

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΟΙΚΟΝΟΜΙΑΣ & ΑΝΑΠΤΥΞΗΣ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΠΡΟΓΡΑΜΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΠΡΟΓΡΑΜΜΑΤΩΝ



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ



ΕΣΠΑ
2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Μαρία Γαβριηλίδου
Ινστιτούτο Επεξεργασίας του Λόγου
ΕΚ Αθηνά



Συμπόσιο για την Ανοικτή Επιστήμη
Αθήνα, 29-30 Νοεμβρίου 2018

#OpenScienceGreece 29/11/2018

- εθνικό σκέλος της ESFRI υποδομής CLARIN ERIC
- παρέχει ψηφιακά γλωσσικά δεδομένα, εργαλεία και διαδικτυακές υπηρεσίες επεξεργασίας γλώσσας
- υποστηρίζει κάθε ερευνητική δραστηριότητα που αφορά επεξεργασία και ανάλυση φυσικής γλώσσας (με έμφαση στα Ελληνικά)
- απευθύνεται στην ακαδημαϊκή και ερευνητική κοινότητα, αλλά και στη βιομηχανία και στο ευρύ κοινό
- υποστηρίζει ολόκληρο τον κύκλο ζωής γλωσσικών δεδομένων (συλλογή, τεκμηρίωση, αποθήκευση, επεξεργασία, αδειοδότηση, διαμοιρασμό)
- διευκολύνει διαδικασίες από τη διατύπωση ερευνητικών ερωτημάτων έως και την επαλήθευση ή διάψευσή τους
- κατανεμημένο δίκτυο οργανισμών και αποθετηρίων
- 9 κόμβοι δικτύου (σύντομα 14), ~500 εγγεγραμμένοι χρήστες στην Ελλάδα
- από τις πρώτες υποδομές με σαφές νομικό σύστημα παροχής πόρων και υπηρεσιών



Δεδομένα (2/2): αρχές FAIR

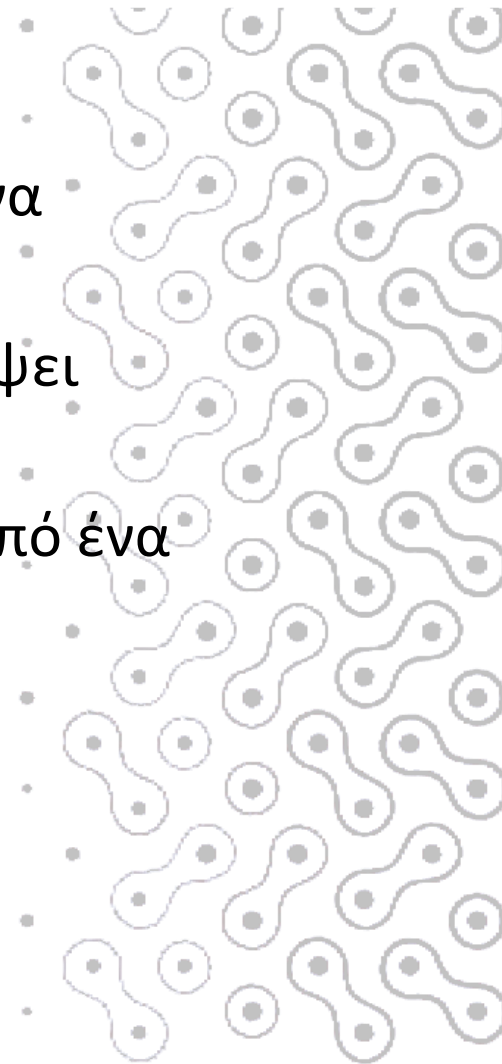
- **Findable:** κοινό σχήμα μεταδεδομένων και μοναδικά αναγνωριστικά (PIDs)
- **Accessible:** μεταδεδομένα και δεδομένα κατανοητά στους ανθρώπους και αξιοποιήσιμα από υπολογιστές, κατατεθειμένα σε ασφαλές δίκτυο αποθετηρίων
- **Interoperable:** διαλειτουργικότητα σε επίπεδο μεταδεδομένων, δεδομένων και εργαλείων/υπηρεσιών επεξεργασίας, συμβατότητα με διεθνή πρότυπα και καλές πρακτικές
- **Reusable:** πληροφορίες για την πηγή κάθε πόρου, σαφής άδεια χρήσης
- 85% των πόρων και των εργαλείων διατίθενται με ανοιχτές άδειες (CC/Apache/BSD/LGPL), από τις οποίες λίγες με περιορισμούς NC / ND
- προωθούνται οι ανοιχτές άδειες χρήσης, σαφής προσανατολισμός της υποδομής προς την ανοιχτότητα (καταστατικό/όροι χρήσης της υποδομής)



Τι προσφέρει το CLARIN:EL στον χρήστη



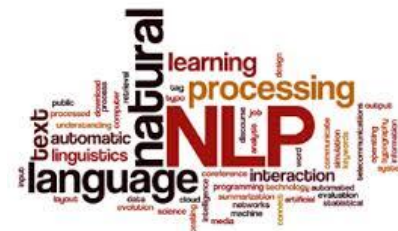
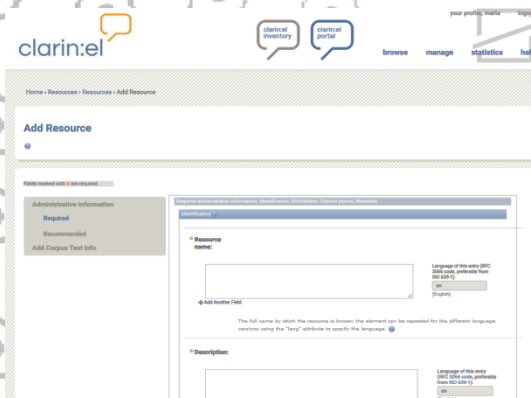
- Ένας ερευνητής από το γραφείο του στην Κέρκυρα μπορεί:
 - με μία εγγραφή (single sign-on) με πιστοποίηση (authentication)
 - να ψάξει, να βρει και να πάρει την έγκριση να χρησιμοποιήσει κείμενα
 - από αποθετήρια στην Αθήνα, τη Ρόδο ή τη Θεσσαλονίκη
 - να επιλέξει το ακριβές σύνολο δεδομένων στα οποία θέλει να δουλέψει
 - ή, εναλλακτικά, να ανεβάσει στην υποδομή δικά του δεδομένα
 - να τρέξει πάνω στα δεδομένα εργαλεία σημασιολογικής ανάλυσης από ένα κέντρο και
 - υπηρεσίες αναγνώρισης λεκτικής επιθετικότητας από ένα άλλο
 - να αποθηκεύσει τα αποτελέσματα της ανάλυσης και
 - να τα μοιραστεί με συνεργάτες του εντός ή και εκτός υποδομής





Διαχείριση δεδομένων

- διαφύλαξη (μεταδεδομένα σε βάση, δεδομένα στον Πίθο)
- μοναδικά αναγνωριστικά (PIDs)
- προσθήκη μεταδεδομένων / επιμέλεια / συντήρηση και επικαιροποίηση (CLARIN:EL metadata editor)
- αδειοδότηση
- έκδοση και διαμοιρασμός (internal-ingested-published)
- επεξεργασία γλωσσικών δεδομένων
 - μέσω εργαλείων και διαδικτυακών υπηρεσιών
 - οργανική ανάπτυξη αποθετηρίων μέσω ενσωμάτωσης παραγώγων
 - σχέσεις δεδομένων και υπηρεσιών

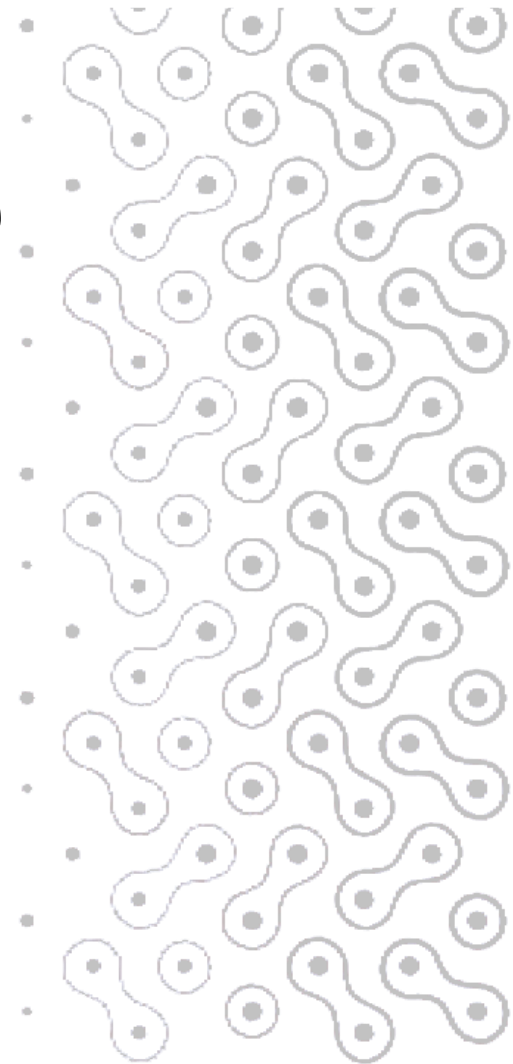


Προκλήσεις

- Τεχνικές προκλήσεις
 - λειτουργία 24/7/365
 - δυσκολία εύρεσης εξειδικευμένου ερευνητικού προσωπικού

- Νομικές προκλήσεις
 - πνευματικά δικαιώματα συγγραφέων
 - GDPR, προσωπικά/ευαίσθητα δεδομένα

- Ανθρώπινος παράγοντας
 - ψηφιακός γραμματισμός / εκπαίδευση χρηστών
 - δυσκολίες στην προσαρμογή παγιωμένων ερευνητικών μεθοδολογιών στη χρήση ψηφιακών υπηρεσιών
 - κουλτούρα διαμοιρασμού



Βιωσιμότητα

- Εκτίμηση ετήσιου κόστους: 300.000 ευρώ
- Συνεργασία με άλλα έργα/υποδομές
 - ΕΔΕΤ
 - CLARIN ERIC
 - DARIAH/DYAS > ΑΠΟΛΛΩΝΙΣ

<https://apollonis-infrastructure.gr/>

clarin:el 



απολλωνις
Εθνική Υποδομή για τις Ψηφιακές Ανθρωπιστικές Επιστήμες και Τέχνες
και για τη Γλωσσική Έρευνα και Καινοτομία



Σας ευχαριστώ!

