Global Under-Resourced MEedia Translation (GoURMET)

H2020 Research and Innovation Action
Number: 825299
D7.2 – Data Management Plan

Nature	Report	Work Package	WP7
Due Date	30/06/2019	**Submission Date**	23/06/2019
Main authors	Alexandra Birch (UEDIN)		
Co-authors	Barry Haddow (UEDIN)		
Reviewers	Felipe Sánchez-Martínez (UA)		
Keywords	data, social media, monitoring, metadata, access, data protection, privacy		
Version Control			
v0.1	**Status**	Draft	16/06/2019
v1.0	**Status**	Final	23/06/2019

Contents

List of Figures

Abstract

The Data Management Plan provides an analysis of the main elements of the data management policy that have been used by the GoURMET consortium with regard to all the datasets collected for or generated by the project. It addresses issues such as collection of data, data set identifiers and descriptions, standards and metadata used in the project, data sharing, property rights and privacy protection, and long-term preservation and re-use, complying with national and EU legislation. This is the first and only data management plan of the project.

1   Introduction

GoURMET (Global Under-Resourced MEedia Translation) is a three-year H2020 project, running from 1 January 2019 to 31 December 2021, under the Research and Innovation Action grant agreement number 825299. GoURMET participates in the H2020 Pilot on Open Research Data. This Data Management Plan (DMP) provides an analysis of the main elements of the data management policy that have been used by the GoURMET consortium with regard to all the datasets collected for or generated by the project.

This deliverable addresses issues such as collection of data, data set identifiers and descriptions, standards and metadata used in the project, data sharing, property rights and privacy protection, and long-term preservation and re-use, complying with national and EU legislation.

In this deliverable we will look at the types of data collected and generated and the formats and standards that they adhere to. We also describe data storage and preservation of the data.

For this project we also address data privacy and the implications of the GDPR regulations on project data, especially looking at protection of personal data.

2   Types of Data Collected

GoURMET is developing a translation service capable of dealing with large volumes of text across many languages. Data will be collected in sixteen languages and English. The first four GoURMET languages are Gujarathi, Swahili, Bulgarian and Turkish.

The project includes two data providers. The BBC and Deutsche Welle (DW) are world broadcasters with a wide range of languages and they act primarily as user partners and content providers in GoURMET.

Three use cases implement the applications and put the data to use:

- Global content creation – managing content creation in several languages efficiently by providing machine translations for correction by humans;
- Media monitoring for low-resource language pairs – tools to address the challenge of monitoring media in strategically important languages;
- International business news analysis – reliably translating and analysing news in the highly specialised financial domain.

This is a broadcaster-focused project, with involvement of world broadcasters with coverage of several languages and thus with a key role for data.

"Collection of data" in this report refers to the acquisition of data by the consortium, primarily through data provision by the participating GoURMET broadcasters. The research partners will also be accessing data generated through crawling data on the internet. Internet data will be accessed either by downloading freely available preprocessed corpora, or by using tools developed by the Paracrawl project[1] to do focussed language specific crawls.

2.1   Languages

The GoURMET project has 4 x 9 month long development, integration and testing cycles. At each cycle four new languages will be added to the project. Only the first four languages have so far been defined. We have already analysed all the languages mentioned in the project proposal and created a language survey. In this survey we describe the challenges related to translation in these languages and the existing resources available for training translation models. Please refer to the deliverable titled "D1.1 Survey of relevant low-resource languages" for details.

2.2   Data Types

Data for GoURMET is being collected at several levels:

- By project target use
    - Ingestion data: text sent to the GoURMET platform once deployed
    - Training data: data for training translation models. This could be monolingual data, parallel translated data, or other data resources such as bilingual dictionaries or morphological taggers.
    - Test data: GoURMET will specify test sets for all languages. Some will be created specifically for the project but many will be selected from existing translated data.

---

[1] https://paracrawl.eu/

- – Usage data: data generated by the use of the GoURMET platform which will be used for evaluation purposes.
- • By type of data
    - – Metadata
    - – Text articles
    - – Social media
    - – Ontologies
    - – Usage data
- • By language
    - – All 16 GoURMET languages
- • By content provider/user partner
    - – BBC
    - – DW
    - – Others

2.3 Translated Data

In all the GoURMET scenarios, the translation is always into or out of English, so GoURMET deals with 16 x 2 language pairs. Current technology in machine translation requires large amounts of training data. Ideally, for *Machine Translation* (MT), 10,000,000 parallel sentences are needed per language combination for a valuable training set. The user partners are unable to provide enough quantity of training data to make a significant difference to the building of MT models, and freely available corpora are used. For many language pairs, we use parallel data from other related languages, for example for Gujarathi-English we used Hindi-English data to improve the model.

2.4 Provision of Global Content Creation Data

Described below is the process for content provision by the broadcasters for the purpose of content creation. This is based on the available infrastructure, content requirements and planned prototyping.

BBC content creation data will be provided via an API through the prototypes that are currently being developed. This will support both the BBC and DW. The data will be in the form of text based documents.

2.5 Provision of Monitoring Data

Described below is the process for content provision by the broadcasters for the purpose of data monitoring. This is based on the available infrastructure, content requirements and planned prototyping.

BBC Monitoring collects and processes content – primarily streaming content – from other broadcasters. BBCM monitors up to 200 channels simultaneously in real life, and has up to 15000 sources including RSS feeds and social media accounts. Much of Deutsche Welle's content provision from internal sources is through APIs and has been implemented already in the SUMMA project (http://summa-project.eu/).

The participating broadcasters have added RSS feeds and podcasting to the channels. DW has granted GoURMET access to all of its RSS feeds and podcasting channels.

3 Types of Data Generated

"Generation of data" in this report refers to the generation of data by the GoURMET platform. This refers to translated broadcast text, either into English or into any of the GoURMET language.

We distinguish between six main categories of data that will be generated during the project:

- Content data generated by BBC or DW or gathered from media monitoring. This is typical broadcast data that remains copyright-protected.
- Specific output formats from GoURMET translation services. This usually also includes broadcast content.
- Software, models, algorithms, lexicons and ontologies, annotations, etc.
- Personalised data generated during field testing and prototype testing
- Social Media Data: Twitter, Facebook, YouTube, etc
- Academic-type research publications

Of course, all this generated data directly exploits and enriches the data collected (Section 2).

4 Data and Metadata Standards

4.1 Metadata

Available metadata at content provider side is considered for inclusion in GoURMET. Metadata formats differ according to content provider.

The original broadcaster metadata format is provided to the integrator and overall preferences and settings are discussed. Within GoURMET, it is agreed that JSON format is preferred; the BBC metadata uses Dublin-core; DW metadata does not use Dublin-core.[2]

The BBC has defined a metadata standard for the GoURMET translation platform, which is a simple JSON format with written text input, and the source and target language codes.

4.2 Dataset Identifiers and Descriptions

Within the GoURMET project, data sets are divided into two main groups: (1) regular content for content creation/monitoring, and (2) specific data for training and testing translation.

In **Group (1)**, "regular content" data files are used within the GoURMET system and are always accompanied by the JSON files. A full specification is described in the GoURMET document "Specification 5.1 – Translation module and version controlspecification". A simple request and response can be shown below:

Query: {"q"=string}. For example: {"q"="hello world"}

Response: { "translation": string/None, "original": string/None, "error": string/None "timeTaken: int/None, "model": string }. For example:{ "translation": "translated text here", "original": "original text here", "error": None, "timeTaken: 100, "model": "0.1" }

The data sets generated in the GoURMET Platform will be stored with date and time stamps.

In **Group (2)**, "specific data" for training and testing the translation tools within the GoURMET project will be generated for specific low-resource language pairs from data available in the BBC and DW. We will also be using training data that is freely available on the internet and data crawled from the internet using Paracrawl tools.

4.3 Social Media Data

Social media – Twitter, Facebook, YouTube, discussion forums etc. – is very time-sensitive and quickly outdated, and also highly relevant to privacy issues. GoURMET does not focus on translating social media content, but it is likely that one of the prototypes developed by the project might contain discussion forums.

In order to ensure the protection of private individuals' personal data (contained in Tweets or Facebook posts), the data providers (User Partners) will ensure that third-party social media content are not transferred to other consortium partners. Of course, social media content authored by the content provider (e.g. @dwnews) can be used if the account holder agrees to this. Thus, protec-

---

[2] http://dublincore.org/specifications/dublin-core/

tion of personal data legislation will be taken into account while developing the platform for the processing of social media.

The potential value of the social media output, anonymisation/pseudonymisation and retention of social media user-related data will be discussed. Social media is an important source, but not the primary content format. Special care will be taken in handling social media content, ensuring privacy protection. Differentiation between different categories of social media posters - e.g. public figures (politicians), political organisations (political parties, NGOs) and private citizens will be taken into consideration.

In Figure 1 you can see examples of DW twitter feeds.

	dw_travel	Tweeting the best of DW's multimedia coverage on travel! Let the journey begin! dw.com/travel
	dw_learngerman	Improve your German – and have fun! We'll deliver tons of exercises straight to your newsfeed.
	dw_futurenow	
	dw_politics	The official account for political news from the DW's political team based in Berlin. Political news, campaign stories and coverage of German politics.
German	dw_deutsch	Wir schreiben, was wichtig ist. Ihr sagt uns, was ihr denkt: Diskutiert mit uns Nachrichten aus Deutschland und der Welt.
	dw_politik	
	dw_sport	
	dw_wirtschaft	Wirtschaftsnachrichten. Berichte. Reportagen. Aus Deutschland und der Welt. Alles, was wichtig ist. Lesen, sehen, hören, mitmachen.
	dw_wissenschaft	Du findest Forschung, Medizin und Technik richtig gut? Wir auch!
	dw_kultur	
	dw_reporter	
	dw_reise	
Arabic	dw_arabic	عربية تقدم من ألمانيا تغطية إعلامية محايدة لأهم الأحداث في العالم العربي وأوروبا والعالم. شاركونا في الحوار وتفاعلوا معناDW!
Russian	dw_russian	Подробно о самом главном. От независимого источника. Вам тоже есть, что сказать! Присоединяйтесь!
Portuguese	dw_brasil	A DW oferece um olhar independente para você formar sua opinião. Discuta conosco os fatos mais importantes do Brasil e do mundo.
Spanish	dw_espanol	Le ofrecemos noticias y trasfondo. Usted nos da su opinión. ¡Dialogue con nosotros acerca de la actualidad alemana y global!
Ukrainian	dw_ukrainian	Ми повідомляємо про те, що важливо для Німеччини та України. Дискутуйте про це разом з нами!
Persian	dw_persian	با ما همراه شوید، در جریان مهم‌ترین تحولات ایران و جهان قرار بگیرید و در بحث‌ها شرکت کنید.

Figure 1: Some DW Twitter Feeds

4.4   RSS Feeds, Podcasting

RSS feeds and podcasts are provided when available and appropriate. The metadata format may differ according to content provider. DW has provided access to all its RSS and podcasting feeds via API. In Figure 2 you can see examples of DW RSS feeds.

RSS Newsfeeds

Language	URL	Description
English	http://rss.dw.de/rdf/rss-en-all	ALL TOP STORIES AND NEWS UPDATES
German	http://rss.dw.com/xml/rss-de-all	DEUTSCHE WELLE GESAMT
Spanish	http://rss.dw-world.de/rdf/rss-sp-all	TODO DW VÍA RSS
Portuguese for Brazil	http://rss.dw.com/rdf/rss-br-all	África
Portuguese for Africa	http://rss.dw.com/rdf/rss-br-africa	Toda a DW Brasil
Persian	http://rss.dw-world.de/xml/rss-per-all_volltext	همه گزارش‌ها
Arabic	http://rss.dw.de/rdf/rss-ar-all	مواضيع
Russian	http://rss.dw.de/xml/rss-ru-all	DEUTSCHE WELLE: ВЕСЬ САЙТ В ОДНОМ
Ukrainian	http://rss.dw-world.de/rdf/rss-ukr-all	DW.DE/ukrainian через RSS

Figure 2: DW RSS Feeds

5 Data Storage, Preservation and Re-Use

The project data repository resides at The University of Edinburgh and it's accessible via SFTP (for all the project members) or via SSH (only for Edinburgh's Valhalla cluster users). The users access this repository via SFTP.

The data repository has the following sub-directories:

- **corpora**
  This directory contains all corpora that have been downloaded or created for this project by name of corpora. Many corpora contain data for multiple languages and require scripts for preprocessing that could be shared across languages.

- **languages**
  System builders require all the data for one particular language pair and as such all relevant corpora will be conveniently symlinked from here. Preprocessed corpora can also be stored here, as long as the pre/post processing tools are also made available. GoURMET test sets will be stored here as well.

- **models**
  All systems developed for the project will be stored here.

- **scripts**
  All necessary scripts for pre and post processing will be stored in this subdirectory.

Each partner which runs the GoURMET platform is responsible for the media ingested and processed on their site.

All data produced during the course of the project is stored and made available in accordance with the Consortium Agreement and licence agreements.

Preservation of data after the project is the responsibility of each of the partners. In a research environment, data is carefully preserved in order to be able to reproduce models and experiments and verify the research undertaken. The partners in the project endevor to keep experimental data stored on their servers for at least a period of five years.

6 Data Privacy

6.1 Different Levels for Access and Sharing

There are different kinds of categories of data that have been collected or generated during the project, with different levels and conditions for access and sharing:

- Original broadcast data is copyright-protected and, as stipulated in the Consortium Agreement, is provided only for use by the consortium partners for the duration of the project. It can therefore not be shared outside the consortium or after the project. Some demo material has been selected for public viewing in agreement with the broadcasters.

- Data generated during content creation/media monitoring. This data is typically owned by the broadcaster; therefore, the consortium does not have the rights to share this as open research data. However, negotiations have been opened with broadcasters with the aim of releasing data sets for specific research use, as has been done in the past by the BBC for the MediaEval and MGB Challenge evaluation campaigns.

- Specific output formats following a particular step in the GoURMET processing chain. This includes translations and usually also includes broadcast content. This data resides on the servers running the platforms and is only shared through the user interface.

- Software, models, algorithms, lexicons and ontologies, annotations, etc. We will make as many of these open source as possible. In cases where data is available, but copyright restrictions do not allow us to publish it, we will released tools to reconstruct it.

- Personalised user-specific data generated during field testing and prototype testing. This is data relating to people's use of GoURMET prototype systems. As this is very specific to the systems being evaluated, and also personal to evaluation subjects, we will not release this data.

- Social Media Data. In order to ensure the protection of private individuals' personal data (contained in Tweets or Facebook posts), the data providers (User Partners) ensured that third-party social media content is not transferred to other consortium partners.

- Academic-type research publications. Academic publications will be made available as "green" open access via institutional repositories and the OpenAire system.

6.2 Planned Measurements for the Protection of Personal Data

This section deals with ethical issues. The consortium will specify how we identify where personal data is involved and how such personal data is protected.

As the GoURMET architecture and information flow is established, security and privacy issues will be taken into consideration. All consortium partners dealing with data, including provision, use, processing and storing, have looked into data protection regulations for their organisation and country.

- For the protection of personal data in social media, we will limit the channels used in the project. Currently there is no pressing need to translate social media content, and we will not consider any user generated data. As the BBC use cases develop, there is a chance that we might be required to translate comments on the bottom of BBC articles. This data is less sensitive than facebook or other social media platform data.

- All consortium partners have sought advice from their respective local data protection authorities (UK, Germany, Spain, Netherlands). It is understood that the consortium as a whole are joint data controllers in this project.

- Security procedures have been established for each partner dealing with data. For instance, all communication by any third parties with the BBC Monitoring web server and storage is secured using SSL via the HTTPS protocol and require a security certificate. The BBC issue these certificates to each of the partners. This is to meet the needs of BBC Information Security and UK Data Protection regulations.

- The consortium is striving for transparency to make the purpose of including social media content and other content containing personal data in the GoURMET project clear. It is clearly stated which partner is responsible for the relevant data.

- During dissemination events, when targetting potential new users, the platforms running with DW content are normally used. This restricts the demo to content which is owned and controlled by our user partners, reducing the risk of displaying sensitive private data. Only when demonstrating at BBC events, are the platforms running content from the external BBC Monitoring channels shown.

7 Conclusion

The present Deliverable D7.2, the Data Management Plan (DMP), provides the basis for the GoURMET project management strategy and planning, as extensively discussed and agreed by all the partners.

D7.2 addresses most identified issues related to the collection and generation of data, data set identifiers and descriptions, standards, data sharing, property rights and privacy protection, and long-term preservation and re-use. Standards and metadata formats have also been agreed upon.

A Personal Data

This appendix catalogues all the personal data sources used in the project. We specify data categories (audio visual, social media etc.), whether or not they are used in GoURMET and protective measures we will take to ensure compliance with data privacy legislation. In GoURMET we have agreed that personal data is not going to be distributed. It will be downloaded by each instance of the GoURMET platform, and data will remain on site.

Content type	Used in GoUR-MET (Y/N)	Protective Measures
AV (Audio Visual) or Text Content		
Videos featuring people	Y	Only material approved for broadcast distribution is used. Analysis results are for research purposes only.
Interviews with people in audio, video or text	Y	Only material approved for broadcast distribution is used. Analysis results are for research purposes only.
Descriptions of people in audio, video or text	Y	Only material approved for broadcast distribution is used. Analysis results are for research purposes only.
Mentions of people in audio, video or text	Y	Only material approved for broadcast distribution is used. Analysis results are for research purposes only.
Social media		
Social media accounts of public figures	N	We consider people who have verified Twitter accounts to be public figures. We do not use social media data.
Social media accounts of non-public figures	N	We do not use social media data.
Social media accounts from the content partners	N	We do not use social media data.
Mentions of people in social media	N	We do not use social media data.
Images of people in social media	N	We do not use social media data.

Links to non-public people's SM accounts	N	We do not use social media data.
Other sources		
Blogs	N	DW Blogs were provided to the technology partners for testing and trialing, but finally not included in the GoURMET platform.
Podcasts	N	DW podcasts were provided to the technology partners for testing and trialing, but finally not included in the GoURMET platform.
RSS feeds	N	DW RSS feeds were provided to the technology partners for testing and trialing, but finally not included in the GoURMET platform.
Analysed data		
Statistical data	Y	The same measures apply as to the original data.
Automated translations	Y	The same measures apply as to the original data.
Knowledge bases	Y	The same measures apply as to the original data.

Table 1: Analysed data personal data sources used in the project

B   EU/National data protection laws

This annex describes the legal responsibilities of partners as relating to EU and their national data protection laws. We first describe the recent changes to EU law, and then we detail all the partners legal responsibilities in a table.

EU regulations as a one-stop shop

On 28 May 2018, the national legislations of 28 countries were replaced by one, simple and clear legal framework and a one-stop-shop for governance and enforcement, as referenced in the EU Data Protection Reform policy. [3]

Thus, within the timeframe of the GoURMET project, this EU Data Protection regulation became applicable and all partners should adhere to the measures in that regulation. Some of the aspects of the new regulation are described below. The objective of this new set of rules is to give citizens back control over of their personal data, and to simplify the regulatory environment for business. The data protection reform is a key enabler of the Digital Single Market which the Commission has prioritised. The reform will allow European citizens and businesses to fully benefit from the digital economy.

In January 2012, the European Commission proposed a comprehensive reform of data protection rules in the EU. On 4 May 2016, the official texts of the Regulation and the Directive were published in the EU Official Journal in all the official languages. While the Regulation entered into force on 24 May 2016, it became applicable from 25 May 2018. The Directive entered into force on 5 May 2016 and EU Member States had to transpose it into their national law by 6 May 2018.

The regulation and the directive are linked to below:

- Regulation (EU) 2016/679[4] of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

- Directive (EU) 2016/680[5] of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data 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, and repealing Council Framework Decision 2008/977/JHA.

A very helpful summary of how these changes affect the activities of the GoURMET project are described in the following article: https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/.

---

[3] https://ec.europa.eu/info/law/law-topic/data-protection_en
[4] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC
[5] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0089.01.ENG&toc=OJ:L:2016:119:TOC

Partner	Legislation	Responsibilities	Action
Deutsche Welle	German and EU-wide* data protection laws	According to section 42 of the German Federal Data Protection Statute, Deutsche Welle needs to appoint a data protection officer. It also needs to publish data and privacy protection policy. DW has appointed a dedicated data protection officer: currently this is Thomas Gardemann, contact: datenschutz@dw.com DW has published its data and privacy protection policy on its distribution channels. For the website, see: http://www.dw.com/en/data-privacy-policy/a-18265246 DW has published its terms of use for interactive content: http://www.dw.com/en/conditions-of-participation/a-16372765 DW has published its terms of use for the DW App 2.1: http://www.dw.com/en/general-conditions-of-use-for-the-dw-app-21/a-18532587	

BBC	UK and EU-wide* data protection laws	Abide by the UK 2018 Data Protection Act which controls how personal information is used by organisations, businesses or the government. Supported by advice and guidelines from the Information Commissioner's Office (ICO), https://ico.org.uk, the UK's independent body set up to uphold information rights.	The BBC provides personal data to any consortium member that needs a licence of our content. Those project partners have a licence in place with us that includes data protection provisions. We have looked at ways of reducing risks, for example instead of providing social media content we are providing our data scraping tool. BBC have worked with our legal team in Research and Development to draft data agreements between ourselves and all consortium members that use BBC-sourced media for training and testing purposes. We are supplying a means to scrape social media to the platform and do not supply any links to social media accounts.
UEDIN	UK and EU-wide* data protection laws	Abides by the UK 2018 Data Protection Act which controls how personal information is used by organisations, businesses or the government. Supported by advice and guidelines from the Information Commissioner's Office (ICO), https://ico.org.uk, the UK's independent body set up to uphold information rights.	UEDIN has a data protection officer who can be contacted at recordsmanagement@ed.ac.uk UEDIN has published its data protection policy: http://www.ed.ac.uk/records-management/data-protection UEDIN has a research data management policy: http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy This is supported by the UEDIN's Research Data Service: http://www.ed.ac.uk/information-services/research-support/research-data-service
UVA	Dutch and EU-wide* data protection laws	Abides by the Dutch Data Protection Authority which controls how personal information is used by organisations, businesses or the government.	The UVA has appointed a data protection officer who can be contacted at fg@uva.nl. The UVA has published a privacy statement: https://www.uva.nl/en/home/disclaimers/privacy.html. This is supported by the UVA's Research Data Management facilities: https://rdm.uva.nl/en

| UA | Spanish and EU-wide* data protection laws | Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). LAW 3/2018 about personal data. | UA has a data protection officer who can be contacted at dpd@ua.es UA has published its data protection policy: https://seuelectronica.ua.es/es/privacidad.html This is supported by the UA's Office of the Vice-President for Academic & Faculty Affairs: https://web.ua.es/en/vr-ordaca/office-of-the-vice-president-for-academic-faculty-affairs.html |

Table 2: Partners legal responsibilities and actions

[6] * Regulation (EU) 2016/679[7] of the European Parliament and of the Council on 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). On 4 May 2016, the official text of the Regulation was published in the EU Official Journal in all the official languages. The Regulation will enter into force on 24 May 2016, it shall apply from 25 May 2018.

---

[6] This regulation has already be mentioned before. Same text.

[7] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

C Roles of GoURMET Partners

Here we describe the role of each partner in terms of data protection, for example who is a data controller, a data provider and a data processor. These follow the definitions of these roles as described by the ICO.[8] Their definition is as follows:

- The GDPR applies to 'controllers' and 'processors'.

- A controller determines the purposes and means of processing personal data.

- A processor is responsible for processing personal data on behalf of a controller.

- If you are a processor, the GDPR places specific legal obligations on you; for example, you are required to maintain records of personal data and processing activities. You will have legal liability if you are responsible for a breach.

- However, if you are a controller, you are not relieved of your obligations where a processor is involved – the GDPR places further obligations on you to ensure your contracts with processors comply with the GDPR.

- The GDPR applies to processing carried out by organisations operating within the EU. It also applies to organisations outside the EU that offer goods or services to individuals in the EU. The GDPR does not apply to certain activities including processing covered by the Law Enforcement Directive, processing for national security purposes and processing carried out by individuals purely for personal/household activities.

In the GoURMET project the BBC and DW take part in decisions about what data is ingested, and how it is processed and stored. The data controller role is therefore shared amongst the user partners. Once the platform is open-source, anyone who downloads and runs the GoURMET platform is both a data processor and a data controller.

Deutsche Welle is a data provider in first instance in the GoURMET project. It supplies textual data from some DW broadcast sources (online, live TV and social media). Its data is subsequently processed in the GoURMET platform and assessed by GoURMET and DW staff. A subselection of the processed content (all originating from previously DW broadcast material) is used for dissemination purposes. As a data provider, Deutsche Welle abides by German and European data protection laws and applies protective measures to protect (personal) data as well as the privacy of its users.

BBC is a provider of Broadcast Media for the purposes of training and testing the GoURMET Platform. The BBC abides by UK and EU data protection laws. The BBC has also drafted their own Data Protection Policies and Information Security Policies and abide by these.

UEDIN, UVA, and UA are primarily data processors. They process data supplied to them by BBC and DW both to train natural language models and to test them. The University of Edinburgh abides by UK and EU data protection laws. The University of Edinburgh has also drafted their own Data Protection Policies and Information Security Policies and abides by these. The Universitat d'Alacant abides by Spanish and European Data Protection laws. The University of Amsterdam abides by Dutch and Europeran Data Protection laws.

---

[8] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/

Partner	Data Provider	Data Processor	Data Controller
BBC	Y		Y
DW	Y		Y
UEDIN		Y	Y
UVA		Y	Y
UA		Y	Y

Table 3: Partners' roles regarding data protection

ENDPAGE

GoURMET

H2020-ICT-2018-2 825299

D7.2 Data Management Plan