Global Under-Resourced MEedia Translation (GoURMET)

H2020 Research and Innovation Action
Number: 825299
D6.3 – Interim Dissemination and Exploitation Report

Nature	Report	Work Package	WP6
Due Date	30/06/2020	**Submission Date**	30/06/2020
Main authors	Kay Macquarrie (DW), Peggy van der Kreeft (DW)		
Co-authors			
Reviewers	Andrew Secker (BBC)		
Keywords	Dissemination, Exploitation		
Version Control			
v0.1	**Status**	Draft	17/06/2020
v1.0	**Status**	Final	28/06/2020

Contents

List of Figures

1 Abstract

This deliverable D6.3 relates to Work Package 6 "Dissemination and Exploitation" and provides an overview of the plans in terms of dissemination and exploitation for GoURMET and reports on the achievements for the reporting period of M0-M18.

This report is divided into two main parts: section 2 (Dissemination) and section 3 (Exploitation). These relate to Tasks 6.1 and 6.2 respectively.

2 Introduction

This deliverable D6.3 relates to Work Package 6 "Dissemination and Exploitation" and provides an overview of the plans in terms of dissemination and exploitation for GoURMET and reports on the achievements for the reporting period of M1-M18.

It is a public document and it will be made available publicly through the GoURMET website.

This report is divided into two main parts: section 2 (Dissemination) and section 3 (Exploitation). These relate to Tasks 6.1 and 6.2 respectively.

Section 2 (Dissemination) shows plans and efforts to inform and inspire other researchers and potential users of the GoURMET components and the platform as well as about the project's intentions and results. GoURMET wants to establish "feedback loops" and engage potential users, early adopters and technology providers to interact with GoURMET products and to continuously provide feedback. The project expects to contribute to and inspire other (low-resource) language translation projects with the goal of building networks and showing that translation technologies can contribute to better and broader news coverage as well as media monitoring activities in under-represented and under-served languages.

Section 3 (Exploitation) outlines the activities aimed at successful exploitation of GoURMET. There are two main ways in which GoURMET can be exploited: Component Based Exploitation (single applications) and Platform Exploitation (GoURMET as a whole). This document outlines progress in both of these areas.

3 Dissemination Plan and Initial Results

This section of the deliverable relates to Task 6.1 (Dissemination), focusing on providing visibility of the project results in the scientific community, the broader community of users and stakeholders and other related research and innovation projects. It provides an initial impact report.

3.1 Overview

This section describes the different dissemination channels, identifies planned dissemination tools, provides a survey of dissemination events that promote related fields of research and suggests journals, press, and mass media channels reaching the targeted audiences.

The purpose of this section of the document is to provide a project dissemination and communication strategy by highlighting targeted groups and communities, define internal dissemination/communication guidelines and procedures, outline the channels, and report on the efforts from M0-M18. Dissemination of the project is a collaborative effort of all project partners and this document describes available tools and partners' responsibilities.

3.2 Dissemination Plan

GoURMET's dissemination and communication strategy is outlined here. It provides the guidelines and procedures for communicating internally and externally and disseminating the results of the project.

3.2.1 Strategy

The project's communication strategy lays down how the communication works internally, with peer researchers and other stakeholders, and with the public at large - and how this communication can be implemented efficiently. It is important to ensure the entire consortium is kept informed of user requirements and priorities, of the development status and details, the challenges and efforts to overcome these, at different levels (technological components, integrated prototype, UX), and user testing and feedback. Equally important is communication towards other stakeholders, the research community, related project groups, the media production and monitoring world, to open up our efforts and achievements to professional communities outside of the consortium and have them profit from the results and provide their feedback. Finally, in order to ensure continued use of the platform, potential end users of the integrated GoURMET solution or any of the technological components (and possibly customisation) need to be made aware of the advantages - and restrictions - of the outcome. The strategy describes what channels and methods will be used for this purpose. It also determines what kind of events will be covered and organised, what materials will be created to describe and show project results and what publications are targeted. The consortium aims to disseminate information regarding the project goals, research, results and experiences to industrial communities (SME and Industry), to academic and research institutions, as well as to the generic audience interested in the project. Project results will be promoted and disseminated during the entire project, as an appropriate prerequisite for a successful exploitation, and at the end of the project in order to engage its stakeholders. The dissemination is both a collective activity managed by the entire consortium and an individual set of actions handled by each single partner

on a local level. All partners are aware that a broad dissemination of results carries a great importance and it is committed to allow access to the results achieved in the project to various kinds of audiences and users: information and research results will be considered to be available to the public unless the consortium decides otherwise.

3.2.2 Target Groups

The consortium plans to target the following diverse range of interested groups as part of the dissemination activities:

- Peer research groups in academia in the EU

- Peer research groups in academic globally

- Peer research groups in other H2020 projects

- Specialist press in research and technology sectors

- Specialist correspondents for science and technology in general press

- International broadcast technology users

- Semantic analysis technology users

- Key stakeholder groups for all project partners

- Diverse industry sectors: broadcasting, media monitoring, translation, business intelligence

- Policy makers and interest groups

- Users of the GoURMET platform, including journalists, editors, monitors, analysts

- Business intelligence consumers

- Educational outreach audiences (the project seeks to encourage young

- people into scientific careers by showcasing the interesting challenges tackled by GoURMET)

3.3 Dissemination: Tools, Materials, Activities

A module-based "dissemination kit" has been developed, adapted to changing requirements of events, target groups and communication channels.

This kit contains a set of key visuals of the GoURMET project, including a project website, providing information about the project mission and objectives, descriptions of the application and use cases, to support the consortium's presence at events and to engage in a dialogue with user group members. Academic and industrial activities will engage stakeholder communities in the project outcome.

3.3.1 Project Identity and Logo

The project identity is expressed by consistent use of the project logo in all its communication. It creates a direct visual recognition of the GoURMET project. Therefore, we aimed to keep the logo is simple, transparent and easy to recognise.

Figure 1: The GoURMET project Logo

The logo appears on the project website, as well as all other external communication, including deliverables, flyers and posters. In addition to the logo, flyers, posters and website have a common style and appearance, to increase project visibility, recognition, and familiarisation.

Figure 2: Variations for light and dark backgrounds

3.3.2 Website and Blog

A project website, www.gourmet-project.eu, was created and made available from the early stages of the project, providing information about the project mission and objectives, descriptions of the application scenarios and use cases, and covering news and results, to support the consortium's presence at events, to promote the project's own events, and to engage in a dialogue with user group members. The website was made such that it is both secure and easy, fast and efficient to add content. It is updated regularly and contains a blog section with short news items or announcements

as well as longer articles. The website also contains links to datasets, links to open source releases as well as public deliverables and publications. The project website is the primary channel of communication towards the outside world.
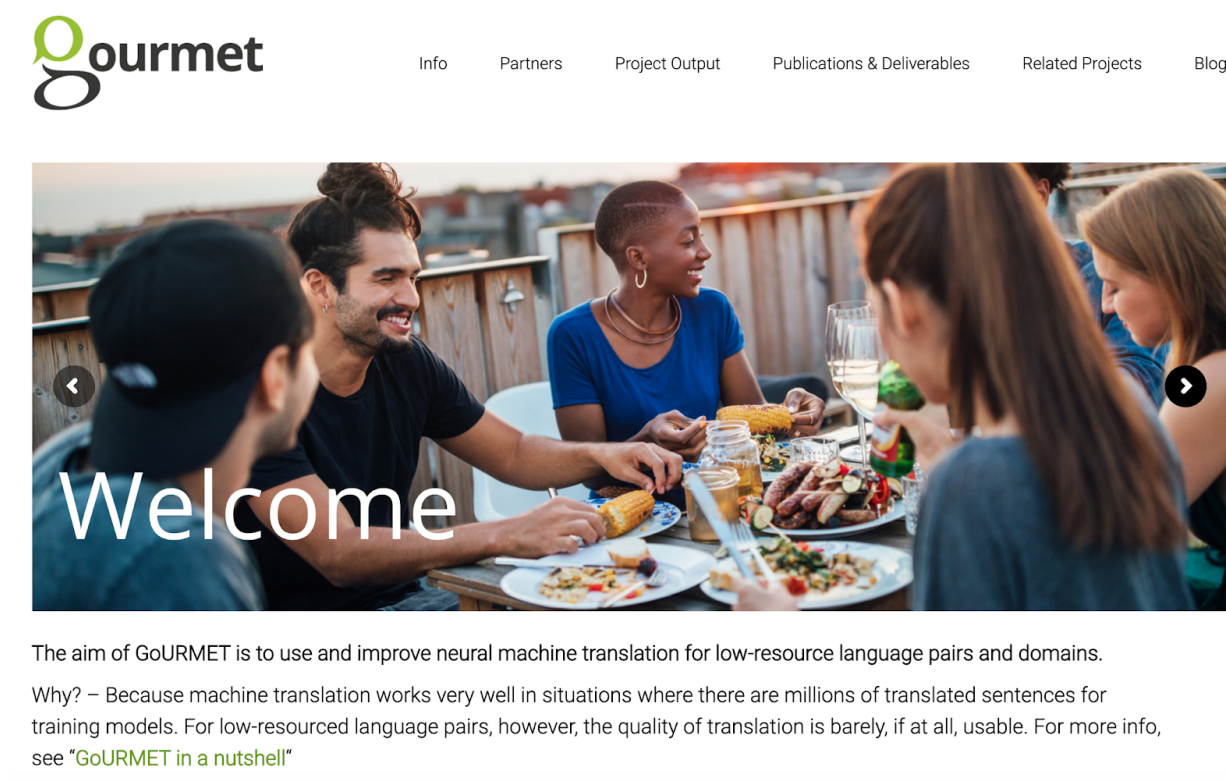
Figure 3: The homepage of the GoURMET project

GoURMET website was jointly designed, launched and enhanced through a collaboration of Deutsche Welle, BBC and University of Edinburgh. A WordPress template is used. The website can be seen at http://gourmet-project.eu. It has a visual image with a diverse set up of people sitting around a table and communicating with each other. The image plays with the double meaning of the project title which also opens association with a "good meal".

Its interactive pages are aimed at engaging the reader. The top navigation offers six main sections: info, partners, output, publications and deliverables, related projects, and blog. The aim is to give the user a good, transparent overview and minimize the number of clicks to arrive at a page. Each partner is represented with a logo and a short description of the organisation. The reader is given an overview of the objectives, the current status of the project and publications. The blog presents news on the technology and application level, announces new releases and describes the project's participation in events. It also regularly features longer articles. The latest GoURMET Tweets also automatically appear on the GoURMET website home page. The section "Data and Model Releases" provides a direct overview to the software that is made available by the project.

The more advanced the technologies and the GoURMET platform are, the more detailed information the "Output" section will contain. The second half of the project will see new enhancements of these pages.

As dissemination leader, Deutsche Welle is in charge of maintaining and updating the GoURMET website, and works closely with all partners to get their input and contributions to make this an

active dissemination channel for the project.

3.3.3 Social Networks

The project also communicates via social networks; currently the focus is on Twitter and LinkedIn.

Twitter Twitter is very useful to inform about activities and achievements and to create ongoing awareness of the developments in the project. Additionally, it can also be used for our followers to easily engage with the GoURMET project, either by following, mentioning, retweeting or commenting on our tweets.

Figure 4: The Twitter presence of GoURMET

LinkedIn LinkedIn is a professional global network where GoURMET expects to meet and address not only target groups from the research communities but also potential users of market areas that we are addressing specifically (translation also in the financial domain). We will also use it as a dissemination channel and syndicate it with articles written for the GoURMET project's website.
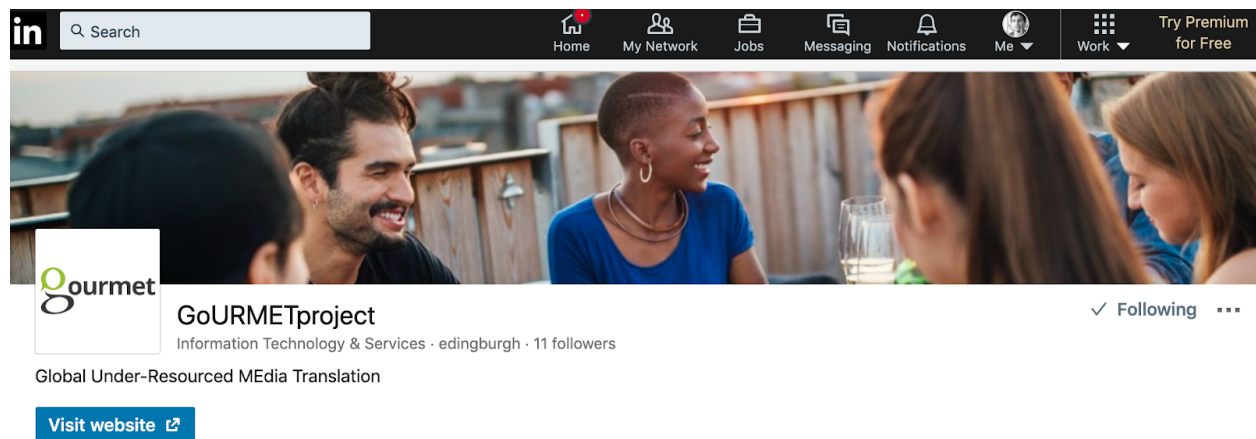
Figure 5: The LinkedIn presence of GoURMET

3.3.4 Printed Media

The project has prepared a set of printed dissemination materials, including Roll-ups, poster and flyers designed and produced specifically for different requirements and target groups.

Roll-up banner A project banner was created and printed to raise visibility at stands during major events.

Two roll-up banners have been produced as a „door-opener" for booth talks. They help to get into contact with interested users. The main objective of the project roll-up is to attract the attention of interested users. In addition with the GoURMET poster, visitors have the chance to easily get a good and basic first impression of the goals and key technologies in GoURMET.

Poster A poster was produced at the very start of the project created by the technical partners, focusing on specific technologies, components, and showing the details of the project. The GoURMET poster primarily is used during conferences and fairs and has two basic functions. In combination with the roll-up (which primarily function is to attract visitors and make them curious to learn more about the project) the function is to enable a first impression of the architecture and the key features of the project.

Flyer The project flyer provides a simple overview of project objectives and achievements suitable for a non-specialist and diverse audience, and to provide overviews of specific project outputs. Flyers are used at smaller and larger events for people to take away and remember the project.

3.3.5 AV Media

Some audiovisual dissemination material will supplement the printed media. These will be produced in the second half of the project. Online walkthroughs / screencasts will be produced at various stages, showing the workflow as used in a media and translation environment. They will be presented online on the website / social media and during events.

3.3.6 Dissemination Events

In the early phase of GoURMET, the consortium outlined which events (conferences, fairs, events, workshops, summer schools) are planned to be attended or organised by partners. Dissemination

Figure 6: GoURMET Banner (Banner 1 "Light version with map", Banner 2 "Dark version focussing the logo, consisting of two units")

will be done throughout the project duration. Topics relate to specific project domains, particular research areas and/or work packages, targeting respective research communities as well as industry. All partners will be required to transfer the knowledge being gained throughout the project to their internal and external communication links.

Described below are the type of dissemination events that are of high interest for the project, and will be covered either through attendance, presentations, publications, panel discussions or poster sessions.

Academic Events Potential academic events include: Scientific workshops in NLP e.g EMNLP (emnlp.org), computational linguistics e.g ACL (aclweb.org/website/acl), machine translation e.g. EAMT (eamt2017.org), speech technology e.g. Interspeech (interspeech2017.org), multimedia indexing e.g. MediaEval (multimediaeval.org)

Industry Events Potential industry events include: Media and broadcast conventions e.g. IBC (ibc.org) Innovation-oriented events, such as LTInnovate (http://www.lt-innovate.org/summit) or STI - Science, Technology and Innovation Indicators (https://sti2017.paris/) Events focusing on journalism and media, e.g. Scoopcamp (http://scoopcamp.de/) and GMF, DeutscheWelle's Global Media Forum (http://www.gmf.de), EBU and ARD meetings, workshops and conferences

Workshops The GoURMET consortium plans two major showcases. In Year 2, a project event

will be either organised separately or as a workshop within one of the relevant community events. This event will showcase the research results in particular. In Year 3 of the project, an event will be jointly coordinated, focusing on application scenarios, use cases and demonstrable project results. It will primarily address the media community and other end users rather than technologists. The focus will be on demonstrating the achievements in the GoURMET application context and discuss planned exploitation and further opportunities.

Hackathon GoURMET will organise a hackathon event in March 2021 with the explicit aim of diverse exploitation of the GoURMET platform and its components with a strong focus on automated monitoring of news in Africa (see DOA 2.2.1). The innovation event will (1) inform participants about the outcome of the GoURMET project, and the software made available for exploitation, (2) provide new technology transfer opportunities through gathering and mentoring a selection of motivated young developers and entrepreneurs around the GoURMET output, and (3) provide a testbed for exploring the flexibility of the GoURMET platform / components.

User Group GoURMET will establish a GoURMET User Group covering a broad range of stake-holders with interests in the GoURMET platform and use cases. An initial set of User Group members has been established, and WP6 will in part be concerned with extending and strengthening the User Group.

Initial members are: Our initial members are Quorate (quoratetechnology.com), Translators without Borders (https://translatorswithoutborders.org/), The Norwegian Broadcasting Corporation (https://www. nrk.no/about/), and Microsoft Research.

Collaboration Activities Concertation meetings will be facilitated where certain themes relevant to GoURMET and to a number of other European projects are discussed. The aim for GoURMET is to communicate with other ongoing European projects on a regular basis, to exchange ideas and concepts, to exploit their results and finding, to broaden the potential user base.

This includes, but is not limited to, some projects which have already been identified as covering related technologies, such as

QT21++ - Research focussed on NMT, informed by the translation needs of the Connecting Europe Facility (CEF), EBU EuroVox, ELG (European Language Grid), The Alan Turing Institute (UK), ParaCrawl Project, Pret-a-LLOD (Ready-to-use Multilingual Linked Language Data for Knowledge Services across Sectors), Bergamot (Browser-based Multilingual Translation), EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media), ELITR (European Live Translator), AI4EU (A European AI On Demand Platform and Ecosystem), COMPRISE (Cost-effective, Multilingual, Privacy-driven voice-enabled Services)

 Engagement with innovation communities We will engage with innovation communities at many levels. At the European level we shall work with CITIA (the Conversational Interaction Technology Innovation Alliance), LT-Innovate (the European Association of the Language Technology Industry) and META (the Multilingual Europe Technology Allicance) through events such as the annual LT-Innvoate Summits, the META forum and the EU ICT Conference.

The GoURMET website will list all dissemination events that will be covered in the project's lifetime and, in addition, announce major events that are closely related to and of particular interest to the project. Whenever possible, events covered will be announced briefly as an upcoming event in the blog section, then reported on in some more detail, be included in the list of events. Presentations or papers presented will be made available on the project website to the extent possible.

3.3.7 Papers and Publications

Both technical and user partners will engage actively in dissemination through the publication of papers to inform the academic and industrial/media world. The goal is to encourage joint publication as much as possible, with collaboration of different project partners and showing joint efforts and results.

References to all publications produced as part of the project will be listed to on the GoURMET website (if possible, they will also be made accessible from there).

Conference Papers Conference papers are a major dissemination channel, in particular for academic partners, to report on their research findings. As mentioned above, joint publication involving several partners is encouraged, to stress the enriched outcome through collaboration. Each partner will suggest and select a number of conferences in their specific area, spread over the entire project duration. All technology covered should be presented in conference papers. **Press Releases** Over the course of the project we envisage several press releases to be issued by different project partners, announcing major achievements. This depends on internal organisational procedures and the authorised use of press releases. Many organisations have moved from press releases to newsletters, blog and other similar information channels. **Blog** The GoURMET blog on the project website provides brief news reports on issues related to GoURMET coverage (technologies, media applications, translation technologies, media monitoring, workflows, etc.), as well as longer technical articles focusing on a specific technology, a component, a research area and user findings. Each partner will contribute to the blog with at least one long article, so that it will cover the entirety of the GoURMET technological base. Blogs hosted on other platforms will also be used to raise visibility and reach a wider audience. This includes for instance Deutsche Welle's Innovation blog (http://blogs.dw.com/innovation/) or the BBC News Labs website (https://bbcnewslabs.co.uk/).

Figure 7: GoURMET Poster

Figure 8: GoURMET Poster (unfolded)

4   Initial Dissemination Report

The current section presents the current status of the different communication channels tools, materials and activities, as well as descriptions of major achievements.

4.1   Website

The website was continuously being updated throughout the first part of the project and serves as the major online communication channel. It has been designed as a central information place with key information about the latest developments and achievements of the project and news about the translation technologies market.

More than 1700 visits were registered throughout the first course of the project (M0-M18). Connections came from all over Europe, especially Germany and Spain, but also from the United States, Africa, South America and Asia.
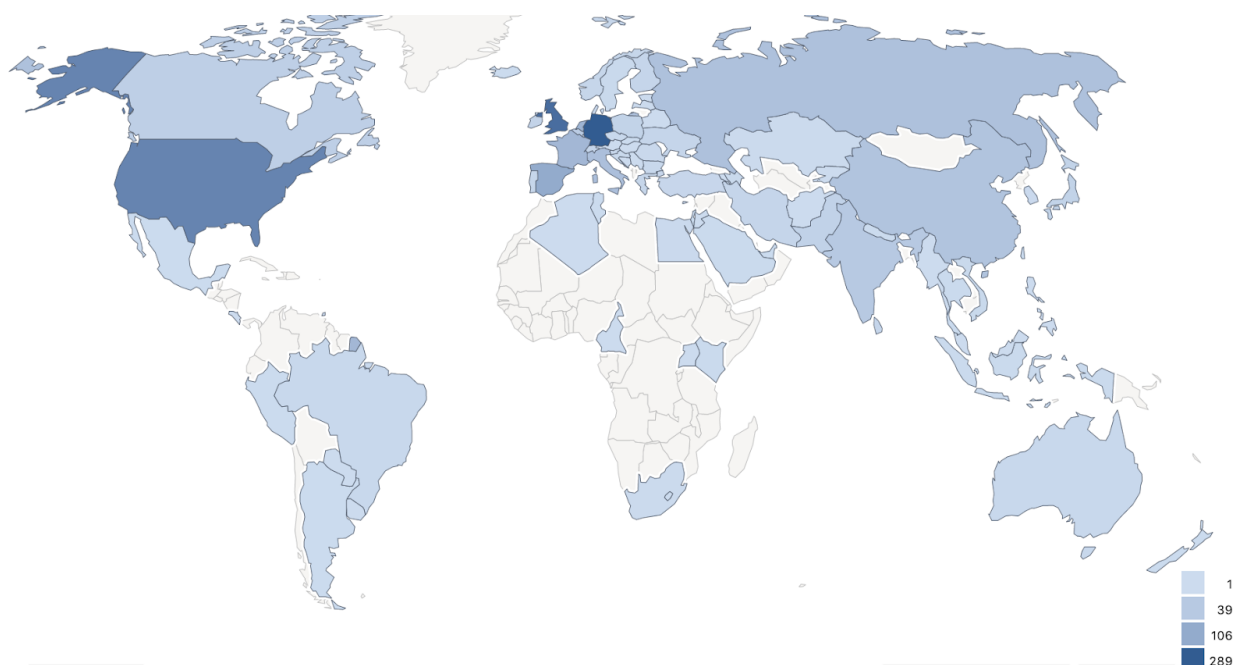
Figure 9: Screenshot showing the global distribution of the website's visits (1699 visits in total)

The page that has been most frequently visited by users, with over 1700 views, was the "GoURMET Project" page. To make the most of this information gained by doing the website's analytics, the content of the page was optimised towards the second half of the first period. Beside the overview of the output of the project - including data, software and models - we also provided their link to GitHub for those available, for immediate use by interested users.

4.2   Twitter

Twitter has been intensively used to inform communities about GoURMET activities and achievements. By selected tweets and retweets, we created an ongoing awareness of the developments in the project in specific target groups. Additionally, Twitter was used for our followers to easily

engage with the GoURMET project, either by following, mentioning, retweeting or commenting on our tweets. By the end of the first period, the Twitter account has more than 125 followers.
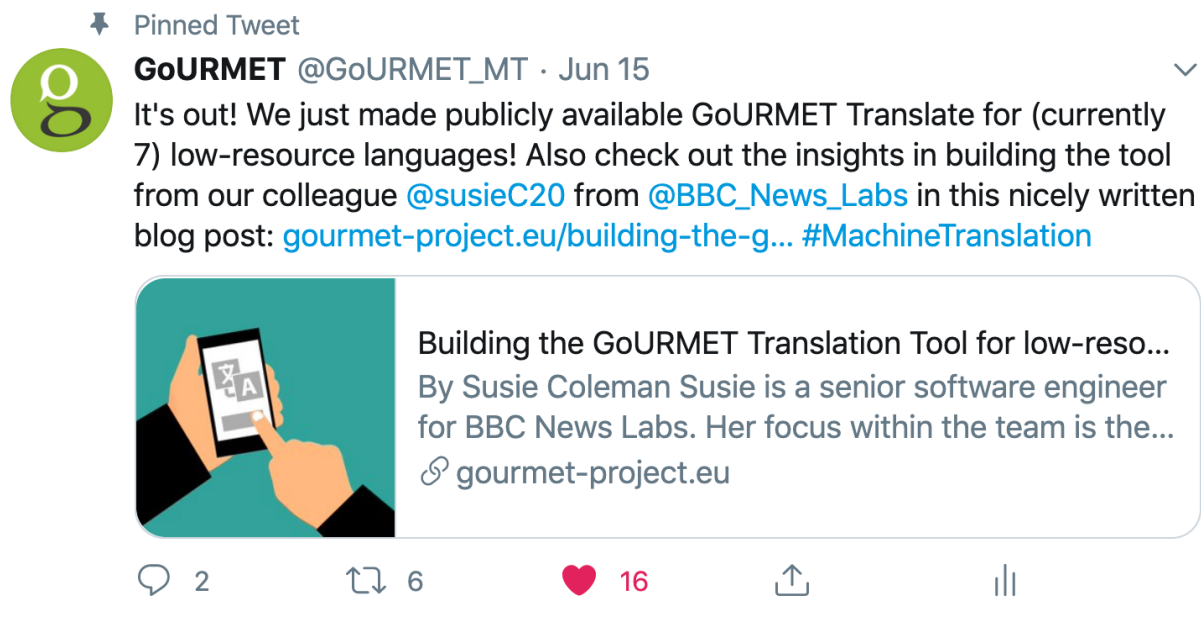
Tweets with a photo - as the one shown below - usually gets more interactions. Whenever a tweet was prepared by a partner, it was highly recommended to attach an image.

Figure 10: One of the most successful GoURMET tweets

The statistical tool also shows the potential reach of the tweet mentioned above. The total reach is calculated by counting together all followers of the accounts that retweeted one ME's tweet. In the case of the GoURMET announcement tweet, the potential reach was 30.777 people.

Figure 11: This tweet has a potential reach of over 30.000 people

The Twitter account is thus a central channel to promote GoURMETs' achievements and engage with an interested audience.

4.3   LinkedIn

The LinkedIn account of GoURMET was primarily established to promote workshops and user days including the hackathon. Although it currently has a rather limited amount of members, it should be noted, that updates in the respective timelines of its members can reach a significantly larger number of people. So, this channel will be used for major news and updates on the project specifically in respect to its output formats for further exploitation purposes.

4.4 GoURMET Representation at Events M0-M18

The following two chapters list the events at which the GoURMET partners have represented the project during the M0-M18 period. It should be noted, that noting that the COVID-19 pandemic severely impacted our attendance at events in 2020.

4.4.1 GoURMET Representation at Academic Events M0-M18

GoURMET has actively participated in the following academic events (see Table 1).

4.4.2 GoURMET Representation at Industry Events M0-M18

GoURMET has actively participated in the following industry events (see Table 2).

Figure 12: Susie Coleman introduces GoURMET at BBC News Labs

DATE	PLACE	EVENT	DISSEMINATION ACTIVITY	PARTNER
30.07.2019	South Africa, University of KwaZulu-Natal	Machine Learning Workshop	Tutorial One day tutorial on machine learning and data curation mainly for neural machine translation	UEDIN Alexandra Birch
2.08.2019	Florence, Italy	RepL4NLP 2019	Presentation	UA Bryan Eikema, Wilker Aziz
1-2.08.2019	Florence, Italy	WMT 2019	Presentation (top constrained system for English-to-Gujarati)	UEDIN Rachel Bawden, Faheem Kirefu, Antonio Valerio, Miceli Barone
1-2.08.2019	Florence, Italy	WMT 2019	Presentation (best performing submission for the English-Kazakh language pair)	University of Alicante Víctor M. Sánchez-Cartagena
2.11.2019	Hong Kong	WNGT2019	Organiser at Workshop on Neural Generation and Translation	UEDIN Alexandra Birch
4-6.11.2019	Hong Kong	EMNLP2019	Conference, Co-presented paper	UEDIN Alexandra Birch
21-23.01.2020	Alcant, Spain	GoURMET tutorial on variational inference	Three day tutorial on variational inference. Tutorial organised by GoURMET and the Universitat d'Alacant's Institute for Computer Research.	UA Bryan Eikema, Wilker Aziz

Table 1: Academic Events 2019 - June 2020

DATE	PLACE	EVENT	DISSEMINATION ACTIVITY	PARTNER
3-5.4.2019	Nairobi, Kenya	Toward a Network of Excellence in Artificial Intelligence for Development (AI4D) in sub-Saharan Africa	Planning for creating data and capacity for natural language processing in African languages	UEDIN
11.06.2019	Alicante	Presentation of the project to EUIPO	Presentation and Workshop	UA
11-12.06.2019	London, UK	newsHACK: Tools for Multilingual Newsrooms and Talk	Hackathon Team	BBC, DW, UEDIN, UA
25.06.2019	London, UK	Managing and Translating Audiovisual Content	Invited talk	BBC
19-23.8.2019	Dublin, IR	Machine Translation Summit 2019	Poster Presentation	DW, UEDIN, UA
08.10.2019	London, UK	Ada Lovelace Day 2019	Talk	BBC
09.10.2019	Brussels, BE	META FORUM - ELG and the European LT Industry	Booth and Presentation	BBC, DW
19.11.2019	Online	European Broadcasting Union's News Report 2019	Report	BBC, DW
27-28.11.2019	Salford, UK	EBU Technology in Production and Distribution Workshop	Workshop	BBC
21-22.02.2020	Bern, AT	SwissInfo Hackathon	Hackathon	DW
08.03.2020	London, UK	International Women's Day special	Podcast	BBC
16.05.2020	Marseille, FR	International Workshop on Language Technology Platforms (Event cancelled, but Proceedings are available)	Workshop	UEDIN

Table 2: Industry Events 2019 - June 2020

4.5 List of Publications M0-M18

See the openAIRE GoURMET page for further details on these publications.

1. Estimating post-editing effort: a study on human judgements, task-based and reference-based metrics of MT quality. Carolina Scarton, Mikel L. Forcada, Miquel Esplà-Gomis, and Lucia Specia. IWSLT 2019, link

2. Block neural autoregressive flow. De Cao, Nicola, Ivan Titov, and Wilker Aziz. Arxiv 2019, link

3. Widening the Representation Bottleneck in Neural Machine Translation with Lexical Short-cuts. Emelin, Denis; Titov, Ivan; Sennrich, Rico. WMT 2019, link

4. Findings of the 2019 Conference on Machine Translation. Loïc Barrault, Ondřej Bojar,Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Shervin Malmasi, Christof Monz, Mathias Müller, Santanu Pal, Matt Post and Marcos Zampieri. WMT 2019, link

5. On the Importance of Word Boundaries in Character-level Neural Machine Translation. Duygu Ataman. WNGT 2019, link

6. Towards a Multi-view Language Representation: A Shared Space of Discrete and Continuous Language Features. Arturo Oncevay, Barry Haddow and Alexandra Birch. TyP-NLP 2019, link

7. Auto-Encoding Variational Neural Machine Translation. Bryan Eikema and Wilker Aziz. RepL4NLP 2019, link

8. Interpretable Neural Predictions with Differentiable Binary Variables. Joost Bastings, Wilker Aziz, and Ivan Titov. ACL 2019, link

9. The Universitat d'Alacant submissions to the English-to-Kazakh to the WMT19 News Translation Task. Víctor M. Sánchez-Cartagena, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez. WMT 2019, link

10. The University of Edinburgh's Submissions to the WMT19 News Translation Task. Rachel Bawden, Nikolay Bogoychev, Ulrich Germann, Roman Grundkiewicz, Faheem Kirefu, Antonio Valerio Miceli Barone, Alexandra Birch. WMT 2019, link

11. Effective Estimation of Deep Generative Language Models. Tom Pelsmaeker and Wilker Aziz. ACL 2020, link

12. A multi-source approach for Breton-French hybrid machine translation. Víctor M. Sánchez-Cartagena, Mikel L. Forcada, Felipe Sánchez-Martínez. EAMT 2020, link

13. An English-Swahili parallel corpus and its use for neural machine translation in the news domain. Felipe Sánchez-Martínez, Víctor M. Sánchez-Cartagena, Juan Antonio Pérez-Ortiz, Mikel L. Forcada, Miquel Esplà-Gomis, Andrew Secker, Susie Coleman, Julie Wall. EAMT 2020, link

14. Improving Massively Multilingual Neural Machine Translation and Zero-Shot Translation. Biao Zhang, Philip Williams, Ivan Titov, Rico Sennrich. ACL 2020, link

15. MultiWord Expression Aware Neural Machine Translation. Andrea Zaninello, Alexandra Birch LREC 2020, link

16. Toward Making the Most of Context in Neural Machine Translation. Zaixiang Zheng, Xiang Yue, Shujian Huang, Jiajun Chen and Alexandra Birch. IJCAI 2020, link

17. A Latent Morphology Model for Open-Vocabulary Neural Machine Translation. Duygu Ataman, Wilker Aziz, Alexandra Birch. ICLR 2020, link

18. Document Sub-structure in Neural Machine Translation. Radina Dobreva, Jie Zhou and Rachel Bawden. LREC 2020, link

19. Document-level Neural MT: A Systematic Comparison. António Lopes, M. Amin Farajian, Rachel Bawden, Michael Zhang and André T. Martins. EAMT 2020, link

20. Architecture of a Scalable, Secure and Resilient Translation Platform for Multilingual News Media. Susie Coleman, Andrew Secker, Rachel Bawden, Barry Haddow and Alexandra Birch. IWLTP 2020, link

4.6 Awards

The paper "Towards a Multi-view Language Representation: A Shared Space of Discrete and Continuous Language Features" (Link) from Arturo Oncevay, Barry Haddow and Alexandra Birch won as Best Paper at TyP-NLP 2019.

Description: The First Typology for Polyglot NLP was held at ACL 2019, aiming to promote the research development in linguistic typology for multilingual NLP tasks, such as machine translation. In our awarded submission, we combine language-level representations from typological knowledge bases, and task-driven learned embeddings of languages from a multilingual machine translation model. Our approach demonstrated that it is possible to fuse both kinds of representations with minimal information loss.

4.7 GoURMET in the Media

This is a list of GoURMET appearance in the media, mainly as articles in newspapers or posts in partner blogs.

1. GoURMET Kickoff

   - Coverage in news agencies and digital newspapers: La Vanguardia, Europa Press, Información, 20 Minutos, Intercomarcal, Network of Valencian Universities for Promoting Research & Innovation, NovaCiencia
   - Blog post on BBC News Labs, https://bbcnewslabs.co.uk/projects/gourmet/
   - Blog post on DW Innovationblog, https://blogs.dw.com/innovation/new-hlt-project-gourmet-to-improve-machine-translation-for-low-resource-languages-and-domains/

2. Provision of a Report to European Broadcasting Union's News Report 2019

    - Report by BBC and DW, November 2019

3. Release of GoURMET Translate

    - Blog post on DW Innovations Blog

Figure 13: Screenshot of an article from Europapress on the GoURMET Kickoff

5 Exploitation Strategy

This section of the deliverable relates to Task 6.2 (Exploitation), focusing on capturing the information required and exploring options to ensure the outputs of the GoURMET project can be exploited by the partners themselves and others. As the exploitation roadmap, this document sets out the activities required for successful exploitation of GoURMET and reports on progress so far.

- Exploitation Committee The role of the Exploitation Committee is to set and execute the exploitation strategy of the consortium as well as to ensure that IPR Management is being carried out appropriately.

- IPR Management The Exploitation Committee has mechanisms in place to manage the record of IPR being contributed to the project by the consortium members.

Building on this foundation there are two main ways in which GoURMET can be exploited:

- Component-Based Exploitation (for other applications) The components and processes that power GoURMET have a high potential value as improvements for existing services, or as the basis of other new services.

- Platform Exploitation (GoURMET as a whole) During the project the partners will investigate the options further into the revenue potential of the best options.

5.1 Exploitation Committee

The project established an Exploitation Committee to coordinate the management of IPR and to set and execute the exploitation strategy of the consortium. It was agreed there would be one representative from each project partner with initial membership confirmed as follows:

- Alexandra Birch, University Edinburgh

- Kay Macquarrie, Deutsche Welle

- Wilker Aziz, University Amsterdam

- Mikel L. Forcada, University Alicante

- Andrew Secker, BBC

The Committee will meet on a monthly basis. All members will be invited to contribute to a workshop to be held in October 2020 that explores potential use of the GoURMET project outputs by using combinations of components to address known and perceived user needs informed by partners' own knowledge and information gathered from user group meetings.

5.2 IPR Management

The Exploitation Committee agreed each work package should be broken down into an agreed set of components and that the following information should be captured for each component / model:

- Component name

- Inputs from

- Outputs to

- Component lead partner

- Component contributors

- Brief description

- What it does (more detail)

- How it works (more detail)

- Key innovative aspects

- Potential applications

- Software & IPR status

- Terms & conditions of use

- Performance requirements

- Further documentation

- Alternatives

- Key contact(s)

In particular the Committee sought to address

- Potential applications of any components aside from GoURMET

- The names and details of any libraries used within components and their IPR status

- Answers to key questions developers wishing to exploit the component are likely to have such as: What spec machines required to run the code?

- The name of and links to alternative open source components if a component will not be part of an open source release

5.3 Component Based Exploitation

The Exploitation Committee agreed the first task for the project towards exploitation will be to gather detailed information on all components to ensure as much information as possible.

As part of GoURMET's efforts to increase resources and tools available for low-resource machine translation, we have released many of the corpora and software created during the project.

NAME	REPOSITORY
English–Swahili	http://data.statmt.org/gourmet/corpora/GoURMET-crawled.en-sw.zip
English–Turkish	http://data.statmt.org/gourmet/corpora/GoURMET-crawled.en-tr.zip
English–Amharic parallel corpus and Amharic monolingual corpus	http://data.statmt.org/gourmet/corpora/GoURMET-crawled.en-am.zip
English–Kyrgyz parallel corpus and Kygryz monolingual corpus	http://data.statmt.org/gourmet/corpora/GoURMET-crawled.en-ky.zip
Kyrgyz–Russian	http://data.statmt.org/gourmet/corpora/GoURMET-crawled.ky-ru.zip
PMIndia – Parallel corpus of languages of India	http://data.statmt.org/pmindia/

Table 3: Parallel Copora - Collection of OpenSource Data and Software Releases

NAME	REPOSITORY
Morphological segmentation using Apertium resources	https://github.com/transducens/smart-segmentation
LASER train (language-agnostic sentence embeddings)	https://github.com/transducens/LASERtrain
Top-level-domain crawler	https://github.com/transducens/linguacrawl/

Table 4: Software - Collection of OpenSource Data and Software Releases

NAME	REPOSITORY	DESCRIPTION
Direct Assessment – Sentence Pairs Evaluation Tool	https://github.com/bbc/gourmet-sentence-pairs-evaluation	The goal of Direct Assessment is to evaluate a translation model by asking a human to compare the quality of a machine translated sentence to a human translated sentence where the human translation is assumed to be the gold standard.
Gap Fill Evaluation Tool	https://github.com/bbc/gourmet-gap-fill-evaluation	The goal of Gap Fill is to evaluate a translation model by asking a human to fill in the gaps in a sentence that has been translated by a human using the machine translation of the same sentence as a guide to what words should go in that sentence.

Table 5: Evaluation Tools (also see D5.3) - Collection of OpenSource Data and Software Releases

5.4   Sustainable Platform Exploitation

There are two main use cases for GoURMET as an integrated product. Further, use cases will be explored at the next full project meeting where the Exploitation Committee will convene to explore potential use of the GoURMET project outputs. The two use cases currently under consideration are as follows:

5.4.1   BBC Integrated Tools

During M0-M18 of the project, the BBC focus has been on the creation of the infrastructure that will ensure the outcomes of the research are usable and exploitable both within the BBC and by the wider project partners (see deliverable D5.3). This is now complete.

The BBC's exploitation focus is threefold. Firstly create tools and experiences to support content publication by allowing the movement of content smoothly between languages, secondly allow enhanced oversight for internal teams across languages, thirdly support media monitoring.

The BBC's first prototype, focussed on the first exploitation use case, is well progressed. This is a tool which will enable the real-time translation of BBC News live pages between languages. Live pages are a particular type of page found on the BBC News website and available via mobile apps which are updated with short pieces of text very regularly. As such, they are most commonly used to cover breaking or fast moving stories. This dynamic nature lends well to automatic translation while the important nature of the breaking stories they typically are used for, tend to ensure the subjects have wide geographical spread, and therefore language spread of interest. The Coronavirus live page pictured below, for example, is published in English but has worldwide interest.
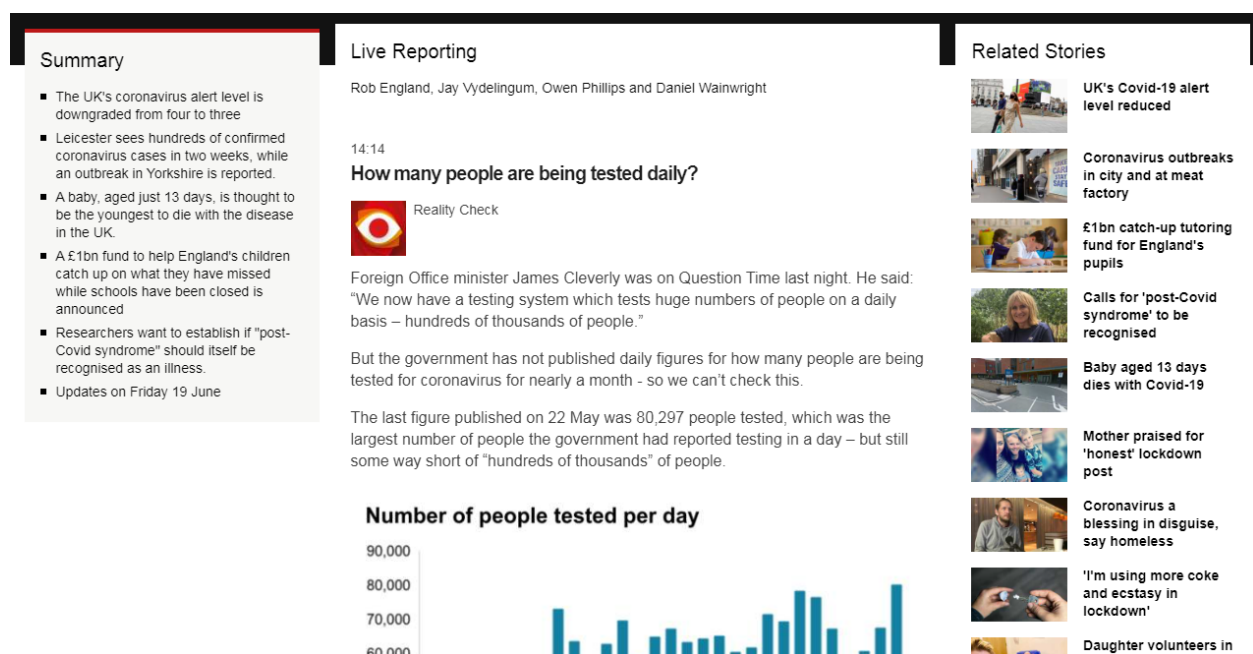

Figure 14: An example of the BBC News Coronavirus live updates page

The prototype tool integrates seamlessly with the live pages production system, allowing journalists to see candidate posts in their language as drafts, ready for any post-editing required and one-click publication to the live website. This prototype will be finished in early Q3 2020.

The next BBC prototype is already under planning and will be a dashboard style tool to allow news teams to see which news stories in other languages are performing particularly well, allowing Editors to make informed editorial decisions regarding areas of interest for their own Journalistic teams. This prototype is likely to be complete in late Q3 or Q4 2020.

The description of work cited integration of GoURMET translation models with the BBC News Labs' Alto tool as a key piece of exploitation (See Task 5.3). In the intervening period, the BBC News Labs has been working with the European Broadcasting Union (EBU) on a project called EuroVox (https://tech.ebu.ch/eurovox). The aim of EuroVox is to create a cloud-based language re-versioning tool for video media in much the same way as the BBC's Alto prototype but on a bigger scale. EuroVox is a shared initiative amongst a number of European broadcasters, including project partner DW. The decision has been taken to focus on integration of the GoURMET translation models with EuroVox rather than Alto. This has the potential to significantly increase impact of the project as instead of GoURMET models being used for the BBC-internal Alto tool, they may be exploited in a Europe-wide tool, potentially accessible to all EBU members - 70 members, representing 116 organisations. The BBC will lead this integration work, resulting in this significant exploitation outcome, to take place in Q4 2020 - Q1 2021.
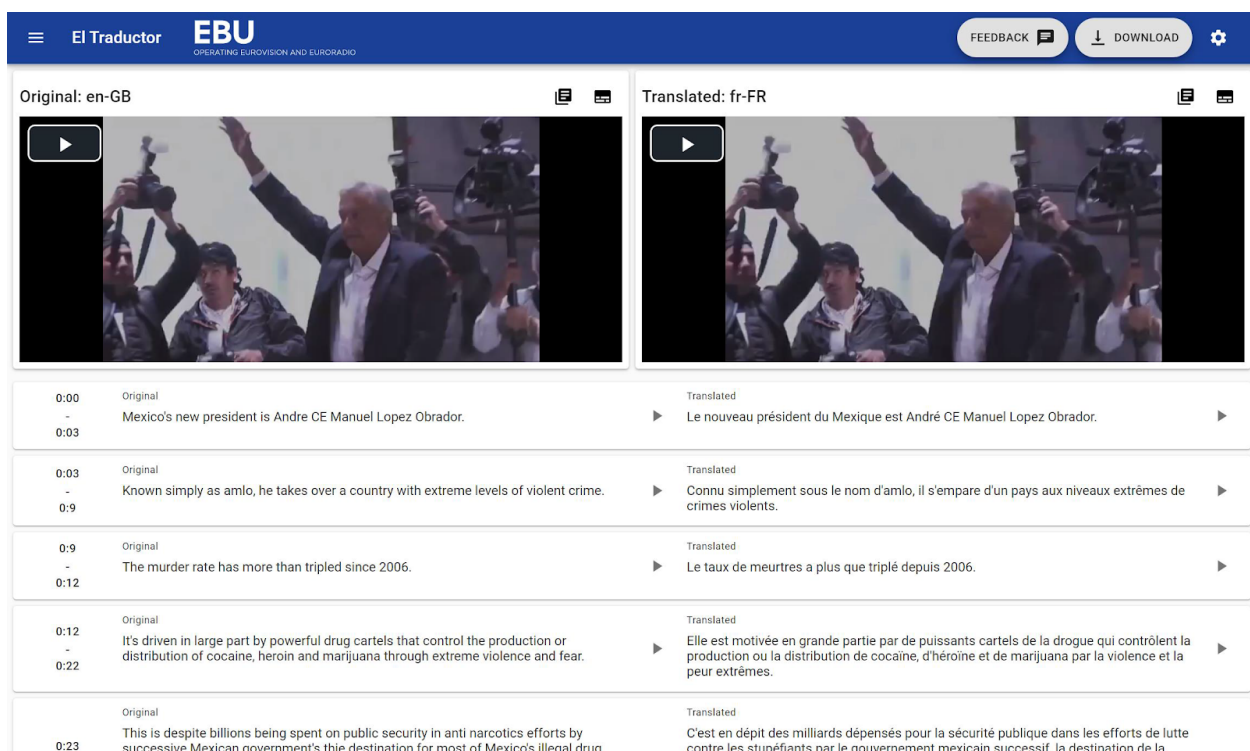
Figure 15: The EBU EuroVox "El Traductor" interface

5.4.2 DW Integrated Tools

New language models from GoURMET will improve DW internal translation and monitoring workflows and increase the coverage of news output in low-resourced languages. Currently, DW is in the process of integrating Bulgarian, Swahili, Turkish, Amharic and Serbian language models into the central HLT platform used in DW. They will be used for publishing more text articles and video items in these languages through automation-supported translation from original

English-source items (English being the primary source language in Deutsche Welle). In addition, a considerable increase in coverage is expected by using the tools for subtitling English-language videos in the target languages listed above. An efficient workflow of semi-automated subtitling can reduce the time needed by over 50 Percent compared to traditional manual translation and subsequent respeaking. The GoURMET tools will be available in the DW HLT platform and appear next to off-the-shelf translation tools, including Google Translate and eTranslation. This enables professional translators to compare the quality output between the different tools and choose the one that fits their target text best. As GoURMET tools are locally installed, it also serves as a cost-saving alternative, compared to paying services, of particular importance for large volumes of data.

6 Conclusion

The dissemination and exploitation goals that the GoURMET project set for itself from the start have been widely covered during the first project period (M0-M18). Based on the dissemination strategy, GoURMET has exploited various communication means and continuously communicated and published news, activities and achievements. Both in online media, including the website, and social media, like Twitter, as well as directly in face-to-face meetings and conferences. All code developed in the course of this period is well-documented and major parts of it are accessible Open Source via the dedicated webpage under https://gourmet-project.eu/data-model-releases/ or for more immediate access via the corresponding GitHub accounts, which are linked from there.

With 10 dissemination events attended and 17 publications produced, the GoURMET project has placed emphasis on dissemination activities during this reporting period. The work on Dissemination and Exploitation during M0-M18 provides a solid foundation on which to build. Within the project the partners have consulted internally as to how output of the project may be best exploited to meet real business needs, with an initial focus on the BBC and DW use cases. During M19-M36 the Exploitation Committee will continue to refine exploitation plans for the GoURMET platform and services while further dissemination activities will continue to ensure GoURMET technology is socialised with the wider community.

ENDPAGE

GoURMET

H2020-ICT-2018-2 825299

D6.3 Interim Dissemination and Exploitation Report