

- EASIER_WP4_01_MUL_DSGS_broadcast

- EASIER_WP4_02_MUL_LSF-CH_broadcast

- EASIER_WP4_03_MUL_LIS-CH_broadcast

- EASIER_WP4_04_MUL_DGS_broadcast

- EASIER_WP4_05_MUL_BSL_broadcast

- EASIER_WP4_06_MUL_DSGS_broadcast

- EASIER_WP4_07_MUL_DSGS_broadcast

- EASIER_WP4_08_MUL_DGS_broadcast_annotation

- EASIER_WP5_03_MUL_qualityEstimationTrainingData

- EASIER_WP6_01_MUL_LREC-Anthology

- EASIER_WP8_01_MUL_VideoRecording

- EASIER_WP8_02_MUL_AudioRecording

- EASIER_WP8_03_PER_UserDataset

- EASIER_WP8_04_MUL_TextMessages

- EASIER_WP8_05_OTH_InternalTranslationRepository

1.3 PUBLISHING INFRASTRUCTURE FOR OPEN ACCESS

While access to data is always guaranteed between the project partners and regulated by the Consortium Agreement (CA) and Description of Action (DoA) in the Grant Agreement, not all data generated within the project can be publicly released, due to privacy and confidentiality reasons e.g. full transcripts of user interviews in WP1.

The data necessary to successfully complete the project Work Packages (WPs) was shared without any restrictions amongst the WP partners either via internal repositories or direct communication. Public data was made available at the project's website or other repositories, as appropriate e.g. individual partner's repository, public repositories, etc. Users were made aware of this data primarily through research publications, patent applications, dissemination activities, invited talks, social networks and the project website. Data was made available to the project consortium as soon as it was available; to standardization bodies when required; and to the public at the due date of the derivable, and, in case a research publication is based on that, as soon as the paper is submitted.

1.4 UNIFIED DOI STRATEGY

In accordance with the FAIR principles, any output of EASIER should be assigned a persistent identifier. Where the assignment of persistent identifiers is not part of the publication procedures (as is often the case when publishing papers in a journal), the partners involved in the creation of the output took care to obtain the persistent identifier(s) unless there were legal reasons not to do so. Where possible, there will be two identifiers, one for the product in general and one for the specific version to be released. If one of the partners involved in creating the output had internal or by-contract access to a persistent identifier issuer, that partner was asked to request the persistent identifier on behalf of the group of authors from EASIER. Where possible or even necessary, this involved submitting the output to the partner's long-term archive. Otherwise, the responsible partner chose a service such as ZENODO to submit the output to a long-term archive and obtain the persistent identifiers.

When submitting the output to an archive, regulations forbidding the output to also be deposited elsewhere should not be accepted. When there is a choice, partners were asked to prefer DOIs over other persistent identifier schemes.

2   DATA

In this section we illustrate the data generated within EASIER, organised by WP. This division allows for better linking the dataset with their respective part of work. Another important remark to make concerns the nature of the data and their purpose. Some of the datasets illustrated below are research datasets that were included as part of the ORDP (e.g. multimedia files, annotations, source code), while others are data related to the project management or dissemination and therefore not part of the ORDP.

2.1 IDENTIFIERS

Each data[set] is identified with the following format:

EASIER_<WP number>_<serial number of dataset>_<data type>_<dataset name>

The *<WP number>* corresponds to those in the DoA, the *<serial number of dataset>* is a progressive number to uniquely identify the dataset within the same WP while the *<data type>* is mapped in the table below (Table 1). The *<dataset name>* is a descriptive name of the dataset assigned by each WP leader.

Table 1 IDENTIFIER TYPES OF DATA

Acronym	Data Type
SRC	Source Code
MUL	Multimedia content (audio, video, text)
PER	Personal data (email addresses, other personal information)
OTH	Other type of data (to be specified in each table below)

2.2 PROPERTIES

In the following Sections, all the datasets are described with the help of the following properties:

- Identifier
- Dataset Summary
- Intellectual Property Rights
- Findability
- Accessibility
- Interoperability
- Security
- Ethical aspects

➲ Other issues

In the APPENDIX B we include the template shared among the WP leaders to collect the dataset's information where these properties are explained in more details.

2.3 DATASETS

The following table summarises the identified datasets and their scientific or administrative nature. Some of the datasets were only used internally for enabling the technical functionality of the software and not for scientific purposes.

Dataset Identifier	Nature
EASIER_WP1_01_PER_FocusGroups	Scientific
EASIER_WP1_02_PER_InterimEvaluation	Scientific
EASIER_WP1_03_PER_FinalEvaluation	Scientific
EASIER_WP3_01_SRC_EMSL	Scientific
EASIER_WP4_01_MUL_DSGS_broadcast	Scientific
EASIER_WP4_02_MUL_LSF-CH_broadcast	Scientific
EASIER_WP4_03_MUL_LIS-CH_broadcast	Scientific
EASIER_WP4_04_MUL_DGS_broadcast	Scientific
EASIER_WP4_05_MUL_BSL_broadcast	Scientific
EASIER_WP4_06_MUL_DSGS_broadcast	Scientific
EASIER_WP4_07_MUL_DSGS_broadcast	Scientific
EASIER_WP4_08_MUL_DGS_broadcast_annotation	Scientific
EASIER_WP4_09_MUL_DGS_translation_test_cases	Scientific
EASIER_WP5_01_OTH_postEditedData	Scientific
EASIER_WP5_02_MUL_verbalisingDiagramCorpus	Scientific
EASIER_WP5_03_MUL_qualityEstimationTrainingData	Scientific

EASIER_WP6_01_MUL_LREC-Anthology	Scientific
EASIER_WP6_02 MUL_Data_Collection_Tasks	Scientific
EASIER_WP6_03_MUL_Corpora	Scientific
EASIER_WP6_04_MUL_Lexical_Resources	Scientific
EASIER_WP6_05_MUL_Interlingual_Index	Scientific
EASIER_WP6_06 SRC_Data_Harmonization_Tools	Scientific
EASIER_WP7_01_MUL_annotated_video_data	Scientific
EASIER_WP8_01_MUL_VideoRecording	Technical
EASIER_WP8_02_MUL_AudioRecording	Technical
EASIER_WP8_03_PER_UserDataset	Technical
EASIER_WP8_04_MUL_TextMessages	Technical
EASIER_WP8_05_OTH_InternalTranslationRepository	Technical
EASIER_WP9_01_PER_ContactList	Scientific
EASIER_WP10_01_PER_NEWSLETTER	Administrative
EASIER_WP10_02_PER_WEBSITE	Administrative
EASIER_WP11_01_PER_Mattermost	Administrative
EASIER_WP11_02_PER_Mailman	Administrative

2.4 DATA GENERATED IN WP1

In WP1, it was planned to collect crucial information to feed the reflections and the understanding of user needs and practices. Thereafter, the plan was to set up performance metrics (components and overall EASIER system) and develop recruitment strategies for evaluation studies. To achieve WP1's objectives, data were needed.

Table 2 EASIER_WP1_01_PER_FocusGroups

Identifier	EASIER_WP1_01_PER_FocusGroups
Dataset Summary	**Responsible partner:** EUD, INT **Purpose**: Data collection during focus groups to feed T1.1 (Analysis of user needs). This helped to achieve the user centred design of EASIER mobile app and end-user tools. The data collected was used by WP2, WP4, WP5, WP7 and WP8 as the user involvement is the key to achieve EASIER's user acceptance. **Type/format**: Transcripts anonymised **Re-use of existing data**: N/A **Data origin**: From EUD/INT focus groups **Expected size**: N/A **Data utility**: To all EASIER partners who want to assess/understand user needs and practices, user expectation, etc.
Intellectual Property Rights	**Owner:** EUD/INT/EASIER **Licensing:** N/A **Dependency:** Only restriction is the use in the framework of EASIER project only **Constraints:** N/A **Timeframe:** Data only in the lifetime of EASIER project
Findability	**DOI:** N/A **Is data discoverable:** No **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** Only for EASIER partners and used to achieve EASIER's objective as the interviewed users agreed that all the data was anonymized and was used only for the EASIER project. **How it will be accessible:** EASIER repository and extracts of data in D1.1

	Methods/software tools to access data: N/A **Repository**: EASIER depository **Restrictions on access:** N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** No, as long as the extract of transcripts are shared. We did not share full transcripts outside of EASIER Consortium to maximise preservation of anonymization. **Is informed consent for data sharing and long term preservation given?** Yes. Only for EASIER purpose and duration of EASIER project lifetime.
Other issues	N/A

Table 3 EASIER_WP1_02_PER_InterimEvaluation

Identifier	EASIER_WP1_02_PER_InterimEvaluation
Dataset Summary	**Responsible partner:** EUD, DCAL, UZH, UHH, RU, ATHENA, STXT, INT **Purpose:** Data collection during interim evaluation T1.3 (End-user evaluation studies) allow partners to work towards user centred design of EASIER mobile app and end-user tools. The data collected was used by WP2, WP4, and WP8 to help develop the EASIER technologies to follow user feedback. **Type/format:** Anonymised transcripts, anonymised rating data **Re-use of existing data:** N/A **Data origin:** From interim evaluations conducted by DCAL, UZH, UHH, RU, ATHENA, STXT, INT **Expected size:** N/A **Data utility:** To all EASIER partners who want to assess/understand user feedback on the technical components evaluated, particularly WP2, WP4 and WP8.

Intellectual Property Rights	**Owner:** EUD/DCAL/UZH/UHH/RU/ATHENA/STXT/INT/EASIER **Licensing:** N/A **Dependency:** Only restriction is the use in the framework of EASIER project only **Constraints:** N/A **Timeframe:** Data only in the lifetime of EASIER project
Findability	**DOI:** N/A **Is data discoverable:** No **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** Only for EASIER partners and used to achieve EASIER's objective as the interviewed users agreed that all the data was anonymized and was used only for the EASIER project. **How it will be accessible:** EASIER repository and extracts of data in D1.3 **Methods/software tools to access data:** N/A **Repository:** EASIER depository **Restrictions on access:** N/A
Interoperability	**Interoperability:** N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** No as long as the extract of transcripts are shared. We did not share full transcripts outside of EASIER Consortium to maximise preservation of anonymization. **Is informed consent for data sharing and long term preservation given?** Yes. Only for EASIER purpose and duration of EASIER project lifetime.
Other issues	N/A

Table 4 EASIER_WP1_03_PER_FinalEvaluation

Identifier	EASIER_WP1_03_PER_FinalEvaluation
Dataset Summary	**Responsible partner:** EUD, DCAL, UZH, UHH, ATHENA, STXT, INT **Purpose:** Data collection during final evaluation T1.3 (End-user evaluation studies) allow partners to work towards user centred design of EASIER mobile app and end-user tools. The data collected was used by WP2, WP4, and WP8 to help develop the EASIER technologies to follow user feedback **Type/format:** Anonymised transcripts, anonymised rating data **Re-use of existing data:** N/A **Data origin:** From interim evaluations conducted by DCAL, UZH, UHH, ATHENA, STXT, INT **Expected size:** N/A **Data utility:** To all EASIER partners who want to assess/understand user feedback on the technical components evaluated, particularly WP2, WP4 and WP8.
Intellectual Property Rights	**Owner:** EUD/EASIER **Licensing:** N/A **Dependency:** Only restriction is the use in the framework of EASIER project only **Constraints:** N/A **Timeframe:** Data only in the lifetime of EASIER project
Findability	**DOI:** N/A **Is data discoverable:** No **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A

Accessibility	**Data openly accessible:** Only for EASIER partners and used to achieve EASIER's objective as the interviewed users agreed that all the data was anonymized and was used only for the EASIER project. **How it will be accessible:** EASIER repository and extracts of data in D1.4 **Methods/software tools to access data:** N/A **Repository:** EASIER depository **Restrictions on access:** N/A
Interoperability	**Interoperability:** N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** No as long as only the extract of transcripts are shared. We did not share full transcripts outside of EASIER Consortium to maximise preservation of anonymization. **Is informed consent for data sharing and long term preservation given?** Yes. Only for EASIER purpose and duration of EASIER project lifetime.
Other issues	N/A

2.5 DATA GENERATED IN WP2

WP2 does not make direct use of human signers' video resources. It receives coded input from WP3 which by no means reveals the identity of persons in videos. Similarly, WP2 work heavily depends on input from WP4 which involves the coded output of MT processes when translation direction is from spoken-to-sign language. It needs to be mentioned that SL representation in WP2 does not face GDPR issues, since it does not involve any human data, but focuses on developing the various aspects of the project signing avatar.

2.6 DATA GENERATED IN WP3

Table 5 EASIER_WP3_01_SRC_EMSL

Identifier	**EASIER_WP3_01_SRC_EMSL**
Dataset Summary	**Responsible partner:** UNIS

	Purpose: All data collected by WP4 was curated to generated poses, cropped/segmented videos, and harmonized subtitles. Features were also extracted using various European Meta Sign Language (EMSL) representations and passed back to WP4 for translation. **Type/format**: The videos in MPEG-4, subtitles are SRT format. EMSL is stored as python/pkl and csv format. **Re-use of existing data**: N/A **Data origin**: WP4 **Expected size**: 9.5TB **Data utility**: Data used primary by WP4.
Intellectual Property Rights	**Owner: Various see WP4 for original sources of data** **Licensing: No** **Dependency: N/A** **Constraints: Internal project use only.**. **Timeframe: N/A**
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** No **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access: N/A**
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Stored at University of Surrey's secure servers.

Ethical aspects	**Possible ethical and legal aspects preventing sharing: Yes** **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	None

2.7 DATA GENERATED IN WP4

Table 6 EASIER_WP4_01_MUL_DSGS_broadcast

Identifier	**EASIER_WP4_01_MUL_DSGS_broadcast**
Dataset Summary	**Responsible partner:** UZH **Purpose**: These are pairings of Swiss German Sign Language videos and corresponding German subtitles from the Swiss Broadcasting Corporation. They are used to train the spoken-to-sign/sign-to-spoken translation models in the project. From this data, WP3 has generated poses, cropped/segmented videos, harmonized subtitles, and European Meta Sign Language (EMSL) representations. **Type/format:** The videos are available in MPEG-TS, the subtitles in SRT format. WP3 has harmonized the videos to MPEG-4 and the subtitles to SRT formats. **Re-use of existing data**: N/A **Data origin**: Interpretations of the German daily news in Switzerland into Swiss German Sign Language **Expected size**: approx. 4TB **Data utility**: The data will primarily be used by WP3 and WP4
Intellectual Property Rights	**Owner:** Swiss Broadcasting Corporation **Licensing:** The data cannot be distributed. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A

	Search keywords: N/A **Versioning:** N/A **Metadata creation:** Internally, metadata is specified in JSON format, storing information on file ID, TV channel, series, episode, date, time, description, duration, size, url, availability of subtitles, video width, video height, video framerate, video duration, biological sex of speakers, and biological sex of signers.
Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. However, we have obtained legal clearance to distribute a subset of the data (see below). **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**: Data cannot be exchanged between partners in the project directly. However, the code for downloading the data is shared among partners.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team will access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** We do not have legal clearance to share the data. **Is informed consent for data sharing and long term preservation given?** no
Other issues	N/A

Table 7 EASIER_WP4_02_MUL_LSF-CH_broadcast

Identifier	EASIER_WP4_02_MUL_LSF-CH_broadcast
Dataset Summary	**Responsible partner:** UZH **Purpose**: These are pairings of French Sign Language of Switzerland videos and corresponding French subtitles from the Swiss Broadcasting

	Corporation. They are used to train the spoken-to-sign/sign-to-spoken translation models in the project. From this data, WP3 has generated poses, cropped/segmented videos, harmonised subtitles, and European Meta Sign Language (EMSL) representations. **Type/format**: The videos are available in MPEG-TS, the subtitles in SRT format. WP3 has harmonised the videos to MPEG-4 and the subtitles to SRT formats. **Re-use of existing data**: N/A **Data origin**: Interpretations of the French daily news in Switzerland into French Sign Language of Switzerland **Expected size**: approx. 3TB **Data utility**: The data was primarily used in WP3 and WP4.
Intellectual Property Rights	**Owner:** Swiss Broadcasting Corporation **Licensing:** The data cannot be distributed. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Internally, metadata is specified in JSON format, storing information on file ID, TV channel, series, episode, date, time, description, duration, size, url, availability of subtitles, video width, video height, video framerate, video duration, biological sex of speakers, and biological sex of signers.
Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A

Interoperability	**Interoperability**: Data cannot be exchanged between partners in the project directly. However, the code for downloading the data is shared among partners.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team will access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** We do not have legal clearance to share the data. **Is informed consent for data sharing and long-term preservation given?** no
Other issues	N/A

Table 8 EASIER_WP4_03_MUL_LIS-CH_broadcast

Identifier	**EASIER_WP4_03_MUL_LIS-CH_broadcast**
Dataset Summary	**Responsible partner:** UZH **Purpose**: These are pairings of Italian Sign Language of Switzerland videos and corresponding Italian subtitles from the Swiss Broadcasting Corporation. They are used to train the spoken-to-sign/sign-to-spoken translation models in the project. From this data, WP3 has generated poses, cropped/segmented videos, harmonized subtitles, and European Meta Sign Language (EMSL) representations. **Type/format**: The videos are available in MPEG-TS, the subtitles in SRT format. WP3 has harmonized the videos to MPEG-4 and the subtitles to SRT formats. **Re-use of existing data**: N/A **Data origin**: Interpretations of the Italian daily news in Switzerland into Italian Sign Language of Switzerland **Expected size**: approx. 2TB **Data utility**: The data was primarily used in WP3 and WP4.
Intellectual Property Rights	**Owner:** Swiss Broadcasting Corporation **Licensing:** The data cannot be distributed.

	Dependency: N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Internally, metadata is specified in JSON format, storing information on file ID, TV channel, series, episode, date, time, description, duration, size, url, availability of subtitles, video width, video height, video framerate, video duration, biological sex of speakers, and biological sex of signers.
Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**: Data cannot be exchanged between partners in the project directly. However, the code for downloading the data is shared among partners.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team will access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** We do not have legal clearance to share the data. **Is informed consent for data sharing and long-term preservation given?** no
Other issues	N/A

Table 9 EASIER_WP4_04_MUL_DGS_broadcast

Identifier	EASIER_WP4_04_MUL_DGS_broadcast
Dataset Summary	**Responsible partner:** UHH **Purpose**: These are pairings of German Sign Language videos and corresponding German subtitles from German TV stations. They are used to train the spoken-to-sign/sign-to-spoken translation models in the project. From this data, WP3 has generated poses, cropped/segmented videos, harmonized subtitles, and European Meta Sign Language (EMSL) representations. **Type/format**: The videos are available in MPEG-4, the subtitles in SRT, VTT, or TTML format. WP3 has harmonized the videos and subtitles to unique formats. **Re-use of existing data**: N/A **Data origin**: Interpretations of German shows into German Sign Language or original German Sign Language shows on German TV stations **Expected size**: approx. 4TB **Data utility**: The data was primarily be used in WP3 and WP4.
Intellectual Property Rights	**Owner:** Various broadcasters. **Licensing:** The data cannot be distributed. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Internally, metadata is specified in JSON format, storing information on file ID, TV channel, series, episode, date, time, description, duration, size, url, availability of subtitles, video width, video height, video framerate, video duration, biological sex of speakers, and biological sex of signers.

Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**: Data cannot be exchanged between partners in the project directly. However, the code for downloading the data is shared among partners.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** We do not have legal clearance to share the data. **Is informed consent for data sharing and long-term preservation given?** no
Other issues	N/A

Table 10 EASIER_WP4_05_MUL_BSL_broadcast

Identifier	**EASIER_WP4_05_MUL_BSL_broadcast**
Dataset Summary	**Responsible partner:** UNIS **Purpose**: These are pairings of British Sign Language videos (with poses extracted) and corresponding English subtitles from the BBC. They are part of the BOBSL dataset (https://www.robots.ox.ac.uk/~vgg/data/bobsl/). From this data, WP3 has generated cropped/segmented videos and European Meta Sign Language (EMSL) representations. **Type/format**: The videos are available in MPEG-4, the subtitles in SRT and TTML format. WP3 has harmonized the videos and subtitles to unique formats. **Re-use of existing data**: N/A

	Data origin: Interpretations of British shows into British Sign Language or original British Sign Language shows on the BBC **Expected size**: approx. 4TB **Data utility**: The data will primarily be used by WP3 and WP4.
Intellectual Property Rights	**Owner:** BBC **Licensing:** https://www.bbc.co.uk/rd/projects/extol-dataset **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Information on the biological sex of speakers and signers was added as part of T6.3.
Accessibility	**Data openly accessible:** Upon signing a form. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**: The data has been downloaded for research purposes by the relevant partners.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long term preservation given?** N/A

Other issues	N/A

Table 11 EASIER_WP4_06_MUL_DSGS_broadcast

Identifier	**EASIER_WP4_06_MUL_DSGS_broadcast**
Dataset Summary	**Responsible partner:** UZH **Purpose**: These are pairings of Swiss German Sign Language videos and corresponding German subtitles from the Swiss Broadcasting Corporation. They represent a subset of EASIER_WP4_01_MUL_DSGS_broadcast made available to the public. **Type/format**: The videos are available in MPEG-4, the subtitles in SRT format. **Re-use of existing data**: This is an extension of the dataset released as part of the Content4All EU project. **Data origin**: Interpretations of the German daily news in Switzerland into Swiss German Sign Language **Expected size**: 12 GB **Data utility**: The data is made publicly available for reproducibility purposes.
Intellectual Property Rights	**Owner:** Swiss Broadcasting Corporation **Licensing:** Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** 10.5281/zenodo.6631275 (poses), 10.5281/zenodo.6637392 (videos, subtitles) **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Internally, metadata is specified in JSON format, storing information on file ID, TV channel, series, episode, date, time, description, duration, size, url, availability of subtitles, video width, video

	height, video framerate, video duration, biological sex of speakers, and biological sex of signers.
Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. However, we have obtained legal clearance to distribute a subset of the data (see below). **How it will be accessible:** videos, subtitles: https://zenodo.org/record/6635710#.Yqaxk5DP3Pg poses: https://zenodo.org/record/6631275#.YqavtJDP3Pg **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**:
Security	**Security measures:**
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** **Is informed consent for data sharing and long term preservation given?** yes
Other issues	N/A

Table 12 EASIER_WP4_07_MUL_DSGS_broadcast

Identifier	EASIER_WP4_07_MUL_DSGS_broadcast
Dataset Summary	**Responsible partner:** UZH **Purpose**: These are pairings of Swiss German Sign Language videos and corresponding German subtitles from the former Web TV channel FocusFive. **Type/format:** The videos are available in MPEG-4, the subtitles in SRT format. **Re-use of existing data**: N/A **Data origin**: signing from teleprompter in Swiss German Sign Language **Expected size**: 20 GB

	Data utility: The data was primarily used in WPs 3 and 4.
Intellectual Property Rights	**Owner:** FocusFilm **Licensing:** The data is publicly distributed for the purposes of a machine translation shared task organized in the context of EASIER. Other uses of the data require the permission of the data owner. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** 10.5281/zenodo.6631159 **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:**
Accessibility	**Data openly accessible:** https://zenodo.org/record/6631159#.YqavrZDP3Pg (only for use in the shared task) **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**:
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** **Is informed consent for data sharing and long term preservation given?** no
Other issues	N/A

Table 13 EASIER_WP4_08_MUL_DGS_broadcast_annotation

Identifier	EASIER_WP4_08_MUL_DGS_broadcast_annotation
Dataset Summary	**Responsible partner:** UHH **Purpose**: In order to compare linguistic annotation with data resulting from subtitles, three shows from EASIER_WP4_04 (spoken with DGS sign-over) have undergone linguistic annotation following the rules of the Public DGS Corpus used in other parts of the project. **Type/format**: The annotation is made available in ELAN and iLex data formats. **Re-use of existing data**: N/A **Data origin**: Stand-off annotation to videos collected in EASIER_WP4_04 **Size**: < 1MB **Data utility**: The data by WP4.
Intellectual Property Rights	**Owner:** The stand-off annotation was done within EASIER and as such is owned by EASIER. However, the videos referenced are owned by ARD (German public broadcaster) and Bundeskanzleramt (Federal Chancellery of Germany). **Licensing:** The data cannot be distributed. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A.
Accessibility	**Data openly accessible:** Copyright restrictions make publishing the data impossible. **How it will be accessible:** N/A

	Methods/software tools to access data: N/A Repository: N/A Restrictions on access: N/A
Interoperability	Interoperability: Data can be exchanged between partners in the project directly as it is stand-off data.
Security	
Ethical aspects	Possible ethical and legal aspects preventing sharing: The annotation is stand-off but can only be interpreted by having access to the original video data. We do not have legal clearance to share the video data. Is informed consent for data sharing and long-term preservation given? no
Other issues	N/A

Table 14 EASIER_WP4_09_MUL_DGS_translation_test_cases

Identifier	EASIER_WP4_09_MUL_DGS_translation_test_cases
Dataset Summary	Responsible partner: UHH Purpose: In order to test performance of the EASIER translation pipeline from signed to spoken language, DGS videos have been recorded both with neutral facial expression as well as with some (staged) affect. This allows this dataset to be used in the data processing pipeline (WP8) to test WP4 and WP7 components. Type/format: The videos have been made available in mp4 format together with a spreadsheet specifying the context. Re-use of existing data: N/A Data origin: Recorded by EASIER staff. Size: < 1MB Data utility: The data by WP4.
Intellectual Property Rights	Owner: EASIER Licensing: The dataset is meant as project-internal test data, so no clearance for distribution was negotiated with the team member filmed.

	Dependency: N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** https://doi.org/10.25592/uhhfdm.13851 **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A.
Accessibility	**Data openly accessible:** No. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability**: Data can be exchanged between partners in the project directly as it is stand-off data.
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The annotation is stand-off but can only be interpreted by having access to the original video data. We do not have legal clearance to share the video data. **Is informed consent for data sharing and long-term preservation given?** no
Other issues	N/A

2.8 DATA GENERATED IN WP5

Table 15 EASIER_WP5_01_OTH_postEditedData

Identifier	EASIER_WP5_01_OTH_postEditedData
Dataset Summary	**Responsible partner:** STXT **Purpose**: Train models in T4.5 from human post-edited translations. **Type/format**: N/A **Signing videos**: mp4, Container format provided by MPEG for MPEG-4 content and standardised in ISO/IEC 14496-12 and -14 **Text**: SRT, Segmented timecoded Text file format. **Re-use of existing data**: No. **Data origin**: Human input collected in the WP tasks. **Expected size**: 5,000 sentences + 5,000 signed entries **Data utility**: WP4 and constitutes D5.5

Intellectual Property Rights	**Owner:**
	
	    ○ **SILAS: rights owned by STXT and the interpreters**
	
	    ○ **Edit: performed by WP5 participants.**
	
	Licensing: N/A
	
	Dependency: original Data of signing and editing (video & text) owned by individual interpreters
	
	Constraints: copyright issues between EU and CH
	
	Timeframe: N/A

Findability	**DOI:** No (confidential) **Is data discoverable:** No (confidential) **Search keywords:** No (confidential) **Versioning:** Two versions, respectively constituting D5.5 and D5.6. **Metadata creation:** N/A
Accessibility	**Data openly accessible:** No (confidential) **How it will be accessible:** N/A

	Methods/software tools to access data: N/A
	Repository: N/A
	Restrictions on access: N/A
Interoperability	**Interoperability**:
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A
	Is informed consent for data sharing and long-term preservation given? N/A
Other issues	N/A

Table 16 EASIER_WP5_02_MUL_verbalisingDiagramCorpus

Identifier	EASIER_WP5_02_MUL_verbalisingDiagramCorpus
Dataset Summary	**Responsible partner:** CNRS **Purpose**: Study spontaneous productions of signers representing their language in a graphical form in T5.3 **Type/format**: Images (PNG or multi-page PDF) + video (MPEG4) **Re-use of existing data**: This is an existing data set, with an option to extend it slightly in WP5 **Data origin**: Previously elicited corpus, as per Michael Filhol, "Elicitation and corpus of spontaneous Sign Language discourse representation diagrams", in *Proceedings of the 9th workshop on the Representation and Processing of Sign Languages*, May 2020. **Expected size**: ~500 GB **Data utility**: T5.3, to specify the format edited in the implemented editor
Intellectual Property Rights	**Owner:** CNRS **Licensing:** N/A **Dependency:** N/A **Constraints:** Data still undisclosed due to option to continue extending it.

	Timeframe: When data complete and after a couple of years of exclusive use, data will be deposited on a research corpus data platform (except for the videos of informants who chose against it).
Findability	**DOI:** Yes. **Is data discoverable:** Yes. **Search keywords:** Verbalising diagrams, graphical SL representation. **Versioning:** No. **Metadata creation:** No standard. A simple field-value list is created to describe the data.
Accessibility	**Data openly accessible:** Yes. **How it will be accessible:** Research data repository (e.g. OrtoLang) **Methods/software tools to access data:** PDF/image viewer and video player. **Repository**: OrtoLang (tbc) **Restrictions on access:** N/A
Interoperability	**Interoperability**: Yes (standard image/video formats)
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** The video content shows people's faces, while some informants might choose not to be displayed online. **Is informed consent for data sharing and long-term preservation given?** Yes.
Other issues	

Table 17 EASIER_WP5_03_MUL_qualityEstimationTrainingData

Identifier	**EASIER_WP5_03_MUL_qualityEstimationTrainingData**
Dataset Summary	**Responsible partner:** UZH **Purpose**: This is the data used to train sentence-based automatic machine translation quality estimation (QE) models for German and German Sign Language.

	Type/format: The data is represented as json with a list of features: system or reference spoken language text, EMSL, video, human quality scores, and several automatically derived features needed to train the quality estimation models. **Re-use of existing data**: N/A **Data origin**: The data consists of pairs of spoken language sentences and their translations in Sign Language, represented as EMSL (as generated in WP3) for the signed-to-spoken direction and pose sequences for the spoken-to-signed direction. **Expected size**: The data comprises 18,500 samples in the language pairs that were considered in the final evaluation (DGS-de, LIS-it, LSF-fr, DSGS-de, BSL-en), in both directions (a total of 10 directions). **Data utility**: The data was used by the partners in WPs 4 and 5.
Intellectual Property Rights	**Owner:** The data is curated by UZH. **Licensing:** The data cannot be distributed. **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** The source sentences involved in the dataset are from a source that has not been legally cleared for release. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: N/A **Restrictions on access:** N/A
Interoperability	**Interoperability:** N/A
Security	**Security measures:** The data is stored on secured file servers. Access to the data within each partner institution is granted to members of the

	project team only and is managed through an identity management system. Any person accessing the data is a registered and verified user in the identity management system. Members of the project team access data on the servers via secured networks.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** We do not have legal clearance to share the data. **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	N/A

2.9 DATA GENERATED IN WP6

WP06 focuses on the harmonization of existing resources, i.e. there is no plan to do (field) data collections within this work package. Instead, WP6 aggregates existing data in novel ways and, where necessary, close gaps. Filling gaps on linguistic detail is achieved by the language intuition of project team members of external consultants.

Table 18 EASIER_WP6_01_MUL_LREC-Anthology

Identifier	EASIER_WP6_01_MUL_LREC-Anthology
Dataset Summary	**Responsible partner:** UHH **Purpose**: The sign-lang@LREC anthology contains all papers published at the LREC workshop on sign languages which is the event where most papers on linguistic resources for sign languages are presented. It also includes select papers from the LREC main conference that address sign languages. The resource indexes the papers not only by conferences and authors, but also by the datasets and tools introduced/used, sign languages covered, and research projects that the papers originate from. **Type/format**: Website **Re-use of existing data**: This resource enriches a collection of existing papers with task-specific metadata that are also of general interest for the sign language data community. **Data origin**: LREC workshop papers **Size**: 423 papers, 2.4 GB **Data utility**: Within EASIER, the data are used as the basis for the overviews of linguistic corpora and lexical resources on European sign languages (EASIER_WP6_03 and EASIER_WP6_04). Beyond the

	project, the resource can be expected to become the major point of orientation for any researcher working with sign language resources.
Intellectual Property Rights	**Owner:** The original papers are copyrighted by ELRA and the authors, the additionally created metadata will be in the public domain. **Licensing:** The papers are licensed under a CC BY-NC license. The site itself is free to use for everyone under CC0 license. **Dependency:** As the website is non-commercial, re-use of papers is permitted by ELDA, the organizing body of the LREC conferences. **Constraints:** None **Timeframe:** N/A
Findability	**DOI:** The resource as a whole has a DOI: https://doi.org/10.25592/dgs.lrec. It is currently under discussion if individual papers in the anthology will receive DOIs. **Is data discoverable:** The website uses established metadata standards to ensure indexing on both general-purpose search engines as well as academic search engines (e.g. Google Scholar). **Search keywords:** The resource as a whole is advertised with appropriate keywording, the individual entries show keywords assigned by the project as part of this data collection. **Versioning:** The resource will be updated for corrections and when new events in the workshop series occur. **Metadata creation:** The content of the resource as bibliographic data is available in bibtex format and via semantic web metadata standards (Highwire Press, Dublin Core, Eprints, Open Graph).
Accessibility	**Data openly accessible:** Data produced in the project will be made openly available. **How it will be accessible:** The data is made available as a website. **Methods/software tools to access data:** N/A **Repository**: Parts of the data reside in a public research data repository, the website providing application-specific access mechanism resides outside the repository. **Restrictions on access:** N/A
Interoperability	**Interoperability**: Data are available in bibtex and semantic web formats (see above).
Security	**Security measures:** N/A

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	none

Table 19 EASIER_WP6_02_MUL_Data_Collection_Tasks

Identifier	**EASIER_WP6_02_MUL_Data_Collection_Tasks**
Dataset Summary	**Responsible partner:** UHH **Purpose**: This resource lists data collection tasks that have been used in known sign language corpus projects to serve as a basis for cross-referencing resources in order to identify shared tasks. **Type/format**: Website **Re-use of existing data**: This resource identifies data collection tasks that have been used in more than one sign language corpus to allow the user to construct meta-collection with specific types of data. **Data origin**: This resource builds on excerpts from dataset EASIER_WP6_01 as well as personal communication with data authors, so it is primarily a bibliographic, i.e. a metadata resource. **Size**: <10 MB **Data utility**: Within EASIER, the data are used as the basis for the overviews of linguistic corpora on European sign languages (EASIER_WP6_03). Beyond the project, the resource can be expected to become a major point of orientation for any researcher interested in cross-linguistic sign language data.
Intellectual Property Rights	**Owner:** EASIER (& other projects at UHH) **Licensing:** The is free to use for everyone **Dependency:** As described, the resource refers to data collections and scientific papers **Constraints:** None. **Timeframe:** N/A
Findability	**DOI:** Website: https://doi.org/10.25592/dgs.sldc-t. The data are also available in PDF format as a chapter in https://doi.org/10.25592/uhhfdm.12016. In addition, a PDF focussing on

	European sign languages is available as EASIER deliverable D6.1 under the DOI https://doi.org/10.25592/uhhfdm.9560. **Is data discoverable:** The website is public. Semantic web standards are used to optimize search engine indexing. **Search keywords:** The resource as a whole is advertised with appropriate keywording. **Versioning:** The first version focussed on European sign languages, while the subsequent versions cover resources worldwide. **Metadata creation:** The resource consists of metadata only.
Accessibility	**Data openly accessible:** Data produced in the project is made openly available. **How it will be accessible:** The data is made available as a website as well as in PDF format. **Methods/software tools to access data:** N/A **Repository**: Parts of the data reside in a public research data repository (cf. DOIs containing "uhhfdm"), the website providing application-specific access mechanism resides outside the repository. **Restrictions on access:** N/A
Interoperability	**Interoperability**: Existing metadata standards will be used where feasible.
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	none

Table 20 EASIER_WP6_03_MUL_Corpora

Identifier	**EASIER_WP6_03_MUL_Corpora**
Dataset Summary	**Responsible partner:** UHH **Purpose**: This resource lists linguistic sign language corpora of European sign languages.

	Type/format: Website **Re-use of existing data**: Metadata referring to the original data when useful. **Data origin**: This resource builds on excerpts from dataset EASIER_WP6_01 as well as personal communication with data authors, so it is primarily a bibliographic, i.e. metadata resource. **Size**: <10 MB **Data utility**: Within EASIER, this resource lists the scope of data harmonization candidates together with the most important features of these corpora from the perspective of making these data high-quality training data for the EASIER machine translation pipeline. Beyond the project, the resource is most useful for researchers looking for existing sign language corpora.
Intellectual Property Rights	**Owner:** EASIER (& other projects at UHH) **Licensing:** The resource is free to use for everyone **Dependency:** As described, the resource provides specific metadata for linguistic sign language corpora and refers to these corpora to view the data. It builds on data from dataset EASIER_WP6_01 and uses dataset EASIER_WP6_02. **Constraints:** None. **Timeframe:** N/A
Findability	Website: https://doi.org/10.25592/dgs.sldc-c. The data are also available in PDF format as a chapter in https://doi.org/10.25592/uhhfdm.12016. In addition, a PDF focussing on European sign languages is available as EASIER deliverable D6.1 under the DOI https://doi.org/10.25592/uhhfdm.9560. **Is data discoverable:** The website is public. Semantic web standards are used to optimize search engine indexing. **Search keywords:** The resource as a whole is advertised with appropriate keywording. **Versioning:** The first version focussed on European sign languages, while the subsequent versions cover resources worldwide. **Metadata creation:** The resource consists of metadata only.
Accessibility	**Data openly accessible:** Data produced in the project is made openly available. **How it will be accessible:** The data is made available as a website as well as in PDF format.

	Methods/software tools to access data: N/A **Repository**: Parts of the data reside in a public research data repository (cf. DOIs containing "uhhfdm"), the website providing application-specific access mechanism resides outside the repository. **Restrictions on access:** N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	none

Table 21 EASIER_WP6_04_MUL_Lexical_Resources

Identifier	EASIER_WP6_04_MUL_Lexical_Resources
Dataset Summary	**Responsible partner:** UHH **Purpose**: This resource lists machine-readable lexical resources for European sign languages. **Type/format**: Website **Re-use of existing data**: Metadata referring to the original data when useful. **Data origin**: This resource builds on excerpts from dataset EASIER_WP6_01 as well as personal communication with data authors, so it is primarily a bibliographic, i.e. metadata resource. **Size**: <10 MB **Data utility**: Within EASIER, this resource lists the scope of candidate data for inclusion in the multi-lingual lexical resource to be built in EASIER.
Intellectual Property Rights	**Owner:** EASIER (& other projects at UHH) **Licensing:** The resource is free to use for everyone

	Dependency: As described, the resource provides specific metadata for sign language lexical resources and refers to these resources to view the data. **Constraints:** None. **Timeframe:** N/A
Findability	**DOI:** Website: https://doi.org/10.25592/dgs.sldc-l. The data are also available in PDF format as a chapter in https://doi.org/10.25592/uhhfdm.12016. In addition, a PDF focussing on European sign languages is available as EASIER deliverable D6.1 under the DOI https://doi.org/10.25592/uhhfdm.9560. **Is data discoverable:** The website is public. Semantic web standards are used to optimize search engine indexing. **Search keywords:** The resource as a whole is advertised with appropriate keywording. **Versioning:** The first version focussed on European sign languages, while the subsequent versions cover resources worldwide. **Metadata creation:** The resource consists of metadata only.
Accessibility	**Data openly accessible:** Data produced in the project is made openly available. **How it will be accessible:** The data is made available as a website as well as in PDF format. **Methods/software tools to access data:** N/A **Repository**: Parts of the data reside in a public research data repository (cf. DOIs containing "uhhfdm"), the website providing application-specific access mechanism resides outside the repository. **Restrictions on access:** N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	none

Table 22 EASIER_WP6_05_MUL_Interlingual_Index

Identifier	EASIER_WP6_05_MUL_Interlingual_Index
Dataset Summary	**Responsible partner:** UHH **Purpose**: This interlingual index binds available lexical descriptions for European sign languages to an existing multi-lingual wordnet. **Type/format**: Website and dataset **Re-use of existing data**: Builds on existing wordnet resources for spoken languages and lexical resources of the included sign languages. **Data origin**: The resources uses information from the wordnet resources Open Multilingual Wordnet and Germanet and from the sign language lexical resources DGS Corpus, DW-DGS (Digital Dictionary of German Sign Language), Polytropon, Noema+, BSL Signbank, Global Signbank, Swiss-German Sign Language Lexicon, DictaSign, Swedish Sign Language Dictionary and Corpus-based Dictionary of Polish Sign Language. **Size**: <1 MB **Data utility**: Within EASIER, this resource was used in the multilingual automatic translation pipeline. Beyond the project, this resource will be of interest for computational sign linguists.
Intellectual Property Rights	**Owner:** EASIER **Licensing:** The public version is free to use for everyone. **Dependency:** The resource provides specific metadata for sign language lexical resources, links to these resources and integrates (video) data of theirs with permission. **The resource uses data from Open Multilingual Wordnet, which is under open licence.** **Constraints:** Non-public data from some lexical resources is excluded from the public release. **Timeframe:** N/A
Findability	**DOI:** https://doi.org/10.25592/dgs.wn for the website and https://doi.org/10.25592/uhhfdm.10168 for the dataset. A detailed discussion of the latest release within the EASIER lifetime (EASIER deliverable D6.5) is available at https://doi.org/10.25592/uhhfdm.13848. **Is data discoverable:** The website is public. Semantic web standards are used to optimize search engine indexing. **Search keywords:** The resource as a whole is advertised with appropriate keywording.

	Versioning: Several versions have been made available through the lifetime of the project, and we expect to produce updates in the future. **Metadata creation:** The resource links existing resources, so is kind of metadata. Resource-specific (Datacite) metadata have been provided when submitting to the data repository.
Accessibility	**Data openly accessible:** Data produced in the project are made openly available as far as possible. **How it will be accessible:** The data is available from the research data repository, cross-linked from the EASIER project webpages. **Methods/software tools to access data:** The data is available as a website as well as in csv files indexing the original sources. **Repository**: The data is deposited in a public research data repository. **Restrictions on access:** N/A
Interoperability	**Interoperability**: CSV data easily usable in any linguistic processing system
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A
Other issues	none

Table 23 EASIER_WP6_06_SRC_Data_Harmonization_Tools

Identifier	**EASIER_WP6_06_SRC_Data_Harmonization_Tools**
Dataset Summary	**Responsible partner:** UHH **Purpose**: Set of tools to allow the integration of language data (corpora, lexical resources) into the EASIER translation pipeline and EASIER_WP6_05. **Type/format**: Software **Re-use of existing data**: Builds on information from EASIER D6.2. Takes public data from existing corpora as input to convert them into a harmonized format.

	Data origin: N/A **Expected size**: <1 MB **Data utility**: EASIER used these tools to integrate the language resources chosen from EASIER_WP6_03 and EASIER_WP6_04. The tools will also be used by external parties to prepare the future integration of additional languages by external parties.
Intellectual Property Rights	**Owner:** EASIER. **Licensing:** CC BY 4.0 **Dependency:** None **Constraints:** No. **Timeframe:** N/A
Findability	**DOI:** https://doi.org/10.25592/uhhfdm.11756. **Is data discoverable:** Code is hosted in a repository for open source code and archived in a research data repository. **Search keywords:** Repositories are given appropriate keywords. **Versioning:** Semantic versioning is used. **Metadata creation:** Metadata is specified as part of code archiving.
Accessibility	**Data openly accessible:** The tools are open source. **How it will be accessible:** Open source code repository. **Methods/software tools to access data:** N/A **Repository**: Github, Hamburg University Research Data Repositoriy **Restrictions on access:** None
Interoperability	**Interoperability**: The code is open-source and produces data files that follow open access standards. Data input supports community-standard file formats.
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A **Is informed consent for data sharing and long-term preservation given?** N/A

Other issues	none

2.10 DATA GENERATED IN WP7

Table 24 EASIER_WP7_01_MUL_annotated_video_data

Identifier	EASIER_WP7_01_MUL_annotated_video_data
Dataset Summary	**Responsible partner:** [DFKI] **Purpose**: These are the datasets used to train the vision-based affect recognition models for German and German Sign Language. **Type/format**: The data consists of video recordings (MPEG-4) with corresponding textual annotations. **Re-use of existing data**: N/A **Data origin**: The data is based on the Content4All dataset released as part of the corresponding EU project. **Expected size**: 36GB **Data utility**: The data served to train the machine-learning-based affect recognition models that are the target of WP7.
Intellectual Property Rights	**Owner:** N/A **Licensing:** N/A **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** N/A **How it will be accessible:** N/A

	Methods/software tools to access data: N/A
	Repository: N/A
	Restrictions on access: N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** N/A
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** N/A
	Is informed consent for data sharing and long term preservation given? N/A
Other issues	N/A

2.11   DATA GENERATED IN WP8

WP8 "Integration: app development" deals with technological implementation of the EASIER translation app and acts as middleware to translation related technical EASIER components developed by WP2, WP3, WP4 and, WP7.

User data is generated and stored after registration in the APP with opt in for processing translations only. Scientific usage was discussed and dismissed, this was decided, by expectation of low quality of user generated data for scientific purpose by high ethical issues on privacy. The user generated data was kept private, only in case of reported failures, for the purpose of debugging, it was accessed by project members in compliance with the data privacy declaration of the APP.

The user data are e-mail address, password, a nickname, and messages for translation (text, audio, video). The purpose of the e-mail address is login, data privacy, and password reset. User input (video, audio, text) recorded by the app is processed by the other WP's components and the translation result is displayed to the user in the app. Data generated by the translation process is stored in an internal archive. Internal usage of the data is mandatory for performing a translation. An opt in for scientific post analysis was not implemented or performed.

The user generated data and user data can be deleted by the user at any time by deleting messages in the archive of the APP or deleting the account with all data. All EASIER APP generated data will be deleted when the project ends, latest mid 2024.

Table 25 EASIER_WP8_01_MUL_VideoRecording

Identifier	**EASIER_WP8_01_MUL_VideoRecording**
Dataset Summary	**Responsible partner:** NURO
	Purpose: Video recordings of Sign Input are used as source materials for the translation. The data is also used in speech input since the videos

	is used for affect recognition. The video files were sent to the Video Recognition Engine and the Multimodal Affect Recognition Engine.
	Type/format: MP4 and other video formats
	Re-use of existing data: No
	Data origin: User
	Expected size: Depends on length and resolution quality of the used hardware but presumably in the range of 10 MB to 1 GB per video recording.
	Data utility: Technical components involved in the translation (sign language recognition, affect detection, translation) process this data.
Intellectual Property Rights	**Owner:** User
	Licensing: No
	Dependency: This is private data which will not be re-used per se.
	Constraints: No
	Timeframe: N/A
Findability	**DOI:** No
	Is data discoverable: No
	Search keywords: N/A
	Versioning: N/A
	Metadata creation: Owner id.
Accessibility	**Data openly accessible:** No, this is private user data.
	How it will be accessible: N/A
	Methods/software tools to access data: N/A
	Repository: Internal
	Restrictions on access: N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data transfer are secured through TSL/SSL encryption. Access is be protected by OAuth login.

Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Since this is private data the consortium did not use this data besides the intended usage of translation. Only in case of reported failures, for the purpose of debugging, it was accessed by project members. **Is informed consent for data sharing and long-term preservation given?** The user must opt in by registration. Long-term preservation is not given.
Other issues	No, in general this data was only processed for translation purpose. User can delete their data in the APP or deleting the account with all data. All EASIER APP generated data will be deleted after the project end, latest mid-2024.

Table 26 EASIER_WP8_02_MUL_AudioRecording

Identifier	EASIER_WP8_02_MUL_AudioRecording
Dataset Summary	**Responsible partner:** NURO **Purpose**: Video recordings of Speech Input are used as source materials for the translation. Video was chosen for a better affect recognition.The video files were sent to the Speech Recognition Engine and the Multimodal Affect Recognition Engine. **Type/format**: MP4 and other video or audio formats. **Re-use of existing data**: No **Data origin**: User **Expected size**: Depends on length and resolution quality of the used hardware but presumably less than 10 MB per audio recording up to 1GB in case of video. **Data utility**: Technical components involved in the translation (speech recognition, affect detection, translation) process this data.
Intellectual Property Rights	**Owner:** User **Licensing:** No **Dependency:** This is private data which will not be re-used per se. **Constraints:** No **Timeframe:** N/A

Findability	**DOI:** No **Is data discoverable:** No **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Owner id
Accessibility	**Data openly accessible:** No, this is private user data. **How it will be accessible:** N/A **Methods/software tools to access data:** N/A **Repository**: Internal **Restrictions on access:** N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data transfer is secured through TSL/SSL encryption. Access is be protected by OAuth
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Since this is private data the consortium did not use this data besides the intended usage of translation. Only in case of reported failures, for the purpose of debugging it was accessed by project members. **Is informed consent for data sharing and long-term preservation given?** The user must opt in by registration. Long-term preservation is not given.
Other issues	No, in general this data was only processed for translation purpose. User can delete their data in the APP or deleting the account with all data. All EASIER APP generated data will be deleted after project end, latest mid-2024.

Table 27 EASIER_WP8_03_PER_UserDataset

Identifier	**EASIER_WP8_03_PER_UserDataset**
Dataset Summary	**Responsible partner:** [NURO] **Purpose**: Purpose is login, data privacy, and password reset. The user data are e-mail address, password, a nickname, and messages for translation (text, audio, video).

	Type/format: Text
	Re-use of existing data: No
	Data origin: User
	Expected size: MB range
	Data utility: User only, login, data privacy, and password reset
Intellectual Property Rights	**Owner:** User
	Licensing: N/A
	Dependency: This is private data which will not be re-used per se.
	Constraints: N/A
	Timeframe: N/A
Findability	**DOI:** N/A
	Is data discoverable: No
	Search keywords: N/A
	Versioning: N/A
	Metadata creation: Owner Id
Accessibility	**Data openly accessible:** No, this is private user data.
	How it will be accessible: N/A
	Methods/software tools to access data: Easier APP
	Repository: Internal Database
	Restrictions on access: N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data transfer is secured through TSL/SSL encryption and OAuth login.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Since this is private data the consortium will not share this data besides the intended usage of translation.
	Is informed consent for data sharing and long-term preservation given? The user must opt in by registration. Long-term preservation is not given.

Other issues	No, in general this data was only processed for translation purpose. User can delete their data in the APP or deleting the account with all data. All EASIER APP generated data will be deleted after project end, latest mid-2024.

Table 28 EASIER_WP8_04_MUL_TextMessages

Identifier	EASIER_WP8_04_MUL_TextMessages
Dataset Summary	**Responsible partner:** NURO **Purpose**: Text messages input is used as source materials for the translation. The message was sent to the Translation Engine and the Multimodal Affect Recognition Engine. **Type/format**: Text **Re-use of existing data**: No **Data origin**: User **Expected size**: Depends on length presumably less than 1 MB per message. **Data utility**: Technical components involved in the translation (affect detection, translation) process this data.
Intellectual Property Rights	**Owner:** User **Licensing:** No **Dependency:** This is private data which will not be re-used per se. **Constraints:** No **Timeframe:** N/A
Findability	**DOI:** No **Is data discoverable:** No **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** Owner id
Accessibility	**Data openly accessible:** No, this is private user data.

	How it will be accessible: N/A **Methods/software tools to access data:** N/A **Repository**: Internal **Restrictions on access:** N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data transfer is secured through TSL/SSL encryption. Access is protected by OAuth login.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Since this is private data the consortium will not use this data besides the intended usage of translation. Only in case of reported failures, for the purpose of debugging it was accessed by project members. **Is informed consent for data sharing and long-term preservation given?** The user must opt in by registration. Long-term preservation is not given.
Other issues	No, in general this data was only processed for translation purpose. User can delete their data in the APP or deleting the account with all data. All EASIER APP generated data will be deleted after the project end, latest mid-2024.

Table 29 EASIER_WP8_05_OTH_InternalTranslationRepository

Identifier	**EASIER_WP8_05_OTH_InternalTranslationRepository**
Dataset Summary	**Responsible partner:** NURO **Purpose**: Translation archive for the user and debugging. Scientific usage was discussed and dismissed. **Type/format**: Mixed, Text, Audio, Video, Metadata (JSON) stored in the internal data base. **Re-use of existing data**: No **Data origin**: User, EASIER Components **Expected size**: Depends on length presumably less than 1 GB per translation. **Data utility**: EASIER App archive.

Intellectual Property Rights	**Owner:** User
	Licensing: No
	Dependency: This is private data which will not be re-used per se.
	Constraints: No
	Timeframe: N/A
Findability	**DOI:** No
	Is data discoverable: No
	Search keywords: N/A
	Versioning: N/A
	Metadata creation: N/A
Accessibility	**Data openly accessible:** No, this is private user data.
	How it will be accessible: N/A
	Methods/software tools to access data: N/A
	Repository: Internal
	Restrictions on access: N/A
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data transfer is secured through TSL/SSL encryption. Access is protected by OAuth login
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Since this is private data the consortium will not share this data besides the intended usage of translation. Only in case of reported failures, for the purpose of debugging it was accessed by project members.
	Is informed consent for data sharing and long-term preservation given? The user must opt in by registration. Long-term preservation is not given.
Other issues	No, in general this data was only processed for translation purpose. User can delete their data in the APP or deleting the account with all data. All EASIER APP generated data will be deleted after project end, latest mid-2024.

2.12   DATA GENERATED IN WP9

No dataset was finally created in WP9.

2.13 DATA GENERATED IN WP10

Table 30 EASIER_WP10_01_PER_NEWSLETTER

Identifier	EASIER_WP10_01_PER_NEWSLETTER
Dataset Summary	**Responsible partner:** MARTEL **Purpose**: The personal data collected as part of a registration to the EASIER newsletter were only be used to verify registrations and distribute the newsletter. **Type/format:** Email addresses of visitors registering to receive the newsletter. During the newsletter registration, we also store the IP address of the computer system assigned by the Internet service provider (ISP) and used by the data subject at the time of the registration, as well as the date and time of the registration. The EASIER newsletter contained the so-called tracking pixels. A tracking pixel is a miniature graphic embedded emails, which are sent in HTML format to enable log file recording and analysis. Based on the embedded tracking pixel, Martel could see if and when an email was opened by the data subject, and which links in the email were called up by data subjects. **Reuse of existing data**: No existing data was reused. **Data origin**: The personal data was collected with previous informed consent upon registration to the newsletter through the dedicated page of the EASIER website. **Expected size**: Between 100 and 200 records. **Data utility**: Data collected expressed the interest of subscribers to receive news from the EASIER project.
Intellectual Property Rights	**Owner:** In compliance with GDPR, data subjects were the only owner of their data and at any time entitled to revoke the respective separate declaration of consent issued by means of the double-opt-in procedure. After a revocation, this personal data was deleted by Martel. Martel automatically regarded a withdrawal from the receipt of the newsletter as a revocation. Once the project completes its final review by the EC, all data will be deleted. **Licensing:** N/A **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A

Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** Data collected are personal data (email address and other contact details) and as such they were never shared with any third party (openly or privately). **How it will be accessible:** N/A **Methods/software tools to access data:** Access to the newsletter subscribers list occurred via username and password. Only authored users had access to it and only admins at Martel could grant access. **Repository**: We used a third-party service provider Mailchimp (The Rocket Science Group, LLC), to process and store data: which EASIER then used to manage the newsletter subscriber lists and send emails to its subscribers. We have not shared the users' data with any third-party. The link to the EASIER Privacy Policy was provided in each newsletter Users data will not be used for marketing purposes after the end of the project. **Restrictions on access:** Only people authorised by the data owner (MARTEL) have access to the collected data. The data owner can authorise as data processors other project partners involved with the dissemination efforts of the project only where necessary.
Interoperability	**Interoperability**: N/A
Security	**Security measures:** Data was stored in a GDPR compliant newsletter service. Data was encrypted, and access to it was possible only via an authentication of previously authorised users by the data owner entity (MARTEL). Authentication and access channels to the data were encrypted as well .
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data collected were personal data, and in compliance with the GDPR were only used within the original scope they have been collected for: informing newsletter subscribers about EASIER and its progress. As such, they were never shared outside the parties in the consortium in charge of providing the newsletter services (so not to any external third party). **Is informed consent for data sharing and long-term preservation given?** All related information is available in detail at the following URL (on section 5 and 6): https://www.project-easier.eu/privacy-policy.

	The subscription to our newsletter could be terminated by the data subject at any time. The consent to the storage of personal data, which the data subject has given could be revoked at any time. For the purpose of revocation of consent, a corresponding link was present in each newsletter. It was also possible to unsubscribe from the newsletter at any time by contacting Martel/EASIER's consortium directly via the Contact section of the website, or through the website's GDPR Requests page: https://www.project-easier.eu/privacy-policy. Furthermore, subscribers to the newsletter were informed by email, as long as this was deemed necessary for the operation of the newsletter service or a registration in question, as this could be the case in the event of modifications to the newsletter offer, or in the event of a change in technical circumstances.
Other issues	N/A

Table 31 EASIER_WP10_02_PER_WEBSITE

Identifier	**EASIER_WP10_02_PER_WEBSITE**

Dataset Summary	**Responsible partner:** MARTEL
	
	Purpose: The data collected from the website visitors was needed for maintaining contact, and as required by law for any legal agreement we have had with them. The data and information collected were used to optimize content and performance of the website and would have also been used in the event of attacks on our information technology systems.
	
	Type/format: The website collects the following data:
	
	● User account data: i.e., account of users authorised to publish content on the website. These are, in general, users who are part of MARTEL and in some cases of other third parties involved in the EASIER project, authorised to publish news on the website. These included:
	
	    ○ Name and relevant titles
	
	    ○ Email
	
	    ○ Job title
	
	    ○ Company name
	
	    ○ Company address (in very rare cases)
	
	● Contact data: i.e., data provided by websites visitors filling in contact forms to request information from the EASIER consortium. This may include:

○ Name and relevant titles

○ Email

○ Job title

○ Company name

○ Company address (in very rare cases)

- Access log data: i.e.data that was collected by servers to control access to the webpages of the web site that may have been used in the event of attacks and other cases requested by the law,including (1) the browser types and versions used, (2) the operating system used by the accessing system, (3) the website from which an accessing system reaches our website (so-called referrers), (4) the sub-websites, (5) the date and time of access to the Internet site, (6) an Internet protocol address (IP address), (7) the Internet service provider of the accessing system, and (8) any other similar data and information that may have been used in the event of attacks on our information technology systems.

Reuse of existing data: No existing data was reused.

Data origin: All data we store has been clearly and voluntarily provided to us by contacts and partners (subscribing to a newsletter, sending us an email, giving us their business card, answering to an online questionnaire etc.).

- The website collected a series of general data and information when a data subject or automated system called up the website.

- The website contained information that enabled a quick electronic contact to the project consortium, as well as direct communication with EASIER, which also included a general address of the so-called electronic mail (e-mail address). If a data subject contacts the project by e-mail or via a contact form, the personal data transmitted by the data subject were automatically stored. Such personal data transmitted on a voluntary basis by a data subject to Martel as data owner were stored for the purpose of processing or contacting the data subject.

As it is common practice with almost all professional websites, this site also used cookies to improve the visitors' experience. In some special cases, we also used cookies anonymized provided by trusted third parties (such as Google Analytics) – more details are provided in the Cookie Policy page of the EASIER website: https://www.project-easier.eu/cookie-policy.

Expected size: We collected minimal data from our site visitors.

Data utility: Data and information collected were needed to (1) deliver the content of our website correctly, (2) optimize the content of our website as well as its advertisement, (3) ensure the long-term viability of our information technology systems and website technology, and (4)

	provide law enforcement authorities with the information necessary for criminal prosecution in case of a cyber-attack.
Intellectual Property Rights	**Owner:** N/A **Licensing:** N/A **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** Data collected were personal data (email address and other contact details) and as such they were never shared with any third party (openly or privately). **How it will be accessible:** The data owner (MARTEL) could authorise, as data processors, other project partners involved in the dissemination efforts of the project only where necessary. **Methods/software tools to access data:** Access to the website log data occurred via username and password. Only authorized users had access and only admins in Martel could grant access. **Repository**: See the Security section. **Restrictions on access:** Only people authorized by the data owner (MARTEL) had access to the collected data.
Interoperability	**Interoperability**: There will be no transfer of personal data collected to any third party.
Security	Users can access the website only via an encrypted connection (https), in order to add a second security layer between the user and the website itself. WordPress (WP) was used to build the project website. This content management system (CMS) used the latest technology about PHP and MariaDB for the business logic and database respectively. WP provided several of plug-ins in order to provide a great level of security for both the content and the users. In fact, plug-ins, such as anti-spam, anti-SQL injection, anti-brute force attack, etc., have helped to prevent spam and

	the most common attacks. Moreover, WP provided different access roles, in order to grant the right permissions to the right users. Users' WP passwords were encrypted through RSA technology, so no one could decrypt them, not even the WP administrator. The visitors were able to adjust the cookies settings directly on their browser. Disabling cookies would usually result in also disabling certain functionalities and features of the site. Therefore, it was recommended for users not to disable cookies. On the website, Martel has integrated components of Twitter/X: Twitter/X received information via the Twitter/X component that the data subject has visited our website, provided that the data subject was logged in on Twitter/X at the time of the call-up to our website. This occured regardless of whether the person clicked on the Twitter/X component or not. If such a transmission of information to Twitter was not desirable for the data subject, they had an option to prevent this by logging off from their Twitter/X account before a call-up to our website was made.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data collected were personal data, and in compliance with GDPR were only be used within the original scope they have been collected for: informing contacts about EASIER. As such, they were never shared outside the parties in the consortium in charge of providing the newsletter services (so not to any external third party). **Is informed consent for data sharing and long term preservation given?** The website respected the latest European privacy laws (GDPR). All users (registered and guests) had to authorise data collection by the data owner (MARTEL) and eventually, in the case of cookies for anonymised tracking – the sharing with a third party. All information is available in detail at the following URLs: - https://www.project-easier.eu/privacy-policy - https://www.project-easier.eu/cookie-policy The aforementioned Privacy Policy page also informed visitors in detail on the data subject's rights (Section 9).
Other issues	Martel shall process and store the personal data of the data subject only for the period necessary to achieve the purpose of storage, or as far as this is granted by the European legislator or other legislators in laws or regulations to which the Martel is subject to. If the storage purpose is not applicable, or if a storage period prescribed by the European legislator or another competent legislator expires, the personal data are routinely blocked or erased in accordance with legal requirements. The criteria used to determine the period of storage of personal data is the respective statutory retention period. After expiration of that period, the corresponding data is routinely deleted.

2.14   DATA GENERATED IN WP11

Table 32 EASIER_WP11_01_PER_Mattermost

Identifier	EASIER_WP11_01_PER_Mattermost
Dataset Summary	**Responsible partner:** MARTEL **Purpose**: The data collected within the Mattermost platform (open source tool installed on a private cloud managed by Martel) were solely used for enabling the user the access to the platform and allowing the usage of the messaging service. Data generated within the platform were private to the consortium. **Type/format**: The Mattermost online application collected the following personal data: ● Username ● Nickname (optional) ● Position (optional) ● Email ● Profile picture (optional) **Re-use of existing data**: No existing data was reused. **Data origin**: All data we stored has been clearly and voluntarily provided to us by consortium partners. Mattermost is an online application which presents itself as a website allowing group and direct communication within EASIER. **Expected size**: We collected minimal data from our site visitors. **Data utility**: Data and information collected were needed to allow the user to use the online tool.
Intellectual Property Rights	**Owner:** N/A **Licensing:** N/A **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A

Findability	**DOI:** N/A **Is data discoverable:** N/A **Search keywords:** N/A **Versioning:** N/A **Metadata creation:** N/A
Accessibility	**Data openly accessible:** Data collected were personal data of consortium members (email address and other contact details) as such they were not shared with any third party (openly or privately). **How it will be accessible:** The data owner (MARTEL) did not disclose any of the provided data. **Methods/software tools to access data:** Only authorised users in Martel had the access and only admins in Martel could grant access. **Repository**: See Security section. **Restrictions on access:** Only authorised users in Martel had the access and only admins in Martel could grant access.
Interoperability	**Interoperability**: There was no transfer of personal data collected to any third party.
Security	Users were able to access the Mattermost service only via an encrypted connection (https), in order to add a second security layer between the user and the website itself. Mattermost was used to build EASIER's dedicated messaging service. This messaging system used the latest technology about GO, Javascript and PostgreSQL for the business logic and database respectively.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data collected were personal data, and in compliance with GDPR could only be used for the original scope they have been collected for: internal consortium communication. As such they were not shared outside the parties in the consortium. **Is informed consent for data sharing and long-term preservation given?** Mattermost respects the latest European laws about Privacy (GDPR). Participants could cancel their subscription at any time.
Other issues	Martel processed and stored the personal data of the data subject only for the period necessary to achieve the purpose of storage, or as far as this is granted by the European legislator or other legislators in laws or regulations to which the Martel is subject to. If the storage purpose was not applicable, or if a storage period prescribed by the European legislator or another competent legislator

	expires, the personal data were routinely blocked or erased in accordance with legal requirements. The criteria used to determine the period of storage of personal data was the respective statutory retention period. After expiration of that period, the corresponding data was routinely deleted.

Table 33 EASIER_WP11_02_PER_Mailman

Identifier	EASIER_WP11_02_PER_Mailman
Dataset Summary	**Responsible partner:** MARTEL **Purpose**: The email collected within the Mailman platform (open source mailing list tool installed on a private cloud managed by Martel) were solely used for enabling the user the access to the platform and allowing the usage of the messaging service. Data generated within the platform were private to the consortium. **Type/format**: The Mailman online application collected the following personal data: ● Username (optional) ● Email **Re-use of existing data**: No existing data was reused. **Data origin**: All data we stored has been clearly and voluntarily provided to us by consortium partners. Mailman is an online application which allows the management of the project's mailing lists and has a light graphical interface for the users to subscribe or unsubscribe. **Expected size**: We collected minimal data from our site visitors, about 50 entries overall, spread and/or duplicated within the different mailing lists (one per WP, general one,, etc) **Data utility**: Data and information collected were needed to allow the user to use the online tool.
Intellectual Property Rights	**Owner:** N/A **Licensing:** N/A **Dependency:** N/A **Constraints:** N/A **Timeframe:** N/A
Findability	**DOI:** N/A **Is data discoverable:** N/A

	Search keywords: N/A
	Versioning: N/A
	Metadata creation: N/A
Accessibility	**Data openly accessible:** Data collected were personal data of consortium members (email address) as such they were not shared with any third party (openly or privately). **How it will be accessible:** The data owner (MARTEL) did not disclose any of the provided data. **Methods/software tools to access data:** Only authorised users in Martel had the access and only admins in Martel could grant access. **Repository**: See Security section. **Restrictions on access:** Only authorised users in Martel had the access and only admins in Martel could grant access.
Interoperability	**Interoperability**: There was no transfer of personal data collected to any third party.
Security	Users could access the Mailman admin interface only via an encrypted connection (https), in order to add a second security layer between the user and the website itself. Mailman has been used to build EASIER's dedicated mailing lists service. This system uses the latest technology about Python.
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Data collected were personal data, and in compliance with GDPR can only be used for the original scope they have been collected for: internal consortium communication. As such they were not shared outside the parties in the consortium. **Is informed consent for data sharing and long term preservation given?** Mailman respects the latest European laws about Privacy (GDPR). Participants can cancel their subscription at any time.
Other issues	Martel processed and stored the personal data of the data subject only for the period necessary to achieve the purpose of storage, or as far as this is granted by the European legislator or other legislators in laws or regulations to which the Martel is subject to. If the storage purpose was not applicable, or if a storage period prescribed by the European legislator or another competent legislator expires, the personal data were routinely blocked or erased in accordance with legal requirements.

	The criteria used to determine the period of storage of personal data was the respective statutory retention period. After expiration of that period, the corresponding data was routinely deleted.

3 CONCLUSIONS

This deliverable provides an updated overview of the data that the EASIER project produced, together with the related data processes and requirements that were taken into consideration.

Most of the generated data were of one of two types: source-code or datasets, which follow standardised formats, and are documented, so that they can be easily used as reference in other research projects or initiatives. The datasets were disseminated through research publications and invited talks, among other means, and will be preserved at the partners repositories for at least three years. For completeness, the main technical (internal) and administrative datasets have also been included in the DMP.

APPENDIX A: LEGAL FRAMEWORK

The collection, use and disclosure of personal data at a European level are regulated in chronological order by the following:

- in 1995 by the 95/46/EC "Data Protection Directive"

- in 2002 by the 2002/58/EC "Privacy and electronic communications Directive"

- in 2009 by the 2009/136/EC "Cookie Directive"

- in 2016 by the 2016/679/EC "General Data Protection Regulation (GDPR)" that repeals Directive 95/46/EC

- in 2016 by the 2016/680/EC "Protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data Directive".

APPENDIX B: TEMPLATE FOR DATA COLLECTION

Each WP Leader should collect the datasets intended to be generated in their WP, collecting information from the partners. In case a dataset might be shared among different WPs, the involved WP leaders should communicate offline and decide a collocation inside a single WP. The dependency should then be indicated in the table below: **Dataset Summary Purpose.**

IDENTIFIERS

Identify each data[set] with the format:

EASIER_<WP number>_<serial number of dataset>_<data type>_<dataset title>

Replace with your WP number as in the DoA, refer to the table below (Table 1) for the data type and assign a meaningful title.

Table 34 IDENTIFIER PATTERN AND TYPES OF DATA

Acronym	Data Type
SRC	Source Code
MUL	Multimedia content (audio, video, text)
PER	Personal data (email addresses, other personal information)
OTH	Other type of data (to be specified in the table below)

DATASETS

Create a table using the provided template (Table 29) for each data[set] following the questions.

➲ Do you plan to generate/gather data (audio, video, text, logs, transcripts, parameters, algorithms, source code, software stacks, APIs, policies, etc.) inside the project or during the project lifetime in your WP?
<YES/NO, why>

➲ If **yes**, please fill in. the table below answering the given questions (in case of not applicability, indicate N/A).

Table 35 DATASET PROPERTIES TEMPLATE

Identifier	EASIER_<WP number>_<serial number of dataset>_<data type>_<dataset title>

Dataset Summary	**Responsible partner:** [MARTEL, ATHENA, UHH, RU, UNIS, UZH, CNRS, DFKI, AUD, NURO, STXT, EUD, INT, UCL] **Purpose**: Short description of data. Also, what is the purpose of data collection/generation (and its relation to project objectives) in the context of EASIER and the Workpackage? Is this dataset being used in others EASIER WPs? **Type/format**: What is the type/format of the data? Do you follow any standards for such data? (NIST, ISO, etc.)? **Re-use of existing data**: Do you use any existing data/datasets, which you can/will reuse for further developments/implementations inside the project or during the project lifetime? **Data origin**: What is the origin/source of the data? **Expected size**: What is the expected data/dataset size? GB? TB? **Data utility**: To whom will this data be useful and how? (inside the project and also to third parties, if applicable)
Intellectual Property Rights	**Owner:** Who owns the data? **Licensing:** Will the data be licensed for reuse? If yes, under which licence? **Dependency:** Are there any restrictions on the reuse of existing/third-party data? **Constraints:** Will data sharing be postponed / restricted e.g. to publish or seek patents? **Timeframe:** Specify the length of time for which the data will remain re-usable or if there is any embargo that prevent immediate publication.
Findability	**DOI:** Will you pursue getting a persistent identifier for your data? **Is data discoverable:** Are the data produced in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? **Search keywords:** Will search keywords be provided that optimize possibilities for re-use? **Versioning:** Will clear version numbers be provided? How many releases are foreseen? **Metadata creation:** Specify standards for metadata creation (if any). If there are no standards in your discipline, describe what type of metadata will be created and how.

Accessibility	**Data openly accessible:** Will data produced in the project be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions. **How it will be accessible:** How will the data be made accessible (e.g. by deposition in an open repository)? **Methods/software tools to access data:** What methods or software tools are needed to access the data? Also, is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? **Repository**: Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. **Restrictions on access:** If there are restrictions on use, how will access be provided?
Interoperability	**Interoperability**: Are the data produced in the project interoperable, i.e. can data be exchanged and re-used between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)? If not, what are the barriers and how to overcome them?
Security	**Security measures:** Security measures implemented for data protection (incl. controlled access, user authentication, firewalls, VPNs, encryption, back-ups, etc.)
Ethical aspects	**Possible ethical and legal aspects preventing sharing:** Are there any ethical or legal issues that can have an impact on data sharing? If yes, describe them and the way they are dealt with. **Is informed consent for data sharing and long-term preservation given?** Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?
Other issues	Refer to other national/funder/sectorial/departmental procedures for data management that you may be using (if any)