

ELRC Update

Andrea Lösch, ELRC Project Manager (DFKI)

6th ELRC Conference (31 March 2022)

ELRC development at a Glance

PHASE 1 (2015-2016):

- Initial Roadshow (Country Workshops)
- Set-up of ELRC Network (T-NAP plus P-NAPs)
- Set-up of data collection & validation workflow
- Technical and Legal Helpdesk

PHASE 2 (2016-2019):

- Sustainability, local ownership
- Extension of ELRC Network on national level (600 data donors)
 - Extended data collection (crawling) & validation workflow (Generic Services Projects)
- Addition: On-site assistance

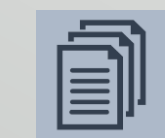
PHASE 3 (2019-2022):

- Sustainability, re-usability, opening (SME)
- Extension of ELRC Network (> 6.000 data donors and users)
- Extended data collection (speech, tools and services, ...)
- New languages (RU, CN, ...)

Language Data Collection



Data currently collected within ELRC



Corpora

**Set of documents or text
in one or more languages**

- official documents / translations in the official administration (decisions, legal acts etc.);
- articles, reports, magazines, newsletters, etc.;
- Translation Memories (i.e. aligned text segments in the source and target language)



Models

**Language
or
translation
models**

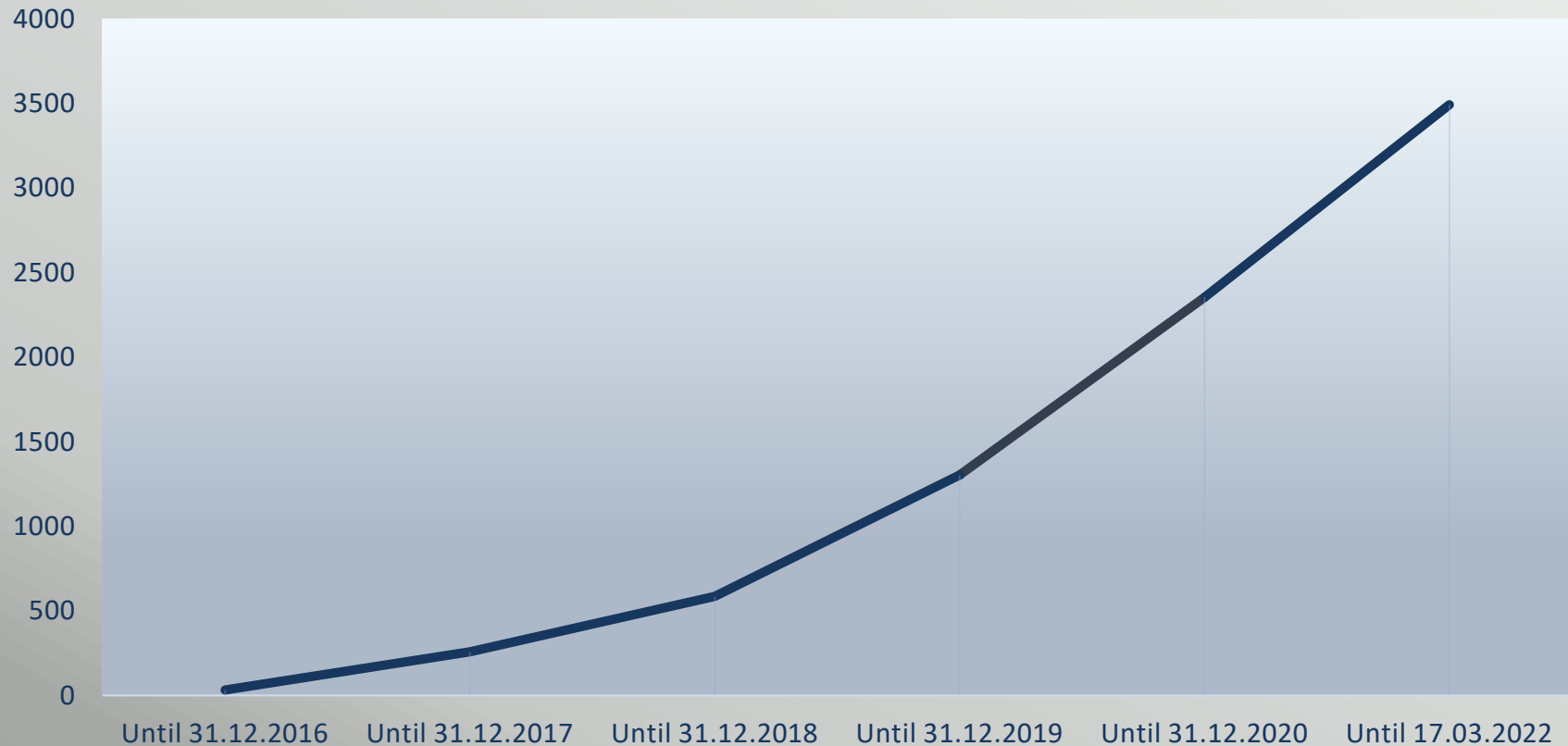


Resources

**Lexical &
Conceptual
Resources**

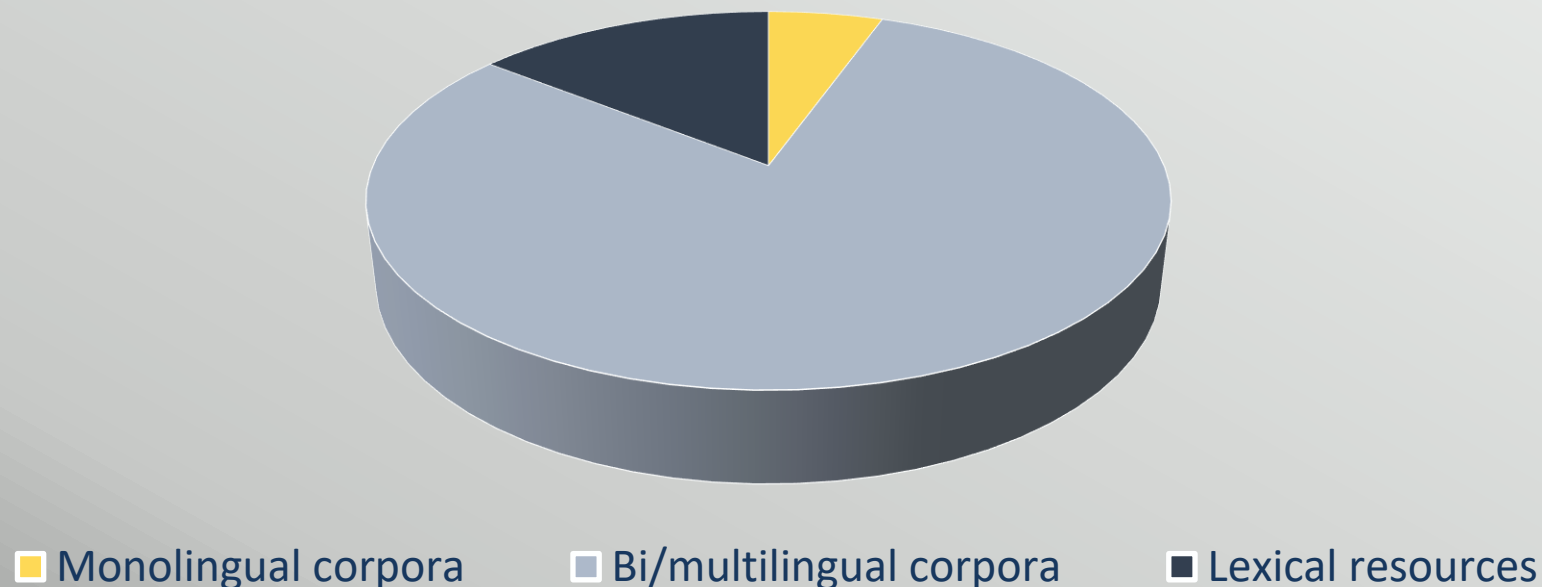
- Terminologies
- Glossaries
- Thesauri
- Wordnets
- Ontologies

Number of LR in the ELRC-SHARE Repository

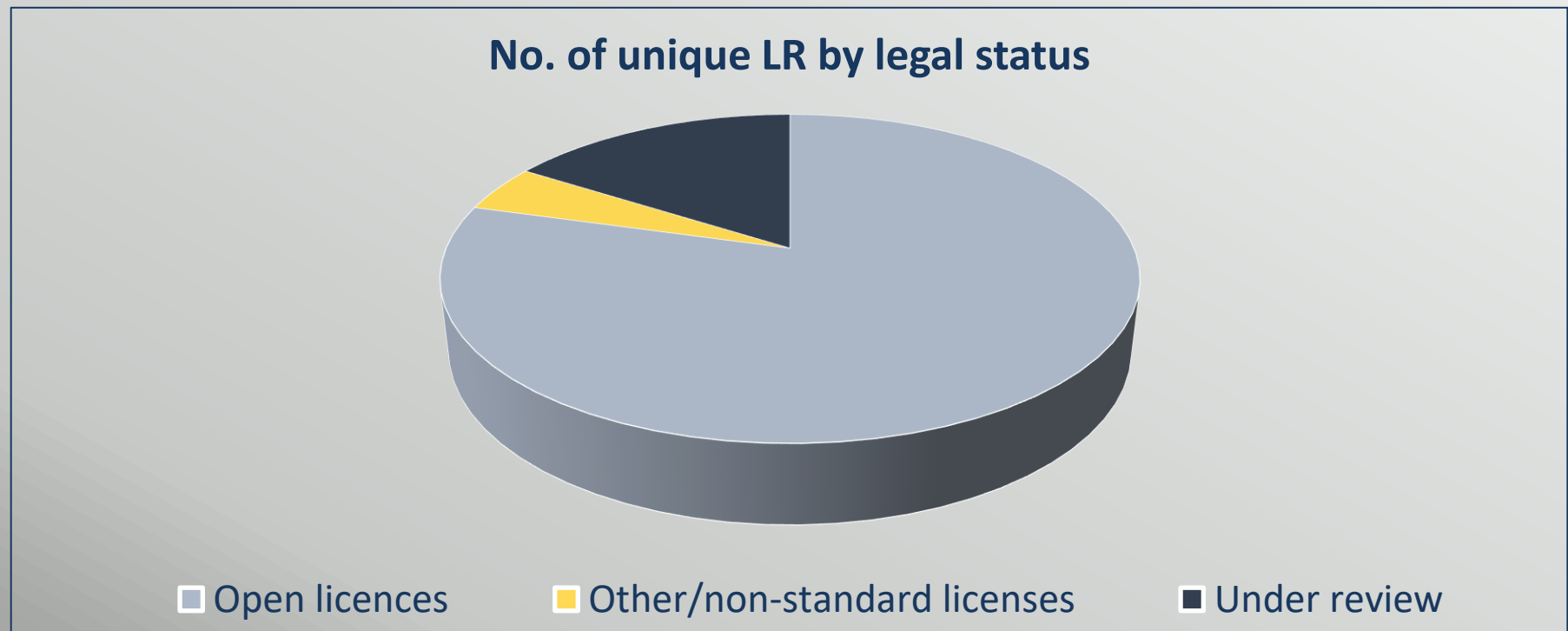


Types of LR in the ELRC-SHARE Repository

No. of unique Language Resources (LR) by type



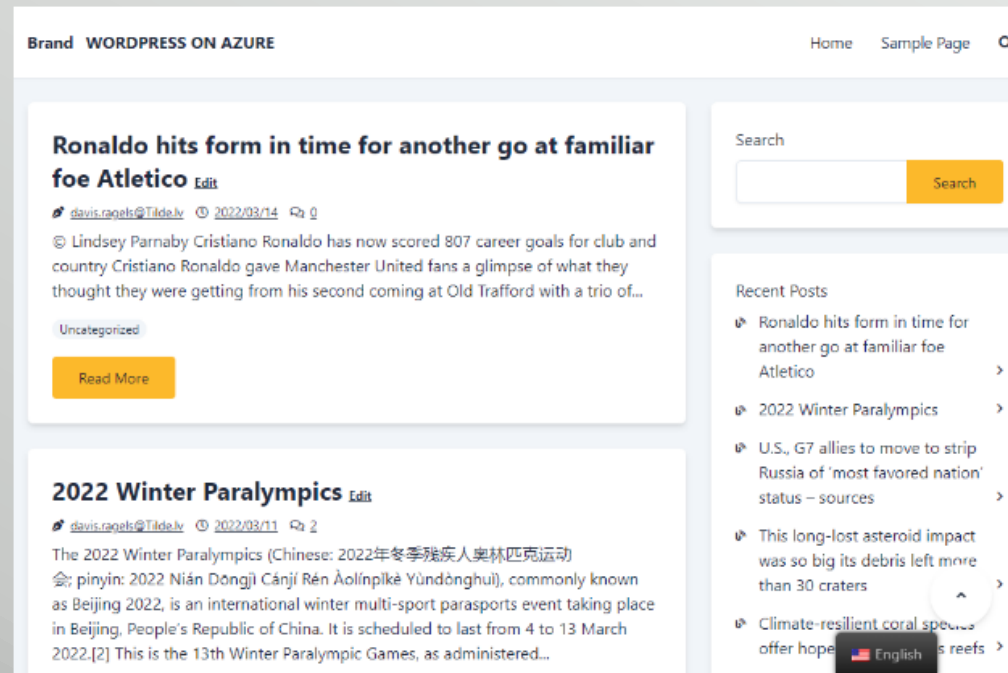
Re-usability of LR in the ELRC-SHARE Repository



LT Specifications



1) Website Translation Tool



Brand WORDPRESS ON AZURE Home Sample Page

Ronaldo hits form in time for another go at familiar foe Atletico [Edit](#)

[davis.ragels@Tilde.lv](#) 2022/03/14

© Lindsey Parnaby Cristiano Ronaldo has now scored 807 career goals for club and country Cristiano Ronaldo gave Manchester United fans a glimpse of what they thought they were getting from his second coming at Old Trafford with a trio of...

Uncategorized

[Read More](#)

2022 Winter Paralympics [Edit](#)

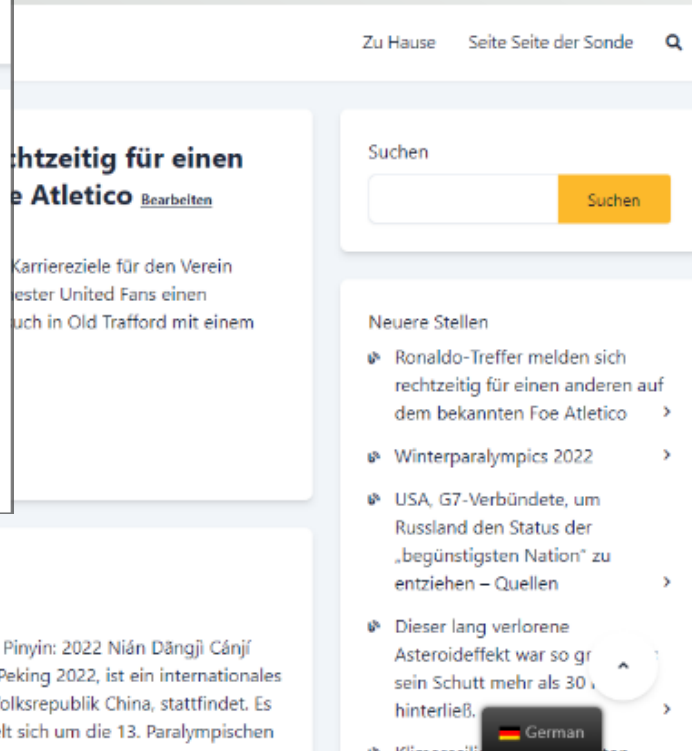
[davis.ragels@Tilde.lv](#) 2022/03/11

The 2022 Winter Paralympics (Chinese: 2022年冬季残疾人奥林匹克运动会; pinyin: 2022 Nián Dōngjì Cánjī Rén Àolínpīkè Yùndònghuì), commonly known as Beijing 2022, is an international winter multi-sport parasports event taking place in Beijing, People's Republic of China. It is scheduled to last from 4 to 13 March 2022.[2] This is the 13th Winter Paralympic Games, as administered...

Recent Posts

- Ronaldo hits form in time for another go at familiar foe Atletico
- 2022 Winter Paralympics
- U.S., G7 allies to move to strip Russia of 'most favored nation' status – sources
- This long-lost asteroid impact was so big its debris left more than 30 craters
- Climate-resilient coral species offer hope

English



Zu Hause Seite Seite der Sonde

Winterparalympics 2022 [Bearbeiten](#)

[davis.ragels@Tilde.lv](#) 2022/03/11

Die Winterparalympiken 2022 (Chinesisch: 2022¹; Pinyin: 2022 Nián Dōngjì Cánjī Rén Àolínpīkè Yùndònghuì, gemeinhin bekannt als Peking 2022, ist ein internationales Winter-Multisport-Parament, das in Peking, der Volksrepublik China, stattfindet. Es soll vom 4. bis 13. März 2022 dauern.[2] Es handelt sich um die 13. Paralympischen

Suchen

Neuere Stellen

- Ronaldo-Treffer melden sich rechtzeitig für einen anderen auf dem bekannten Foe Atletico
- Winterparalympics 2022
- USA, G7-Verbündete, um Russland den Status der „begünstigsten Nation“ zu entziehen – Quellen
- Dieser lang verlorene Asteroideffekt war so groß sein Schutt mehr als 30 hinterließ.

German

2) Anonymisation



- Masking/replacing personal data is needed to make such data sets sharable (see e.g. GDPR limitations)
 - Problems:
 - Correctly identifying personal information in unstructured data
 - Ensuring re-use of “anonymised” data sets for LT training
 - Dockers for four major tools (TM-Anonymiser, Text Transformer, Biroamer, MAPA)
- <https://github.com/CrossLangNV/elrc3-anonymisation-specs>

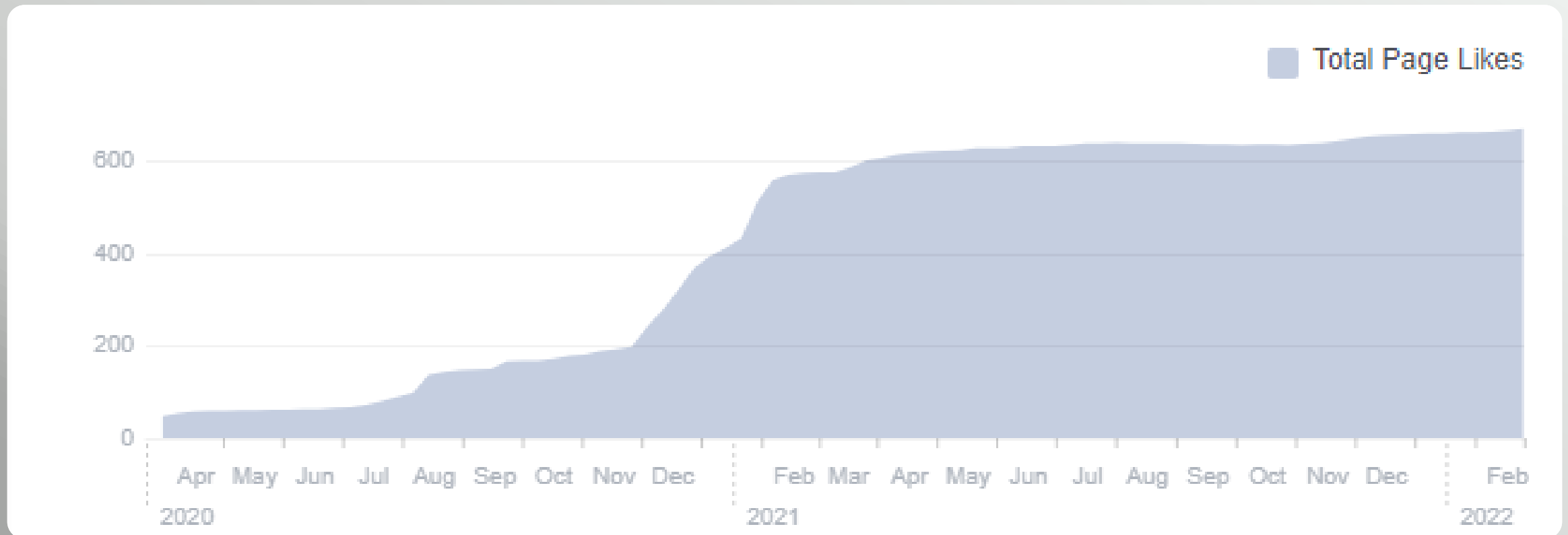
3) Fake News Detection



- Investigation of best practices for supporting fact-checkers / detecting potential fake news
- Two types of experiments (training and testing) conducted:
 - Supervised fake news detection: deep learning systems trained from true and false articles
 - Unsupervised fake news detection: systems trained from true articles only
- Key questions:
 - What kind of fact-checking activities can / should be supported through automation ?
 - What type of fake news detection modules would be useful ?
 - What is the best strategy to evaluate the output of a fake news detection tool ?
 - What languages should be focused on in the area of fact-checking / fake news detection ?

ELRC Initiatives and Campaigns

Facebook: page “likes” development



Facebook: User activities

More than
2.980.000



Full video views

More than
8.400.000



User reach

More than
216.000



Link clicks generated

Language data for Ukrainian

- In March, the regular Social Media campaign was suspended
- Promotion of the EC initiative on data donation for Ukrainian & the eTranslation Ukrainian engine



	reach	impressions	link clicks	link clicks (paid)	video plays (min 3 sec.)	user actions
Facebook						
Call for data	121.967	208.990	70.732	2.095	128.181	20
eTranslation	22.514	46.384	1.880	623	3.752	82
Twitter						
Call for data		5.496	23		1.095	198
eTranslation		6.844	42		1.261	204
LinkedIn						
Call for data					5.206	209
eTranslation					2.693	110

Stay tuned! Follow us on:



<https://www.lr-coordination.eu>



@LR_Coordination



EuropeanLanguageResourceCoordination



Ircoordination



ELRC

THANK YOU!

Email: info@lr-coordination.eu