

Η δράση Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων: γιατί και πώς να συνδράμετε και να επωφεληθείτε

Στέλιος Πιπερίδης
Ινστιτούτο Επεξεργασίας του Λόγου / Ε.Κ. «Αθηνά»
ELRC

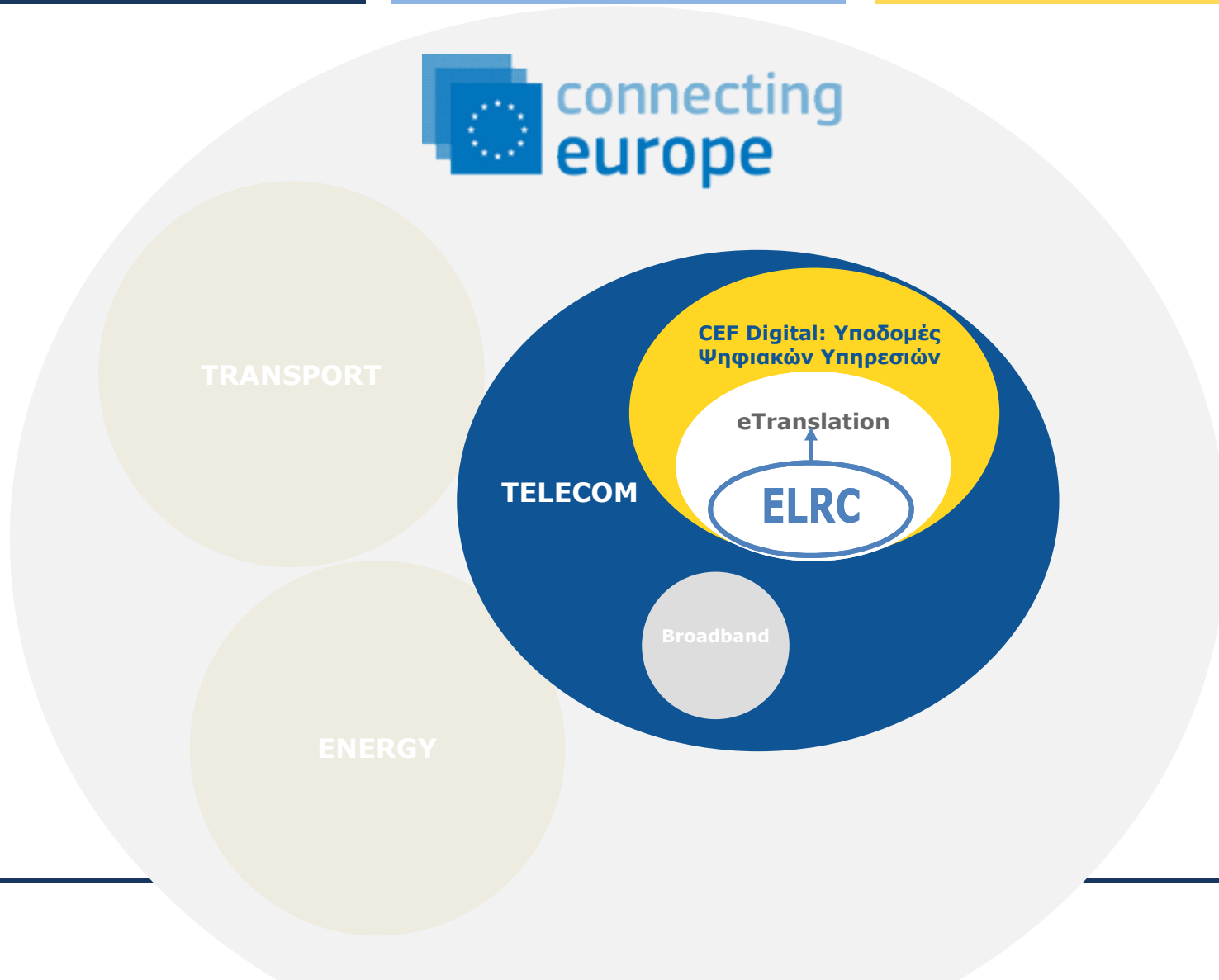




TRANSPORT

TELECOM

ENERGY





Η κοινοπραξία:



German
Research Center
for Artificial
Intelligence





60 Εθνικά σημεία επαφής:

- Τεχνικοί Εμπειρογνώμονες:
(Τεχνολογικοί Εθνικοί Εκπρόσωποι)
- Εμπειρογνώμονες του Δημοσίου (Εθνικοί Εκπρόσωποι Δημοσίων Υπηρεσιών)



Φρύνη Κακογιάννη-Ντοά
Πανεπιστήμιο Κύπρου



Νατάσα Χαρατσή-Αβρααμίδα
Γραφείο Τύπου και
Πληροφοριών

Τι κάνει ο Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων



Συλλέγει

- Γλωσσικούς πόρους

Εντοπίζει

- Τις ανάγκες των δημόσιων υπηρεσιών

Συνεργάζεται

- Με τον δημόσιο τομέα για τον εντοπισμό γλωσσικών πόρων

Βοηθά

- Σε τεχνικά ή νομικά θέματα

Υλοποιεί

- Ευρωπαϊκό παρατηρητήριο γλωσσικών πόρων

Τι κάνει το ELRC;



- Διευκολύνει την πολυγλωσσική επικοινωνία και την ανταλλαγή πληροφορίας στο πλαίσιο των ψηφιακών υπηρεσιών
- Πώς; Με τη βοήθεια κατάλληλων γλωσσικών πόρων!
- Παραδείγματα ψηφιακών υπηρεσιών στα οποία μπορεί να βοηθήσει η Αυτόματη Μετάφραση:
 - Δικαιώματα Καταναλωτή (ODR),
 - Κοινωνική Ασφάλιση (EESSI),
 - Δικαιοσύνη (eJustice),
 - Εμπόριο (BRIS),
 - Ανοιχτά Δεδομένα (ODP),
 - ...

Slide 8

LS2

This slide contains the same message as the video and should be deleted if the video is shown

Lilli Smal; 31/07/2017

Γλωσσική ισότητα στην ψηφιακή εποχή



11 Σεπτεμβρίου 2018:
Πρόταση ψηφίσματος
του Ευρωπαϊκού Κοινοβουλίου
**σχετικά με τη γλωσσική ισότητα στην
ψηφιακή εποχή**

Γλωσσική ισότητα στην ψηφιακή εποχή



«18. καλεί την Επιτροπή και τα κράτη μέλη να αναπτύξουν στρατηγικές και δράσεις πολιτικής οι οποίες θα διευκολύνουν την **πολυγλωσσία στην ψηφιακή αγορά**. καλεί την Επιτροπή και τα κράτη μέλη να ορίσουν **τους ελάχιστους γλωσσικούς πόρους, τους οποίους θα πρέπει να διαθέτουν όλες οι ευρωπαϊκές γλώσσες**, όπως σύνολα δεδομένων, λεξικά, αρχεία προφορικού λόγου, μεταφραστικές μνήμες, σχολιασμένα σώματα κειμένων και εγκυκλοπαιδικό περιεχόμενο, **προκειμένου να αποφύγουν την ψηφιακή εξαφάνιση**.»

Γλωσσική ισότητα στην ψηφιακή εποχή



«21. καλεί την Επιτροπή να αναδείξει σε προτεραιότητα της γλωσσικής τεχνολογίας **τα κράτη μέλη που έχουν μικρό μέγεθος και έχουν δική τους γλώσσα**, ώστε να δοθεί σημασία στις γλωσσικές προκλήσεις που αυτά αντιμετωπίζουν.»

Γλωσσική ισότητα στην ψηφιακή εποχή



Γνωμοδότηση της Επιτροπής Βιομηχανίας, Έρευνας και Ενέργειας: «υπογραμμίζει τον ρόλο της ΕΕ, των κρατών μελών, των πανεπιστημίων και άλλων δημόσιων οργανισμών στη συνεισφορά για τη **διατήρηση των γλωσσών τους στον ψηφιακό κόσμο** και στην **ανάπτυξη βάσεων δεδομένων και μεταφραστικών τεχνολογιών για όλες τις γλώσσες της ΕΕ**, συμπεριλαμβανομένων των γλωσσών που δεν ομιλούνται ευρέως· ζητεί τον συντονισμό μεταξύ έρευνας και βιομηχανίας με τον κοινό στόχο της ενίσχυσης των ψηφιακών δυνατοτήτων για τη γλωσσική μετάφραση, και με ελεύθερη πρόσβαση σε δεδομένα η οποία απαιτείται για την τεχνολογική εξέλιξη.»

Cracking the Language Barrier Federation (2016):

«Συστήνουμε η Ευρώπη να κάνει μια ενεργή προσπάθεια να γίνει ανταγωνιστική στη διεθνή σκηνή στον τομέα της E&A στη Γλωσσική Τεχνολογία, διότι **δεν μπορούμε να περιμένουμε από τρίτους, εκτός Ευρώπης, να λύσουν τα προβλήματά μας που σχετίζονται με τη μετάφραση και τη διαχείριση γνώσης** με τρόπο που να ανταποκρίνεται στις ιδιαίτερες επικοινωνιακές, κοινωνικές και πολιτισμικές μας ανάγκες. **Το κλειδί είναι Γλωσσική Τεχνολογία που παράγεται στην Ευρώπη για την Ευρώπη.** Αυτό θα διευκολύνει στο μέλλον τη διασυνοριακή και διαγλωσσική επικοινωνία στην Ευρώπη, την οικονομική ανάπτυξη και την κοινωνική σταθερότητα.»

Τι κάνει ο Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων



Συλλέγει

- Γλωσσικούς πόρους

Εντοπίζει

- Τις ανάγκες των δημόσιων υπηρεσιών

Συνεργάζεται

- Με τον δημόσιο τομέα για τον εντοπισμό γλωσσικών πόρων

Βοηθά

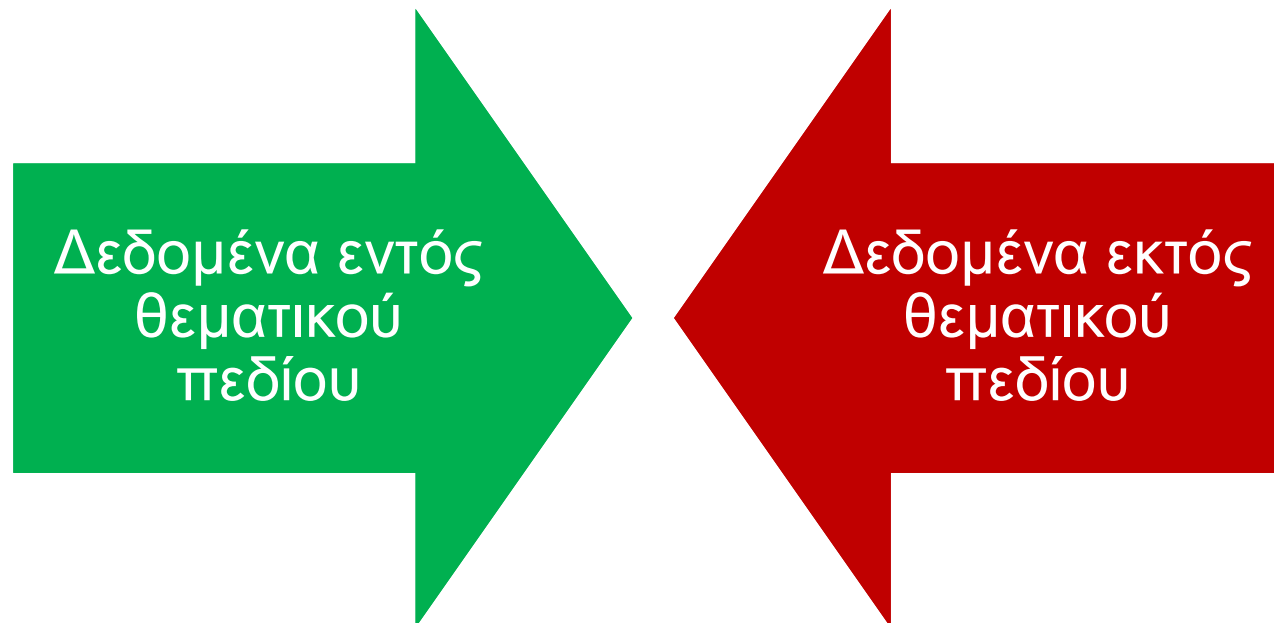
- Σε τεχνικά ή νομικά θέματα

Υλοποιεί

- Ευρωπαϊκό παρατηρητήριο γλωσσικών πόρων

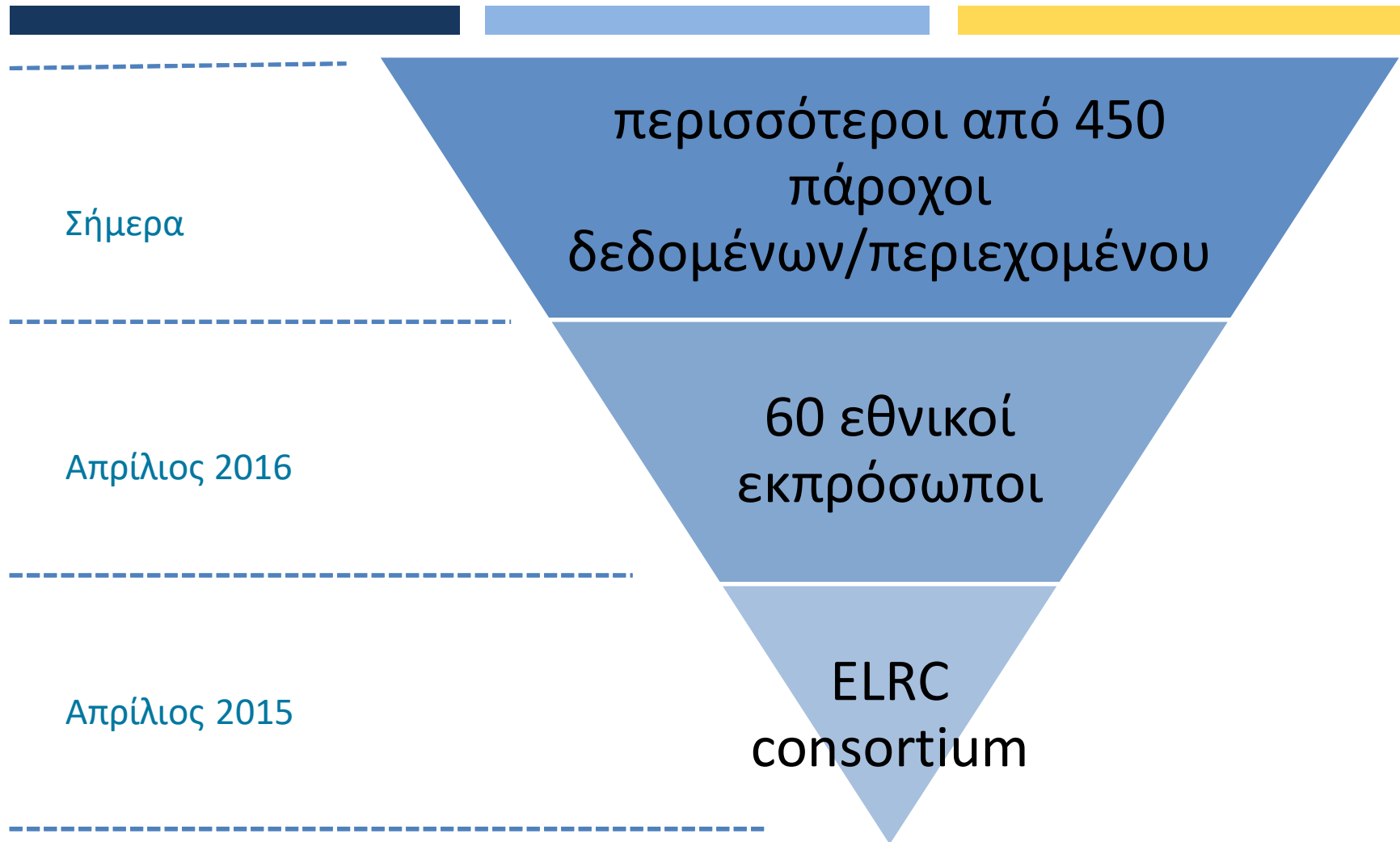
Πώς διευκολύνεται η Αυτόματη Μετάφραση;





➔ Τα δικά σας γλωσσικά δεδομένα κάνουν τη διαφορά!

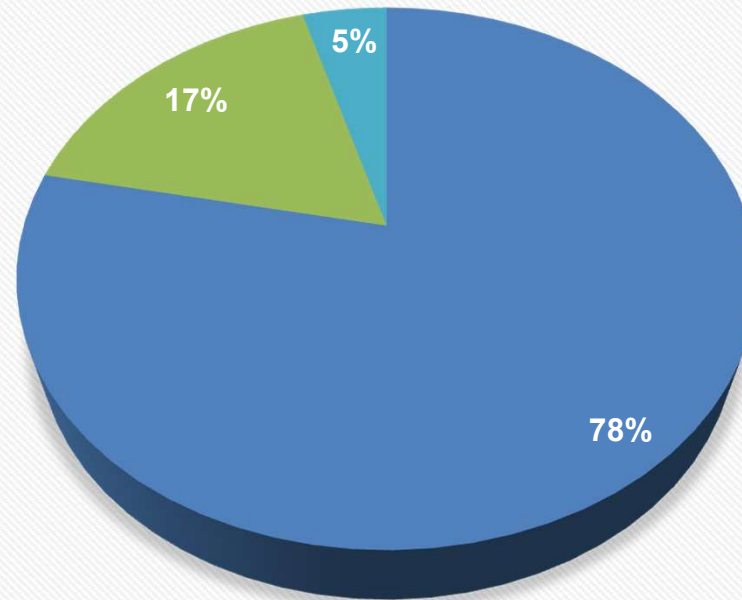
Τι έχουμε πετύχει μέχρι στιγμής: Το δίκτυο του ELRC



Τι έχουμε πετύχει μέχρι στιγμής: Γλωσσικοί πόροι



225 Γλωσσικοί πόροι τον Απρίλιο 2017

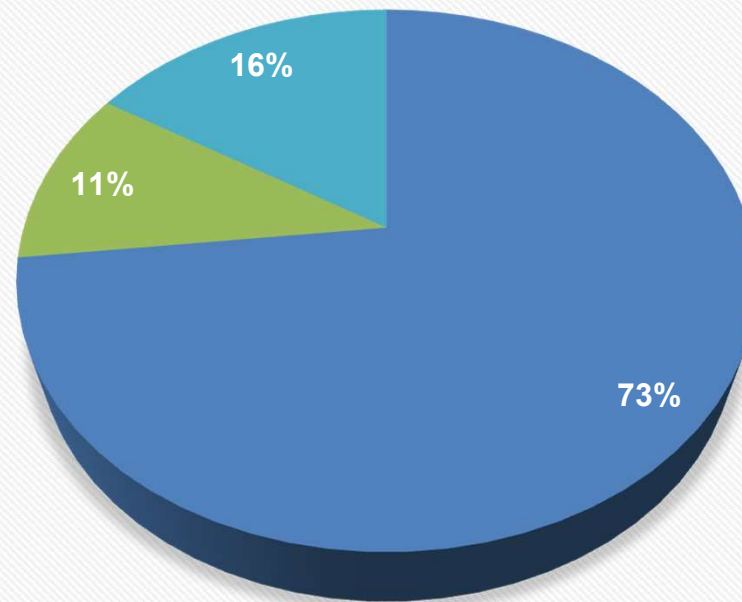


■ Ανοιχτά δεδομένα / CC άδειες ■ Μη τυποποιημένες / άλλες άδειες ■ Υπό εξέταση

Τι έχουμε πετύχει μέχρι στιγμής: γλωσσικοί πόροι



>570 γλωσσικοί πόροι τον Δεκέμβριο 2018

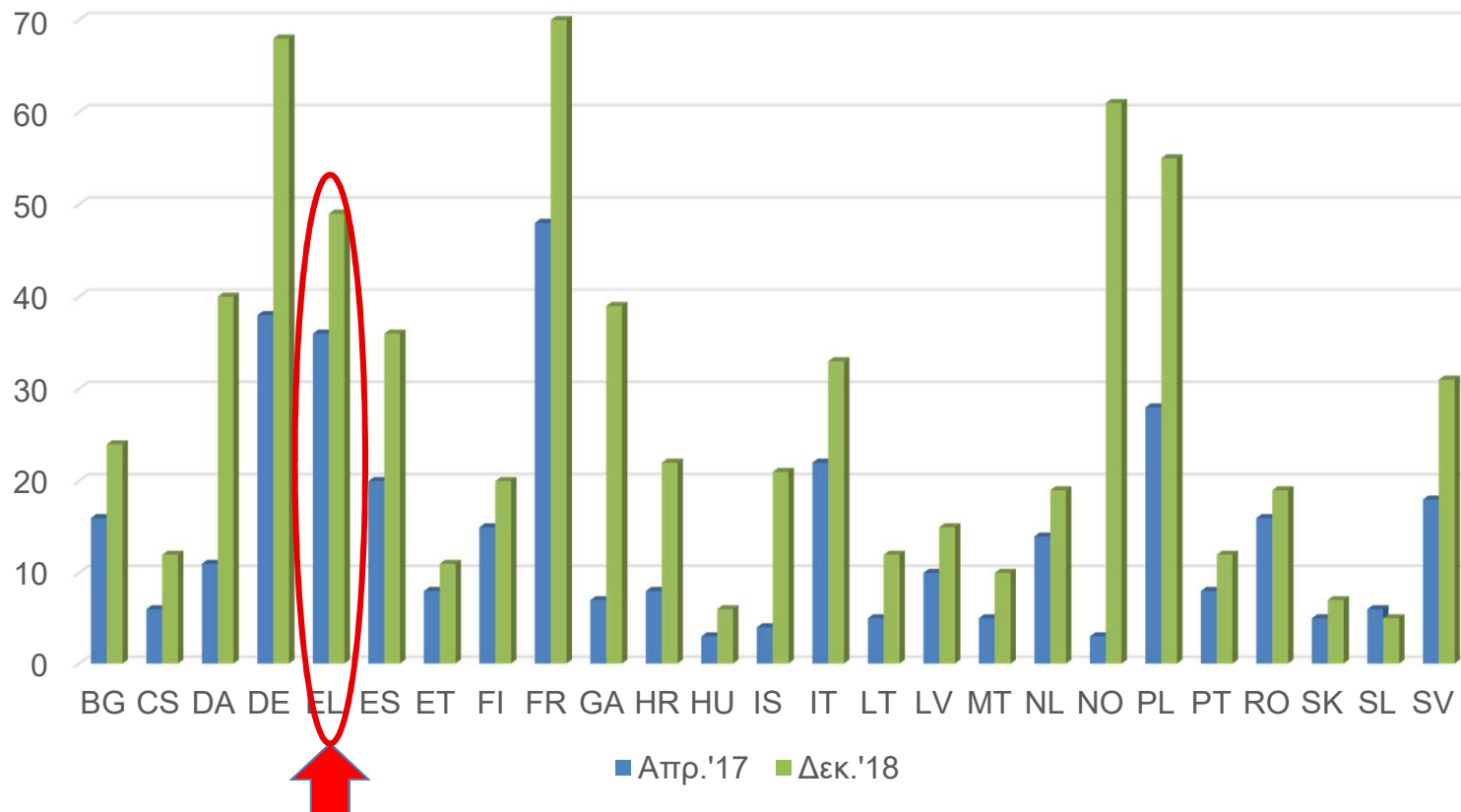


■ Ανοιχτά δεδομένα / CC άδειες ■ Μη τυποποιημένες / άλλες άδειες ■ Υπό εξέταση

Τι έχουμε πετύχει μέχρι στιγμής: Γλωσσικοί πόροι



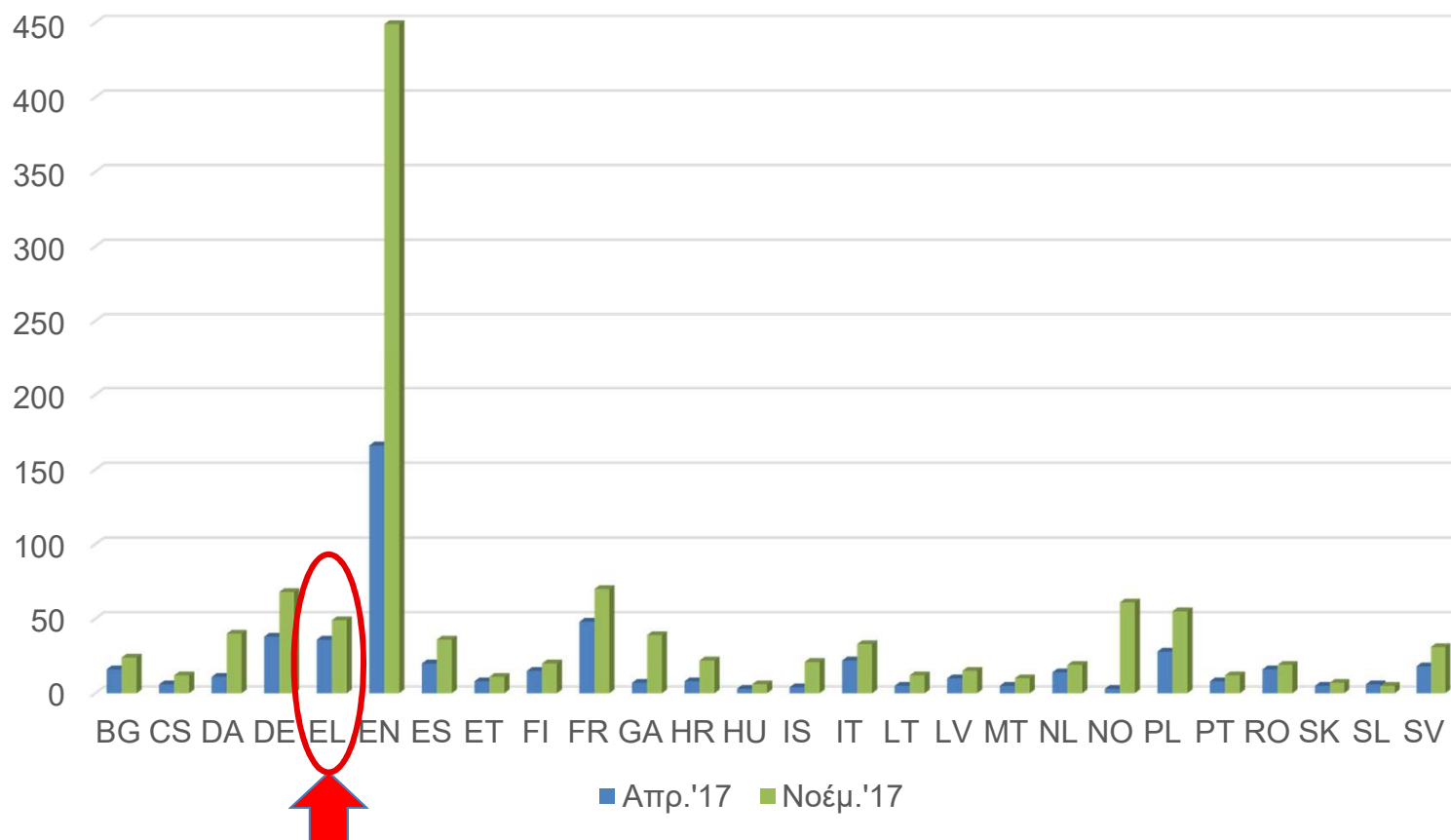
Εξέλιξη της συλλογής γλωσσικών πόρων ανά γλώσσα



Τι έχουμε πετύχει μέχρι στιγμής: Γλωσσικοί πόροι



Εξέλιξη της συλλογής γλωσσικών πόρων ανά γλώσσα



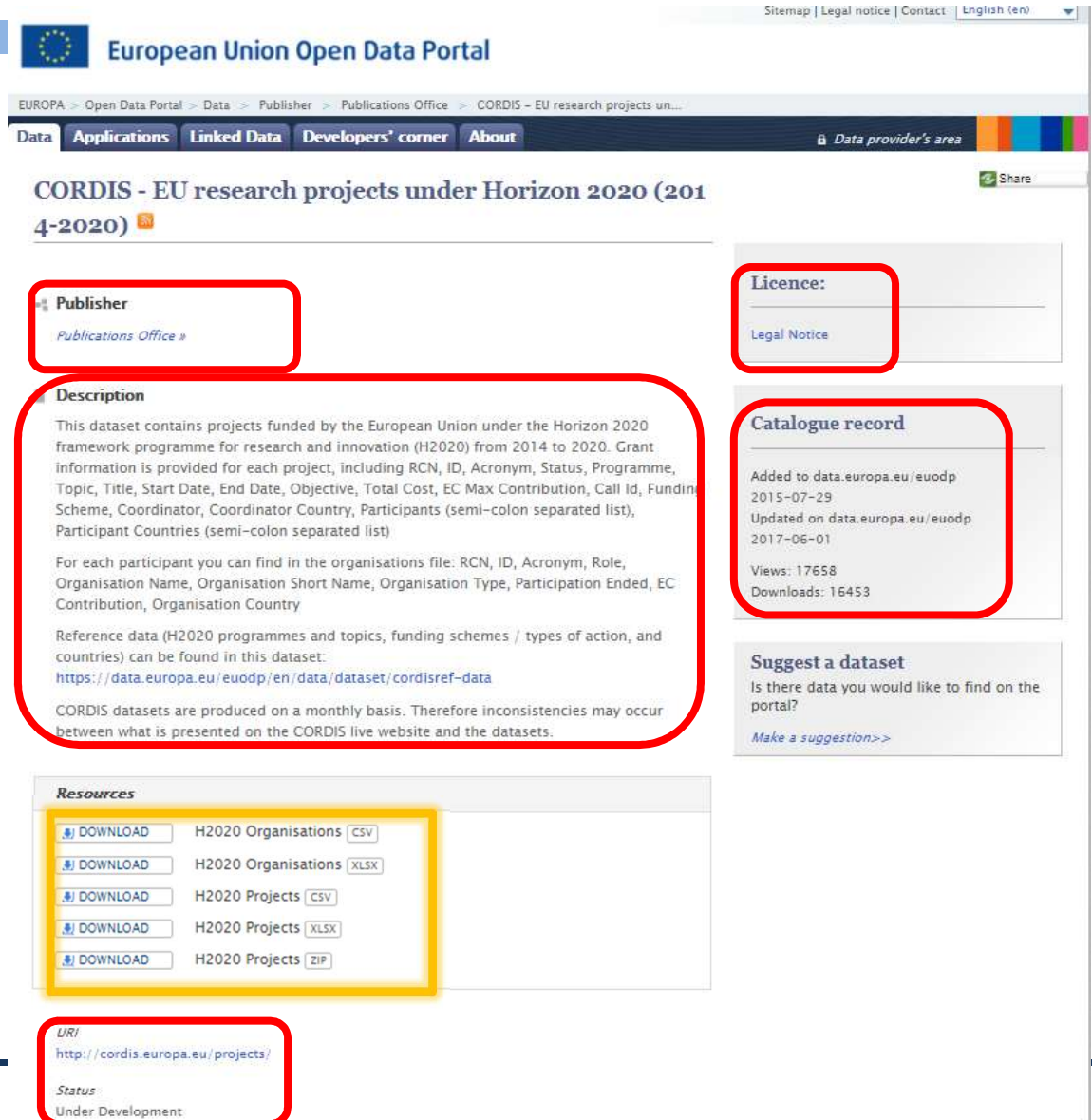


Το πετρέλαιο του 21^{ου} αιώνα



Βασικές έννοιες:

- **Δεδομένα:** οποιοδήποτε ηλεκτρονικά αποθηκευμένο περιεχόμενο
- **Σύνολο δεδομένων (ή πόρος):** συλλογή αρχείων δεδομένων ομαδοποιημένων σύμφωνα με συγκεκριμένα κριτήρια
- **Μεταδεδομένα:** δεδομένα σχετικά με τα δεδομένα, δηλ. τα χαρακτηριστικά που περιγράφουν ένα σύνολο δεδομένων (π.χ. τίτλος, εκδότης, περιεχόμενο και URL)



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

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/>

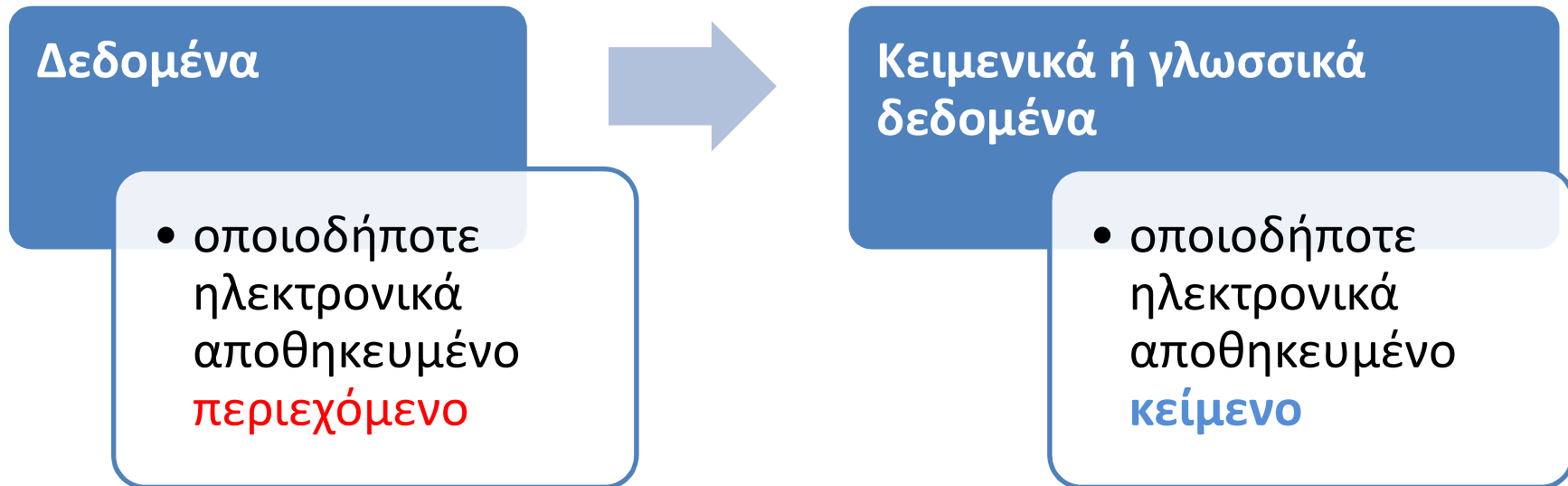
Status
Under Development

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>>](#)

Γλωσσικά Δεδομένα για το eTranslation μέσω του ELRC



EL

Η ΕΕ αποτελεί ένα σημαντικό εγχείρημα που αφορά περισσότερους από 500 εκατομμύρια πολίτες, οι οποίοι μοιράζονται περίπου 80 διαφορετικές γλώσσες. Ενώ η πολυγλωσσία αποτελεί ένα από τα σημαντικότερα πλεονεκτήματα της Ευρώπης, συνιστά σημαντικότερη πρόκληση για τη δημιουργία μιας πολιτιστικά και κοινωνικά ολοκληρωμένης ΕΕ, αποτελεί δε εμπόδιο για την εκπλήρωση των στόχων της ψηφιακής ενιαίας αγοράς.

EN

The EU is a unique endeavour involving more than 500 million citizens sharing about 80 different languages. While multilingualism is one of the biggest assets of Europe, it also poses one of the most substantial challenges for the creation of a culturally and socially integrated EU as well as an obstacle for the fulfilment of the goals of the Digital Single Market.

Γλωσσικά δεδομένα που χρησιμοποιεί ήδη το eTranslation



αλλά...
δεν είναι αρκετά

- Γλωσσικά δεδομένα δημόσιων φορέων που είτε παράγονται από τους ίδιους τους φορείς ή από εξωτερικούς συνεργαζόμενους φορείς, π.χ.
 - Αναφορές
 - Κείμενα δημοσιότητας
 - Νέα / Ανακοινώσεις
 - Πολύγλωσσο διαδικτυακό περιεχόμενο
 - Κείμενα πολιτικής, στρατηγικής, σχεδιασμού
 - Ορολογίες
 - Αρχεία
 - Φόρμες
 - Οδηγίες, συχνές ερωτήσεις

Τύποι γλωσσικών δεδομένων που είναι χρήσιμοι | 1



- Οποιοδήποτε κείμενο σε ηλεκτρονική μορφή σε μία ή περισσότερες Ευρωπαϊκές γλώσσες
- Ιδανικά, κείμενα σε μια γλώσσα συνοδευόμενα από την μετάφρασή τους σε μια άλλη ευρωπαϊκή γλώσσα (δηλ. παράλληλα δίγλωσσα ή πολύγλωσσα)

Κείμενο στα Ελληνικά

Διαφθορά είναι η εκμετάλλευση μιας θέσης εξουσίας απ' αυτόν που την κατέχει προς ίδιον όφελος. Είναι η κατάχρηση της εξουσίας από κομματικούς και κυβερνητικούς παράγοντες, από υπαλλήλους, από δημόσια πρόσωπα.

...

Το ίδιο κείμενο στα Αγγλικά

Corruption is the exploitation of a position of power by a person for their own benefit. It is the abuse of power by political party and government officials, employees and public figures.

...

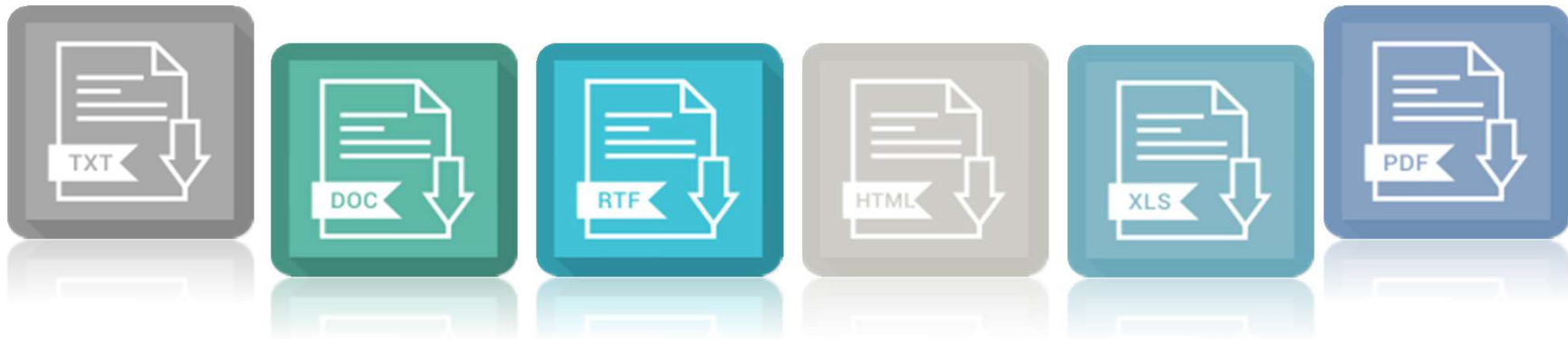
Τύποι γλωσσικών δεδομένων που είναι χρήσιμοι | 2



- Λίστες όρων και οι μεταφράσεις τους, δηλ. μια **ορολογία**

	A	B
	English	Ελληνικά
1		
2	FISCAL TERMS	ΔΗΜΟΣΙΟΝΟΜΙΚΟΙ ΟΡΟΙ
3	Accountability.	Υποχρέωση λογοδοσίας.
4	Accounting	Λογιστική
5	accounting basis	λογιστική βάση
6	accounting controls	λογιστικοί έλεγχοι
7	accounting system	λογιστικό σύστημα
8	accounts payable / receivable	λογαριασμοί πληρωτέοι / εισπρακτέοι
9	accrual accounting	λογιστική με βάση τα δεδουλευμένα έσοδα και έξοδα
10	accrual-based budget (accrual budgeting)	κατάρτιση προϋπολογισμού με βάση τα δεδουλευμένα έσοδα και έξοδα
11	acquis communautaire	κοινοτικό κεκτημένο
12	Activity	Δραστηριότητα
13	activity-based costing (ABC)	κοστολόγηση βάσει δραστηριότητας (ABC)
14	administrative controls	διοικητικοί έλεγχοι
15	Agenda 2000	Πρόγραμμα Δράσης 2000
16	aid-in-kind	αρωγή σε είδος
17	allocative efficiency	επιμεριστική αποδοτικότητα

Μορφότυποι κειμένων



- Κάθε κείμενο σε ηλεκτρονική μορφή, μηχανικά αναγνώσιμο
- Ορισμένοι μορφότυποι είναι πιο κατάλληλοι, γιατί απαιτούν λιγότερη χειρωνακτική ή αυτόματη προεπεξεργασία
- Περισσότερη προεπεξεργασία → μεγαλύτερη πιθανότητα για λάθη
→ λιγότερο χρήσιμο για το eTranslation

United Nations Convention against Corruption

Preamble

The States Parties to this Convention,

Concerned about the seriousness of problems and threats posed by corruption to the stability and security of societies, undermining the institutions and values of democracy, ethical values and justice and jeopardizing sustainable development and the rule of law,

Concerned also about the links between corruption and other forms of crime, in particular organized crime and economic crime, including money-laundering,

Concerned further about cases of corruption that involve vast quantities of assets, which may constitute a substantial proportion of the resources of States, and that threaten the political stability and sustainable development of those States,

Convinced that corruption is no longer a local matter but a transnational phenomenon that affects all societies and economies, making international cooperation to prevent and control it essential,

Convinced also that a comprehensive and multidisciplinary approach is required to prevent and combat corruption effectively



- Οι ακόλουθοι μορφότυποι είναι ιδιαίτερα χρήσιμοι (σε φθίνουσα σειρά προτίμησης):
 - Για παράλληλα κείμενα
 1. Μεταφραστικές μνήμες (.tmx)
 2. XML μεταφραστικά αρχεία (.xliff)
 3. Απλό κείμενο (.txt, .csv)
 4. Λογιστικά φύλλα (e.g. xlsx)
 - Για ορολογίες
 1. TermBase eXchange (.tbx)
 2. Απλό κείμενο (.txt, .csv)
 3. Λογιστικά φύλλα (e.g. xlsx)
 - Για μονόγλωσσα κείμενα
 1. Απλό κείμενο (.txt, .csv)

Χειρισμός των παράλληλων κειμένων σας



OXI



This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph.

Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου.

This is another English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph. This is an English paragraph.

Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου. Αυτή είναι η Ελληνική μετάφραση της παραπάνω παραγράφου.

This is an English sentence.

Αυτή είναι η μετάφραση της παραπάνω πρότασης.



OXI



English	Greek
<p>This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document.</p> <p>This is the content of an English document.</p> <p>This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document. This is the content of an English document.</p>	<p>Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου.</p> <p>Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου.</p> <p>Αυτή είναι η Ελληνική μετάφραση του κειμένου.</p> <p>Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου. Αυτή είναι η Ελληνική μετάφραση του κειμένου.</p>



- filename01_EL.txt
- filename01_EN.txt
- filename02_EL.txt
- filename02_EN.txt
- filename03_EL.txt
- filename03_EN.txt
- filename04_EL.txt
- filename04_EN.txt
- filename05_EL.txt
- filename05_EN.txt
- filename06_EL.txt
- filename06_EN.txt
- filename07_EL.txt
- filename07_EN.txt
- filename08_EL.txt
- filename08_EN.txt
- filename09_EL.txt
- filename09_EN.txt
- filename10_EL.txt
- filename10_EN.txt

Χρησιμοποιήστε **όμοια ονόματα αρχείων** για κάθε ζεύγος κειμένων (πρωτότυπο – μετάφραση)



- filename01_EL.txt
- filename01_EN.txt
- filename02_EL.txt
- filename02_EN.txt
- filename03_EL.txt
- filename03_EN.txt
- filename04_EL.txt
- filename04_EN.txt

Ενσωματώστε στο όνομα
αρχείου **αναγνωριστικούς
κωδικούς της γλώσσας**
των αρχείων



Όνομα

- τεκστ.txt
- Μετάφραση αναφοράς.txt
- κείμενο.txt
- traduction française.txt
- texte.txt
- text.txt
- report.txt
- report in Bulgarian.txt
- README.txt

Αρχείο	Επεξεργασία	Μορφή	Προβολή	Βοήθεια
English	Bulgarian	French	Greek	
text.txt	текст.txt	texte.txt	κείμενο.txt	
report.txt	report in Bulgarian.txt	traduction française.txt	Μετάφραση αναφοράς.txt	

Κριτήρια ομαδοποίησης των αρχείων σε σύνολα δεδομένων (πόρους)



- Θυμηθείτε! Σύνολο δεδομένων (ή πόρος): συλλογή αρχείων δεδομένων **ομαδοποιημένων** σύμφωνα με συγκεκριμένα **κριτήρια**
- Για τους πόρους που θα χρησιμοποιηθούν από το σύστημα μηχανικής μετάφρασης, δύο κριτήρια είναι κρίσιμα:
 - **Γλώσσα**: κάθε γλωσσικός πόρος προσδιορίζεται από τη γλώσσα ή τις γλώσσες που περιλαμβάνονται στα δεδομένα του, π.χ.
 - *Collection of texts in English – German*
 - *Documents in English – Norwegian - Finnish*
 - **Θεματική περιοχή**: κάθε γλωσσικός πόρος ιδανικά ανήκει σε μια θεματική, π.χ.
 - *Collection of texts in English – German in the culture domain*
 - *Social security documents in English – Norwegian - Finnish*

- Διοικητικά / κανονιστικά κείμενα και
- Θεματικές σχετικές με τις Υποδομές Ψηφιακών Υπηρεσιών του CEF (DSIs)

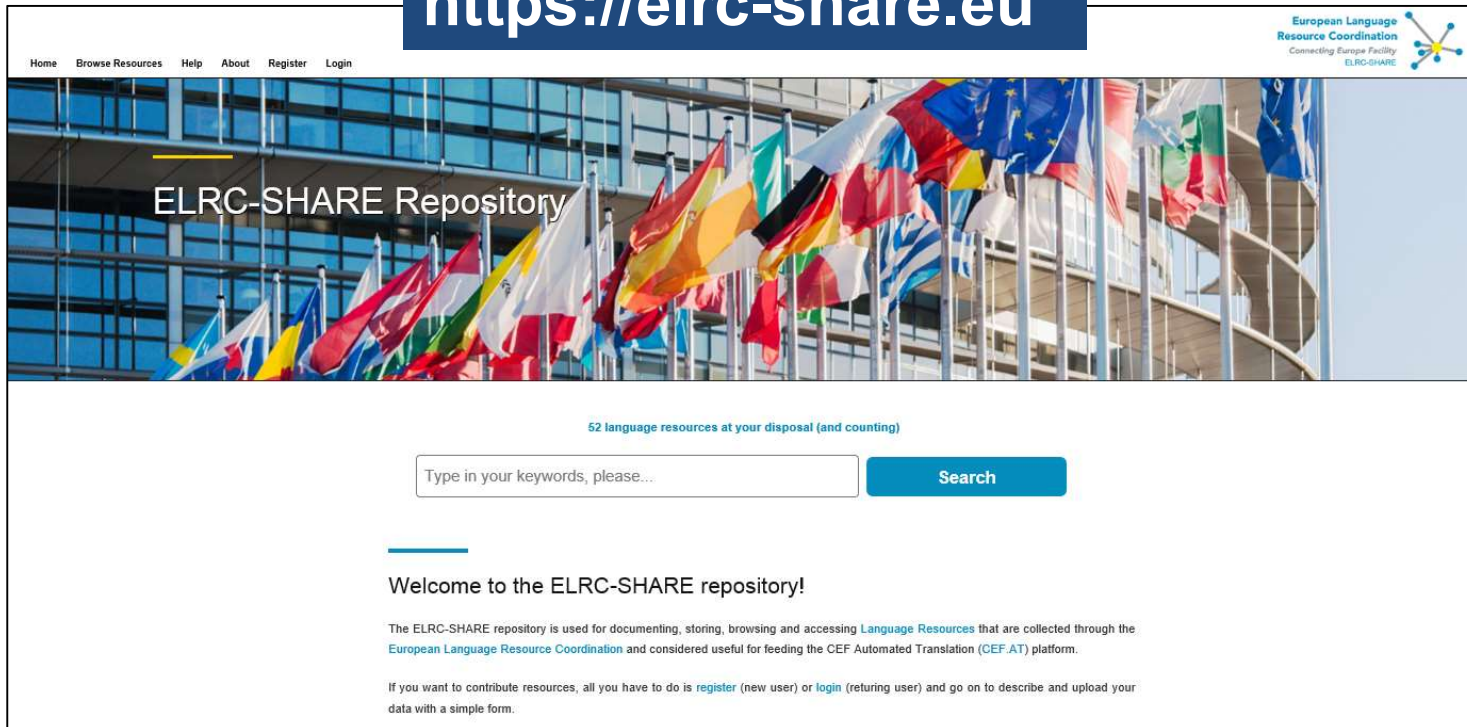
DSI	Θέματα
Online Dispute Resolution	Δικαιώματα καταναλωτών
Electronic Exchange of Social Security Information	Κοινωνική ασφάλιση, απασχόληση, ανεργία
eProcurement	Δημόσιες συμβάσεις, προκηρύξεις
European e-Justice Portal	Δικαιοσύνη, νόμοι
eHealth	Υγεία, ιατρική
Business Registers Interconnection System	Επιχειρήσεις, αγορά
Safer Internet	
Cybersecurity	
Public Open Data	
Europeana	Πολιτισμός

Πώς να συνεισφέρετε τα γλωσσικά δεδομένα σας
βήμα-προς-βήμα

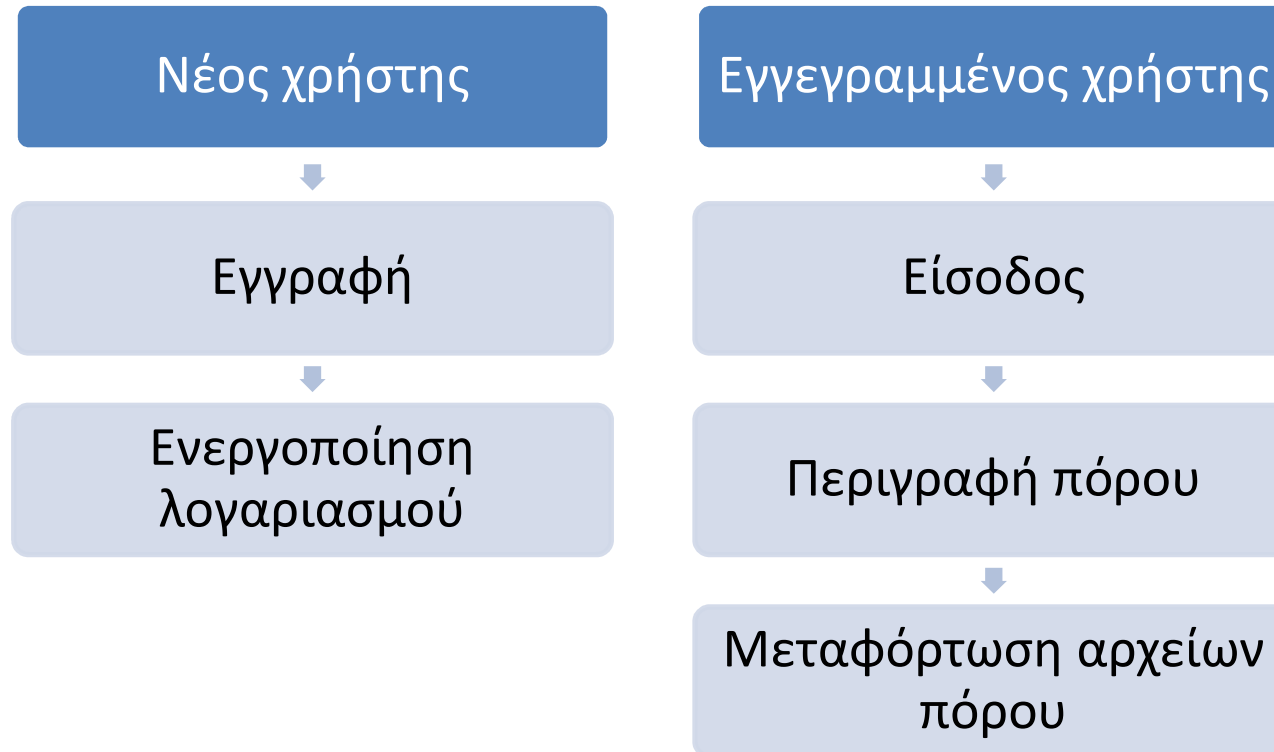
Τι είναι: αποθετήριο γλωσσικών πόρων

Τι προσφέρει: πρόσβαση σε, συνεισφορά και διαμοιρασμό γλωσσικών πόρων

<https://elrc-share.eu>



The screenshot shows the homepage of the ELRC-SHARE repository. At the top, there is a navigation menu with links for Home, Browse Resources, Help, About, Register, and Login. The main header features the text "ELRC-SHARE Repository" overlaid on a background image of various European national flags. Below the header, a status bar indicates "52 language resources at your disposal (and counting)". A search bar with the placeholder text "Type in your keywords, please..." and a blue "Search" button is positioned below the status bar. The main content area begins with a welcome message: "Welcome to the ELRC-SHARE repository!". This is followed by a paragraph explaining that the repository is used for documenting, storing, browsing, and accessing language resources collected through the European Language Resource Coordination, which are considered useful for feeding the CEF Automated Translation (CEF.AT) platform. A final paragraph states that users can contribute resources by registering (new user) or logging in (returning user) and then describing and uploading their data using a simple form.





 Register

[Home](#) [Browse Resources](#) [Help](#) [About](#) [Register](#) [Login](#)

ELRC-SHARE Repository



Welcome to the ELRC-SHARE repository!

Create Account

Desired account name*

First name*

Last name*

E-mail*

Country*

Organization*

Phone number

Password*

Password confirmation*

I accept the [ERLC-SHARE Terms of Service for registered users.](#)



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"

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)



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*

Choose File 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

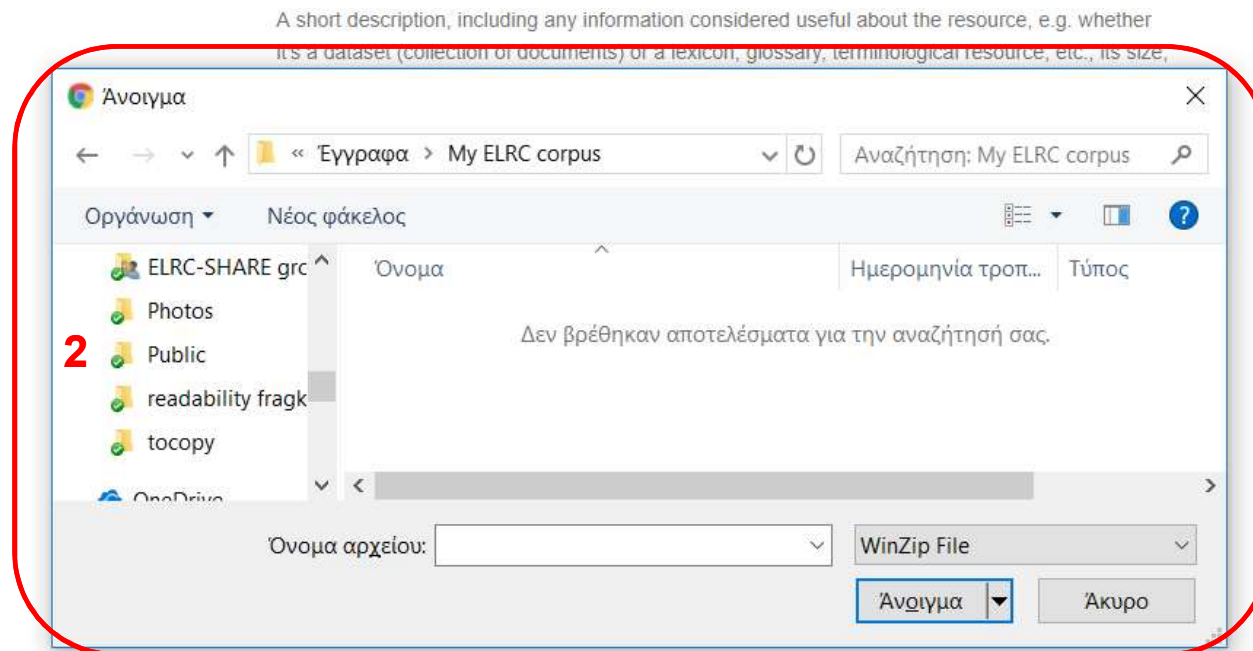
Submit

Reset



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

Συνεισφορά γλωσσικών δεδομένων (4/5)



Upload Resource **1** Choose File 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 Submit Reset



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*

Please provide a URL containing the files you wish to contribute

Submit

Reset



Help

Documentation on the ELRC-SHARE editor

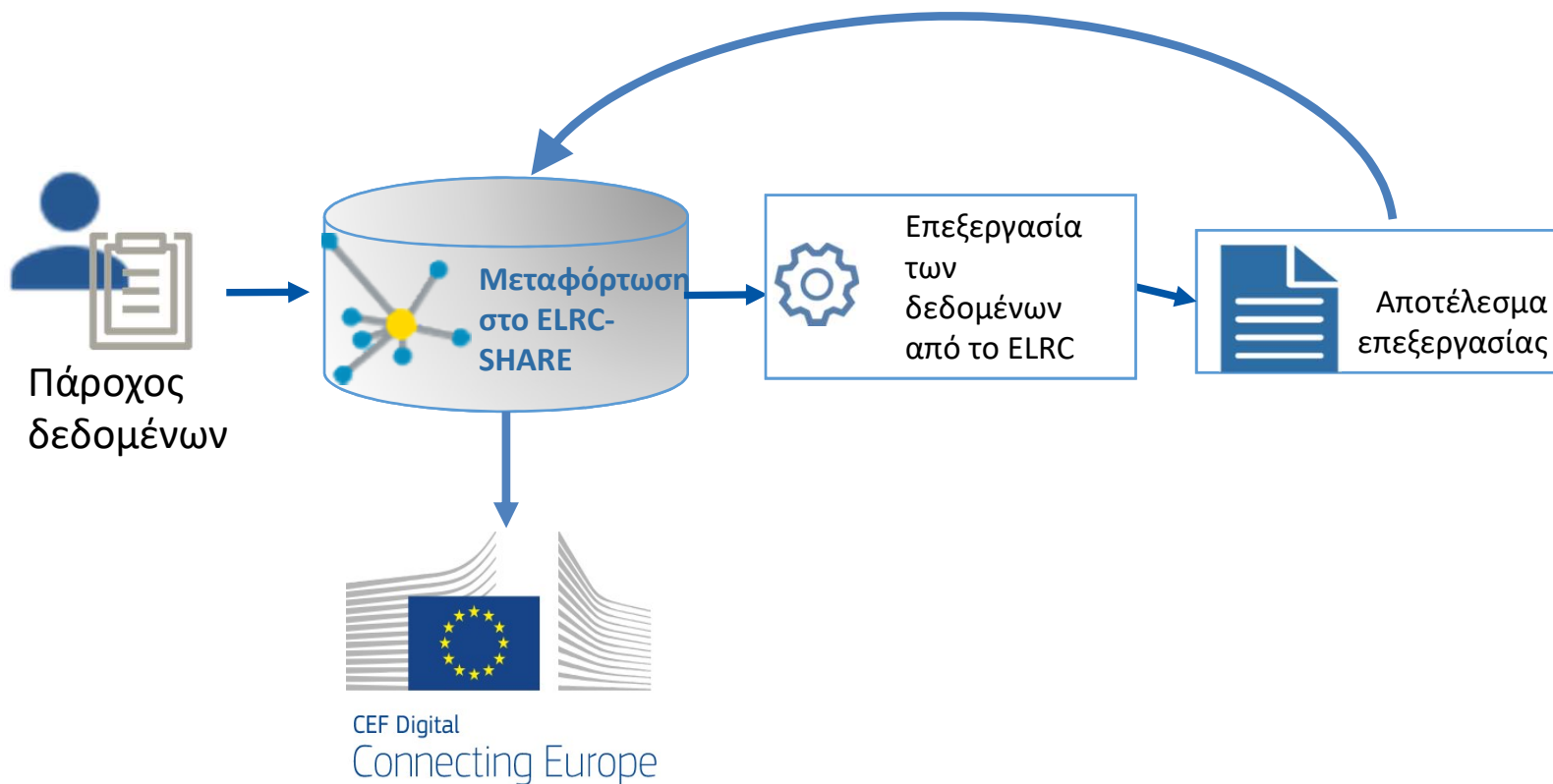
The following guidelines provide detailed information on how to use the editing facility for documenting and uploading LRs:

- [Walkthrough for contributors](#)
- [Walkthrough for editors](#)

ELRC-SHARE schema

- [ELRC-SHARE schema XSD](#) (based on the META-SHARE Schema)
- [Documentation about the schema](#)

Τι συμβαίνει μετά;





Εξαγωγή δεδομένων

Κείμενα
«παγιδευμένα» σε
ένα συμπιεσμένο
αρχείο ή μια βάση
δεδομένων;



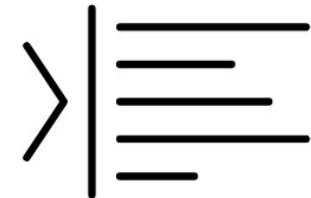
Ανωθυμοποίηση

Περιέχουν τα
δεδομένα σας
ευαίσθητες
πληροφορίες;



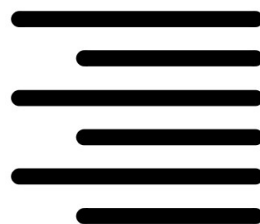
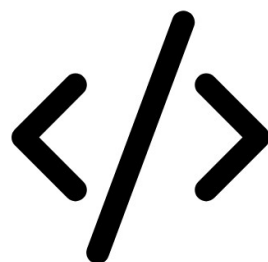
Καθαρισμός

Περιέχουν τα
δεδομένα σας
θόρυβο (π.χ.
πολλές λανθα-
σμένες εγγραφές);



Επαναμορφοποίηση

Πρέπει να επανα-
μορφοποιήσετε
αρχεία DOCX σε
XML, ή αρχεία PDF
σε WORD;



Μετατροπή δεδομένων

Αν τα δεδομένα σας δεν είναι στον κατάλληλο μορφότυπο (φορμάτ), μπορούμε να σας βοηθήσουμε να τα μετατρέψετε

Αφαίρεση ετικετών

Τα δεδομένα σας περιέχουν περιττή επισήμειωση;

Στοίχιση

Δεν είναι σωστά στοιχισμένες οι μεταφράσεις σας; Θα το κάνουμε εμείς για εσάς με ειδικό λογισμικό!

Μεταδεδομένα

Τα μεταδεδομένα είναι ζωτικής σημασίας για τη χρήση των πόρων σας! Μπορούμε να τα επικυρώσουμε και να τα οργανώσουμε για την ομάδα σας



File01_el.txt
File01_en.doc
File02_el.pdf
File02_en.txt
File03_el.doc
File03_en.doc
...

After
processing

```
<tu tuid="3931">  
  <tuv xml:lang="en">  
    <seg>Corruption is the exploitation of a position of  
power by a person for their own benefit.</seg> </tuv>  
  <tuv xml:lang="el">  
    <seg>Διαφθορά είναι η εκμετάλλευση μιας θέσης  
εξουσίας απ' αυτόν που την κατέχει προς ίδιον  
όφελος.</seg></tuv></tu>  
<tu tuid="3932">  
  <tuv xml:lang="en">  
    <seg>It is the abuse of power by political party and  
government officials, employees and public  
figures.</seg>  
  </tuv>  
  <tuv xml:lang="el">  
    <seg>Είναι η κατάχρηση της εξουσίας από κομματικούς  
και κυβερνητικούς παράγοντες, από υπαλλήλους, από  
δημόσια πρόσωπα.</seg></tuv></tu>
```

Cyprus at a glance (Processed) 🗨️

Attribution details: 'Cyprus at a glance (Processed)' language resource was created for the European Language Resources Coordination Action (ELRC) (<http://lr-coordination.eu/>) with primary data copyrighted by Press and Information Office, Republic of Cyprus (www.pio.gov.cy/).

Press and Information Office (P.I.O.) publication providing information on Cyprus (number of pages 44).

DSI Relevance: Europeana

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

Distribution

Availability: Available

Licences

Non-standard/ Other Licence/ Terms

Conditions: Attribution, Non Commercial Use

Distribution Details

Distribution Medium: Data Downloadable

Attribution Details: 'Cyprus at a glance (Processed)' language resource was created for the European Language Resources Coordination Action (ELRC) (<http://lr-coordination.eu/>) with primary data copyrighted by Press and Information Office, Republic of Cyprus (www.pio.gov.cy/).

IPR Holders

Press and Information Office, Republic of Cyprus 🗨️

Contact Person

Ioannis Soulos 🗨️

text 🗨️

Multilingual text corpus

Languages

Spanish; Castilian (es)

French (fr)

English (en)

Modern Greek (1453-) (el)

Italian (it)

Linguality

Linguality type: Multilingual

Multi-linguality type: Parallel (10 TMX files, one for each language pair)

Text Format

TMX

Size

3,733 Translation Units

Character encoding

UTF-8

Domains

POLITICS

Political Framework (Eurovoc 0408)

SOCIAL QUESTIONS

Demography And Population (Eurovoc 2818)

GEOGRAPHY

Europe (Eurovoc 7208)

Eurovoc Classification

Text type: Encyclopaedic Texts

Text genre: Information

Creation

Creation mode details: Text extraction from pdf files, document alignment on document and sentence level, generation of a TMX file for each language pair.

Resource Creation

Created using ELRC Services

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.lr-coordi...>

Funding Type: Service Contract

Funder: European Commission

Funding Country: European Union (EU)

Project duration: 13/12/2016 - 12/02/2020

Metadata

Created: 01/07/2016

Last Updated: 12/12/2016

Metadata Language: English (en)

Metadata Creator

Ioannis Soulos 🗨️

Vassilis Papavassiliou 🗨️

Version

Version: 2.0

Relations

Related Resource: [Cyprus at a glance](#)

Relation Type: Is Aligned Version Of

Όλα τα στάδια επεξεργασίας είναι δυνατό να πραγματοποιηθούν μέσω της υπηρεσίας επιτόπιας υποστήριξης, σε όλους τους παρόχους δεδομένων του δημόσιου τομέα, δωρεάν

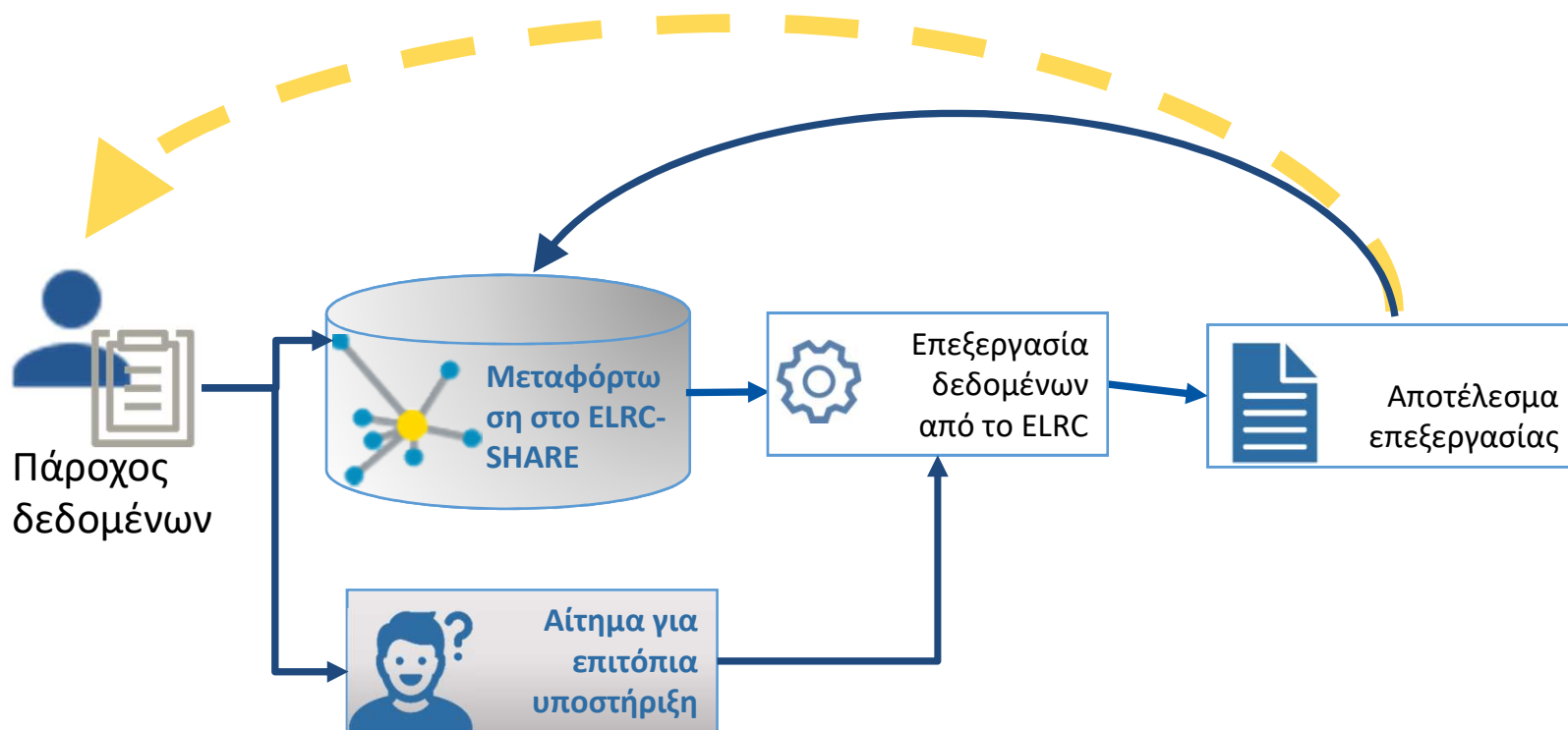




Η ομάδα των ειδικών μας θα σας επισκεφθεί για να σας παράσχει επιτόπια υποστήριξη στον χώρο εργασίας σας

Θα διορθώσουμε τα προβλήματα στα δεδομένα σας και θα επιστρέψουμε τα αποτελέσματα της επεξεργασίας απευθείας σε εσάς. Μπορούμε επίσης να σας βοηθήσουμε να βελτιώσετε τις διεργασίες διαχείρισης των δεδομένων σας. Απλώς ρωτήστε μας!

Τι συμβαίνει στα δεδομένα σας;



Πώς μπορείτε να ζητήσετε
τις υπηρεσίες του ELRC;

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-cooridantion.eu

Ευχαριστώ!

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