



**European Language
Resource Coordination**
Connecting Europe Facility

Deliverable D3.2.19

Task 10

ELRC Workshop Report for Romania



Author(s): Laura MIHĂILESCU (European Institute of Romania)
Dissemination Level: Public
Version No.: <V1.0>
Date: 2019-01-14



Contents

1	Executive Summary	3
2	Workshop Agenda	4
3	Summary of Content of Sessions	6
3.1	Welcome and introduction	6
3.2	Connecting public services across Europe: ambition and results so far	6
3.3	National initiatives for digital public services and (open) data	7
3.4	CEF in Romania: an outlook into current and future challenges – Panel session	7
3.5	The CEF eTranslation platform @ work	8
3.6	The European Language Resource Coordination (ELRC) action	9
3.7	ELRC in Romania	10
3.8	Can language data be shared and how? National and European legal framework	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	Conclusions	12
4	Synthesis of Workshop Discussions	12
4.1	ELRC and Open language Data in Romania	12
4.2	Success stories and lessons learnt	12
5	Workshop Presentations	13

1 Executive Summary

The second ELRC workshop in Romania took place at the Representation of the European Commission in Bucharest on 1st November 2018.

The workshop was jointly opened by Ms Toma, DGT Field Officer, European Commission, Mr Tufiş, ELRC Technology National Anchor Point (Research Institute for Artificial Intelligence “Mihai Drăgănescu”, Romanian Academy) and by Ms Mihăilescu, ELRC Public Services National Anchor Point (European Institute of Romania).

The first presentation of the day focused on the connection of digital public services in Europe and was given by Mr Ceauşu, Programme Officer - EU policies, DG Connect, European Commission.

Insights into national open data initiatives were presented by Mr Cazacu, Cabinet Director for the Secretary of State Coordinator of IT, Secretariat-General of the Government.

The panel “CEF in Romania: an outlook into current and future challenges” was moderated by Ms Mihăilescu. Each of the five panellists presented a Digital Service Infrastructure (DSI) and the challenges faced by public services in Romania. Ms Panait presented the National Open Data Portal, Mr Bena presented SOLVIT, Mr. Crăciunescu e-Justice and Ms Roia presented Europeana.

The CEF eTranslation platform was jointly presented by Mr Tufiş and Mr Ceauşu.

Mr Piperidis (ILSP/Athena R.C., ELRC) presented the European Language Resource Coordination initiative and Mr Tufiş presented ELRC in Romania.

Mr Kamocki, legal expert, Evaluations and Language resources Distribution Agency, presented the European and the Romanian legal framework on the re-use of public sector information and on copyright. He underlined that, according to the Romanian Copyright Act, official acts of political, judicial, administrative or legislative nature, as well as their official translations are not protected by copyright.

Ms Giagkou, ILSP/Athena R.C., ELRC gave a practical illustration of how to share data using the ELRC-SHARE repository.

The workshop ended with a session of questions & answers, followed by conclusions.

Overall, the workshop raised awareness not only about ELRC, the DSIs and their interconnection, but also of the importance and value of language resources.

2 Workshop Agenda

09:00 – 09:30 **Registration**

09:30 – 09:45 **Welcome and introduction**

Carmen Toma, DGT Field Officer, European Commission

Dan Tufiş, ELRC Technology National Anchor Point (Research Institute for Artificial Intelligence “Mihai Drăgănescu”, Romanian Academy)

Laura Mihăilescu, ELRC Public Services National Anchor Point (European Institute of Romania)

Session 1. Connecting a multilingual Europe: European context and local needs

09:45 – 10:05 **Connecting public services across Europe: ambition and results so far**

Alexandru Ceauşu, Programme Officer - EU policies, Directorate-General for Communications Networks, Content and Technology, European Commission

10:05 – 10:25 **National initiatives for digital public services and (open) data**

Mihail Cazacu, Cabinet Director for the Secretary of State Coordinator of IT, General Secretariat of the Government

10:25 – 11:10 **CEF in Romania: an outlook into current and future challenges – Panel session**

Moderator: *Laura Mihăilescu*, head of Translation Coordination Unit, European Institute of Romania & ELRC Public Services National Anchor Point

Panelists:

- *Larisa Panait*, adviser, Information Technology Department, General Secretariat of the Government
- *Bena Romulus-Andrei*, European Affairs adviser, Solvit Centre, Ministry of Foreign Affairs
- *Răzvan Crăciunescu*, Information Technology Department, Ministry of Justice
- *Cristina Ioana Roiu*, head of unit at the Romanian Academy Library, Councillor of European Network Association

11:10 – 11:30 **The CEF eTranslation platform @ work**

Dan Tufiş, director, Research Institute for Artificial Intelligence “Mihai Drăgănescu”, Romanian Academy & ELRC Technology National Anchor Point

Alexandru Ceauşu, Programme Officer - EU policies, Directorate-General for Communications Networks, Content and Technology, European Commission

11.30 – 12:00 **Coffee Break**

Session 2. Engage: hands-on data

12:00 – 12:25 **The European Language Resource Coordination (ELRC) action**

Stelios Piperidis, ILSP/Athena R.C., ELRC

- 12:25 – 12:45 **ELRC in Romania**
Dan Tufiş, director, Research Institute for Artificial Intelligence “Mihai Drăgănescu”,
Romanian Academy & ELRC Technology National Anchor Point
- 12:45 – 13:20 **Can language data be shared and how? National and European legal framework**
Pawel Kamocki, legal expert, Evaluations and Language resources Distribution
Agency
- 13:20 – 14:20 **Lunch Break**
- 14:20 – 14:45 **Preparing and sharing data with the ELRC repository – and what happens next**
Maria Giagkou, ILSP/Athena R.C., ELRC
- 14:45 – 15:15 **Identifying and managing your data: Questions & Answers**
Moderator: *Laura Mihăilescu*, head of Translation Coordination Unit, European
Institute of Romania & ELRC Public Services National Anchor Point

All speakers
- 15:15 – 15:30 **Conclusions**
Laura Mihăilescu, head of Translation Coordination Unit, European Institute of
Romania & ELRC Public Services National Anchor Point
- 15:30 – 16:00 **Coffee Break and networking**

3 Summary of Content of Sessions

3.1 Welcome and introduction

Carmen Toma, DGT Field Officer, European Commission

Dan Tufiş, ELRC Technology National Anchor Point (Research Institute for Artificial Intelligence “Mihai Drăgănescu”, Romanian Academy)

Laura Mihăilescu, ELRC Public Services National Anchor Point (European Institute of Romania)

The workshop was opened by Ms Toma, DGT Field Officer, European Commission, who welcomed the participants to the second ELRC workshop, at the same premises as the first workshop. In the context of the EU motto “United in diversity”, an important EU objective is to create better connection among citizens, mainly through the Connecting Europe Facility (CEF). In order to overcome the linguistic barrier, one of the main obstacles in creating the Digital Single Market, the EC has initiated and supports the ELRC. The ELRC collects language resources from public institutions from all Member States in order to improve the performance, the quality and the geographic coverage of eTranslation, a service offered free of charge to public institutions.

Mr Tufiş expressed his pleasure to meet again several participants who attended the first ELRC workshop two years ago and informed the participants that progress in automated translation was quite significant.

Ms Mihăilescu expressed hope that participants working in public institutions would better understand the link between the technical and the linguistic side of their work in order to further deepen cooperation in the future. She also encouraged participants to spread the word about what they’d learn at this ELRC workshop.

The video about ELRC and the objectives of the European Commission in the field of language technologies, with Claire Bury (Deputy Director General at DG Connect) and Rytis Martikonis (Director General at DG Translation), was played in loop at the beginning of the workshop and during all breaks.

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

Alexandru Ceauşu, Programme Officer - EU policies, Directorate-General for Communications Networks, Content and Technology, European Commission

Mr Ceauşu started his presentation by reminding participants about the recent European Parliament resolution of 11 September 2018 on language equality in the digital age.

He briefed the participants on the CEF Regulation, CEF Telecom guidelines and CEF Work Programme. He presented the main characteristics of a Digital Service Infrastructure (DSI), such as: cross-border use, delivery of services by digital means, contribution to EU policies, sustainability. Examples of DSIs are BRIS, ODR, eHealth, EESSI, eJustice, Open Data, eProcurement.

The speaker then elaborated on the CEF AT building block, which includes the eTranslation service, the ELRC action, a group of related projects and language technology tools. Mr Ceauşu briefly explained how officials working in public institutions could register and use eTranslation free of charge.

In the end, he presented the next Multi-annual Financial Framework (2021-1027) and developed two components where multilingualism will play an important role: Digital Europe and Horizon Europe

R&D&I. While the latter will cover research and innovation activities to achieve digital language transparency and to support language preservation, the Digital Europe Programme will focus on deployment of language technologies (LT) and will support modernization of public administrations and areas of public interest.

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

Mihail Cazacu, Cabinet Director for the Secretary of State Coordinator of IT, Secretariat-General of the Government (SGG)

Mr Cazacu presented an overview of the open data in Romania:

- legal framework: the PSI Directive was transposed by Law no 109/2007 and amended by Law no 299/2015; in 2011 Romania joined the Open Government Partnership and at present is implementing the fourth action plan (2018-2020);
- public institutions, under Government Ordinance no 41/2016 and using the methodology for publishing datasets as open data (SIPOCA 36), upload data to data.gov.ro portal;
- in October 2018, the Secretariat-General of the Government signed a partnership agreement with Code 4 Romania (community of volunteers) which will improve the quality of open data and will reuse them in order to create open source digital tools).

Although public institutions invested significantly in IT equipment, a lot of IT systems are not interoperable. However, in 2016 the position of Secretary of State Coordinator of IT (SGG) was created and in 2018 an e-government committee was set up. Significant improvements are expected with the existing National Strategy for the Digital Agenda, the feasibility studies for the government cloud, the electronic single point of contact, and the electronic identity card.

Mr Cazacu also commented on Romania's ranking in the Digital Economy and Society Index 2017 (number of employed IT experts versus freelancers, banking and online shopping versus 45% rural population and degree of poverty).

3.4 CEF in Romania: an outlook into current and future challenges – Panel session

The panel was moderated by Ms. Mihăilescu, head of Translation Coordination Unit, European Institute of Romania & ELRC Public Services National Anchor Point.

Ms Larisa Panait (panellist), adviser, Information Technology Department, Secretariat-General of the Government

After presenting the legal framework of open data, Ms Panait talked about the Secretariat-General of the Government, which coordinates the open data in Romania and manages the national portal data.gov.ro – the central point of access for the open datasets published by public institutions in Romania and contact point for the europeandataportal.eu.

The speaker defined open data and described their main characteristics. She followed on with information on data.gov.ro portal (launched in 2013, 97 institutions registered, more than 1400 datasets published, government licence adapted for the Romanian public institutions, methodology for publishing open datasets).

Ms Panait also presented the European Data Portal, the services it offers and how eTranslation is integrated into the portal. In this case, the automated translation for Romanian is not of sufficient quality.

Ms Mihăilescu (moderator) added that eTranslation significantly improved in comparison with two years ago and that it has been consistently used within the Translation Coordination Unit. eTranslation proved useful for certain types of documents and very useful for large documents as it preserves formatting. Quality will improve once public institutions feed it with the right data.

Bena Romulus-Andrei (panellist), European Affairs adviser, Solvit Centre, Ministry of Foreign Affairs

Mr Bena presented SOLVIT (functioning at European level since July 2002 and in Romania since January 2007). At present, the Solvit Centre operates within the Ministry of Foreign Affairs, Government Agent for the Court of Justice of the European Union.

The speaker presented the Solvit operating criteria, how to submit a request, the Solvit database, and the steps to take in order to solve a case. In the end, he thanked the institutions for their collaboration.

Răzvan Crăciunescu (panellist), Information Technology Department, Ministry of Justice

Mr Crăciunescu presented the European e-Justice Portal, e-Codex, and CEF Telecom 2018-4 call (still open at the time of the workshop).

Since 2016, e-Codex has interlinked national and European IT systems in the e-Justice domain. The speaker presented the e-Codex high-level system architecture, as well as the e-Justice current context (strategic documents, final draft of the 2019-2023 Strategy on European e-justice and the e-Justice Action Plan for 2019-2023, where CEF is mentioned for e-Justice projects).

In Romania, funds were accessed by the National Trade Register Office for interconnection with BRIS (2 CEF Telecom financed projects).

Mr Crăciunescu also presented the future revamped IT system in Romania (case management IT system, business analysis project, system development project) and several current challenges (mainly related to financing, human resources, IT projects).

Cristina Ioana Roiu (panellist), head of unit at the Romanian Academy Library, Councillor of Europeana Network Association

Ms Roiu presented Europeana (one of the DSIs under CEF) and the Europeana Network Association. Europeana, Europe's digital platform for cultural heritage, currently provides access to over 58 million digital resources. In the main portal 'Europeana Collections' (europeana.eu) there are items from more than 50 languages and the portal has functionalities in 28 languages.

Europeana offers services such as data publication, discovery and funding to data partners, end-users and re-users.

The Europeana Network Association, with over 1700 cultural heritage, creative and technology professionals, supports the values and work of Europeana, exchanges and promotes best practice on digitisation and metadata standardisation and stimulates capacity building in the cultural heritage community. She stressed that in Romania a lot of work had been done through the experts' collaboration.

She reminded the recommendations of the EC:

- to strengthen multilingual aspects for users to explore, enjoy and re-use content regardless of the original language of the cultural materials;
- to improve its digital offering by making use of various artificial intelligence-based approaches, natural language processing and big data to automate -to the extent possible- the creation, management and presentation of content and metadata.

3.5 The CEF eTranslation platform @ work

Mr Dan Tufiş, Romanian Academy & ELRC Technology National Anchor Point

Mr Alexandru Ceauşu, Directorate-General for Communications Networks, Content and Technology, European Commission

This session was jointly presented by Mr Dan Tufiş and Mr Alexandru Ceauşu.

At the beginning of the presentation, Mr Ceauşu stressed that eTranslation was the result of the collaboration between DGT and DG Connect and that the CEF AT platform has as its main objective the support of the public institutions in the CEF affiliated countries. The main challenges are to adapt to different needs and to collect language resources.

Mr Tufiş continued the presentation by underlining the advantages of using automated translation. He then informed the participants on the developments in automated translation, from MT@EC (the European Commission's Machine Translation system), to eTranslation and to CEF Automated Translation (the platform which includes eTranslation and it will additionally integrate other NLP tools).

Mr Tufiş briefed participants about the difference between statistical and neural automated translation and emphasized that language resources are the key to success.

He also presented the steps to be followed in order to connect to the eTranslation service and in the end called upon participants to share language resources in order to improve the quality of the automated translation in Romanian.

3.6 The European Language Resource Coordination (ELRC) action

Mr Stelios Piperidis, ILSP/Athena R.C., ELRC

Mr Piperidis presented the position of eTranslation within the Connecting Europe Facility (Telecom/CEF Digital).

He briefed participants on the ELRC Consortium (DFKI – the German Research Centre for Artificial Intelligence; ELDA – the Evaluations and Language Resources Distribution Agency; TILDE – Language Technology and Services Provider and ILSP – the Institute for Language and Speech Processing) and on the network of National Anchor Points (one Technological NAP and one Public Services NAP per CEF affiliated country).

ELRC collects language resources, identifies translation needs of public services, engages with the public sector in the identification of language resources, helps with associated legal and technical issues and acts as observatory for language resources across Europe.

Mr Piperidis talked about the importance of the European Parliament resolution of 11 September 2018 on language equality in the digital age, which:

- "calls on the Commission and the Member States to develop strategies and policy action to facilitate multilingualism in the digital market;
- requests [...] that the Commission and the Member States define the minimum language resources that all European languages should possess, such as data sets, lexicons, speech records, translation memories, annotated corpora and encyclopaedic content, in order to prevent digital extinction;
- [...] calls on the Commission to make as a priority of language technology those Member States which are small in size and have their own language, in order to pay heed to the linguistic challenges that they face; [...]
- stresses the role of the EU, the Member States, universities and other public institutions in contributing to the preservation of their languages in the digital world and in developing databases and translation technologies for all EU languages, including languages that are less widely spoken;
- calls for coordination between research and industry with a common objective of enhancing the digital possibilities for language translation and with open access to the data required for technological advancement".

Mr Piperidis presented the ELRC network's achievements so far (i.e. in October 2018 more than 500 languages resources were contributed to the ELRC). Most of the language resources available in the ELRC-SHARE repository can be used by anyone. He also invited participants to resort to ELRC in case they need assistance with language resources.

3.7 ELRC in Romania

Mr Dan Tufiş, Romanian Academy & ELRC Technology National Anchor Point

Mr Tufiş started his presentation with an overview of the ELRC (number of countries, of NAPS, types of events) and encouraged the participants to refer to the ELRC website.

He presented graphics with data by language and the Romanian's ranking. He thanked for the languages resources that had been shared after the first workshop in Romania (23 March 2016) and ensured participants that the quality of the Romanian automated translation had improved thanks to their contributions. The speaker went on by presenting two programmes in which the Romanian Academy was involved as a partner: Marcell (Multilingual Resources for CEF.AT in the Legal Domain) and CEF EU Presidency Translation Toolkit for Romania, Finland and Croatia.

Mr Tufiş finished his presentation with the ELRC motto: *Supporting our languages is supporting Europe and supporting Europe is supporting our languages!*

3.8 Can language data be shared and how? National and European legal framework

Pawel Kamocki, legal expert, Evaluations and Language Resources Distribution Agency

Mr Kamocki presented Directive 2003/98/EC on the re-use of public sector information and the Romanian legal framework transposing the PSI Directive.

He underlined that, according to the Romanian Copyright Act, official acts of political, judicial, administrative or legislative nature, as well as their official translations are not protected by copyright.

The speaker explained the stages to follow when releasing data under the PSI Directive, i.e.:

- exclude confidential information;
- obtain prior informed consent, find a legal basis, anonymize or exclude personal data;
- ensure there is no 3rd party copyright, that the material is in the Public Domain or that the necessary licences have been obtained;
- follow the national PSI transposition rules.

Mr Kamocki gave some tips to follow by the resource provider and by the data user. In the end, he presented some success stories and recommended the participants to contact the ELRC department for legal assistance.

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

Ms Maria Giagkou, ILSP/Athena R.C., ELRC

Ms Giagkou started her presentation with examples of different types of data and she emphasized that language data are very valuable for language technologies, especially for machine translation. In fact, any piece of electronically stored text and its equivalent in another language is useful for training eTranslation.

She presented the appropriate file formats for different types of language datasets and listed the preferred domains for CEF eTranslation (i.e.: ODR, EESSI, eProcurement, eJustice, eHealth, BRIS, Safer Internet, Cybersecurity, Public Open Data and Europeana).

Ms Giagkou provided examples of file formats of parallel texts and commented both on the steps to be taken when preparing data and the criteria for grouping data into meaningful datasets.

Participants were also informed about how to contribute data to the ELRC-SHARE repository (how to access the ELRC-SHARE, how to upload the data etc.). She provided examples of processing services applied to the contributed data in order to convert them to MT-ready training datasets (i.e. data

extraction, anonymization, cleaning, re-formatting, data conversion, tag removal, alignment, completing the metadata).

Ms Giagkou invited the participants to use the freely available ELRC technical and legal helpdesk and the on-site assistance.

3.10 Identifying and managing your data: Questions & Answers

This session was moderated by Ms Mihăilescu, who informed the participants about some common issues that public institution in Romania are dealing with:

- very few (under 5) public institutions use translation memories;
- the public institutions outsourcing translations do not ask for translation memories because they do not have the translation software to open them;
- a lot of institutions do not keep a standardized detailed translations record (name of the document in the source and target language, name of the translator/reviser, date of translation/revision, domain, number of pages etc.) and there is the risk of doubling the work.

She underlined that if public institutions receive translations in doc formats (or equivalent), these documents could be transformed in translation memories by ELRC and the public institution who gave them could receive the translation memory for internal use.

Question: In case the translation is outsourced, can the public institution give the translation if it does not own the copyright for the translation?

Answer: It depends on what is stipulated in the contract with that language service provider. Depending on the nature of the document to be translated, there could be a clause stating that copyright is transferred to the public institution.

Question: How is the risk of uploading legal quality translation covered? Who checks the content of the translations?

Answer: Unfortunately, due to the limited human resources available at the Translation Coordination Unit at the European Institute of Romania, translations done by other public institutions cannot be revised.

ELRC leaves the content as it is, it only transforms documents in terms of the format.

There are two filters: at the ELRC for the file format and at the EC where translations are checked by human translators before being uploaded in order to train the translation machine.

Question: Are you satisfied with the awareness of the ELRC action? What measures are taken to raise awareness (apart from organizing workshops)? Is there a systematic approach?

Answer: Success is measured by the number of people using eTranslation, by the number and quality of languages resources uploaded to the ELRC-SHARE. As eTranslation was not designed to compete with online, free services such as Google Translate, but it is a service available for the public sector, workshops and other types of events in this field focus on officials working in public institutions. ELRC is open to receive any suggestion of improvement in this area.

Question: Is there a platform where all DSI can collaborate?

Answer: There is an architectural committee where each DSI is represented and this represents a top-down collaboration. There is also collaboration between some DSIs (such as eTranslation and Europeana for German).

There is also a related contract, CEF eTranslation Tools and Services, which investigates the multilingual and language technology needs of the eTranslation stakeholders, i.e. of the DSIs among others. It is important to encourage the local level representation of the DSI.

3.11 Conclusions

The workshop was considered successful as it contributed to raising awareness about ELRC, CEF, several DSIs and how they are interconnected, the value of language data – which can also be open data, and about the importance of sharing data and information.

Even if language resources already available in the public sector are not officially verified by persons specialized in the translation area, these resources could improve eTranslation, if shared.

It was also underlined that there is a need to standardize and modernize the way public institutions work with language data.

Participants were invited to spread the word about ELRC and what they learned during the workshop, to use eTranslation and to share language resources. As most of the audience belonged to the public sector, experts working in different public institutions were invited to deepen collaboration among themselves.

4 Synthesis of Workshop Discussions

The main concern was about the quality of the language resources that would be uploaded to the ELRC-SHARE by people who are not trained as translators.

It was also underlined that every public institution should open up the language resources they possess, even if they decide not to share them only with ELRC. The public sector should modernize the way they work with language data.

4.1 ELRC and Open language Data in Romania

In Romania, the national open portal (data.gov.ro) is coordinated by the Secretariat-General of the Government. Although there are 97 institutions registered, and more than 1400 datasets published, most datasets are numerical.

When presenting the legal framework for sharing language resources, it was underlined that, unlike in many other countries, according to the Romanian Copyright Act, official acts of political, judicial, administrative or legislative nature, as well as their official translations are not protected by copyright.

4.2 Success stories and lessons learnt

The following points can be considered success stories for Romania in the context of ELRC:

- the widespread dissemination of information on the ELRC workshop as well as the presentation of the ELRC project objectives will lead to an increase in the number of people using eTranslation and hopefully in the number of shared resources from and into Romanian;
- sending invitations to all central administrations widely promoted the initiative, e.g. the need for a large number of language resources in order to train the eTranslation platform from and into Romanian;
- some of the participants have already provided information about translations done in their organisations where relevant documents could be found.

Some of the major conclusions from the Romanian workshop are the following:

- some contributors are reluctant to give their language resources because their translations are not officially reviewed by professional translators;
- there is a need for raising awareness of the fact that language data created by public bodies are also open data that can and should be shared;
- there is a need to standardize and modernize the way public institutions work with language data.

5 Workshop Presentations

All presentations and videos are available on the website of the European Institute of Romania (<http://ier.gov.ro/event/al-doilea-atelier-elrc-european-language-resource-coordination/>) and on the ELRC website: (http://www.lr-coordination.eu/ro/l2Romania_agenda).