ELRC Workshop in Ireland

New services provided by the ELRC consortium

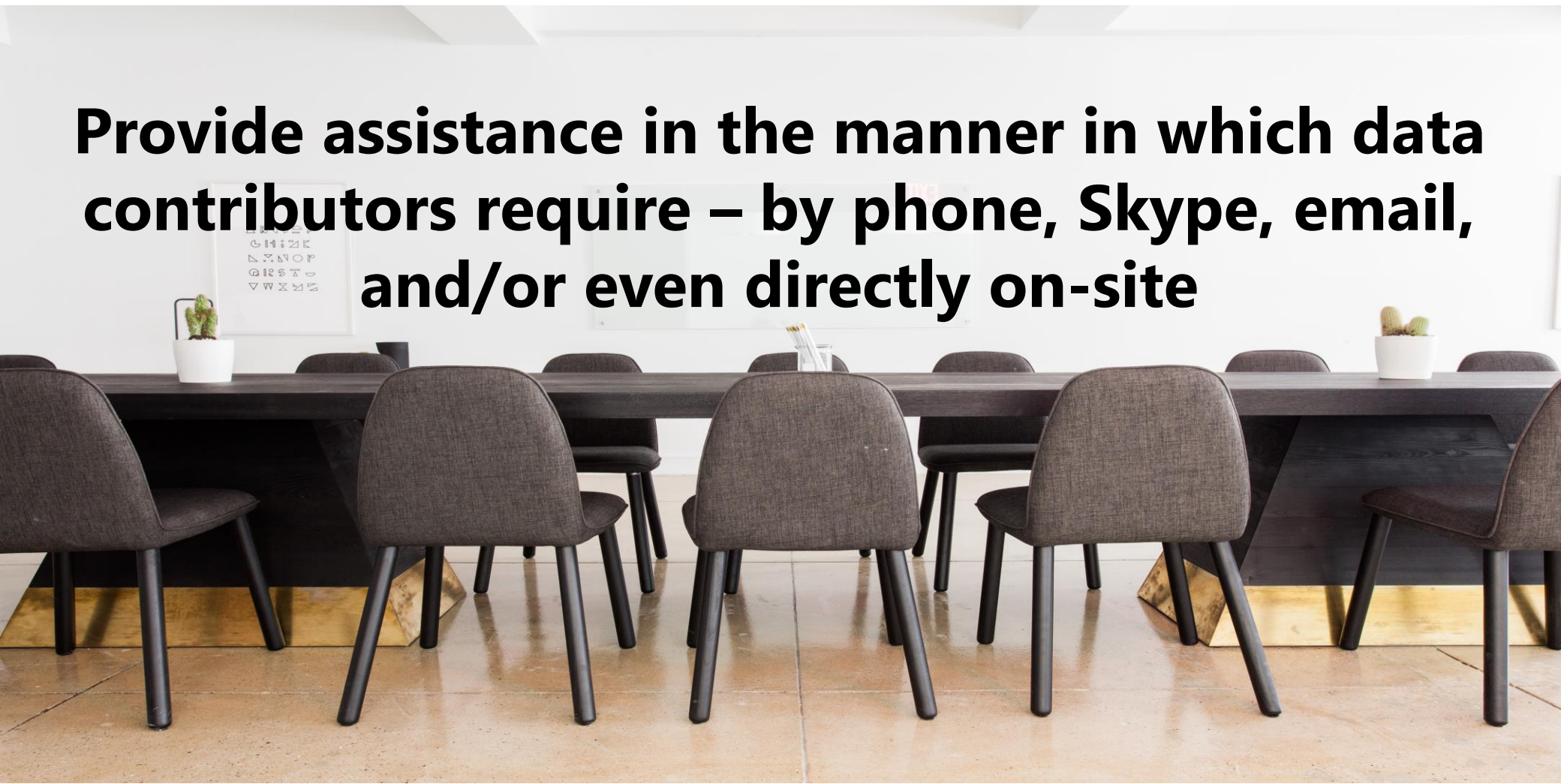**Khalid Choukri, Hélène Mazo**
ELRA/ELDA

Help data contributors to provide language resources for building better MT systems (CEF eTranslation), benefiting all of Europe

Remove <u>all obstacles</u> to contributing language resources – both legal and technical (i.e., data processing issues)

Provide assistance in the manner in which data contributors require – by phone, Skype, email, and/or even directly on-site

New ELRC services (2017 and beyond)

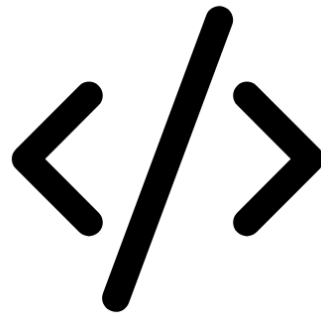**Available to all data contributors in CEF countries**

If you encounter any legal and/or technical issues with your language resources, ELRC and our broad network of experts can help!
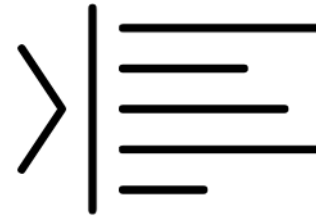
Language processing services

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!

Re-formatting

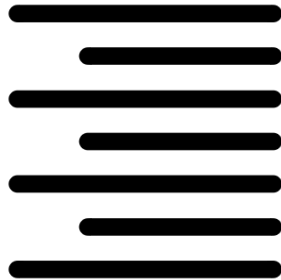Need to re-format DOCX to XML, or PDF to WORD? Let us do it for you!

Data extraction

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

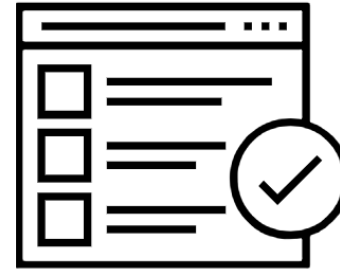# Language processing services

Cleaning

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

Alignment

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

Metadata validation

Metadata is crucial! We can validate and organize metadata for your team

Anonymisation

Does your data contain private info? We can help "**identify**" private elements and fully anonymise

Language processing services

Phone

Pick up the phone and call the ELRC helpdesk
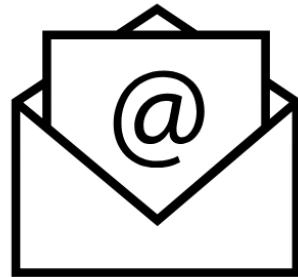
Email

Send us off an email briefly explaining your technical/legal issue
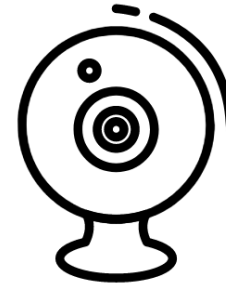
Skype

Skype us! We are online and ready to help with any questions

On-site

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

ELRC consortium provides language processing services and assistance <u>directly on-site</u> to data contributors

In providing services, ELRC utilizes its wide range of skills and expertise in language processing

Assistance will be provided in close cooperation with a broad network of language experts

Data requires
processing on-site

**Request on-site
assistance**

ELRC processes
the data at
your premises

Processed data

Data contributor

No issues with data

**Upload to
ELRC-SHARE**

CEF Digital
Connecting Europe
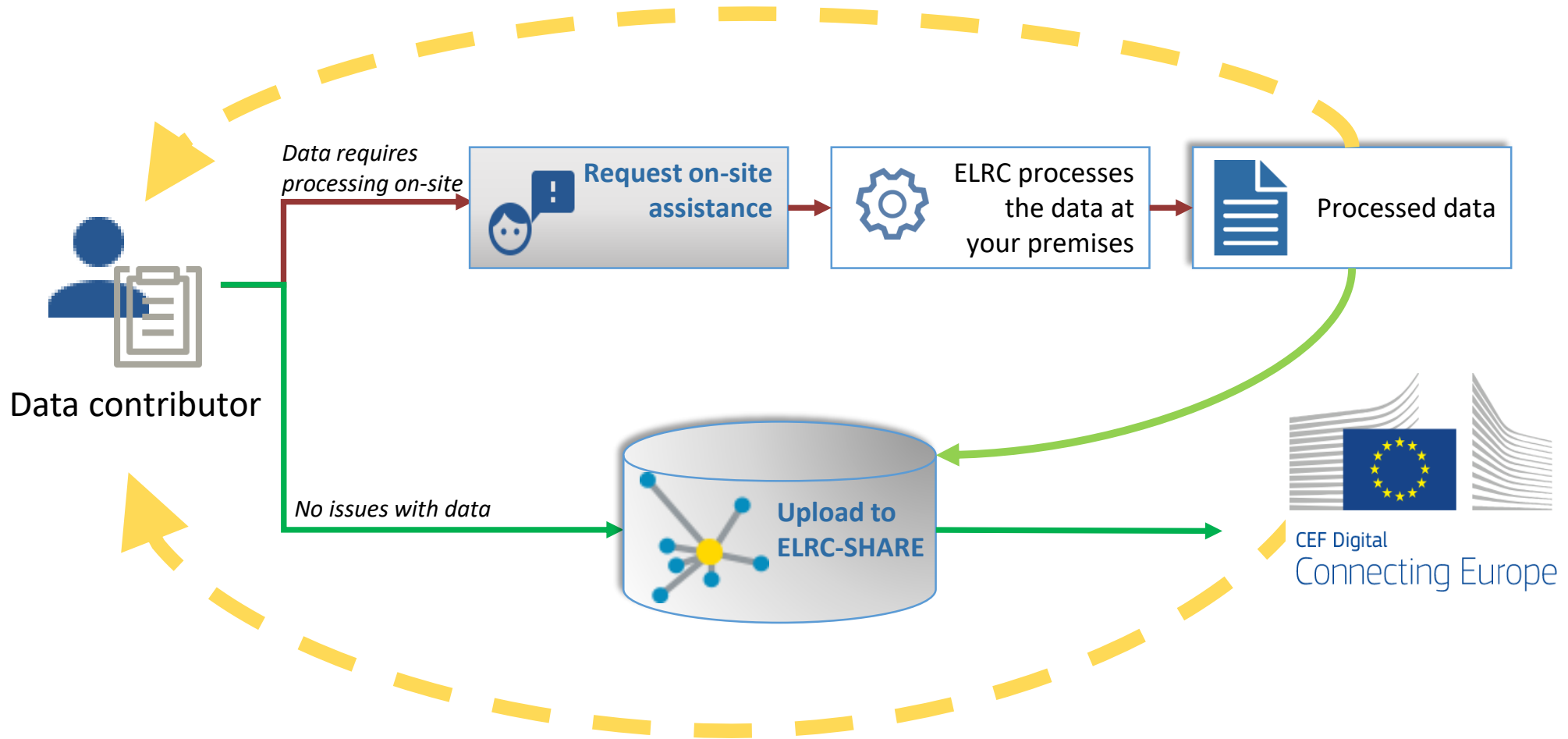
> **We will fix your data issues and return the processed data directly to you.**

> **We can also help improve your data management processes. Just ask!**

How to request services

ELRC website

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

lr-coordination.eu/request-onsite-assistance

ELRC Helpdesk

Helpdesk for Language Resources

We are happy to answer any questions on the technical or legal aspects related to the use, production, collection, processing, and sharing of 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-cooridantion.eu

lr-coordination.eu/helpdesk

Case studies (2015-2016)

Spain

Problem: Data provider didn't store translations as <u>related documents</u>, therefore source/target translation weren't paired

Solution: ELRC helped crawl a local system to find, relate documents, and pair source/target translations

Spain

Problem: In some Spanish governmental departments, archives were only available in PDF

Solution: ELRC helped provide good converters to get usable documents

Germany

Problem: Data owner needed help with <u>anonymization</u>, as databases contained personal info. Another need: cleaning up 'junk' data (URLs, numbers, fragments)

Solution: ELRC helped provide anonymization services and data cleaning

Estonia

Problem: Data donor found that legal acts in EN, ET, RU couldn't be <u>aligned on a document level</u> (no common machine-readable cross-language ID)

Solution: ELRC helped provide alignment services for documents

Questions?

www.lr-coordination.eu