



# ΔΗΜΙΟΥΡΓΙΑ, ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΔΙΑΜΟΙΡΑΣΜΟΣ ΓΛΩΣΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ: ΥΠΑΡΧΟΥΣΕΣ ΠΡΑΚΤΙΚΕΣ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ ΣΤΗΝ ΕΛΛΑΔΑ

Μαρία Γαβριηλίδου  
15/12/2020



## ΓΛΩΣΣΙΚΑ ΔΕΔΟΜΕΝΑ

- Το καύσιμο της Γλωσσικής Τεχνολογίας και της Τεχνητής Νοημοσύνης είναι τα **γλωσσικά δεδομένα** (σύνολα δεδομένων λόγου σε κάθε μορφή: κειμενικά, ηχητικά, λεξικά/εννοιολογικά δεδομένα, βίντεο...)
- **Γλωσσικά** δεδομένα: ειδική κατηγορία δεδομένων
- Υπάρχουν παντού, παράγονται διαρκώς, για πολλούς λόγους
  - αλλά όχι με στόχο τη γλωσσική επεξεργασία
- Τα συστήματα γλωσσικής τεχνολογίας χρειάζονται επεξεργάσιμα δεδομένα σε ψηφιακή μορφή
  - Σε μεγάλες ποσότητες – και κατά το δυνατόν καλύτερης ποιότητας
  - Σε διάφορες γλώσσες, θεματικές, κειμενικά είδη....



κάθε άλλο παρά εύκολη διαδικασία

## ΚΥΚΛΟΣ ΖΩΗΣ ΓΛΩΣΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ



## ELRC WHITE PAPER: ΣΥΜΠΕΡΑΣΜΑΤΑ ΓΙΑ ΤΗΝ ΕΛΛΑΔΑ



### • Μετάφραση

- χαρτογράφηση μεταφραστικών αναγκών του Δημοσίου
- μεταφραστική διαδικασία
  - επίσημες μεταφράσεις μέσω μεταφραστικής υπηρεσίας ΥΠΕΞ
  - για τις καθημερινές μεταφραστικές ανάγκες δεν υπάρχει θεσμοθετημένη διαδικασία μετάφρασης, κοινή για τις υπηρεσίες του Δημοσίου
  - δεν υπάρχει κουλτούρα ψηφιοποίησης της διαδικασίας

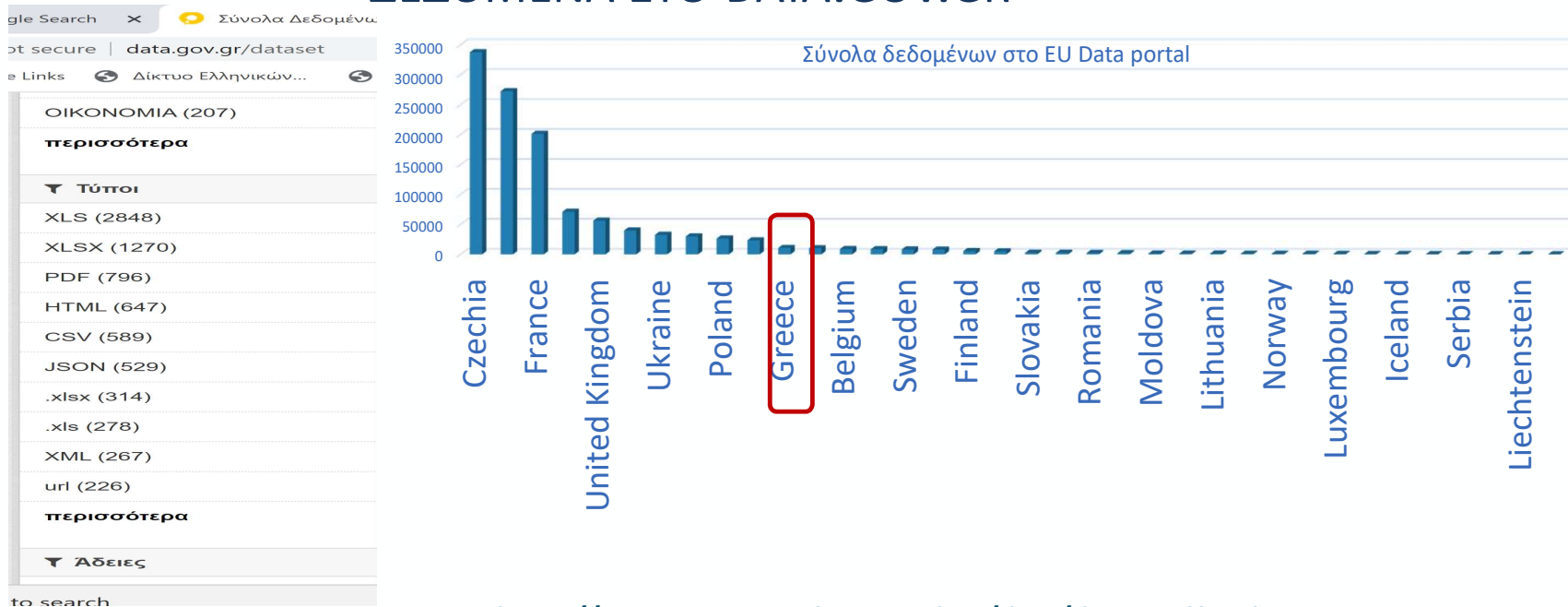


## ΔΙΑΘΕΣΙΜΟΤΗΤΑ ΑΝΟΙΧΤΩΝ ΓΛΩΣΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

- Υποδομές Δημοσίου
  - Διαύγεια (αποφάσεις Δημόσιων Υπηρεσιών)
  - Data.gov.gr
- Ερευνητικές υποδομές γλωσσικών δεδομένων
  - ELRC-SHARE
  - CLARIN
  - European Language Grid



## ΔΕΔΟΜΕΝΑ ΣΤΟ DATA.GOV.GR



<https://www.europeandataportal.eu/data/datasets?locale=en&minScoring=0&page=1> / 14.12.2020

## ΓΛΩΣΣΙΚΗ ΠΟΛΙΤΙΚΗ ΚΑΙ ΓΛΩΣΣΙΚΑ ΔΕΔΟΜΕΝΑ

- Ο ψηφιακός γραμματισμός και η κουλτούρα της ψηφιακότητας δεν είναι ακόμη ικανοποιητικά διαδεδομένα στην Ελλάδα
- Η Γλωσσική Τεχνολογία δεν συμπεριλαμβάνεται στη γλωσσική πολιτική της χώρας
- Η σημασία της στην ανάπτυξη της Τεχνητής Νοημοσύνης μόνο πολύ πρόσφατα άρχισε να αναγνωρίζεται
- Για να αξιοποιήσουμε τα υπάρχοντα γλωσσικά δεδομένα, να μεριμνήσουμε για τη διατήρησή τους και να διασφαλίσουμε τη συνεχή παραγωγή τους μελλοντικά, απαιτούνται
  - Θεσμικό και νομικό πλαίσιο διαχείρισης και διάθεσης γλωσσικών δεδομένων
  - Τεχνικό πλαίσιο διαχείρισης της παραγωγής, αποθήκευσης και διαμοιρασμού τους
  - Ευαισθητοποίηση και εκπαίδευση των παρόχων



**Θανάσης Σκλαπάνης**  
Υπ. Ψηφιακής  
Διακυβέρνησης



**Δημήτρης Γιουτίκας**  
ΕΣΔΔΑ



**Δημήτρης Καπόπουλος**  
Υπ. Ψηφιακής  
Διακυβέρνησης



**Αλέξανδρος Νούσιας**  
ΕΚΕΦΕ Δημόκριτος



**Βασίλης Παπαβασιλείου**  
ΙΕΛ / Ε.Κ. «Αθηνά»



## ΠΟΙΟ ΕΙΝΑΙ ΤΟ ΠΛΑΙΣΙΟ ΓΙΑ ΤΟΝ ΔΙΑΜΟΙΡΑΣΜΟ ΨΗΦΙΑΚΩΝ ΓΛΩΣΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΕΛΛΑΔΑ;



- Νέα οργάνωση Δημόσιας Διοίκησης – Υπουργείο Ψηφιακής Διακυβέρνησης
- Η κύρια εφαρμοστική εγκύκλιος είναι σε ισχύ από 8.1.2015  
[https://www.culture.gov.gr/DocLib/efarmostiki\\_egiklios\\_N.4305-2014.pdf](https://www.culture.gov.gr/DocLib/efarmostiki_egiklios_N.4305-2014.pdf)
- Ειδική εγκύκλιος για τη διάθεση γλωσσικών δεδομένων του Δημοσίου στο ELRC (Απρ. 2016)
- Ήταν ικανοποιητική η εφαρμογή τους; Τι μπορούμε να πούμε κοιτώντας πίσω;
- Υπάρχουν νέες προβλέψεις για τη διαχείριση και διάθεση των **ψηφιακών γλωσσικών** δεδομένων;
  - μεταξύ υπηρεσιών του Δημοσίου τομέα
  - μεταξύ Δημοσίου τομέα και πολιτών;

## ΝΟΜΙΚΟ ΠΛΑΙΣΙΟ



- Η συζήτηση για το άνοιγμα και διαμοιρασμό των δεδομένων δεν είναι καινούργια. Ποια είναι η κατάσταση που έχει διαμορφωθεί στη νομοθεσία σχετικά με τα ανοιχτά δεδομένα αλλά και ευρύτερα σε σχέση με το διαμοιρασμό;
- Υπάρχει κάποια ειδική πρόβλεψη στον ελληνικό (ευρωπαϊκό) νόμο σχετικά με τα γλωσσικά δεδομένα;
- Ποιος ο ρόλος της Πνευματικής Ιδιοκτησίας και των Προσωπικών Δεδομένων στην εφαρμογή πολιτικών ανοιχτών δεδομένων;

## ΤΕΧΝΙΚΟ ΠΛΑΙΣΙΟ

- data.gov.gr: ο καθρέφτης της τεχνικής υλοποίησης και εφαρμογής της πολιτικής της ΔΔ για τα ανοιχτά δεδομένα
- λίγα στατιστικά στοιχεία από το EU dataportal  
<https://www.europeandataportal.eu/catalogue-statistics/CurrentState>
  - τα 10.600 Ελληνικά σύνολα δεδομένων = το 1% του συνόλου των δεδομένων του EU Data Portal
  - όμως, από αυτά, σύμφωνα με την αξιολόγηση του EU Data Portal, μόνο 2.800 έχουν επαρκή μεταδεδομένα (= σύμφωνα με τα FAIR principles)
- Δεδομένης της ανάγκης γλωσσικών δεδομένων για την ΓΤ και τη ΤΝ, πώς θα εξασφαλίσουμε μεγαλύτερη ποσότητα και καλύτερη ποιότητα δεδομένων και μεταδεδομένων;



## ΠΡΟΚΛΗΣΕΙΣ



- Πώς αντιμετώπισε η υπηρεσία σας το εγχείρημα της συλλογής και του διαμοιρασμού των δεδομένων; Τι δυσκολίες συναντήσατε;
- Ποιες είναι οι βασικές προκλήσεις για τον διαμοιρασμό των γλωσσικών δεδομένων στην Ελλάδα;
- Τι πιστεύετε ότι πρέπει να γίνει για να βελτιωθεί η κατάσταση και να διευκολυνθεί ο διαμοιρασμός των γλωσσικών δεδομένων στην Ελλάδα;
- Υπάρχει πρόβλεψη για άλλες υποδομές διάθεσης γλωσσικών δεδομένων του Δημοσίου;

- Πώς μπορεί να ανταποκριθεί η Γλωσσική Τεχνολογία στην ανάγκη
  - διαχείρισης μιας κρίσης που αντιμετωπίζει μια χώρα
  - με άμεση και γρήγορη παροχή υπηρεσιών
  - για άμεση χρήση από την κοινότητα (Δημόσιο και ιδιωτικό τομέα);





ΕΥΧΑΡΙΣΤΟΥΜΕ ΠΟΛΥ ΓΙΑ ΤΗΝ ΠΡΟΣΟΧΗ ΣΑΣ!

Website: [www.lr-coordination.eu](http://www.lr-coordination.eu)

Twitter: @LR\_Coordination

Email: [info@lr-coordination.eu](mailto:info@lr-coordination.eu)

