



The European Language Resource Coordination (ELRC): why and how you can engage and benefit

Stelios Piperidis
Institute for Language and Speech Processing / Athena R.C.
ELRC





The ELRC context

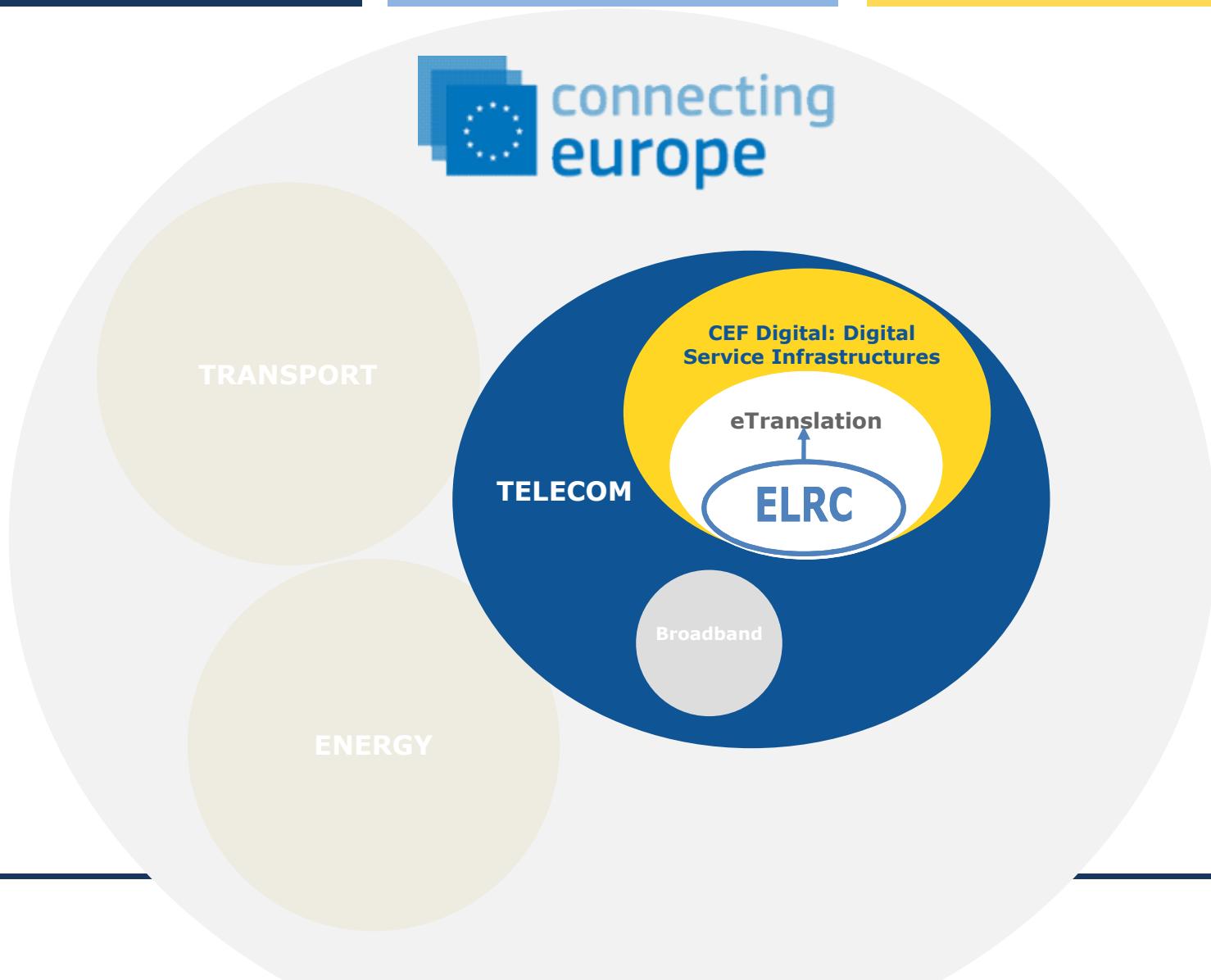


TRANSPORT

TELECOM

ENERGY

The ELRC context





Heads & organisations behind the ELRC



The ELRC Consortium:



German
Research Center
for Artificial
Intelligence



Heads & organisations behind the ELRC



30 ELRC Technological
National Anchor Points

30 ELRC Public Services
National Anchor Points



Tamás Váradi

Research Institute for Linguistics, Hungarian Academy
of Sciences

Zoltán Bódi

Institute for Hungarian Language Strategy

Janos Szente

Department of EU Law at the Ministry of Justice

The ELRC anchor points in Hungary



Tamás Váradi

Research Institute for Linguistics, Hungarian
Academy of Sciences



Zoltán Bódi

Institute for Hungarian Language Strategy

Janos Szente

Department of EU Law at the Ministry of
Justice

Why ELRC...?



- To facilitate multilingual communication and exchange of information in key public service scenarios
- How? With the right language resources!
- Examples of public services scenarios to which eTranslation may be applied:
 - Consumer Rights (ODR),
 - Social Security (EESSI),
 - Justice (eJustice),
 - Commerce (BRIS),
 - Open Data (EDP),
 - ...

Language equality in the digital age



THE EUROPEAN PARLIAMENT VOTED
IN FAVOR OF THE RESOLUTION ON
**"LANGUAGE EQUALITY IN THE
DIGITAL AGE"** ON 11. SEPTEMBER



“18. Calls on the Commission and the Member States to develop strategies and policy action to **facilitate multilingualism in the digital market**; requests, in this context, 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**;”



“21. 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;”



Committee on Industry, Research and Energy (11.9.18): 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".

Why ELRC...?



- Cracking the Language Barrier Federation (2016): “We recommend that Europe actively makes an effort to compete in the global landscape for research and development in language technology since **we cannot expect third parties from other continents to solve our translation and knowledge management problems** in a way that suits our specific communicative, societal and cultural needs. **Language Technology made for Europe in Europe is the key.** It will contribute to future European crossborder and cross-language communication, economic growth and social stability.”

What does ELRC do?



Collect

- Language resources

Identify

- Needs of public services

Engage

- With the public sector in the identification of LR

Help

- With any technical or legal issues

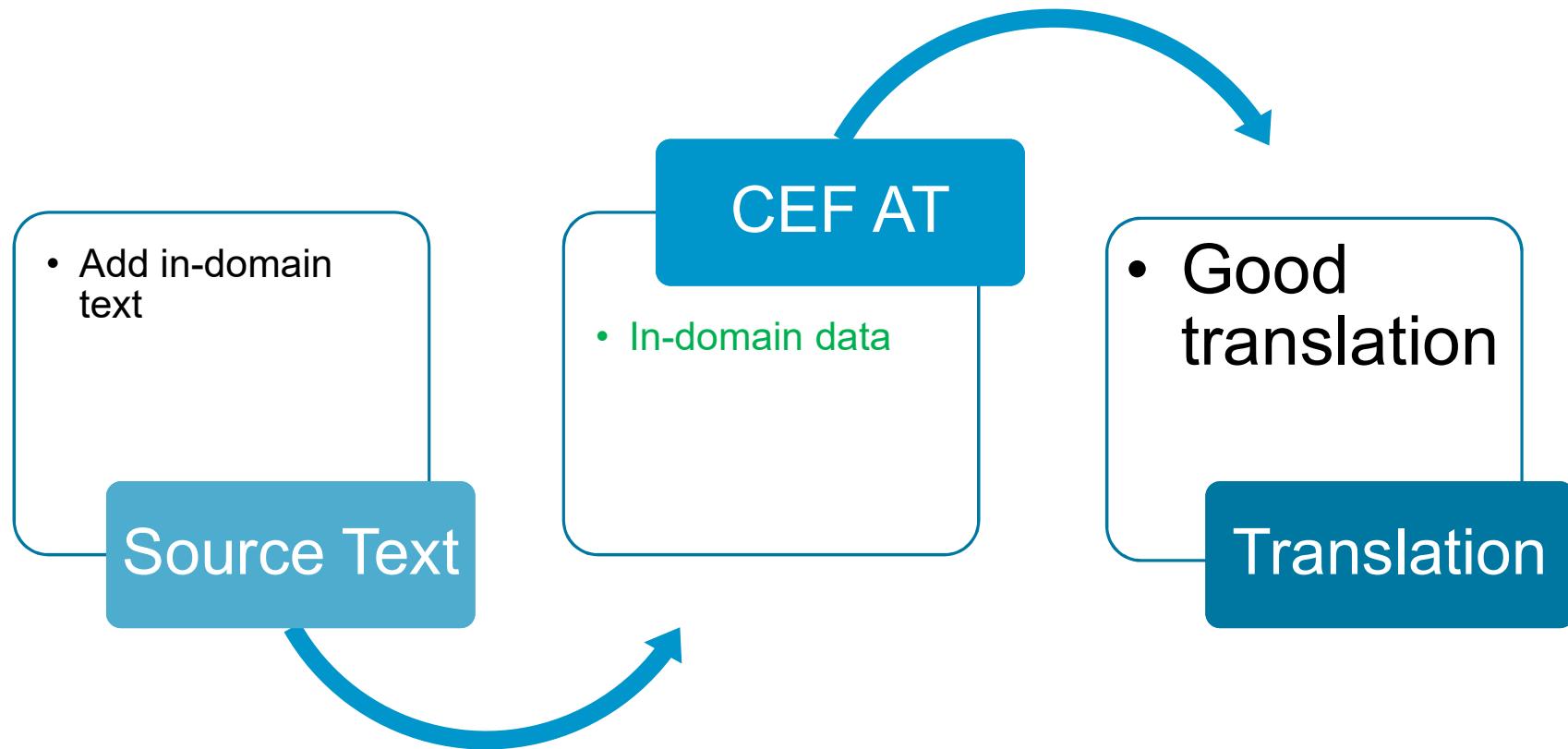
Act

- Observatory for language resources across Europe

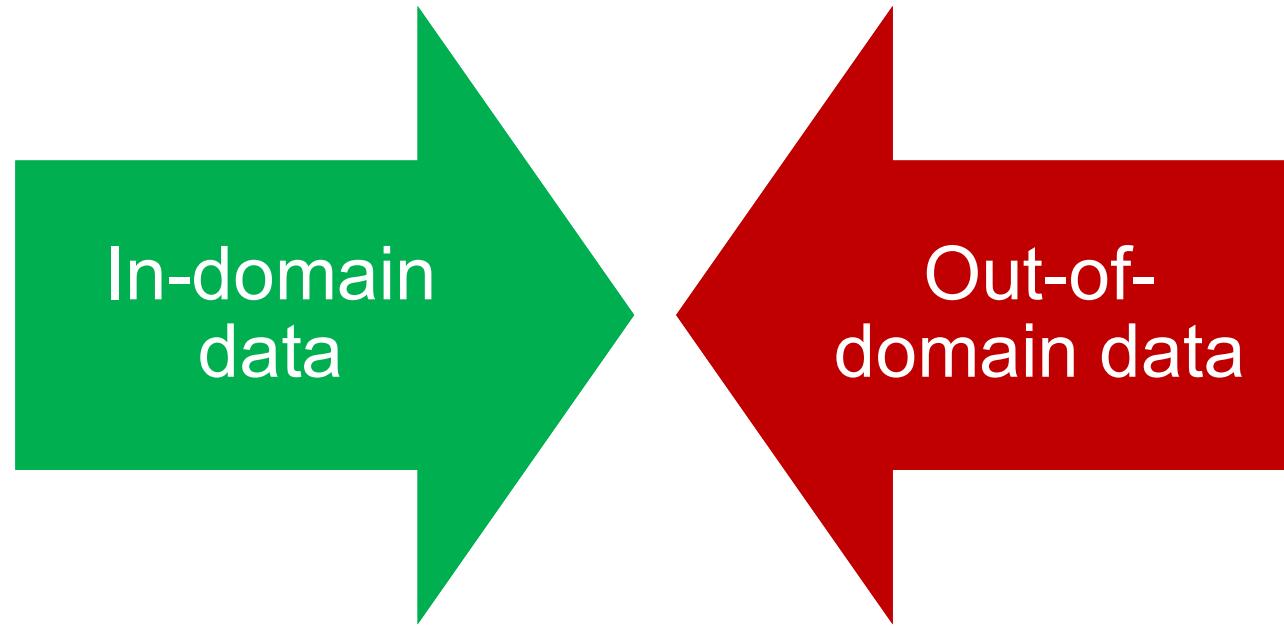
How to make it work...?



With the right language resources!



How to make it work...?



→ Your language resources make a difference!

What has been achieved so far...?

The ELRC network



Today

More than 400 identified
data holders

April 2016

60 National
Anchor Points

April 2015

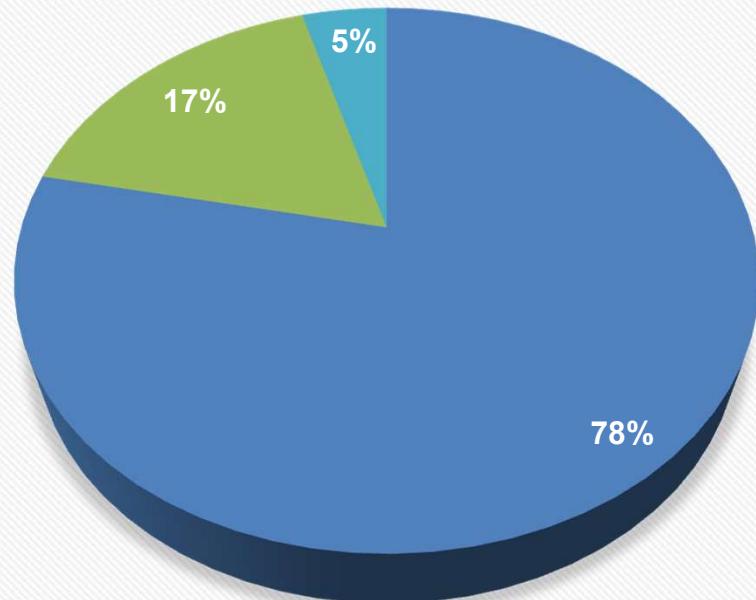
ELRC
consortium

What has been achieved so far...?

ELRC Data in April 2017



A total of 225 LRs provided by ELRC



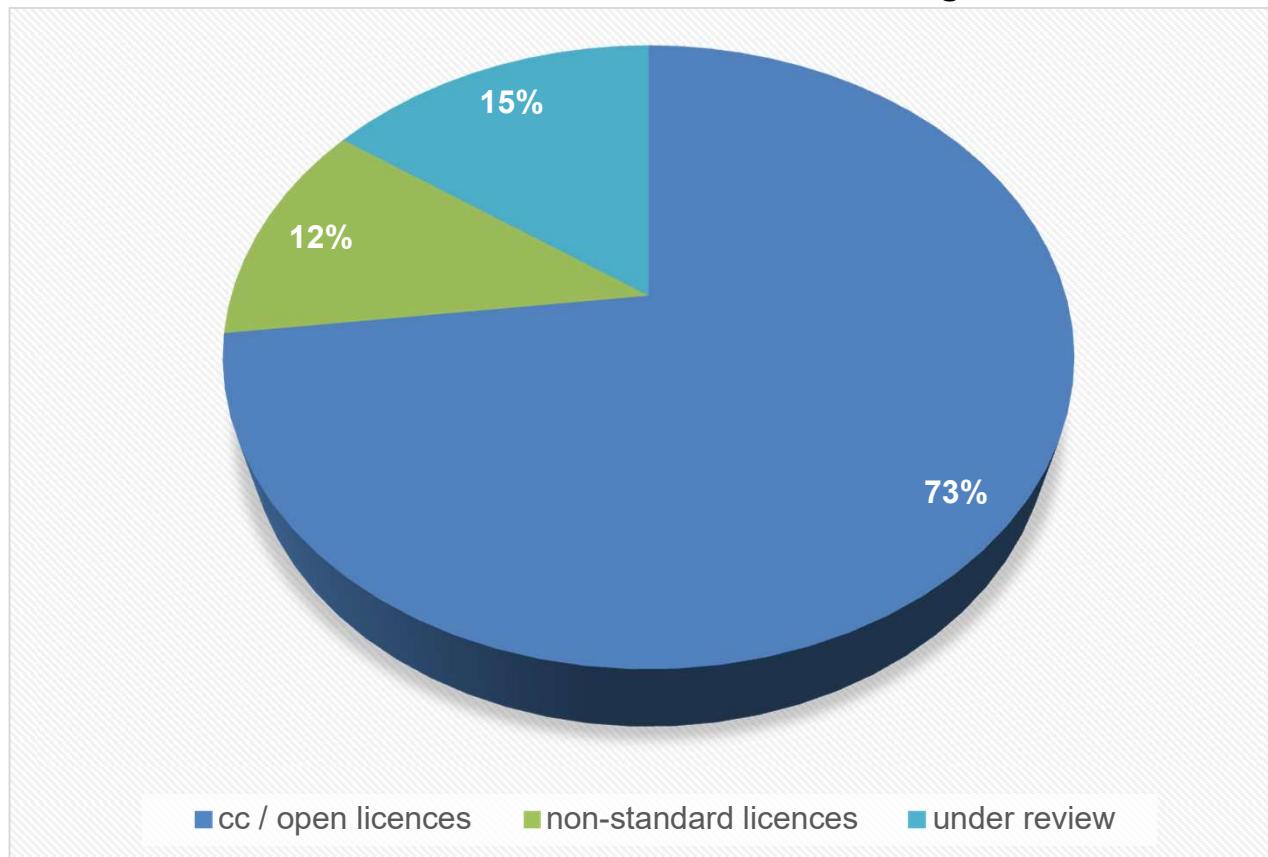
■ Open data / cc licenses ■ Non-standard/other licenses ■ Under review

What has been achieved so far...?

ELRC Data in October 2018



More than 500 LRs contributed through ELRC

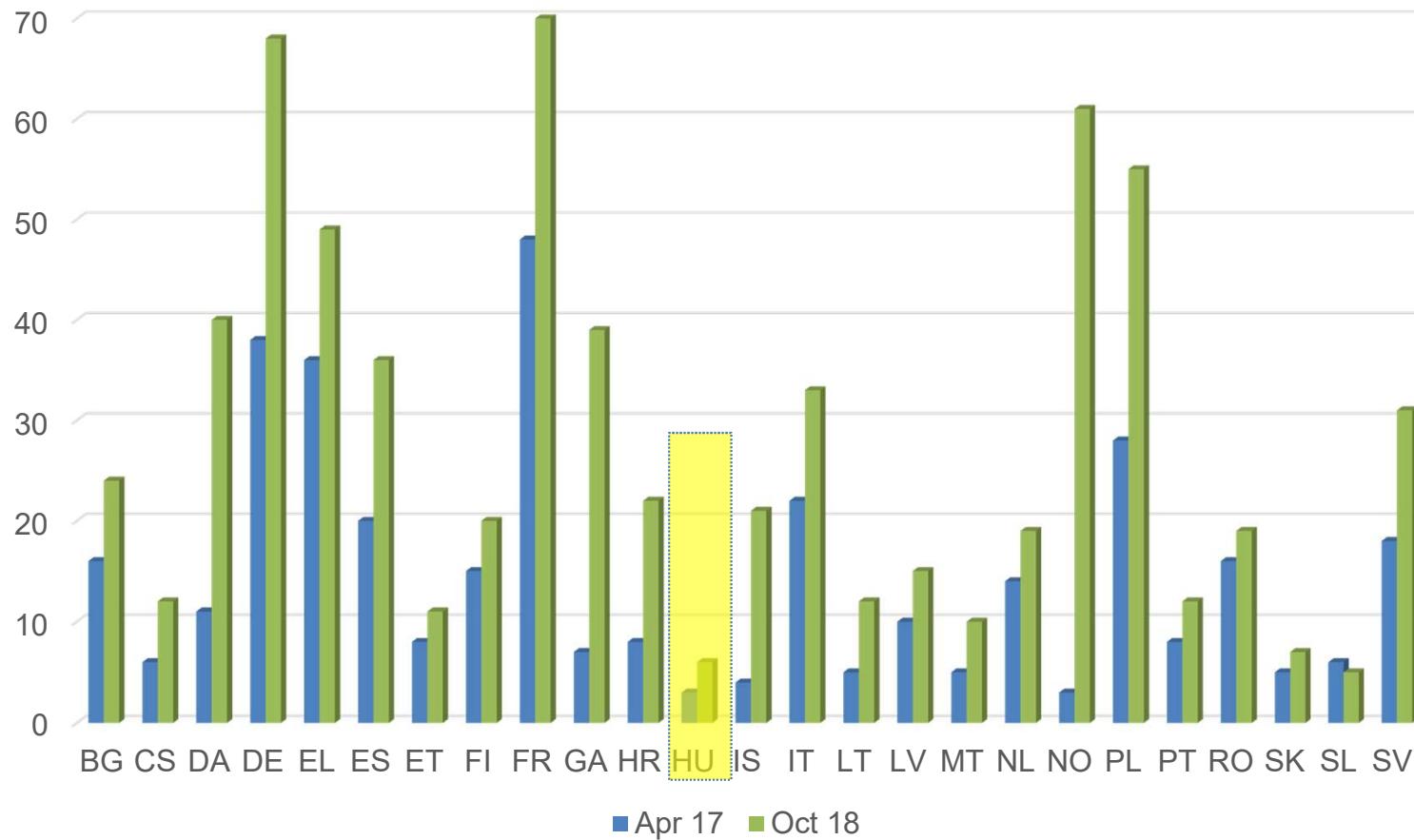


What has been achieved so far...?

ELRC Data by language



Development of LR collection by language

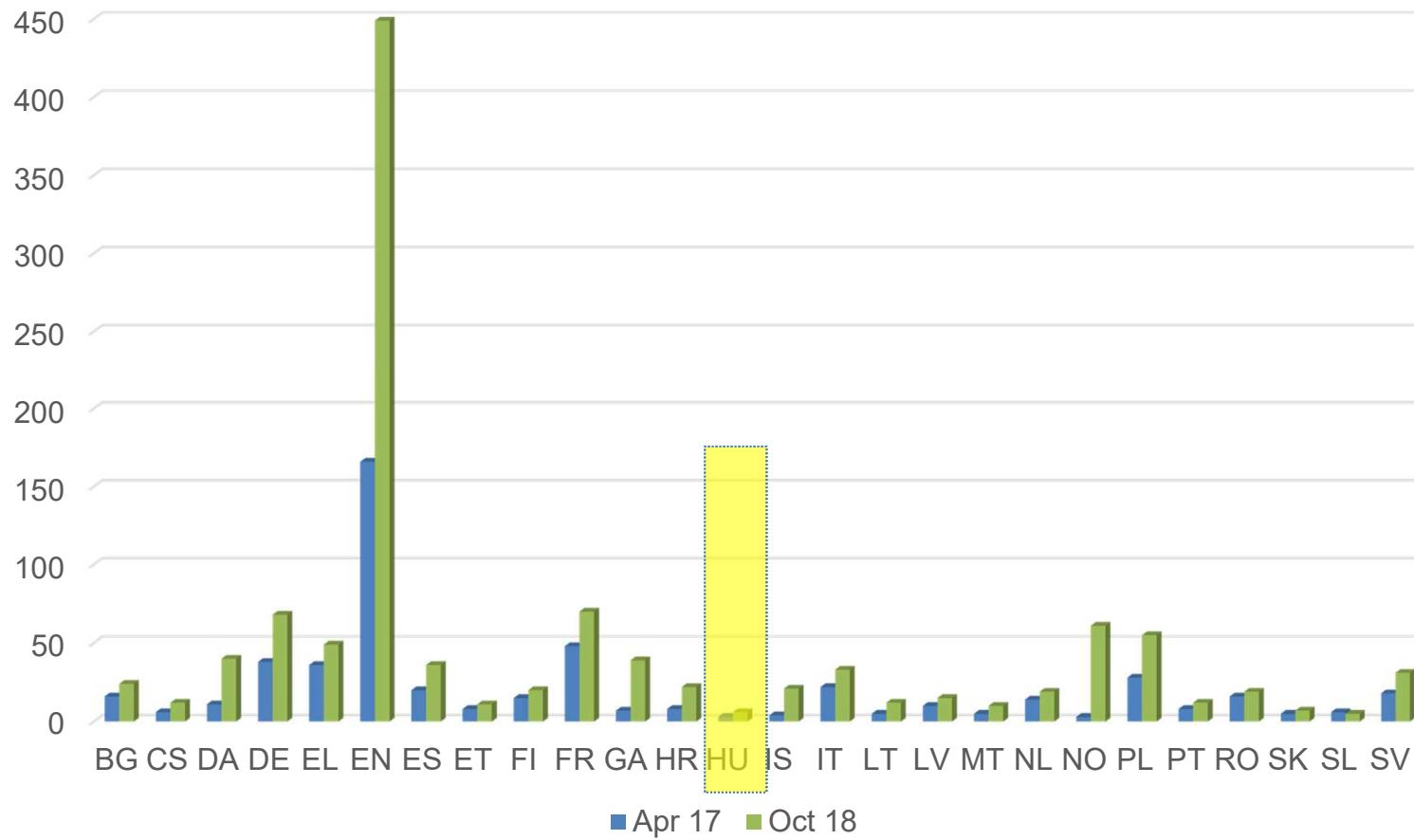


What has been achieved so far...?

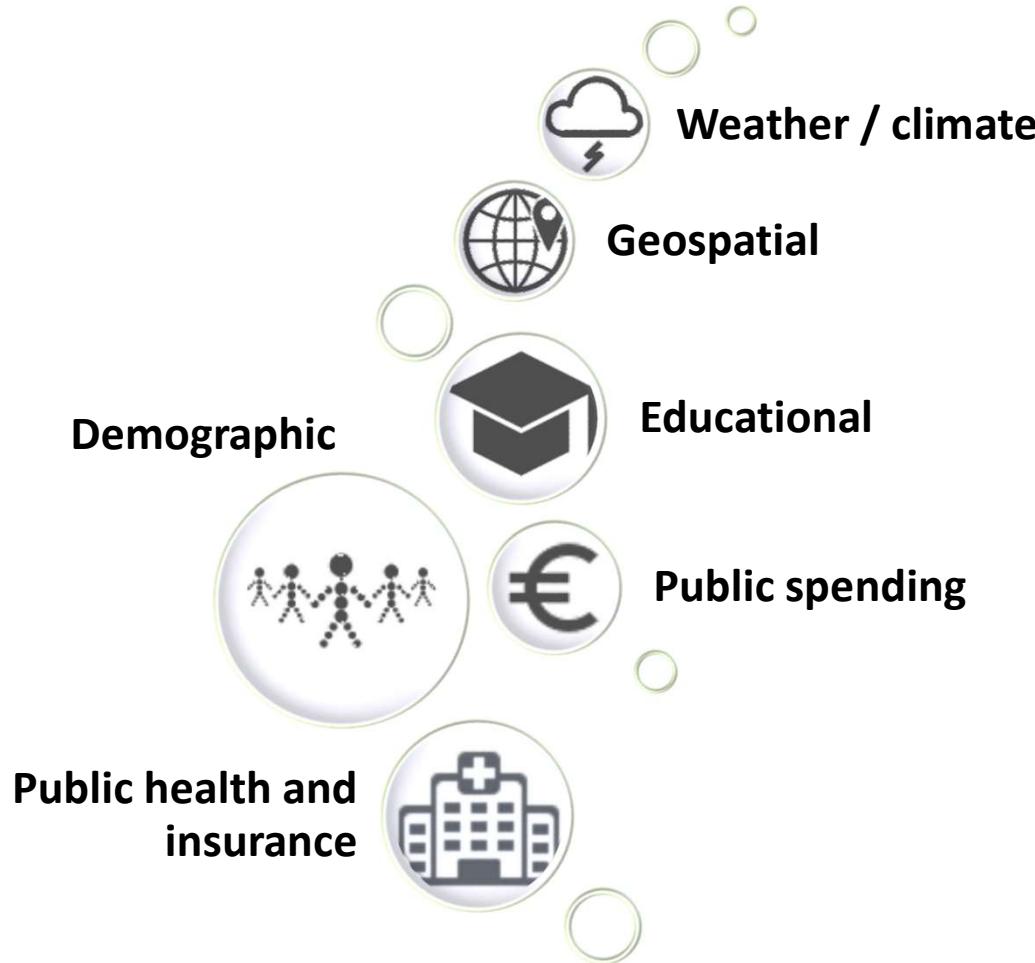
ELRC Data by language



Development of LR collection by language



The notion of data

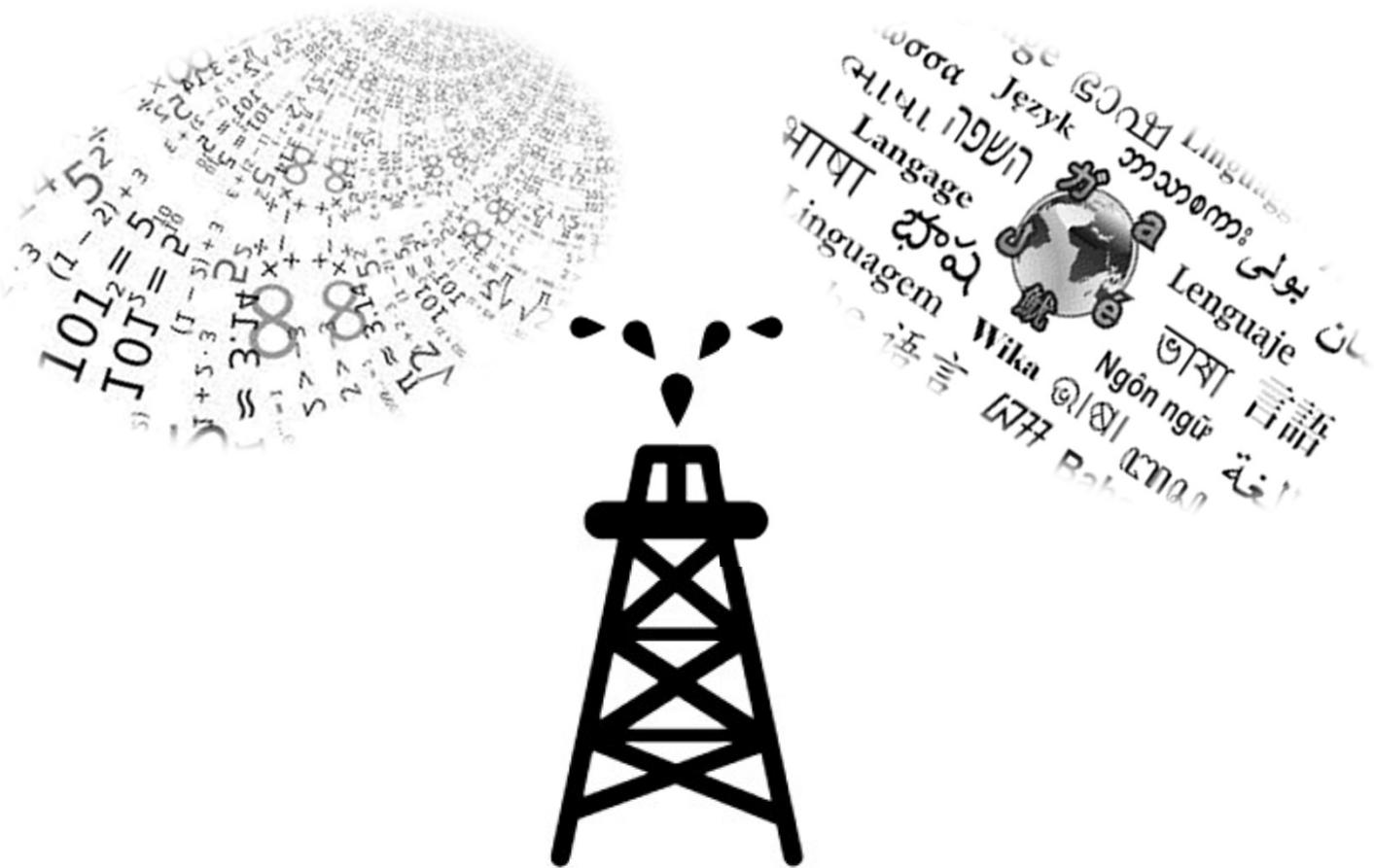


The notion of data

Data: the oil of the 21st century



The notion of data



The notion of data



Basic concepts:

- **Data:** any piece of electronically stored content
- **Dataset (or resource):** the collection of one or many data files **grouped** according to certain **criteria**
- **Metadata:** *data about the data*, i.e. description of a dataset with properties (e.g. title, publisher, description of the content and URL)

Sitemap | Legal notice | Contact | English (en) ▾

 European Union Open Data Portal

EUROPA > Open Data Portal > Data > Publisher > Publications Office > CORDIS - EU research projects un...

Data Applications Linked Data Developers' corner About Data provider's area Share

CORDIS - EU research projects under Horizon 2020 (2014-2020)

Publisher
[Publications Office](#)

Description
 This dataset contains projects funded by the European Union under the Horizon 2020 framework programme for research and innovation (H2020) from 2014 to 2020. Grant information is provided for each project, including RCN, ID, Acronym, Status, Programme, Topic, Title, Start Date, End Date, Objective, Total Cost, EC Max Contribution, Call Id, Funding Scheme, Coordinator, Coordinator Country, Participants (semi-colon separated list), Participant Countries (semi-colon separated list)
 For each participant you can find in the organisations file: RCN, ID, Acronym, Role, Organisation Name, Organisation Short Name, Organisation Type, Participation Ended, EC Contribution, Organisation Country
 Reference data (H2020 programmes and topics, funding schemes / types of action, and countries) can be found in this dataset:
<https://data.europa.eu/euodp/en/data/dataset/cordisref-data>
 CORDIS datasets are produced on a monthly basis. Therefore inconsistencies may occur between what is presented on the CORDIS live website and the datasets.

Resources

DOWNLOAD	H2020 Organisations	CSV
DOWNLOAD	H2020 Organisations	XLSX
DOWNLOAD	H2020 Projects	CSV
DOWNLOAD	H2020 Projects	XLSX
DOWNLOAD	H2020 Projects	ZIP

URI
<http://cordis.europa.eu/projects/>

Status
Under Development

Licence:
[Legal Notice](#)

Catalogue record

Added to data.europa.eu/euodp
2015-07-29
Updated on data.europa.eu/euodp
2017-06-01
Views: 17658
Downloads: 16453

Suggest a dataset
Is there data you would like to find on the portal?
[Make a suggestion>>](#)



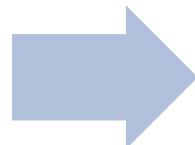
Language data (resources) for eTranslation through ELRC

The notion of language data



Data

- any piece of electronically stored **content**



(Textual) Language Data

- any piece of electronically stored **text**

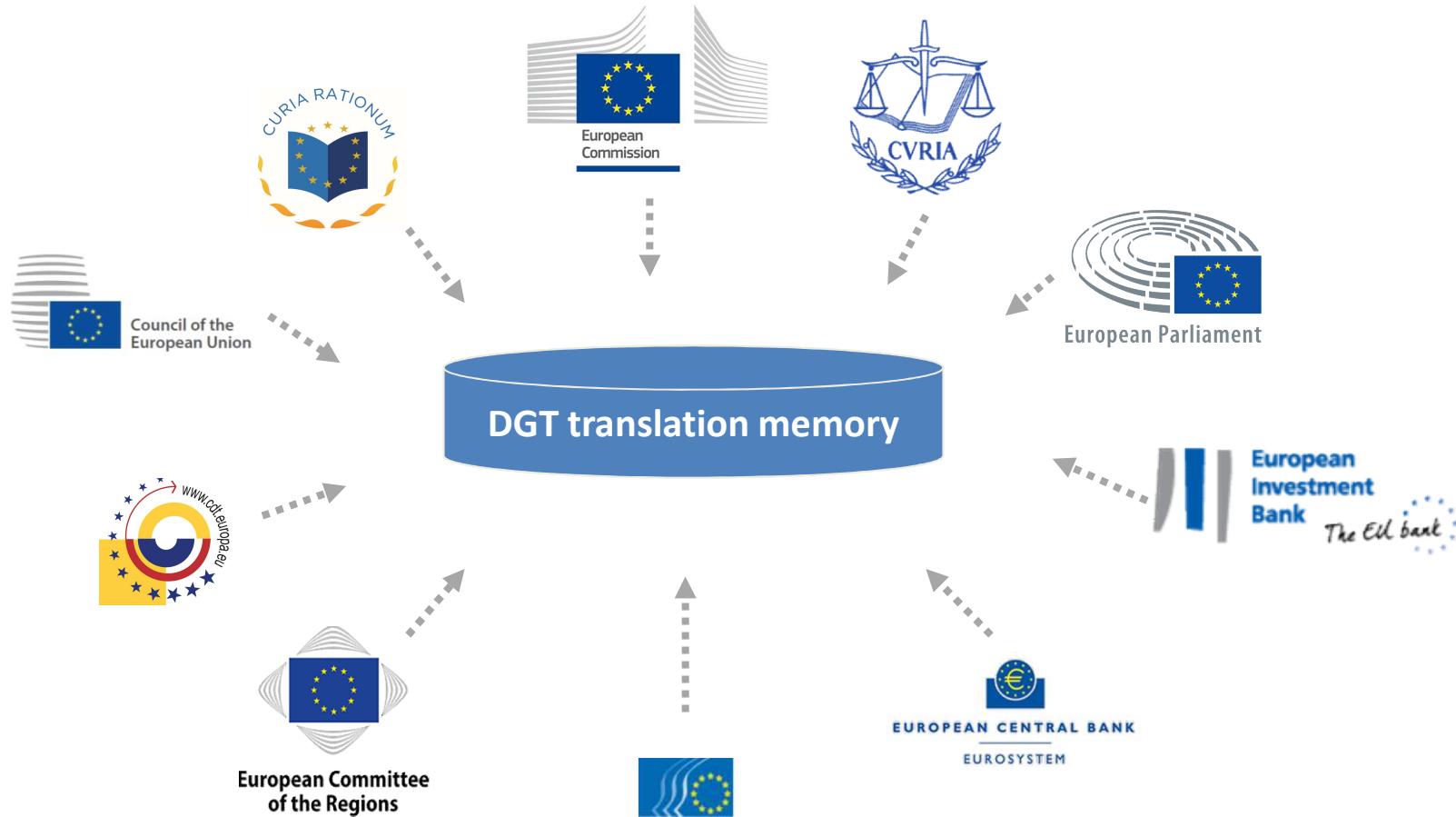
HU

Az emberi jogokra vonatkozó nemzetközi törvény abban az alapvetésben gyökerezik, hogy minden egyes embernek joga van az Emberi Jogok Egyetemes Nyilatkozatában lefektetett minden joghoz és szabadsághoz. Az emberi jogok szexuális orientációtól vagy a nemi azonosságtól függetlenül minden egyes emberre egyformán vonatkoznak.

EN

International human rights law is grounded on the premise that all individuals are entitled to the same rights and freedoms as set out in the Universal Declaration of Human Rights. Human rights shall apply equally to every human being, regardless of sexual orientation or gender identity.

Data used by eTranslation



Such data are already available
BUT
they are not enough...

What data are useful for eTranslation as per type | 1



- Any **electronically stored text** in an EU language plus NO and IS
- **Texts and their translations** (i.e. parallel bilingual or multilingual)

Hungarian text

A feladatok harmadik csoportját az Országgyűlés saját felkészülése képezi, amelynek során a Parlament felkészül az elnökség alatt megrendezendő parlamenti bizottsági találkozókra. Ennek irányítására létrejött egy Koordinációs csoport, amelynek vezetője az Országgyűlés Külügyi Hivatalának vezetője, tagjai az EU Főoszál vezetője, a Protokoll Főosztály vezetője, valamint a Főtitkárság, a Sajtóiroda, a Gazdasági Főigazgatóság és az Országgyűlés Kutatószolgálatának képviselői, illetve az Országgyűlés brüsszeli állandó irodájának vezetője.

Translation in English

The third group of the preparatory tasks is the National Assembly's own preparation, where the Parliament prepares for the meetings of the parliamentary committees to be held during its presidency. To coordinate these meetings, a Coordination Group was established that is led by the head of the Office for Foreign Relations of the National Assembly, its members are the head of the EU Department, the head of the Protocol Department and the representatives of the General Secretariat, the Press Office, the Economic Directorate and the Research Service of the National Assembly, as well as the Permanent Representative of the National Assembly to the EU in Brussels.

What data are useful for eTranslation as per type | 2



- List of terms and their translations, i.e. a **terminology**

English	Hungarian
accident insurance	balesetbiztosítás
health insurance	betegbiztosítás, egészségbiztosítás
household insurance	háztartási biztosítás
insurance broker	biztosítási ügynök
medical insurance	betegbiztosítás
sickness insurance	betegbiztosítás
unemployment insurance	munkanélküliségi biztosítás
health insurance card	egészségbiztosítási kártya
nursing care insurance	gondozási biztosítás
...	...

What data are useful for eTranslation as per format | 1



- In principle, any text in machine readable format
- But, some formats are more “MT-ready” than others, i.e. they require less manual or automatic processing
- More processing introduces more errors in the final output, making it less useful for eTranslation

File formats for parallel texts



A pokoli komédia még egyre tartott a bőrzén. A halálra ítélt papírok, a bondavári gyártelep s a bondavári vasút részvényei egyik kézből a másikba repültek. Most már a komikumig vitték a tragédiát, s kezdett humor vegyülni a szerencsétlenségbe. Ez a szó: "Bondavár" csak arra való volt, hogy derültséget idézzen elő a bőrzeemberek között. Aki az utolsó részvényén túladhatott nagy veszteséggel, nevetett azon, aki megszerezte azt. Kezdték a részvényeket megfoghatlan becsű tárgyakért cserébe kínálgatni. Ráadásul egy új esernyőre egy ócska esernyőért. Használták előfizetési ekvivalensül olyan lapokért, amiket valakinek a szerkesztő a nyakára köt erővel. Ajándékozgatták jótékony célokra.

What data are useful for eTranslation as per format | 2



- The following formats are particularly useful (in descending order):
 - For bilingual/multilingual parallel texts
 1. Translation memories (.tmx)
 2. XML translation files (.xliff)
 3. Plain text (.txt, .csv)
 4. Spreadsheets (e.g. .xlsx)
 - For terminologies
 1. TermBase eXchange (.tbx)
 2. Plain text (.txt, .csv)
 3. Spreadsheets (e.g. .xlsx)
 - For monolingual texts
 1. Plain text (.txt, .csv)



File formats of parallel texts and their manipulation

Preparing your data | 1



Don'ts



This is a paragraph in English. This is a paragraph in English.

Ez a fenti bekezdés magyar fordítása. Ez a fenti bekezdés magyar fordítása.

A second paragraph in English. A second paragraph in English.

Ez a második bekezdés magyar fordítása. Ez a második bekezdés magyar fordítása.

Preparing your data | 2



Don'ts



This·is·a·paragraph·in·English.·This·is·a·
paragraph·in·English.·This·is·a·paragraph·in·
English.·This·is·a·paragraph·in·English.·This·is·
a·paragraph·in·English.·This·is·a·paragraph·
in·English.·This·is·a·paragraph·in·English..
This·is·a·paragraph·in·English.·This·is·a·
paragraph·in·English.·This·is·a·paragraph·in·
English.·This·is·a·paragraph·in·English..¶

A·sentence·in·English.¶

¶

A·second·paragraph·in·English..A·
second·paragraph·in·English..A·second·
paragraph·in·English..A·second·paragraph·in·
English..A·second·paragraph·in·English..A·
second·paragraph·in·English..A·second·
paragraph·in·English..¶

¶

Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..¶

¶

Ez·a·mondat·magyar·fordítása·balra..¶

¶

Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..Ez·a·bal·oldali·bekezdés·magyar·
fordítása..¶

Preparing your data | 3



Don'ts

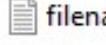
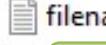
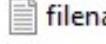
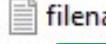
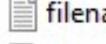
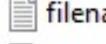
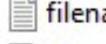
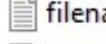
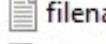
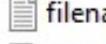
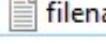


English	Magyar
<p>This is a paragraph in English. This is a paragraph in English.</p>	<p>Ez a bal oldali bekezdés magyar fordítása. Ez a bal oldali bekezdés magyar fordítása.</p>
<p>A second paragraph in English. A second paragraph in English.</p>	<p>Ez a bal oldali bekezdés magyar fordítása. Ez a bal oldali bekezdés magyar fordítása.</p>

Preparing your data | 4



Do's

-  filename01_EN.txt
-  filename01_HU.txt
-  filename02_EN.txt
-  filename02_HU.txt
-  filename03_EN.txt
-  filename03_HU.txt
-  filename04_EN.txt
-  filename04_HU.txt
-  filename05_EN.txt
-  filename05_HU.txt
-  filename06_EN.txt
-  filename06_HU.txt
-  filename07_EN.txt
-  filename07_HU.txt
-  filename08_EN.txt
-  filename08_HU.txt
-  filename09_EN.txt
-  filename09_HU.txt
-  filename10_EN.txt
-  filename10_HU.txt

Use **identical filenames** for each document pair (source – translation)

Preparing your data | 5



Do's

- filename01_EN.txt
- filename01_HU.txt
- filename02_EN.txt
- filename02_HU.txt
- filename03_EN.txt
- filename03_HU.txt
- filename04_EN.txt
- filename04_HU.txt
- filename05_EN.txt
- filename05_HU.txt
- filename06_EN.txt
- filename06_HU.txt
- filename07_EN.txt
- filename07_HU.txt
- filename08_EN.txt
- filename08_HU.txt
- filename09_EN.txt
- filename09_HU.txt
- filename10_EN.txt
- filename10_HU.txt

Include language
identifiers in the filename

Preparing your data | 5



Όνομα

-  [τεκτ.txt](#)
-  [Μετάφραση αναφοράς.txt](#)
-  [κείμενο.txt](#)
-  [traduction française.txt](#)
-  [texte.txt](#)
-  [text.txt](#)
-  [report.txt](#)
-  [report in Bulgarian.txt](#)
-  [README.txt](#)

Αρχεία σε ειδική γλώσσα θωρηψη τημέρου παραγόντων				
English	Bulgarian	French	Greek	
text.txt	текст.txt	текст.txt	текст.txt	κείμενο.txt
report.txt	report in Bulgarian.txt	traduction française.txt	traduction française.txt	Μετάφραση αναφοράς.txt

Criteria for grouping your data into meaningful datasets



- A dataset is a collection of data **grouped according to certain criteria**
- For the purpose of enhancing and adapting CEF eTranslation, two criteria are critical:
 - **Language(s)**: each collection is defined by the language or language pairs of its data, e.g.
 - *Collection of texts in English – Hungarian*
 - *Documents in English – Hungarian - French*
 - **Domain**: each collection ideally belongs to a single domain, e.g.
 - *Collection of texts in English – Hungarian in the culture domain*
 - *Social security documents in English – Hungarian- French*

Preferred domains



- Administrative/regulatory domain and
- Topics relevant to the CEF DSIs

CEF DSI	Domain
Online Dispute Resolution	Consumers' rights, complaints
Electronic Exchange of Social Security Information	Social security, insurance
eProcurement	Public procurement, contractual agreements
European e-Justice Portal	Justice, Law
eHealth	Health, Medicine
Business Registers Interconnection System	Business, market
Safer Internet	
Cybersecurity	
Public Open Data	
Europeana	Culture



How to contribute your data to CEF eTranslation

A step-by-step guide



The ELRC-SHARE Repository

- Access to, sharing and contribution of language resources
- Access to tools and services catalogue (upcoming)



The screenshot shows the homepage of the ELRC-SHARE repository. At the top, there is a dark blue header bar with the URL <https://elrc-share.eu>. Below the header is a navigation bar with links: Home, Browse Resources, Help, About, Register, and Login. On the right side of the header, there is a logo for European Language Resource Coordination, Connecting Europe Facility, and ELRC-SHARE, accompanied by a network icon. The main content area features a large image of many flags of different countries flying outside a modern building with a glass facade. Overlaid on this image is the text "ELRC-SHARE Repository". Below the image, there is a search bar with the placeholder "Type in your keywords, please..." and a "Search" button. A message at the bottom left says "52 language resources at your disposal (and counting)". At the very bottom, there is a welcome message: "Welcome to the ELRC-SHARE repository!" followed by a paragraph of text about the repository's purpose and how to contribute.

ELRC-SHARE Repository

52 language resources at your disposal (and counting)

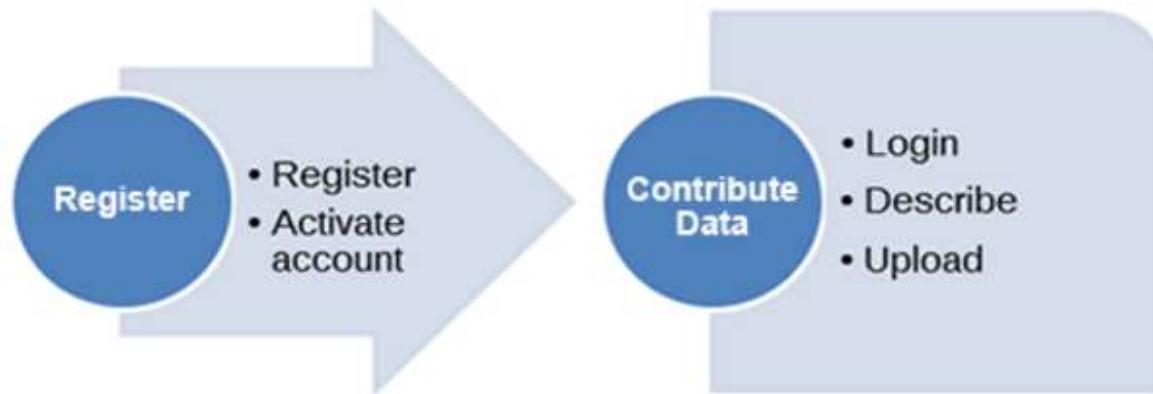
Type in your keywords, please...

Welcome to the ELRC-SHARE repository!

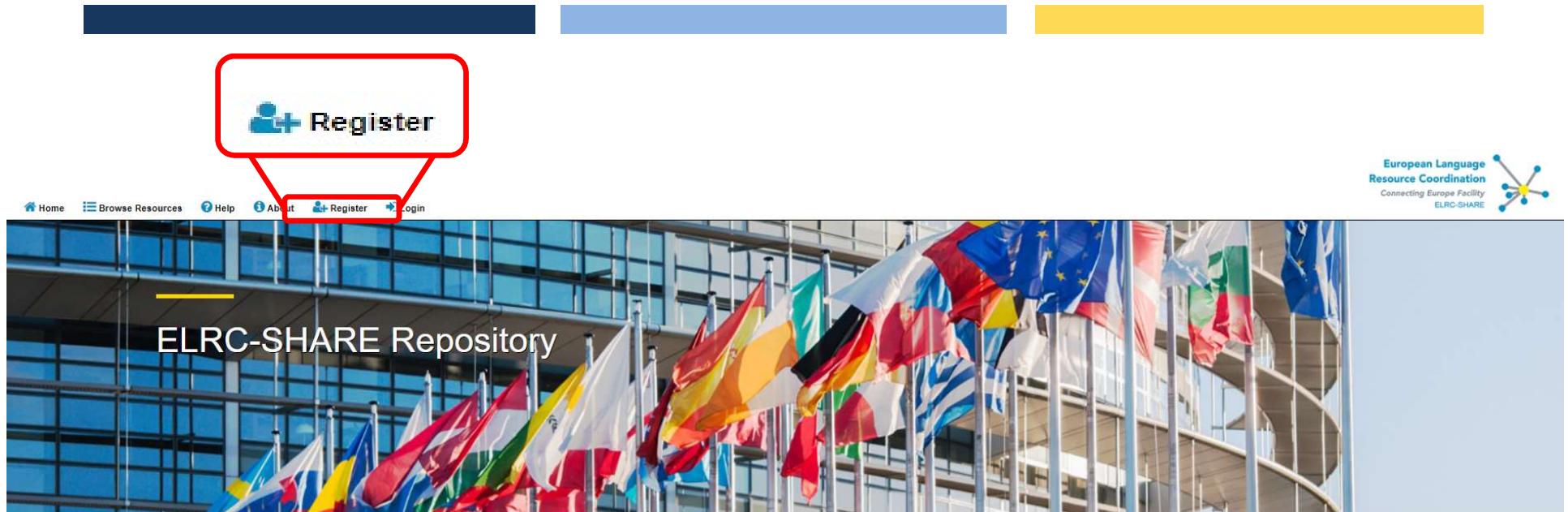
The ELRC-SHARE repository is used for documenting, storing, browsing and accessing [Language Resources](#) that are collected through the European Language Resource Coordination and considered useful for feeding the CEF Automated Translation ([CEF.AT](#)) platform.

If you want to contribute resources, all you have to do is [register](#) (new user) or [login](#) (returning user) and go on to describe and upload your data with a simple form.

How to Contribute Data



How to Register (1/2)



The screenshot shows the homepage of the ELRC-SHARE Repository. At the top, there is a navigation bar with links: Home, Browse Resources, Help, About, Register, and Login. The 'Register' link is highlighted with a red box and a red arrow pointing to it from a callout bubble containing the text 'Register' and a user icon. Below the navigation bar is a large banner featuring a modern building with many windows and a row of European flags flying in front of it. The text 'ELRC-SHARE Repository' is displayed prominently on the left side of the banner. To the right of the banner, there is a search bar with the placeholder text 'Type in your keywords, please...' and a magnifying glass icon. In the top right corner of the page, there is a logo for 'European Language Resource Coordination Connecting Europe Facility ELRC-SHARE'.

Welcome to the ELRC-SHARE repository!

How to Register (2/2)

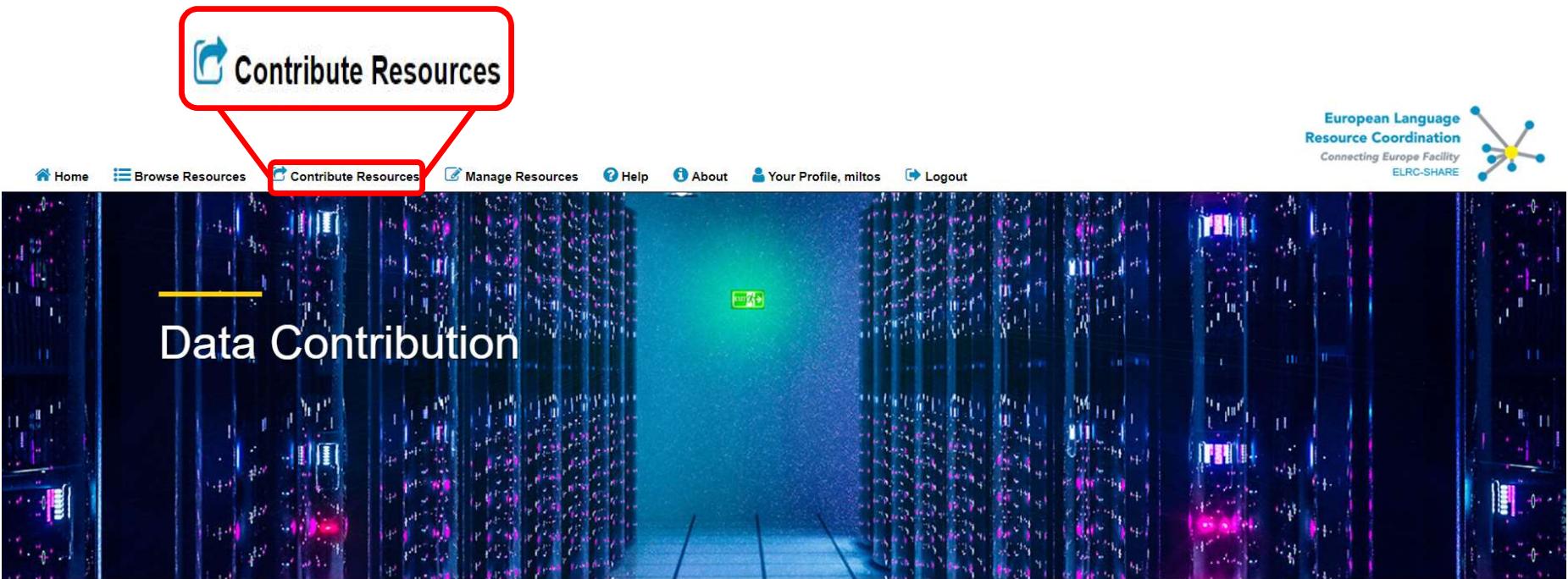


- Fill in the required info
- Read the *Terms of Service* and click *Accept*, if you agree
- Click the *Create Account* button
- Activate your account according to the guidelines emailed to you

*All fields are required

Desired account name*	MyAccountName
First name*	FirstName
Last name*	LastName
E-mail*	myemail@myemail.com
Country*	Greece
Organization*	MYORG
Phone number*	123456789
Password*	****
Password confirmation*	
<input checked="" type="checkbox"/> I accept the ELRC Terms of Service for registered users.	
<input type="button" value="Create Account"/>	

How to Contribute Data (1/6)



The screenshot shows the ELRC website's main navigation bar with several options: Home, Browse Resources, Contribute Resources (which is highlighted with a red box), Manage Resources, Help, About, Your Profile (mitos), and Logout. Below the navigation bar, a large banner features the text "Data Contribution" over a background of abstract blue and purple server racks. In the top right corner of the banner, there is a small "Edit" icon. The overall theme is digital and technical.

New Resource

Resource Title*

The name by which the resource is already known or by which you would like it to be known; e.g. "The GSRT bilingual corpus of Greek-English bulletins"

How to Contribute Data (2/6)



- Fill in the details of the dataset

Resource Title* The name by which the resource is already known or by which you would like it to be known; e.g. "The GSRT bilingual corpus of Greek-English bulletins"

Resource short description* A short description, including any information considered useful about the resource, e.g. whether it's a dataset (collection of documents) or a lexicon, glossary, terminological resource etc., its size, language(s), classification information (e.g. health reports, news bulletins, lexicon of sports terminology etc.)

Language(s)

Danish
Dutch Flemish
English
Estonian
Finnish
French
German
Hungarian

How to Contribute Data (3/6)



- Three modes for contributing your data

Contribution Mode*

- Upload ZIP archive
- Provide URL of resources
- eDelivery (Generate XML file to attach to your eDelivery contribution)

Please select the way you wish to contribute your data. Uploading a ZIP archive is recommended.

Upload Resource*

No file chosen

Please upload a **.zip file** up to 100MB.

In case the **.zip file** file you wish to upload is larger than 100MB, please contact elrc-share@ilsp.gr

How to create a .zip file



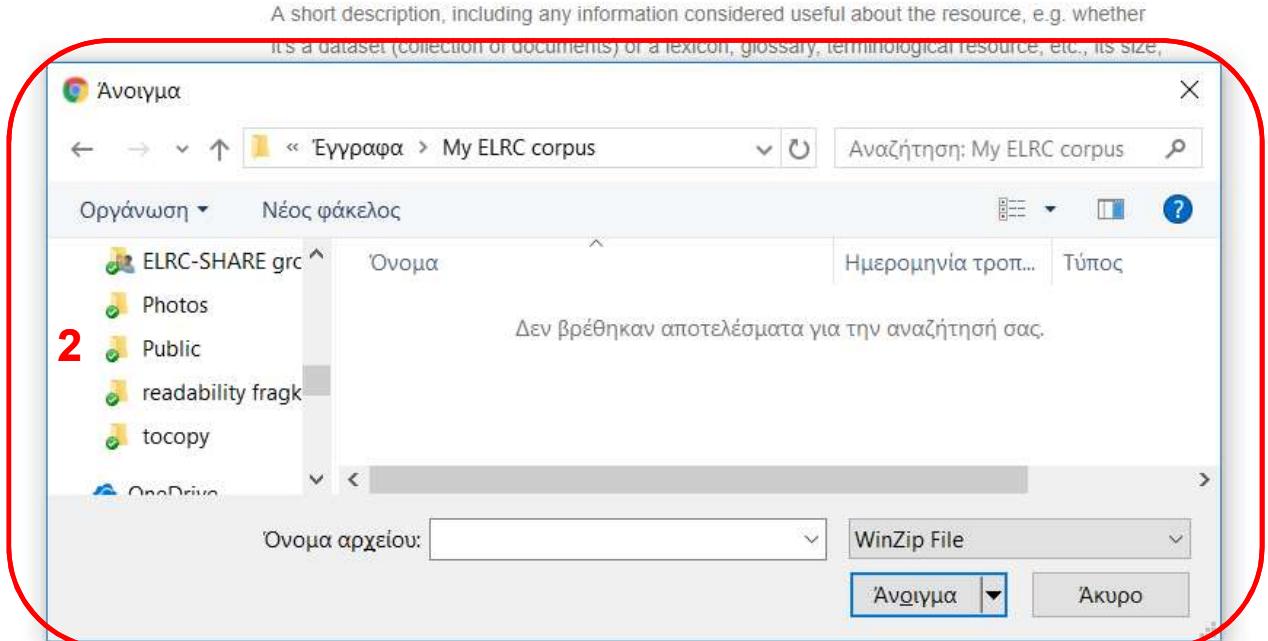
- Free file compression tools (indicative):
 - 7zip
 - PeaZip
 - Hamster Free Zip Archiver
 - Universal Extractor
 - ZipltFree
- Windows embedded compression functionality

How to Contribute Data (4/6)



1. Click on Choose file
2. Locate your resource in your hard disk
3. Click on Submit

A short description, including any information considered useful about the resource, e.g. whether it's a dataset (collection of documents) or a lexicon, glossary, terminological resource, etc., its size,



Ανοιγμα

← → ⌂ ↑ « Έγγραφα > My ELRC corpus

Αναζήτηση: My ELRC corpus

Οργάνωση ▾ Νέος φάκελος

	Όνομα	Ημερομηνία τροπ...	Τύπος
	ELRC-SHARE grc		
	Photos		
	Public		
	readability fragk		
	tocopy		
	OneDrive		

Δεν βρέθηκαν αποτελέσματα για την αναζήτησή σας.

Όνομα αρχείου: WinZip File

Άνοιγμα ▾ Άκυρο

Upload Resource **1** No file chosen

Please upload a **.zip** file up to 100MB.

In case the **.zip** file you wish to upload is larger than 100MB, please contact elrc-share@ilsp.gr

3

How to Contribute Data (5/6)



- Alternatively indicate a url (directory listing)

Language(s)*

- Bulgarian
- Czech
- Croatian
- Danish
- Dutch; Flemish
- English
- Estonian
- Finnish
- French
- German
- Hungarian

The language(s) of the resource; for resources with multiple languages, hold down CTRL key to select multiple values

Contribution Mode*

Upload ZIP archive
 Provide URL of resources

Please select the way you wish to contribute your data. Uploading a ZIP archive is recommended.

Resource URL*

www

Please provide a URL containing the files you wish to contribute

Submit **Reset**

How to Contribute Data (6/6)

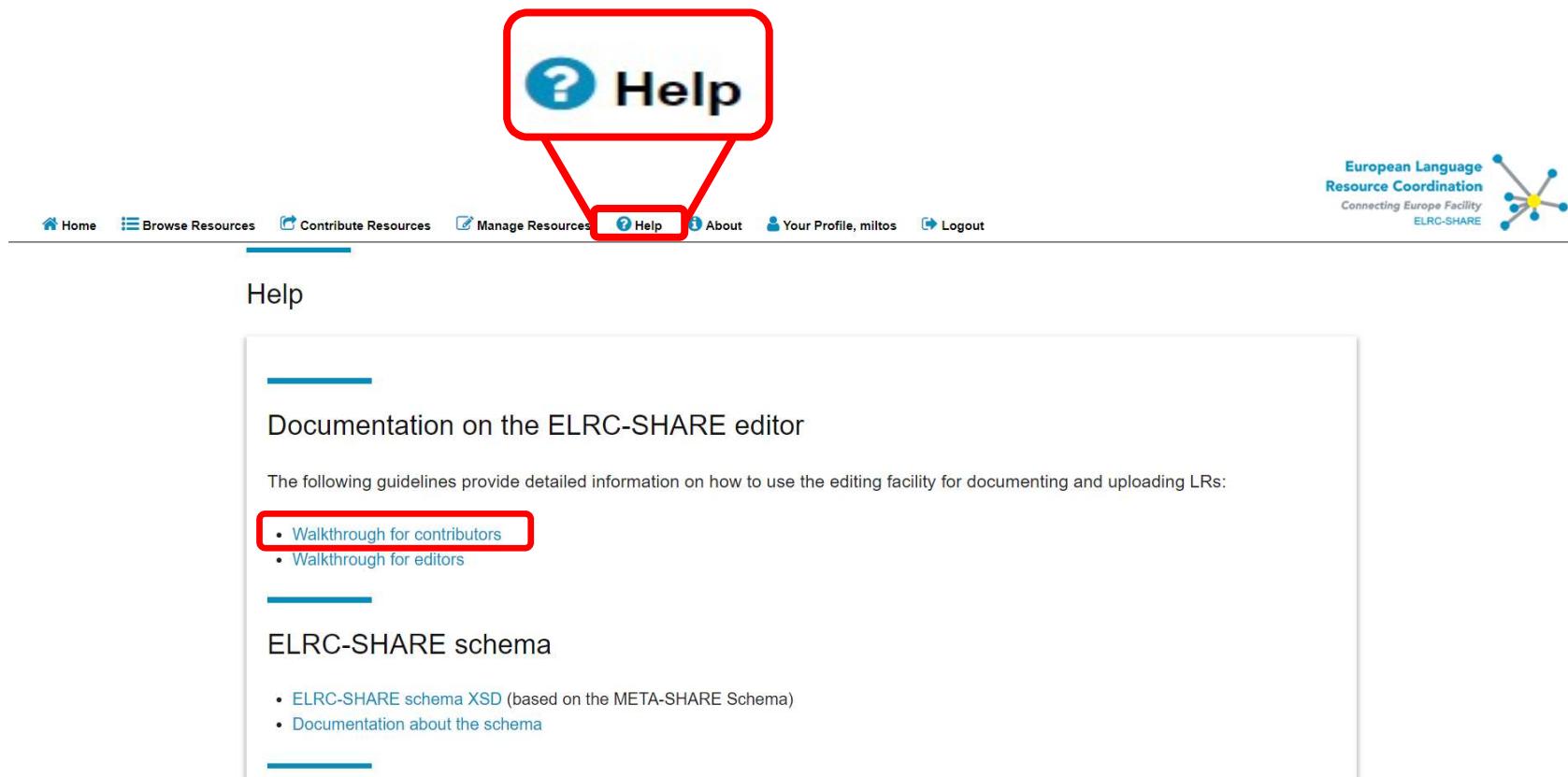


- Repeat the process if you want to contribute another resource, or log out



The screenshot shows the ELRC SHARE interface with a red box highlighting the "Logout" button in the top navigation bar. The "Logout" button is accompanied by a blue arrow icon pointing right. The rest of the interface includes a banner with "Data Contribution" text, a sidebar with various menu items like Home, Browse Resources, Contribute Resources, Manage Resources, Help, About, Your Profile, and a user icon, and the European Language Resource Coordination logo in the top right corner.

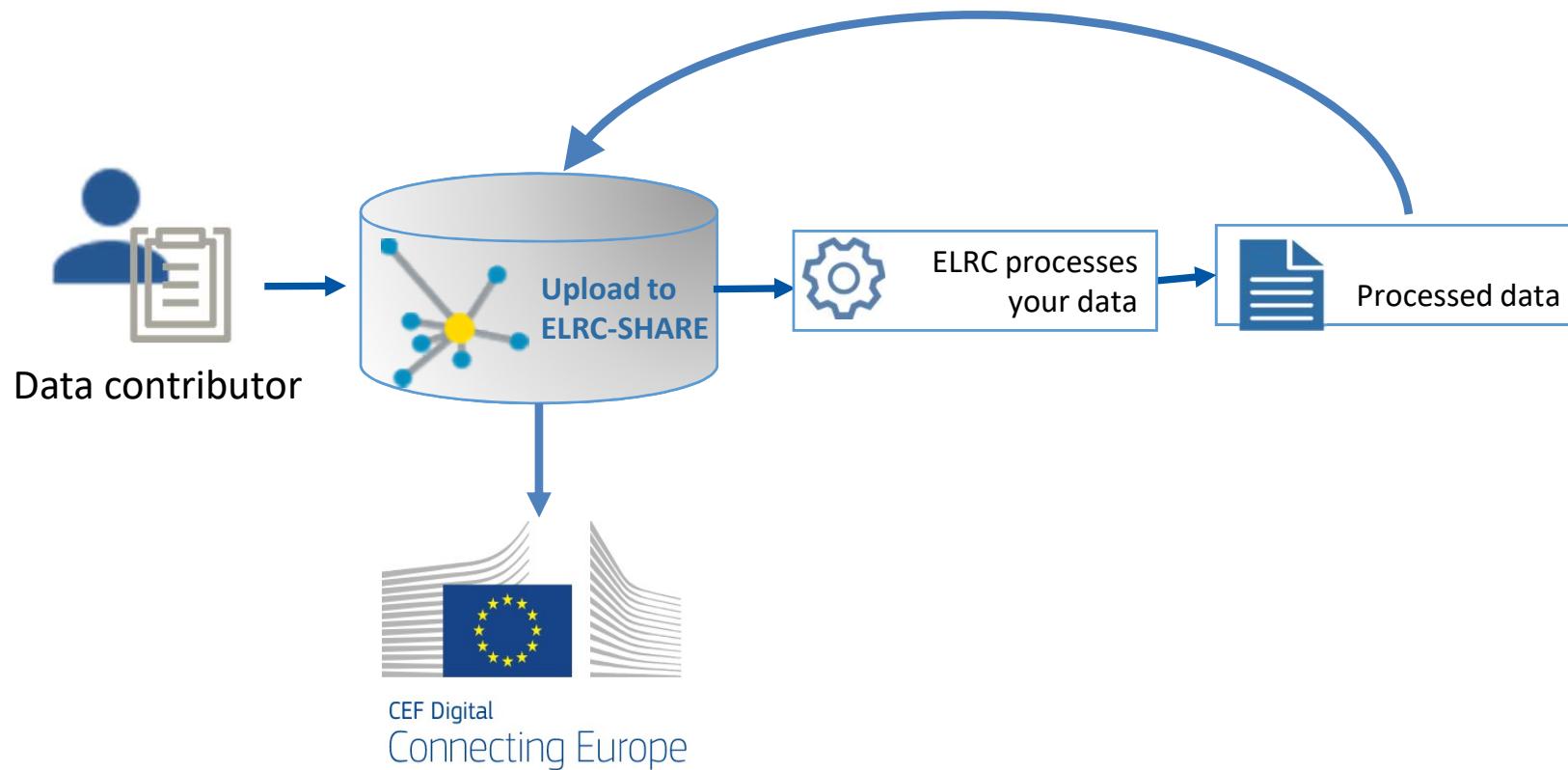
Guidelines for contributors



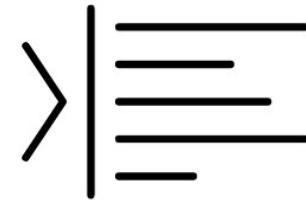
The screenshot shows the ELRC-SHARE editor interface. At the top, there is a navigation bar with links: Home, Browse Resources, Contribute Resources, Manage Resources, Help, About, Your Profile, miltos, and Logout. The 'Help' link is highlighted with a red box and a red speech bubble containing a question mark icon and the word 'Help'. Below the navigation bar, the page title is 'Help'. The main content area has a heading 'Documentation on the ELRC-SHARE editor' and a list of items: 'Walkthrough for contributors' and 'Walkthrough for editors', where 'Walkthrough for contributors' is also highlighted with a red box. There is another section titled 'ELRC-SHARE schema' with a list of items: 'ELRC-SHARE schema XSD (based on the META-SHARE Schema)' and 'Documentation about the schema'.

What happens next?

What happens to your data?



Language processing services



Data extraction

If your data is trapped in archives and databases, we can help extract it

Anonymisation

Does your data contain private info? We can help to anonymise

Cleaning

If your data is messy (i.e., lots of noise), we will clean it up

Re-formatting

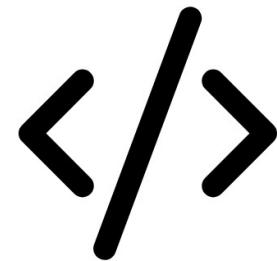
Need to re-format DOCX to XML, or PDF to WORD? Let us do it for you!

Language processing services



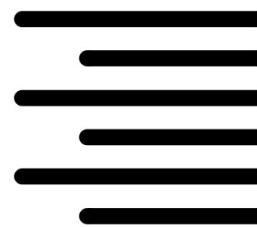
Data conversion

If your data isn't converted to the proper formats, we can help convert it



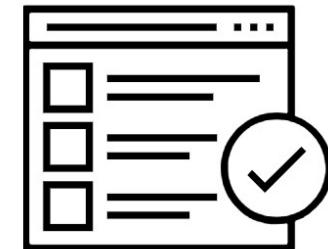
Tag removal

Does your data contain unneeded tags? We can assist in removing them!



Alignment

Translations aren't aligned? We'll do it for you with our tools!



Metadata

Metadata are crucial! We can organise and validate metadata for your team

What has happened to your data?

File01_hu.txt
File01_en.doc
File02_hu.pdf
File02_en.txt
File03_hu.doc
File03_en.doc
...

After
processsing

```
<tu tuid="3931">
  <tuv xml:lang="en">
    <seg>Weights: they express the typical
percentages the groups of different goods or services
represent within the total expenditures of
households.</seg> </tuv>
  <tuv xml:lang="hu">
    <seg>Súlyarány: kifejezi, hogy az egyes termékek és
szolgáltatások csoportjai a lakosság összes kiadásán
belül milyen jellemző százalékarányt
képviselnek.</seg></tuv></tu>
<tu tuid="3932">
  <tuv xml:lang="en">
    <seg>Weights are revised annually, based on household
consumption in the year two years before the reference
year.</seg>
  </tuv>
  <tuv xml:lang="hu">
    <seg>A súlyok felülvizsgálata évente történik, a
tárgyévet két évvel megelőző év lakossági fogyasztása
alapján.</seg></tuv></tu>
```

How your dataset is described



Translations of Hungarian from public websites (Processed)

Attribution details: Research Institute for Linguistics, Hungarian Academy of Sciences.

A webcrawl of 14 different websites covering parallel corpora of Hungarian with Polish, Czech, Swedish, Finnish, French, German, Italian, English and Slovenian (Processed)

[Back](#) [Download](#) [Edit Resource](#)

Distribution
Availability: Available
Licences
Terms for PSI-compliant resources
Open Under-PSI
Distribution Details
Attribution Details: Research Institute for Linguistics, Hungarian Academy of Sciences
Contact Person
Tamas Varadi 

text 
Multilingual text corpus
Languages
Dutch, Flemish (nl)
Slovenian (sl)
German (de)
Czech (cs)
Hungarian (hu)
English (en)
Polish (pl)
Finnish (fi)
Swedish (sv)
French (fr)
Linguality
Linguality type: Multilingual
Text Format
TMX
Size
13 Files
11,087 Translation Units
Character encoding
UTF-8

Resource Creation
Resource Creator
Tamas Varadi 
Funding Project
European Language Resource Coordination LOT3 (ELRC Data - Tools and Resources for CEF Automated Translation - LOT3 (SMART 2015/1091 - 30-CE-0816766/00-92))
URL: http://www.lrc-coordi...
Funding Type: Service Contract
Funder: European Commission
Funding Country: European Union (EU)
Project duration: 13/12/2016 - 12/02/2020
Metadata
Created: 20/07/2017
Metadata Language: English (en)
Metadata Creator
Fraser Bowen 
Version
Version: 2.0
Relations
Related Resource: Translations of Hungarian from public websites
Relation Type: Is Aligned Version Of

On-site assistance



**All these services can also be offered on-site to
all data contributors free of charge**



On-site assistance



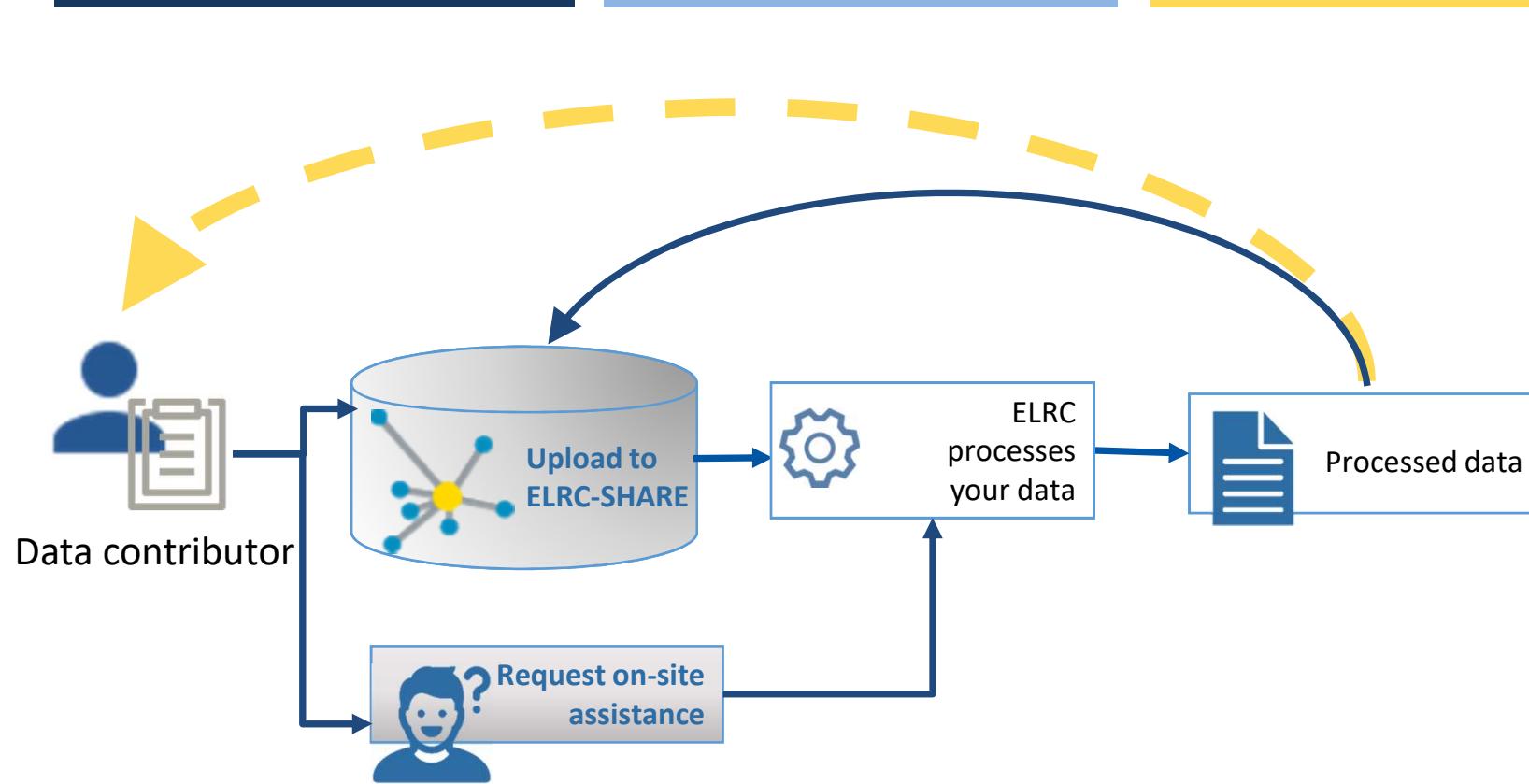
**Our team of experts will travel
directly to assist you
at your own offices**

On-site assistance



**We will fix your data issues and return the processed data directly to you. We can also help to improve your data management processes.
Just ask!**

What happens to your data?



How to request services and help

ELRC onsite assistance



www.lr-coordination.eu/request-onsite-assistance

Submit a request for on-site assistance by filling out the form below. See a list of services [here](#).

First name *

Last name *

Institution *

Country *

Email *

Types of assistance required *

- Legal assistance
- Data processing
- Anonymisation
- Other

Description of assistance required

Submit



www.lr-coordination.eu/helpdesk

[Home](#) [Discover](#) [Resources](#) [Services](#) [Events](#) [Anchor Points](#) [News](#) [Helpdesk](#)



Helpdesk for Language Resources

Please feel free to contact us through one of the following channels:

Telephone* **+33 970 440 522**

Secretariat Support **+49 681 857 7552 85**

Skype **ELRC Helpdesk**

E-mail **help@lr-coordination.eu**



Köszönöm!

Email: info@lr-coordination.eu

Website: www.lr-coordination.eu



Icons used in this presentation



- By [Michael Mellon](#), GB, , CC-BY 3.0 US
- By [Joana Pereira](#), BR, CC-BY 3.0 US
- By [Becca O'Shea](#), NZ, CC-BY 3.0 US
- By [Creative Stall](#), Basic licence www.iconfinder.com
- By [Creative Stall](#), PK, CC-BY 3.0 US
- By [Arthur Shlain](#), IL, CC-BY 3.0 US
- By [Shmidt Sergey](#), US, CC-BY 3.0 US
- By [Gregor Cresnar](#), CC-BY 3.0 US
- By [anbileru adaleru](#), CC-BY 3.0 US
- By [Vectors Market](#), CC-BY 3.0 US