



Deliverable D3.2.28 Task 8

ELRC Workshop Public Report for Malta



Author(s):	Michael Rosner, Donatienne Spiteri, Mark Vella
Dissemination Level:	Public
Version No.:	V2.5
Date:	2020-02-25



Contents

1	Executive Summary	3
2	Workshop Agenda	4
3	Summary of Content of Sessions	6
3.1	Welcome and introduction	6
3.2	Welcome by the EC	6
3.3	Connecting public services across Europe: ambition and results so far	6
3.4	National initiatives for digital public services and (open) data	7
3.5	CEF in Malta: an outlook into current and future challenges – Panel session	8
3.6	The CEF eTranslation platform @ work	9
3.7	ELRC in Malta	9
3.8	Can language data be shared and how?	10
3.9	Preparing and sharing data with the ELRC repository – and what happens next	10
3.10	Identifying and managing your data: Questions & Answers	11
3.11	Concluding Remarks	11
4	Synthesis of Workshop Discussions	12
4.1	ELRC and Open language Data in Malta	12
4.2	Success stories and lessons learnt	12
5	Workshop Participants	13
5.1	Workshop Presentations	13

ELRC Workshop Report for Malta

1 Executive Summary

This document reports on the ELRC+2 seminar in Malta, which took place in Valletta, on November 1st, 2019 at the Valetta Campus of the University of Malta. It includes the agenda of the event (section 2) and briefly sums up the content of each presentation and of the panel workshop session (sections 3 & 4).

The seminar was attended by 98 people, amongst whom 80 came from Public Administrations. The remainder came from a variety of sectors including the EC, the Council of the European Union, the Maltese Permanent Representation to the EU in Brussels, the ELRC consortium, academia, research staff, and secretarial support staff. The dedicated event page can be found at http://lr-coordination.eu/l2malta_agenda.

The discussion panel on *Current and Future challenges to the role of CEF in Malta* was moderated by representatives from the University's Department of Artificial Intelligence and the Department of Maltese, whose research work is highly relevant to the topic of "CEF in Malta: an outlook into current and future challenges".

The workshop ended with a short discussion session, followed by conclusions.

2 Workshop Agenda

<i>Internal reference number</i>	ELRC Malta seminar – 1 November Aula Prima, Valletta Campus	
	08:00 – 09:00	Registration
S0.1	09:00 – 09:10	Welcome and introduction <i>Dr Donatienne Spiteri, Legislation Unit - Office of the Attorney General ELRC Public Services National Anchor Point</i>
S0.2	09:10 – 09:15	Welcome by the EC <i>Dr Elena Grech - Head of the European Commission Representation in Malta</i>
	Session 1. Connecting a multilingual Europe: European context and local needs	
S1.1	09:15 – 09:35	Connecting public services across Europe: ambition and results so far <i>Ms Gudrun Stock, Deputy Head of Unit, DG CONNECT - European Commission Unit G3 - Accessibility, Multilingualism & Safer Internet</i>
S1.2	09:35 – 09:55	Actions and policies for digital public services in Malta <i>Ms Joyce Dimech - Permanent Secretary (Strategy & Implementation), OPM</i>
	09:55 – 10:15	Open data initiatives in Malta <i>Dr Joseph S. Azzopardi, Enterprise Data Manager at Malta Information Technology Agency (MITA) and Data Governance Manager for the Government of Malta</i>
S1.3	10:15 – 11:00	CEF in Malta: an outlook into current and future challenges – Panel session Moderator: University of Malta <i>Dr Michael Spagnol, Department of Maltese Dr Claudia Borg, AI - Faculty of Information & Communication Technology</i>
		Panelists: <ul style="list-style-type: none"> • <i>Andre Arciola, Head, Quality and Training, Servizz.gov Agency</i> • <i>Keith Cortis, Senior Solutions Architect - MITA</i> • <i>Albert Vella, CIO – Ministry for Finance</i> • <i>Donald Mangion, CIO – Ministry for Justice, Culture and Local Government</i> • <i>Dr Joseph S. Azzopardi, Enterprise Data Manager (MITA) and Data Governance Manager for the Government of Malta</i> • <i>Greta Attard, Software Developer - Malta Digital Innovation Authority (MDIA)</i>
S1.4	11:00 – 11:30	Coffee Break

Session 2. Engage: hands-on data

S2.1	11:30 – 12:10	The CEF eTranslation platform @ work <i>Michael Jellinghaus, Directorate General for Translation (DGT) - European Commission</i> <i>Mark Vella, Language Officer – European Commission Representation in Malta</i>
S2.2	12:10 – 12:25	ELRC in Malta <i>Michael Rosner, Department of Artificial Intelligence - University of Malta</i> <i>ELRC Technology National Anchor Point</i>
S2.3	12:25 – 12:50	Can language data be shared and how? State of play in Europe and beyond <i>Dr Prodromos Tsiavos, ELRC legal expert</i>
	12:50 – 13:50	Lunch Break
S2.4	13:50 – 14:25	Preparing and sharing data with the ELRC repository – and what happens next <i>Khalid Choukri, ELDA CEO - ELRC Consortium</i>
S2.5	14:25 – 15:00	Question time <i>All available for answers: speakers, panelists, NAPS etc.</i>
S2.6	15:00 – 15:10	Conclusions <i>Donatienne Spiteri, ELRC PA NAP</i> Coffee Break and networking

3 Summary of Content of Sessions

3.1 Welcome and introduction

Donatienne Spiteri, Malta's public services NAP, welcomed participants to the 2nd ELRC Malta seminar. She explained that *eTranslation* is an automatic translation service, offered by the European Commission to all officials working in any of the public administrations across the EU. As an integral part of the Connecting Europe Facility, *eTranslation* has the aim of pulling down language barriers that could hinder the creation of an effective digital single market across Europe.

She emphasised the need of qualitative and quantitative data for translation engines to be trained and for translation results to be constantly fine-tuned. The public administration could play a crucial role in ensuring that a variety of styles and sector-specific terminologies are injected into these 'data hungry' engines and for translation results to be constantly improved.

The message is clear: let's give a new life to our documents. There is a lot of awareness about maximising the value of 'resources' generally, but not enough awareness about *language resources*. The public administration's policy documents, correspondence, websites etc. constitute the language resources we are after. She emphasised the need for a joint effort to share and collect data, which could also contribute to the evolution of the Maltese language itself.

Ms Spiteri also stated that one of the aims of *eTranslation* is to ensure 'language equality', that is, the need to provide all languages, large or small, with effective tools that could address their current needs including their digital presence. Maltese is a small language, at risk of digital extinction. She referred to a report issued in 2018 by the European Parliament entitled "Languages and the Digital Single Market" which stated that although new technologies can facilitate language learning, translation and interpretation, however paradoxically, the smaller languages that could benefit most from these technologies are the least resourced in data.

She concluded by saying that she hoped the seminar would not only be of interest to participants, but also of practical relevance to their daily work.

3.2 Welcome by the EC

Dr Elena Grech, Head of the European Commission Representation in Malta, welcomed the audience and mentioned that EU policies strive to reconcile the preservation of Europe's rich cultural diversity against the demands for increasing levels of harmonisation in the service of high standards, low costs and greater efficiency. Technology in general and Machine Translation in particular have a very important role in helping to bring this reconciliation about when it comes to preserving language and cultural diversity whilst lowering language barriers.

3.3 Connecting public services across Europe: ambition and results so far

Ms Gudrun Stock, Deputy Head of Unit, DG CONNECT - EC Unit G3 - Accessibility, Multilingualism & Safer Internet, covered 5 topics: the Connecting Europe Facility (CEF); CEF Automated Translation as a building block; Uptake of CEF Automated Translation; Multilingualism in the next Multiannual Financial Framework (MFF); Beyond CEF Automated Translation.

She began by describing the CEF as a legal instrument to connect public services in Transport, Energy and Telecommunications. The last of these includes Digital Service Infrastructures (DSIs) which are either *sector specific* - based upon mature technical solutions in diverse areas, or *sector independent building blocks* which provide basic and re-usable digital services that can be integrated into other DSI and IT projects. Automated translation (AT), which acts as a multilingual enabler for a variety of public services (e.g online dispute resolution, business registration, eHealth, eJustice, etc.), falls into the

ELRC Workshop Report for Malta

latter category. She distinguished between two complimentary services for which eTranslation is used: DGT eTranslation, for translators and digital services of the EU Institutions and CEF AT, which offers eTranslation to CEF Digital Service Infrastructures, pan-European digital public services, and Member State public administrations.

She then presented some figures on the take-up of eTranslation, emphasizing how much this has increased recently and the key role played by language resources for better quality translation in different sectors, indicating how those working in the public administration can get involved. Turning to the future, Ms. Stock gave a brief overview of the role of multilingualism in the next MFF, presenting the interplay between the four main pillars: Digital Europe, Horizon Europe, Connecting Europe and Creative Europe. She explained that multilingualism figured prominently in areas such as human-centric next generation internet, cultural heritage, public administration capacity building.

She concluded with remarks on the positioning of Language Technology at a nexus between Artificial Intelligence, Language Resources, and High Performance Computing and mentioned the European Language Grid (ELG) as one of the first projects currently running aimed at improving access to this multi-faceted technology through a portal.

3.4 National initiatives for digital public services and (open) data

Ms. Joyce Dimech, Permanent Secretary of Strategy and Implementation within the Office of the Prime Minister, emphasised the Government's support for the idea that public services should be available in both official languages. Government would therefore promote use of the eTranslation service. She emphasised that the efficient exchange and sharing of data is a constant challenge and pointed to the need to avoid systems that impact citizens negatively – for example by requiring repeated disclosure of same information to different government agencies. New technologies should be adopted and in particular AI should be integrated into services. She added that the Government looks forward to the day where eForms can be in spoken as well as written form. She concluded by reporting that a new public services portal is currently being designed which aims to offer a single, citizen-centric digital gateway which will be comfortable for the citizen to use. She added that language had an important role to play in contributing to comfortable interaction.

Following this presentation, Dr Joseph S. Azzopardi, Enterprise Data Manager at Malta Information Technology Agency (MITA) and Data Governance Manager for the Government of Malta gave a short talk entitled *Open Data Initiatives in Malta*. He began by referring to official sources that deal with open data. [Cap 546](#) of the Laws of Malta handles the transposition of the PSI directive into national law, and the National Data Strategy, a vision for the management of data across the whole Public Administration. The main layers are (i) an Authorisation and Representation Platform, (ii) a Foundation Data Layer, (iii) Metadata Platform (Register of Registers) (iv) National Data Portal with links to the European Data Portal and other local and international portals and websites. Further details can be found in a draft version of the strategy that is available online¹.

He then considered some of the challenges related to open data. One problem was the initial discovery of documents from the statute book (over 4000 documents in PDF format), and the correct linking of such documents to their originating government entity. A related challenge is the establishment – and acceptance - of official ownership by that entity, since this is a legal prerequisite for registration, custodianship, and releasing of open data stipulated by the strategy.

¹ <https://mita.gov.mt/en/nationaldatastrategy/Pages/National-Data-Strategy.aspx>

ELRC Workshop Report for Malta

A second issue is that any scheme has to recognize that openness of data is not absolute but part of a continuum admitting different levels of access. To cater for this, the strategy foresees a “traffic lights” classification system for data, whereby each document is assigned a colour: red – sharing not considered; amber/yellow – controlled sharing allowed; green – sharing without conditions. However, it is thought that putting this classification scheme into practice will not be very straightforward.

Dr Azzopardi concluded with a few reflections on language-related issues. Traditionally, language data has not been considered as data in the same way as e.g. CSV files. Admitting language data engenders new considerations – concerning for example the quality of documents versus mere quantity of data. The strategy mentions multilinguality as an underlying principle of the European Interoperability Framework, so one need is metadata in both official languages concerning content, source, type and intended usage. Dr Azzopardi expressed the opinion that to avoid confusion, this should first be developed in English followed by a translation into Maltese.

3.5 CEF in Malta: an outlook into current and future challenges – Panel session

The panel was chaired by Dr. Michael Spagnol (Dept Maltese) and Dr. Claudia Borg (Dept AI), both from the University of Malta. The five panel members were Mr. Andre Arciola, Head, Quality and Training, Servizz.gov Agency, Mr. Keith Cortis, Senior Solutions Architect – MITA, Mr. Albert Vella, CIO – Ministry for Finance, Mr. Donald Mangion, CIO – Ministry for Justice, Culture and Local Government, Dr Joseph S. Azzopardi, Enterprise Data Manager (MITA) and Data Governance Manager for the Government of Malta, Ms. Greta Attard, Software Developer - Malta Digital Innovation Authority (MDIA).

Dr. Spagnol opened by stating that he was very pleased to be discussing the issue of legal equality between Malta’s official languages and asked members of the panel to describe the systemic challenges posed by the use of Maltese from their individual perspectives.

Dr Borg mentioned that as a consequence of the AI Strategy, Servizz.gov was planning to introduce AI-inspired online government services capable of operating bilingually and that an email assistant operating in both English and Maltese was being developed as a building block for use throughout the public service.

Mr Arciola pointed out that Servizz.gov is bilingual although some webpages are currently missing the full set of Maltese orthographic symbols.

Mr Cortis said that most off-the-shelf systems come with English interfaces built in by default, and that the creation of a Maltese interface is costly, one-off, and potentially lengthy. For these reasons the cost-benefits have to be very carefully evaluated before deciding to develop bilingual interfaces.

Amongst the more general problems, Mr Vella indicated that business language in general is English and many terms do not sound natural in Maltese.

Mr Mangion added that the challenge is that of finding correct Maltese terminology in different specialised subdomains.

Ms Attard noted that small screens posed particular problems for Maltese language interfaces because the words were typically longer to write than their English equivalents.

Mr. Azzopardi discussed the challenges pertaining to the planned project for consistent document management throughout the public service. One difficulty was that of establishing appropriate metadata. Other challenges were in the area of privacy, where there must be clear guidelines on what employees can feed into the system. He agreed with the idea of AI integrated services. Noting that AI consists of algorithms + data, he warned that problems arise when either is of low quality.

ELRC Workshop Report for Malta

Ms Attard asked what language resources were available for Maltese. Dr. Borg responded that Named Entity Recognition, crucial for automatic anonymisation of documents, was not yet available for Maltese.

3.6 The CEF eTranslation platform @ work

This talk was given jointly by Michael Jellinghaus, Directorate General for Translation (DGT) - European Commission and by Mark Vella, Language Officer – European Commission Representation in Malta.

Dr. Jellinghaus first gave an overview of the eTranslation platform, explaining how it evolved from the phrase-based-translation MT@EC, to forming part of CEF.AT which provides other services such as transliteration, and named entity recognition besides translation. After identifying the main users of eTranslation, he gave a view of how a user actually submits a translation job, indicating the various options available such as language choice, interactive vs batch mode etc., explaining how the system is set up to protect the privacy and IPRs of the user. He showed how the EU approached multilingual access to a multilingual website - the EU Committee of the Regions - where original contributions arrive in different languages, by integrating on-the-fly eTranslation of dynamic content with human translation of fixed elements such as menus. To avoid confusion, users always have to explicitly request eTranslation. He then gave an example of machine-to-machine translation-as-a-service as used by the European Data Portal. Mr Jellinghaus continued with a brief summary of the differences between Neural Machine Translation (NMT) and Statistical Machine Translation (SMT) and gave examples of the different kinds of errors typically generated by each of these architectures. He concluded by listing key factors of success such as more data for all languages, better lexical coverage for all languages, machine translation adapted to different domains. He listed some future objectives including better translation quality, handling of more file formats, translation of texts outside the EU domain, and new languages (Russian, Chinese, Arabic and Japanese).

He concluded with a list of benefits offered by eTranslation including reliable and trustworthy translation for public administrations, support for languages with fewer speakers, opportunities for the private sector, a higher profile for language technologies, and public availability of data collected by ELRC.

The second part of the session was a demonstration of eTranslation by Mr. Mark Vella. He showed the system from the point of view of a translator showing exactly how the interface works, and how easy it is to use. He gave several examples of translations of previously unseen documents between various language pairs including Maltese. Although some translations were very good, there were some errors and the audience was able to judge where the main difficulties lay.

3.7 ELRC in Malta

This presentation was given by Michael Rosner, Technical NAP for Malta, who began by providing some information about the background and context for ELRC in terms of the Departments at University that have been active in the field of language resources and the projects with which they have been involved so far. He then turned to ELRC itself which was first manifest locally in the first workshop in 2016 whose main aim was to identify contacts who might provide some language data from the public sector. This contrasts with the second workshop which is focused on achieving adoption of eTranslation technology by Public Service employees. The next part of the talk presented the current contents of the ELRC Share repository for resources involving Maltese, concluding that this was for somewhat limited by the small size of the available corpora (e.g. the Laws of Malta contains 3000 sentence pairs). The one exception is the automatically harvested Paracrawl corpus containing 3 million such pairs; its quality, however, is as yet unclear.

ELRC Workshop Report for Malta

The presentation then enumerated some of the obstacles to the collection of bilingual resource in Malta. These were divided into three categories: legacy, awareness and procedural. Legacy problems include the predominant role of English in official communications, the difficulty experienced by many of writing correct Maltese, and a certain reluctance to adopt state-of-the-art IT support tools. A lack of perception of the potential of eTranslation and of the value of bilingual data for improving it counts among the problems of awareness. Finally, there are procedural problems stemming from the lack of a coherent translation workflow that inhibit the retention of resources generated as a by product of translation, such as translation memories.

The talk concluded with an assessment of actions critical for future success, including (i) the need for decisions concerning the implementation of a bilingual national language policy (ii) the recognition that Language Technology has a crucial role to play in the rolling out of the proposed AI Strategy, (iii) the establishment of coherent translation workflow in the public administration that leads to the acquisition of high quality language resources, and finally the adoption of the eTranslation platform on a broad scale within the Government service. To this end the onsite assistance service offered by ELRC to help with the takeup of eTranslation is a good first step.

3.8 Can language data be shared and how?

Next, Dr. Prodromos Tsiavos, ELRC expert on legal matters, gave an overview of the EU perspective on how language data can be shared.

He began by distinguishing language data (e.g. official documents in different languages) from traditionally conceived data in the form of tables. He then pointed to various legal perspectives on data in general concerning e.g. copyright, sui generis database rights, protection of personal data (versus free-flow of non-personal data), public sector information, and confidentiality. He explained different legal regimes: those regarding data as property versus those that treat it as an object for protection and other regimes governing openness pertaining to cultural data and open science, patent publications and metadata.

Distinguishing between the roles of producer or collector versus disseminator of language data, he then pointed out that in order to share data legally, disseminators must adopt the legal measures which depend on the kind of protection the data enjoys (e.g. copyright, GDPR) or whether it is of a confidential nature or whether a cultural heritage licence is required.

He concluded by describing how to share in terms of (i) the kind of information that has to be recorded by disseminators, (e.g. ownership, limitations (e.g. classified/ confidential), legal basis of any personal data, licensing information, notice and takedown procedures), (ii) the range of licences available (from the least to the most restrictive) and (iii) the kind of infrastructure required (e.g. standard open source technologies for repositories, use of appropriate open data standards concerning data/metadata and services, interoperability with initiatives such as ELRC/Europeana, and incorporation of FAIR (Findable, Accessible, Interoperable, Reusable) principles into sharing practice

3.9 Preparing and sharing data with the ELRC repository – and what happens next

Dr. Khalid Choukri, CEO of ELDA, a member of the ELRC Consortium, began by explaining to the audience the notion of language data and its role in eTranslation, giving several examples of the kind of data already used by the eTranslation platform from e.g. the EU Parliament, EU committee of the Regions, ECB etc. He then identified the kind of data now being sought - namely, data residing in local public organisations, produced in-house or outsourced, e.g. reports, communications, news, multilingual web content, policies, terminologies, archives, forms, FAQs etc. He indicated which data formats were preferred and some basic do's and don'ts concerning the preparation of data, criteria

ELRC Workshop Report for Malta

for grouping data and preferred domains. He then gave a demonstration of the steps to follow in order to register an account with the ELRC share platform and submit such prepared data.

As the title of the talk suggests, he continued by explaining what happens to data after it is submitted: cleaning; format conversion; sentence alignment and metadata completion before being processed for inclusion amongst the translation memories of the eTranslation platform.

He continued by listing the wide variety of services and processes available free of charge and explaining that expert help was available to fix problems with data including e.g. anonymisation, format conversion, tag removal, alignment, metadata generation, and concluded by explaining how help and onsite visits can be requested online.

3.10 Identifying and managing your data: Questions & Answers

A brief QA session followed Dr Choukri's presentation.

Many questions concerned the mechanics of logistics of obtaining and operating the eTranslation platform and participants were informed about the various help services available from ELRC including onsite assistance. One participant from the commercial sector asked whether the eTranslation platform was available by the EC to SMEs. The audience was informed that although it was not feasible to make the service available to all SMEs, some initiatives in this respect were under consideration and SMEs would be informed about it in the near future.²

3.11 Concluding Remarks

Ms. Donatienne took the floor once again to thank all participants for their presence and their contributions, and re-emphasised that the preservation and the management of language data are therefore key factors and a common responsibility.

She added that for this purpose, one-to-one meetings will be held throughout 2020 with Ministries, including at granular levels, to spread the word about *eTranslation* and to encourage the contribution of language resources, a process for which ELRC will provide both technical and legal assistance. Ms Spiteri emphasised that the ELRC seminar was not intended as a one-off event but as part of a process and she looks forward for the dialogue to continue in the coming months through onsite visits. For this purpose ELRC Malta will be in contact for the organisation of one-to-one meetings with the public administration.

² At the time of publishing this report, the eTranslation service is being tested among a group of SMEs with so far mainly positive results.

4 Synthesis of Workshop Discussions

4.1 ELRC and Open language Data in Malta

The open data policy in Malta is based on the PSI directive (2003/98/EC) which was transposed into Maltese Law in 2015 and appears as [Cap 546](#) of the Laws of Malta. A National Data strategy which defines an infrastructure for handling national data is in preparation and a [draft version](#) was released in 2016: <https://mita.gov.mt/en/nationaldatastrategy/Pages/National-Data-Strategy.aspx>.

A BETA version of the Malta Data Portal is designed to enable a shared platform for the management and support of the Foundation Data Layer and its main Administrative Registers. It will eventually serve as the Data Governance workbench for the management of metadata, requests for data and data sharing and re-use authorisations in respect of the national official registers emanating from the laws of Malta.

This portal and its contents are currently work in progress and unless otherwise indicated should not be considered as providing official records.

4.2 Success stories and lessons learnt

- The main success story of the second workshop was the size of the audience (almost twice that of the first workshop) and the very high proportion of participants from the public sector. This bodes well for the future of ELRC in Malta. The lesson learnt here is the importance of having a Public NAP in a position to communicate effectively, personally and persistently with those working in the many government entities that could potentially benefit from the eTranslation platform.
- A surprise for the local organisers was the level of enthusiasm for the idea of eTranslation shown by many personnel in public administration who would not normally count translation amongst their everyday activities. One reason for this could be that eTranslation offers the possibility of a mechanism to save effort when it comes to producing written documents in Maltese. Currently, many native speakers of Maltese still tend to use English for written official communications.
- The organisers carefully considered the simultaneous interpretation requirements for this workshop. Bearing in mind that the vast majority of participants were Maltese native speakers who understood English, bidirectional interpretation was deemed unnecessary. Instead, it was decided to offer simultaneous interpretation from Maltese to English only to the relatively small number of non-Maltese speakers (c. 6). This arrangement worked well and so the organisers suggest that given the costs associated with simultaneous interpretation, a flexible approach be adopted in future events.

5 Workshop Participants

The distribution among workshop participants is as follows. The list of those who attended the workshop is provided below in Section 6.4.

	Number of participants	Percentage
Public sector	80	81.6%
Other (EC and ELRC consortium representatives, academics, research staff, secretariat support staff, etc)	18	18.4%
Total	98	100%

5.1 Workshop Presentations

All presentations are available online on the ELRC website: http://www.lrc-coordination.eu/l2malta_agenda