

Η Ερευνητική Υποδομή CLARIN:EL

Γλωσσικά δεδομένα & εργαλεία γλωσσικής
επεξεργασίας στην υπηρεσία της έρευνας



Η Υποδομή CLARIN:EL

- εθνικό σκέλος της ευρωπαϊκής 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)
- από τις πρώτες υποδομές με σαφές νομικό σύστημα παροχής πόρων και υπηρεσιών
- κατανεμημένο δίκτυο οργανισμών και αποθετηρίων
- 10 κόμβοι δικτύου (σύντομα 14), ~600 εγγεγραμμένοι χρήστες στην Ελλάδα



Η αποστολή του CLARIN

το CLARIN ενσωματώνει

- **Γλωσσικούς Πόρους:** ψηφιακό περιεχόμενο κάθε είδους (κείμενο, ήχο, εικόνα, βίντεο), λεξικά, οντολογίες, ορολογικά γλωσσάρια κλπ. και
- **Εργαλεία Γλωσσικής Τεχνολογίας:** εργαλεία αναγνώρισης φωνής, λημματοποιητές, συντακτικούς αναλυτές, εργαλεία αυτόματης εξαγωγής περίληψης, εργαλεία εξαγωγής πληροφορίας κτλ.
- σε ένα συστηματικά οργανωμένο **δίκτυο αποθετηρίων** το οποίο είναι διαθέσιμο σε ερευνητές όλων των επιστημών
- που διαθέτει **εθνικά υπο-δίκτυα** που μεριμνούν για την έρευνα και την ψηφιακή προσαρμογή και ετοιμότητα των διαφόρων γλωσσών

Το όραμα του CLARIN



- ένας ερευνητής από το γραφείο του στην Κέρκυρα θα μπορεί:
- με μία εγγραφή με πιστοποίηση
- να ψάξει, να βρει και να πάρει την έγκριση να χρησιμοποιήσει κείμενα
- από την Οξφόρδη, τη Θεσσαλονίκη και το Λέιντεν
- να επιλέξει το ακριβές σύνολο δεδομένων στα οποία θέλει να δουλέψει και να αποθηκεύσει την επιλογή του
- να τρέξει πάνω στην επιλογή του εργαλεία σημασιολογικής ανάλυσης από την Αθήνα και
- στατιστικά εργαλεία από τη Βουδαπέστη
- να χρησιμοποιήσει την υπολογιστική ισχύ ενός άλλου υπολογιστικού κέντρου, όπου και όποτε απαιτείται
- να αποθηκεύσει τη διαδικασία και τα αποτελέσματα της ανάλυσης και
- να τα μοιραστεί με συνεργάτες του στο Παρίσι, στη Βιέννη και στο Ελσίνκι

Δεδομένα CLARIN:EL - Γλωσσικοί πόροι κάθε τύπου και μέσου



367 σώματα κειμένων / συλλογές



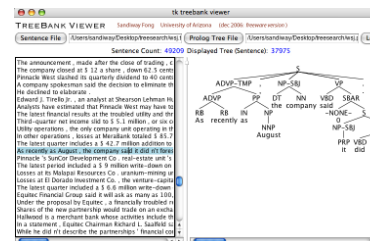
Επισημαίνεται ότι η παρούσα έκθεση περιλαμβάνει πληροφορίες που προέρχουν από διάφορες πηγές και μπορεί να μην είναι εντελώς ακριβείς. Η CLARIN:EL δε φέρει ευθύνη για τυχόν λάθη ή παραλείψεις. Η έκθεση αυτή είναι διαθέσιμη ως κείμενο PDF στην ιστοσελίδα της CLARIN:EL.



48 λεξικοί/εννοιολογικοί πόροι



38 εργαλεία και διαδικτυακές υπηρεσίες επεξεργασίας φυσικής γλώσσας

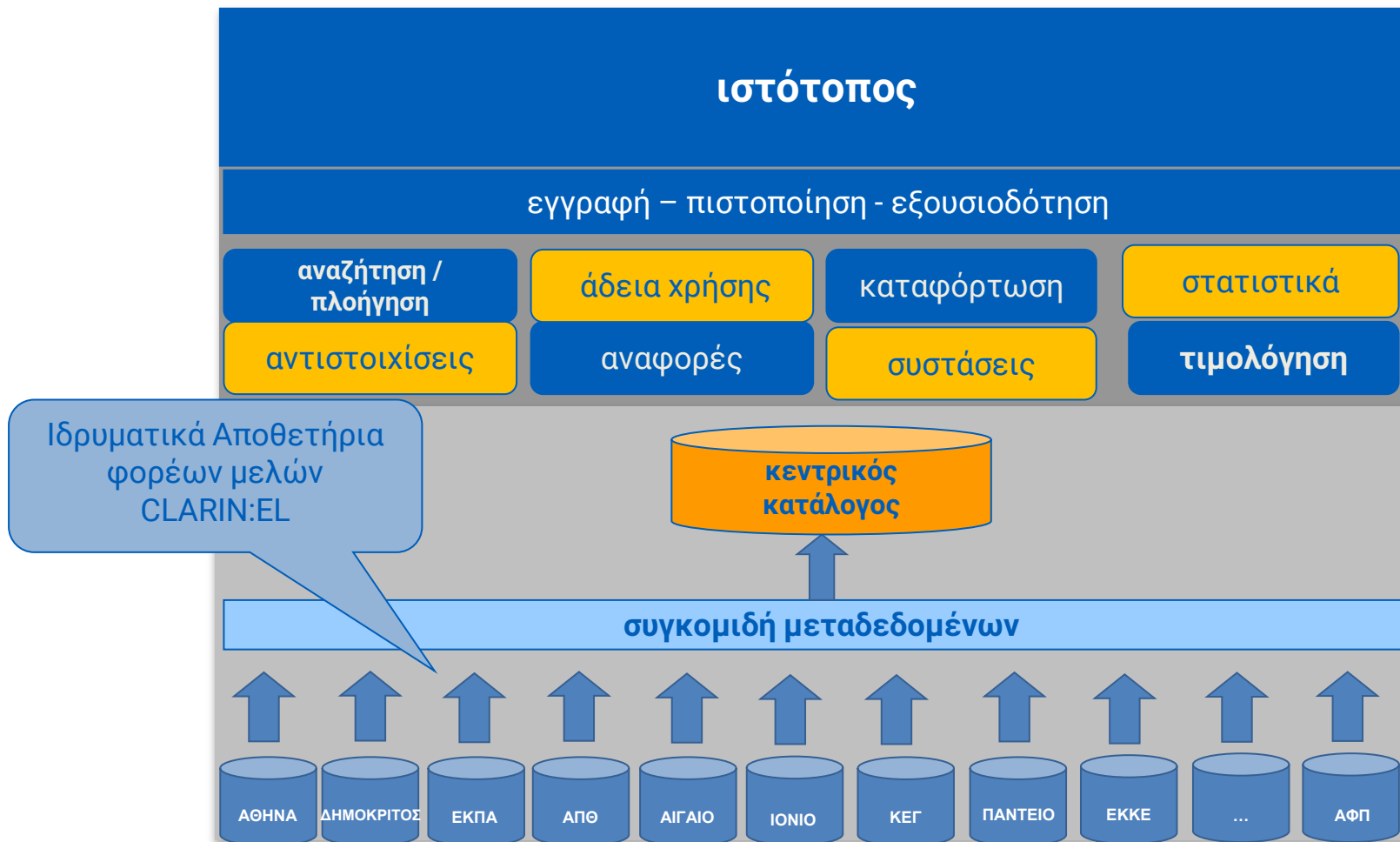


Τι προσφέρει το CLARIN:EL



- δύο στενά διασυνδεδεμένα υποσυστήματα:
 - **σύστημα πόρων (τεκμηρίωσης, αποθήκευσης, διαμοιρασμού + αναζήτησης, ανάκτησης, καταφόρτωσης** γλωσσικών δεδομένων (resources infrastructure)
 - **σύστημα επεξεργασίας** γλωσσικών πόρων μέσω διαδικτυακών υπηρεσιών γλωσσικής επεξεργασίας και παραγωγή νέων δεδομένων (processing infrastructure)
- **διαθέσιμα στους ερευνητές όλων των επιστημών**

Αρχιτεκτονική δικτύου CLARIN:EL



Ποιοι μπορούν να χρησιμοποιούν την υποδομή



- η υποδομή CLARIN:EL είναι **διαθέσιμη σε όλους**, δηλαδή
 - στα μέλη του δικτύου,
 - στους εγγεγραμμένους χρήστες, και
 - σε απλούς επισκέπτες χρήστες (χωρίς εγγραφή)
- ωστόσο, τα μέλη του δικτύου και οι εγγεγραμμένοι χρήστες έχουν πρόσβαση στο σύνολο των λειτουργιών της υποδομής, ενώ οι επισκέπτες χρήστες σε περιορισμένες λειτουργίες

Πλεονεκτήματα για τους χρήστες



- στο CLARIN:EL μπορεί κανείς να βρει
 - **γλωσσικό υλικό διαφόρων ειδών** (π.χ. κείμενα, προφορικό λόγο, λεξικά, γλωσσάρια ορολογίας, βίντεο κ.λπ.)
 - από ποικίλα επιστημονικά πεδία
- αλλά το σημαντικότερο:
 - για πρώτη φορά, **στην ίδια θέση** με το **γλωσσικό υλικό** διατίθενται και **γλωσσικά εργαλεία** με τα οποία μπορεί κανείς να επεξεργαστεί το γλωσσικό υλικό
 - χωρίς να χρειαστεί να εγκαταστήσει κάτι στον υπολογιστή του
 - χωρίς να ανησυχεί για τη νομιμότητα της διαδικασίας



Πλεονεκτήματα για τους παρόχους clarin:el

- οι πόροι, μέσω της πλούσιας τεκμηρίωσης (με βάση διεθνή πρότυπα)
 - αποκτούν **μοναδική ταυτότητα** και καταγράφονται στα μητρώα γλωσσικών πόρων διεθνώς,
 - με συγκεκριμένο όνομα, μοναδικό μόνιμο αναγνωριστικό (Persistent Identifier),
 - **μόνιμη σύνδεση πόρου-δημιουργού**,
 - **σαφή άδεια διάθεσης** και
 - ρητά δικαιώματα χρήσης
- τόσο οι πόροι όσο και οι δημιουργοί τους
 - αποκτούν αναγνωρισιμότητα, **δημοσιότητα** και διεθνή προβολή (μέσω της σύνδεσης με την ευρωπαϊκή υποδομή)

Ασφάλεια στη διάθεση των πόρων clarin:el



- σημαντικό πλεονέκτημα και για πάροχο και για χρήστη
- η νόμιμη χρήση των διαθέσιμων πόρων εξασφαλίζεται
 - μέσω ενός ενιαίου σχήματος αδειοδότησης που περιλαμβάνει ευρέως καθιερωμένες άδειες που συνοδεύουν κάθε πόρο ή εργαλείο της υποδομής (νομική κάλυψη), και
 - μέσω του συστήματος πιστοποίησης χρηστών, το οποίο ταυτοποιεί κάθε χρήστη και τα δικαιώματά του στην υποδομή (τεχνική κάλυψη)

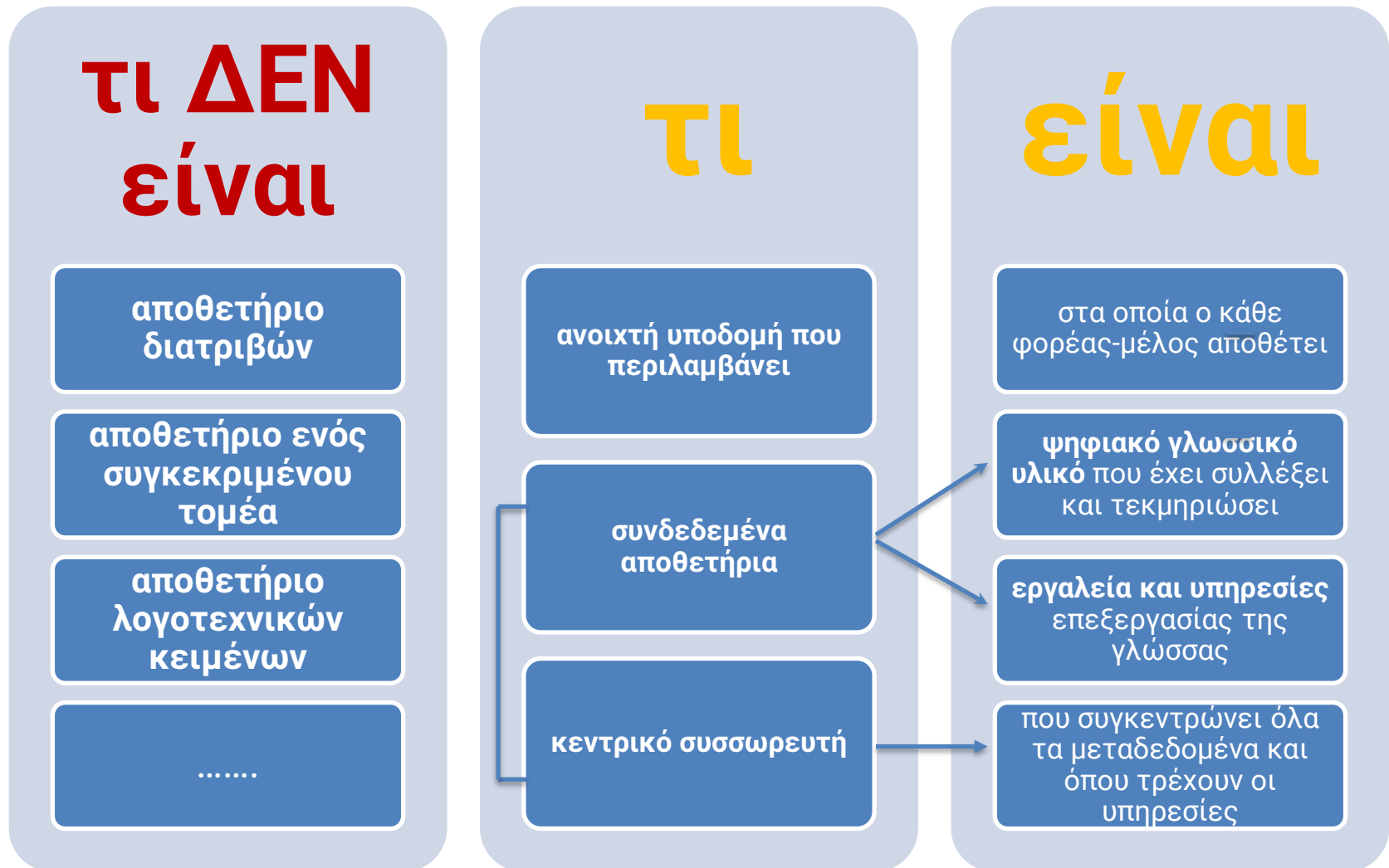
Ανοιχτότητα δεδομένων & αρχές FAIR (Fair Data Principles) (1/2)

- προώθηση της "**κουλτούρας αμοιβαιότητας**" στην ερευνητική κοινότητα
- ενημέρωση για οφέλη από τον διαμοιρασμό πόρων και υπηρεσιών
- προώθηση των ανοιχτών δεδομένων (επιστημονικών και μη) και της διάθεσής τους στην ερευνητική κοινότητα και στο ευρύ κοινό
- προγραμματισμός σχετικών εκπαιδευτικών και ενημερωτικών δράσεων

Ανοιχτότητα δεδομένων & αρχές FAIR (Fair Data Principles) (2/2)

- **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

Τελικά, τι είναι η υποδομή CLARIN:EL;





www.clarin.gr
<http://inventory.clarin.gr>



info@clarin.gr



[clarin.gr](https://www.facebook.com/clarin.gr)



[@CLARIN_el](https://twitter.com/CLARIN_el)



[CLARIN EL](https://www.youtube.com/CLARIN_EL)