



EUROPEAN LANGUAGE RESOURCE INFRASTRUCTURE



Co-financed by the European Union
Connecting Europe Facility

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

Initiative

- ▶ Co-financed by the European Union - Connecting Europe Facility (CEF) Programme
- ▶ From 2017/10/02 to 2019/09/30
- ▶ Provision of an infrastructure to help collect, prepare and share language resources that can improve translation services in Europe
- ▶ 7 partners, 4 Member States: France, Ireland, Portugal and Spain



Project participants

- ▶ Vicomtech, Spain:
 - ▶ Project coordinator
 - ▶ Research centre with strong expertise in language technologies and technology transfer
- ▶ Agencia para a Modernizaçao Administrativa (AMA), Portugal:
 - ▶ Public institute in charge of administrative modernization and digital transformation
- ▶ Dublin City University (DCU), Ireland:
 - ▶ Among the top centres in language technologies in Europe
- ▶ Evaluations and Language resources Distribution Agency (ELDA), France:
 - ▶ European leaders in LR collection, processing and distribution
- ▶ University of Lisbon, Faculty of Sciences (FCUL), Portugal:
 - ▶ Experts in language technology and language resource sharing infrastructures
- ▶ Linkare, Portugal:
 - ▶ Experts in professional software development and CEF building blocks
- ▶ Secretary of State for Digital Advancement (SEAD), Spain:
 - ▶ Public institution in charge of the Spanish National Plan for Language Technologies



Why ELRI?

- ▶ It's not easy to share data:
 - ▶ Am I allowed to share my data?
 - ▶ With whom?
 - ▶ It takes time and effort: why should I do it?
- ▶ It's not easy to prepare and distribute data:
 - ▶ Requires expertise in language technology
 - ▶ Requires expertise in administrative and legal issues
 - ▶ Requires expertise in developing software infrastructure
 - ▶ It takes time to prepare useful multilingual language resources
- ▶ But...
 - ▶ We still need language resources from Member States to support a truly multilingual Europe
 - ▶ We can build upon current expertise to offer a flexible environment that can benefit all stakeholders



Objectives

- ▶ Complement other EU initiatives such as ELRC
- ▶ Develop additional means to collect and share language resources
- ▶ Improve human and automated translation services at national and European level
- ▶ Provide both short-term and long-term benefits for stakeholders:
 - ▶ Member States' public institutions and translation centres
 - ▶ European institutions and Digital Service Infrastructure
 - ▶ Users and developers of European translation services
- ▶ Take into account current difficulties:
 - ▶ Practical
 - ▶ Technical
 - ▶ Legal and administrative



Approach

- ▶ Build an infrastructure to simplify resource collection from the different Member States:
 - ▶ National Relay Stations for each Member State
 - ▶ Enable local collection, processing and handling of LRs
 - ▶ Start local, build global!
- ▶ Integrate language processing technology to:
 - ▶ Automate the creation of translation memories from raw user data
 - ▶ Automate the preparation of clean language resources
- ▶ Involve public translation centres as well as public institutions from the Member States
- ▶ Provide flexible means to share data:
 - ▶ At different levels: National, European and as Open Data
 - ▶ Sharing according to the actual data
 - ▶ Support and clearance prior to sharing beyond the Member State level



Issues and answers

- ▶ Not easy to share data:
 - ▶ Can I? *We provide a dedicated team of experts in your Member State to help you decide*
 - ▶ With whom? *National Relay Stations provide flexible means to share at different levels*
 - ▶ Why should I do it? *You gain automatically prepared resources to help optimise your own translation processes*
- ▶ Not easy to prepare and distribute data:
 - ▶ *We automate the creation of language resources*
 - ▶ *We provide an infrastructure accessible to members of public administrations and translation centres*
 - ▶ *Distribution is handled according to your sharing conditions*



Benefits

- ▶ For national translation centres and public institutions, as well as institutional EU translation services
- ▶ Short term:
 - ▶ Gain clean translation memories from in-house document collections or existing translation memories
 - ▶ Gain flexible means to share data: national user groups, European Commission repository or Open Data
 - ▶ Gain access to other resources shared by other institutions in your Member State or beyond
- ▶ Medium and long term:
 - ▶ Benefit from adapted eTranslation services for your language(s)
 - ▶ More efficient services for multilingual Europe



LR collection expertise

- ▶ ELRC National Anchor Points (NAP):
 - ▶ Experts who support the language resource collection effort in the EU
 - ▶ 2 per Member State
 - ▶ Public Services NAP: representative from public services administration
 - ▶ Technology NAP: representative for language technologies
- ▶ 4 National Anchor Points in ELRI:
 - ▶ Paulo Vale, Portugal (Public NAP)
 - ▶ António Branco, Portugal (Technology NAP)
 - ▶ Andy Way, Ireland (Technology NAP)
 - ▶ David Pérez Fernández, Spain (Public NAP)
- ▶ Leading institutions in LR collection and sharing in Europe



Technical expertise

- ▶ Top-performing language technology tools:
 - ▶ Create language resources from raw data such as document collections in multiple languages
 - ▶ Clean up existing resources such as translation memories
 - ▶ Integrated in the ELRI network as processing pipelines
- ▶ Strong experience in dealing with actual user data:
 - ▶ From public administrations
 - ▶ In various languages and domains
- ▶ Developers of language resource sharing infrastructures
- ▶ Legal and administrative issues related to data sharing



The process

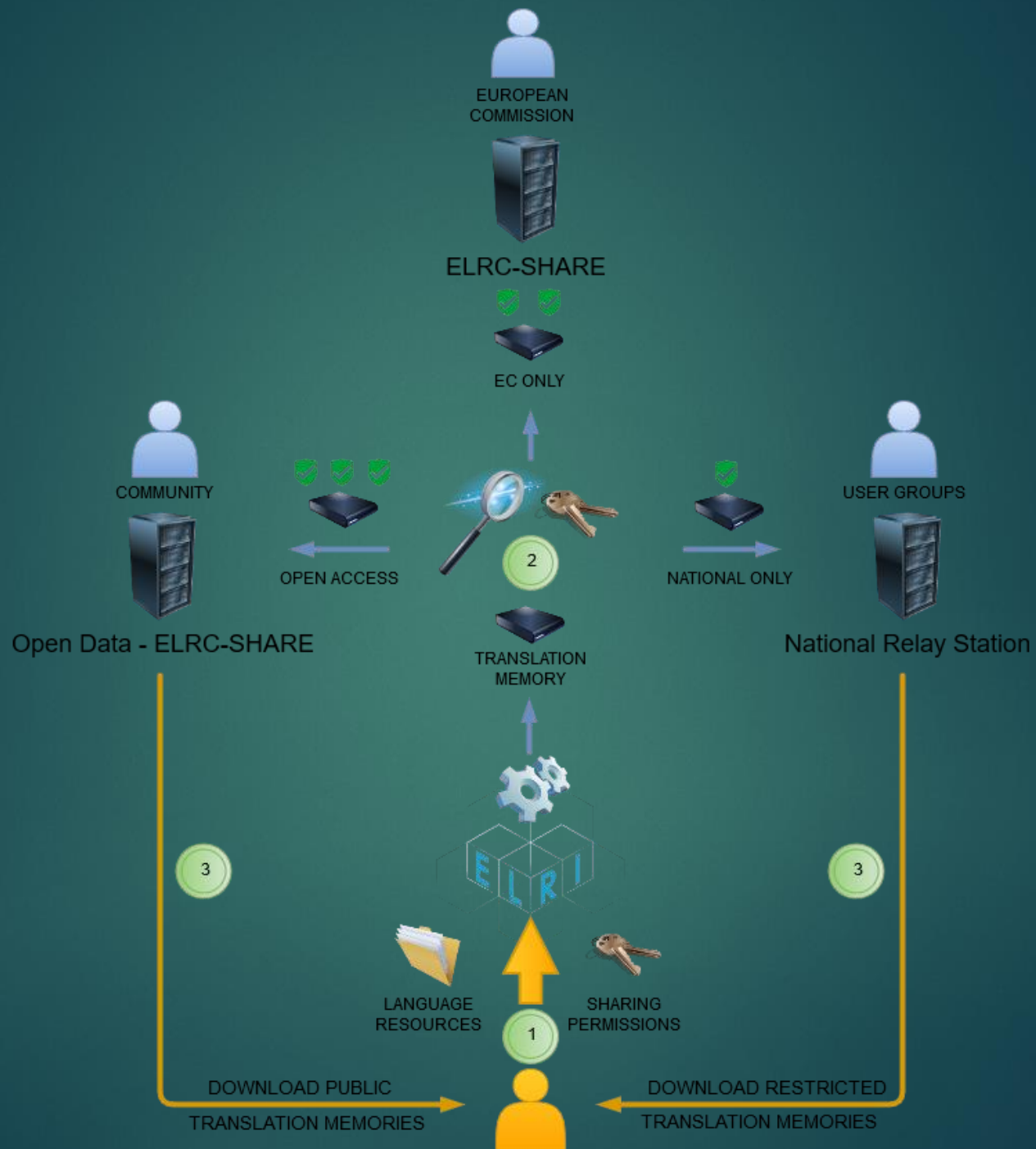
- ▶ Join a National Relay Station in your Member State:
 - ▶ Register on the web site
 - ▶ ELRI personnel will authorise members of public institutions and translation centres
- ▶ Contribute your data:
 - ▶ Multilingual documents, etc.
 - ▶ Specify with whom you would like to share your data
- ▶ Browse and download resources:
 - ▶ Your own automatically prepared resources
 - ▶ Other resources shared by other institutions



Beyond Member States

- ▶ National Relay Stations
- ▶ Data shared at the European level:
 - ▶ Benefits even more stakeholders
 - ▶ Essential to improve European translation services
- ▶ Resources prepared via ELRI can be shared:
 - ▶ At the European level via the ELRC-SHARE repository:
 - ▶ Only with the European Commission to help improve eTranslation services available to all EU public institutions
 - ▶ With the community at large via Open Data repositories:
 - ▶ Some types of data can be shared with everyone
 - ▶ Benefits the language processing research community
 - ▶ ELRI will place in these repositories the data specified as open by the data holders





Timeline

- ▶ 2018:
 - ▶ Design and development of the infrastructure
 - ▶ Integration of early adopters
 - ▶ Dissemination of initiative
- ▶ 2019:
 - ▶ Informational and hands-on workshops:
 - ▶ At least one workshop per Member State in the project
 - ▶ Spring 2019
 - ▶ Extension to more participants
 - ▶ Active LR collection



FAQ

- ▶ Can I share some of my data within a smaller group?
 - ▶ *Yes. Although the goal is to share with as many users as possible, to benefit everyone, we can create restricted groups of users*
- ▶ Are National Relay Stations open?
 - ▶ *Coming soon*
 - ▶ *Early adopters can contact us to go through the resource preparation and sharing process prior to NRS opening.*
- ▶ Do you handle data anonymisation?
 - ▶ *We are redirecting this type of work to ELRC, who already provides on-site assistance*
- ▶ My questions aren't listed here...
 - ▶ *Question time coming soon :)*
 - ▶ *More sessions during the day with more details!*



Follow us

- ▶ Project website:
 - ▶ www.elri-project.eu
 - ▶ Project info and resources
 - ▶ Past events, links and more
- ▶ Twitter:
 - ▶ @ElriProject
 - ▶ Latest news



Contact

- ▶ Project coordination:
 - ▶ elri-coordination@vicomtech.org
 - ▶ Contact list @ <http://www.elri-project.eu/contact-us.html>
- ▶ Spain:
 - ▶ Vicomtech: Thierry Etchegoyhen
 - ▶ SEAD: David Pérez Fernández, Maite Melero
- ▶ France:
 - ▶ ELDA: Victoria Arranz, Khalid Choukri
- ▶ Ireland:
 - ▶ DCU: Jane Dunne, Andy Way
- ▶ Portugal:
 - ▶ AMA: Paulo Vale
 - ▶ FCUL: António Branco
 - ▶ Linkare: Pedro Neiva, Rui Neto



THANK YOU!

QUESTIONS or COFFEE?

