
CLARIN

Δημοσίευση 1

CLARIN Technical Team

26 Ιουλίου 2023

1	Σχετικά με το CLARIN:EL	3
2	Πώς να χρησιμοποιήσετε αυτό το εγχειρίδιο	5
3	Επικοινωνία & Βοήθεια	7
4	Εγγραφή/Σύνδεση	9
4.1	Register (Εγγραφή)	9
4.2	Sign in (Σύνδεση)	12
5	Χρήστες	15
5.1	Εγγεγραμμένοι χρήστες έναντι μη εγγεγραμμένων	15
5.2	Τύποι εγγεγραμμένων χρηστών	15
6	Τι μπορώ να βρω στον κεντρικό κατάλογο;	17
7	Πότε είναι προσβάσιμο το περιεχόμενο ενός πόρου;	19
8	Αρχική σελίδα κεντρικού καταλόγου	21
9	Μενού για εγγεγραμμένους χρήστες	23
9.1	Dashboard (Πίνακας ελέγχου)	24
9.2	Help (Βοήθεια)	36
9.3	Your name (Το όνομά σας)	36
9.4	Exit (Έξοδος)	36
10	Περιήγηση	37
11	Αναζήτηση	41
12	Θέαση & Καταφόρτωση	45
12.1	Το ανώτερο τμήμα	45
12.2	Το κατώτερο τμήμα	47
13	Επεξεργασία	53
13.1	1. Ξεκινώντας από τα σώματα κειμένων	54
13.2	2. Ξεκινώντας από τη λειτουργία	57

14 Ο κύκλος ζωής του πόρου	65
15 Σημαντικές πληροφορίες	67
16 Πώς να δημιουργήσετε έναν πόρο χρησιμοποιώντας το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων	69
16.1 Βήμα 1: Δώστε ένα όνομα στον πόρο σας	70
16.2 Βήμα 2: Μεταφορτώστε τα αρχεία περιεχομένου του πόρου	71
16.3 Βήμα 3: Συμπληρώστε τα υποχρεωτικά μεταδεδομένα	77
16.4 Βήμα 4: Δείτε πως φαίνεται ο πόρος που δημιουργήσατε	79
17 Πώς να δημιουργήσετε έναν πόρο μεταφορτώνοντας ένα αρχείο μεταδεδομένων XML	81
18 Πώς να προετοιμάσετε τα δεδομένα πριν τη μεταφόρτωση	85
18.1 I. Τα δεδομένα ως περιεχόμενο του πόρου	85
18.2 II. Τα δεδομένα ως τροφοδοσία μίας υπηρεσίας	90
19 Συνιστώμενα μορφότυπα (formats) αρχείων	93
19.1 Συστάσεις για την επιλογή μορφότυπων αρχείων που διασφαλίζουν μακροπρόθεσμα την προ- σβασιμότητα και διαλειτουργικότητα των δεδομένων	93
20 Τι είναι τα μεταδεδομένα και γιατί είναι τόσο σημαντικά;	97
21 Οι αρχές FAIR	101
21.1 Ευρεσιμότητα (Findability)	101
21.2 Προσβασιμότητα (Accessibility)	102
21.3 Διαλειτουργικότητα (Interoperability)	103
21.4 Επαναχρησιμοποίηση (Reuse)	104
22 Υποχρεωτικά μεταδεδομένα	105
22.1 Κοινά υποχρεωτικά μεταδεδομένα	105
22.2 Υποχρεωτικά μεταδεδομένα ανά τύπο πόρου	107
23 Γενικές οδηγίες για τα μεταδεδομένα	111
23.1 Γλώσσα	111
23.2 Συνέπεια	112
23.3 Πληρότητα	112
23.4 Επεξεργασία	112
23.5 Δημιουργία νέας έκδοσης (versioning)	113
24 Συγκεκριμένες οδηγίες για τα υποχρεωτικά μεταδεδομένα	115
24.1 1. <code>resourceName</code>	115
24.2 2. <code>description</code>	116
24.3 3. <code>version</code>	116
24.4 4. <code>keyword</code>	117
24.5 5. <code>additionalInformation</code>	118
24.6 6. <code>distribution</code> και σχετικά μεταδεδομένα	118
24.7 7. <code>licenceTerms</code> και σχετικά μεταδεδομένα	120
24.8 8. <code>data</code>	121
24.9 9. <code>personalData, sensitiveData & anonymized</code>	121
24.10 10. <code>Subclass</code> και σχετικά μεταδεδομένα	123
24.11 11. <code>encodingLevel</code>	126
24.12 12. <code>function</code>	127
24.13 13. <code>inputContentResource</code>	127

25 Παραδείγματα μεταδεδομένων	131
25.1 resourceName	131
25.2 resourceCreator	132
25.3 isPartOf	132
25.4 annotationType	133
25.5 multilingualityType	134
25.6 isDocumentedBy	134
25.7 fundingProject	135
25.8 inputContentResource	136
25.9 outputResource	136
25.10 attributionText	137
26 Πλήρεις περιγραφές μεταδεδομένων σε XML	139
26.1 1. Σώματα κειμένων	139
26.2 2. Λεξικοί/Εννοιολογικοί πόροι	158
26.3 3. Εργαλεία/Υπηρεσίες	172
26.4 4. Γλωσσικές περιγραφές	181
27 Το σχήμα μεταδεδομένων	187
28 Τι ενέργειες μπορώ να κάνω πάνω στους πόρους;	189
28.1 I. Ταξινόμηση ενεργειών βάσει του σταδίου στο οποίο βρίσκεται ο πόρος & του τύπου χρήστη . .	189
28.2 II. Παρουσίαση και ανάλυση των ενεργειών	192
29 Πληροφορίες νομικού χαρακτήρα	205
30 Δημοσιεύσεις	207
31 Indices and tables	209

Welcome to the user manual of CLARIN:EL!

Cite this version:

Pouli Kanella, Bakagianni Juli, Galanis Dimitris, Labropoulou Penny, Tsiouli Iro, Gavriilidou Maria. 2023. The CLARIN:EL User Manual, v.1.0.

Βιβλιογραφική αναφορά:

Πουλή Κανέλλα, Μπακαγιάννη Γιούλη, Γαλάνης Δημήτρης, Λαμπροπούλου Πένυ, Τσιούλη Ηρώ, Γαβριηλίδου Μαρία. 2023. Εγχειρίδιο χρήσης της υποδομής CLARIN:EL, v.1.0.

Σχετικά με το CLARIN:EL

Το **CLARIN:EL** είναι μια Ερευνητική Υποδομή για *Γλωσσικούς πόρους & Τεχνολογίες* που αποτελεί το ελληνικό τμήμα της ευρωπαϊκής υποδομής **CLARIN ERIC**. Παρέχει ένα πλήθος **πόρων που σχετίζονται με τη γλωσσική τεχνολογία** για και από τις **Κοινωνικές και Ανθρωπιστικές Επιστήμες** (και όχι μόνο), εστιάζοντας κυρίως αλλά όχι αποκλειστικά στα **ελληνικά**. Η ερευνητική υποδομή CLARIN:EL λειτουργεί ως ένα κατακευματισμένο δίκτυο αποθετηρίων που αποτελείται από:

- τα Ιδρυματικά Αποθετήρια (που δημιουργούνται για κάθε Οργανισμό που συμμετέχει στο δίκτυο CLARIN:EL) και
- το Αποθετήριο Φιλοξενούμενων Πόρων (ΑΦΠ), το οποίο υποστηρίζεται από το Ερευνητικό Κέντρο ΑΘΗΝΑ.

Ακολουθήστε τους συνδέσμους¹ για να δείτε τα περιεχόμενα του κάθε αποθετηρίου:

- Ερευνητικό Κέντρο Αθηνά,
- Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,
- Οικονομικό Πανεπιστήμιο Αθηνών,
- Πανεπιστήμιο Αιγαίου,
- Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών,
- Κέντρο Ελληνικής Γλώσσας,
- Αποθετήριο Φιλοξενούμενων Πόρων,
- Πανεπιστήμιο Κρήτης,
- Ionian University,
- Εθνικό Κέντρο Κοινωνικών Ερευνών (ΕΚΚΕ),
- Εθνικό Κέντρο Επιστημονικών Ερευνών «Δημόκριτος» (ΕΚΕΦΕ),
- Πάντειο Πανεπιστήμιο, και
- Πανεπιστήμιο Δυτικής Αττικής.

¹ Η απουσία υπερσυνδέσμου σημαίνει ότι ο εν λόγω οργανισμός δεν έχει ακόμη πόρους στο αποθετήριό του.

Στον κεντρικό κατάλογο θα βρείτε όλους τους γλωσσικούς πόρους και εργαλεία/υπηρεσίες που έχουν δημοσιευτεί στα αποθετήρια αυτά.

Πρακτική συμβουλή: Δείτε [εδώ](#) πώς συνδέονται οι χρήστες με τα αποθετήριά τους και την υποδομή.

Πώς να χρησιμοποιήσετε αυτό το εγχειρίδιο

Αυτό το εγχειρίδιο¹ έχει ως στόχο να σας βοηθήσει να εξερευνήσετε ή/και να χρησιμοποιήσετε την υποδομή **CLARIN:EL** για να διαθέσετε τους πόρους σας στην ευρύτερη ερευνητική κοινότητα των **Ανθρωπιστικών και Κοινωνικών Επιστημών** (και όχι μόνο). Δεν χρειάζεται να το διαβάσετε σειριακά (αν και μπορείτε). Είναι γραμμένο με τέτοιο τρόπο ώστε να σας βοηθήσει να βρείτε συγκεκριμένες πληροφορίες ανάλογα με τις ανάγκες σας. Υπάρχουν κεφάλαια με γενικές πληροφορίες σχετικά με *βασικές έννοιες*, άλλα που περιγράφουν τη *διαδικασία*² μέσω της οποίας ένας πόρος αποκτά ζωή και κεφάλαια που θα σας βοηθήσουν συγκεκριμένα:

- να *περιηγηθείτε* και να *αναζητήσετε* πόρους στον κεντρικό κατάλογο, ώστε να βρείτε αυτούς που θέλετε να *καταφορτώσετε* ή/και να *επεξεργαστείτε*,
- να δημιουργήσετε πόρους μέσω του *περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων* ή με *μεταφόρτωση αρχείου μεταδεδομένων XML*, και
- να κάνετε *ενέργειες πάνω σε πόρους* ανάλογα με το ρόλο σας.

Αν ψάχνετε κάτι συγκεκριμένο, παρακαλούμε, χρησιμοποιήστε το πλαίσιο αναζήτησης στην επάνω αριστερή πλευρά της σελίδας, κάτω από το λογότυπο του CLARIN:EL.

¹ Η τρέχουσα έκδοση του εγχειριδίου συνοδεύει την τρίτη επίσημη έκδοση της υποδομής CLARIN:EL, που μπήκε σε λειτουργία στις 31 Μαΐου 2021. Καθώς προστίθενται συνεχώς νέες λειτουργίες στην υποδομή, το παρόν εγχειρίδιο επικαιροποιείται ακολουθώντας την εξέλιξη της πλατφόρμας CLARIN:EL.

² Πριν ξεκινήσετε, ανατρέξτε στην ενότητα *Ο κύκλος ζωής του πόρου* για να μάθετε τι πρέπει να κάνει κάθε τύπος χρήστη κατά τη διάρκεια της διαδικασίας. Προϋπόθεση για την ανάληψη οποιουδήποτε ρόλου είναι η *εγγραφή*.

ΚΕΦΑΛΑΙΟ 3

Επικοινωνία & Βοήθεια

Υπάρχουν τρία γραφεία υποστήριξης για να απαντήσουν σε οποιαδήποτε ερώτημά σας: το Γραφείο τεχνικής υποστήριξης, το Γραφείο νομικής υποστήριξης, και το Γραφείο υποστήριξης τεκμηρίωσης.

Σύνδεσμοι για τα γραφεία υποστήριξης (αλλά και τις συχνές ερωτήσεις και τη βιβλιογραφία) υπάρχουν στο κάτω μέρος σε κάθε σελίδα της υποδομής.

The screenshot shows the CLARIN:EL portal homepage. At the top left is the 'clarin:el' logo with a 'BETA' badge. To the right are links for 'CLARIN:EL portal', 'Help', and 'Sign in'. The main section is titled 'Join the CLARIN:EL community!' and includes a paragraph about sharing language data and accessing resources, with a 'Learn more' button. Below this is a dark blue footer bar containing social media icons (Facebook, Twitter, YouTube), copyright information '© CLARIN:EL 2021 Terms of service | Privacy Policy', and a 'Need Help ?' section with links to 'CLARIN:EL Helpdesks', 'FAQs', and 'Bibliography'. At the bottom, there is a row of logos for partner organizations: 'απολλωνις', 'CLARIN' (Common Language Resources and Technology Infrastructure), 'ΑΘΗΝΑ' (Εθνικα & Κεντρικα Πανεπιστημια Θεσσαλονικης), a logo with the text 'ΑΕΡΟΠΟΡΕΥΣΗ', 'grnet' (National Research and Education Network), 'ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ' (Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών), 'ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ', and a right-pointing arrow. Below these are logos for the 'Ευρωπαϊκή Ένωση' (European Union), 'ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ' (ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ & ΕΚΣΥΓΧΡΟΝΙΣΜΟΣ), 'ΕΣΠΑ 2014-2020', and a 'Co-funded by Greece and the European Union' text.

Εγγραφή/Σύνδεση

Αυτό το κεφάλαιο παρέχει πληροφορίες στους χρήστες:

1. που **ήδη έχουν** ακαδημαϊκό λογαριασμό και μπορούν να τον χρησιμοποιήσουν για να *συνδεθούν*,
2. που **δεν έχουν** ακαδημαϊκό λογαριασμό και **πρέπει πρώτα** να *εγγραφούν* για να δημιουργήσουν έναν προσωπικό λογαριασμό, τον οποίο θα χρησιμοποιήσουν αργότερα για να *συνδεθούν*.

4.1 Register (Εγγραφή)

Προσοχή: Παραλείψτε αυτή την ενότητα, αν έχετε ήδη ακαδημαϊκό λογαριασμό, τον οποίο μπορείτε να χρησιμοποιήσετε για να *συνδεθείτε*.

Για να εγγραφείτε στο CLARIN:EL, ακολουθήστε τα επόμενα βήματα:

- Κάντε κλικ στο κουμπί **Sign in** (Σύνδεση) στο επάνω δεξί μέρος της σελίδας.



CLARIN:EL portal >

Help

Sign in

clarin:el

Inventory of language resources and services

Browse

or

Search for services, tools, datasets...

Search



- Στο επόμενο παράθυρο επιλέξτε **Register with personal account** (Εγγραφή με προσωπικό λογαριασμό).

CLARIN:EL

Log In

with academic account

Greek academic login

CLARIN Service Provider Federation login

with personal account

Email as Username

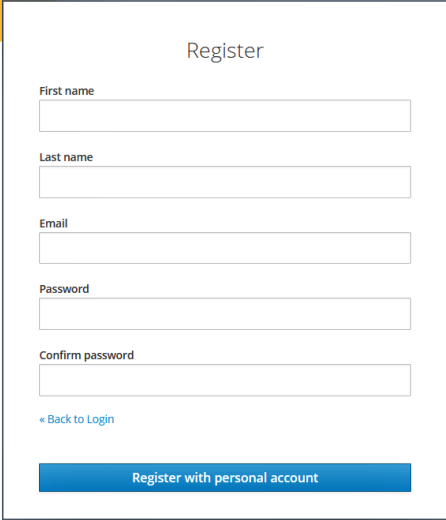
Password

[Forgot Password?](#)

Log In

New user? [Register with personal account](#)

- Στη συνέχεια, συμπληρώστε όλα τα απαιτούμενα στοιχεία στη φόρμα που εμφανίζεται και κάντε κλικ στο **Register with personal account** (Εγγραφή με προσωπικό λογαριασμό).



CLARIN:EL

Register

First name

Last name

Email

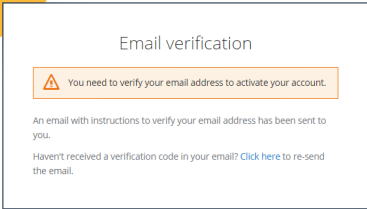
Password

Confirm password

[« Back to Login](#)


[Register with personal account](#)

- Θα λάβετε ένα email με έναν σύνδεσμο προκειμένου να επιβεβαιώσετε την ηλεκτρονική διεύθυνση που δηλώσατε αλλά και για να δηλώσετε ότι συμφωνείτε με τους Όρους Χρήσης και την Πολιτική Απορρήτου του CLARIN:EL.



CLARIN:EL

Email verification

 You need to verify your email address to activate your account.

An email with instructions to verify your email address has been sent to you.

Haven't received a verification code in your email? [Click here](#) to re-send the email.

- Μόλις γίνει η επιβεβαίωση, ενεργοποιείται ο λογαριασμός σας. Κάθε φορά από εδώ και στο εξής που θέλετε να χρησιμοποιήσετε την υποδομή, **συνδέστε** με τον προσωπικό σας λογαριασμό κάνοντας κλικ πάνω στο κουμπί **Sign in**.

4.2 Sign in (Σύνδεση)

Μπορείτε να συνδεθείτε χρησιμοποιώντας είτε τον **προσωπικό** είτε τον **ακαδημαϊκό** σας λογαριασμό (ελληνικό ή ευρωπαϊκό) για να μεταφερθείτε στο αποθετήριο του φορέα σας. Εάν δεν είστε μέλος κάποιου οργανισμού/ακαδημαϊκού ιδρύματος, θα οδηγηθείτε αυτόματα στο **Αποθετήριο φιλοξενούμενων πόρων**.

Για να χρησιμοποιήσετε τον ακαδημαϊκό σας λογαριασμό κάντε κλικ στο **Greek academic login** (Ελληνική ακαδημαϊκή σύνδεση) που βρίσκεται κάτω από το **with academic account**. Θα μεταφερθείτε σε μια νέα σελίδα όπου θα πρέπει να πληκτρολογήσετε στο πλαίσιο το όνομα του οργανισμού/ακαδημαϊκού ιδρύματος του οποίου είστε μέλος και στη συνέχεια να επιλέξετε **Confirm** (επιβεβαίωση).

Στο νέο παράθυρο που ανοίγει, χρησιμοποιήστε τα ακαδημαϊκά σας διαπιστευτήρια και τέλος κάντε κλικ στο **Login** (σύνδεση).



Username (user@athenarc.gr ή
user@*.athena-innovation.gr)

> Forgot your password?

Password

> Need Help?

☐ Don't Remember Login

☐ Clear prior granting of
permission for release of your
information to this service.

Login

Έχετε συνδεθεί!

Χρήστες

Οι χρήστες έχουν ποικίλα δικαιώματα ανάλογα με το αν είναι εγγεγραμμένοι ή όχι. Οι *μη εγγεγραμμένοι* χρήστες μπορούν να πλοηγηθούν στην υποδομή με **περιορισμένα** δικαιώματα. Τα **προηγμένα** δικαιώματα που σχετίζονται με την *δημιουργία*, τη *διαχείριση* και την *επεξεργασία* των πόρων δίνονται μόνο σε *εγγεγραμμένους* χρήστες σύμφωνα με την πολιτική της υποδομής.

5.1 Εγγεγραμμένοι χρήστες έναντι μη εγγεγραμμένων

Ο κεντρικός κατάλογος περιλαμβάνει δημοσιευμένους γλωσσικούς πόρους και εργαλεία/υπηρεσίες από όλα τα αποθετήρια και είναι προσβάσιμος με ή χωρίς εγγραφή/σύνδεση. Όσον αφορά στην **κατανάλωση** των πόρων, υπάρχει μόνο ένα πρόσθετο δικαίωμα που έχει ένας εγγεγραμμένος χρήστης, το οποίο είναι η χρήση εργαλείων και υπηρεσιών, όπως φαίνεται στον παρακάτω πίνακα.

Πρακτική συμβουλή: Κάντε κλικ σε κάθε μια από τις χρήσεις της υποδομής για να μάθετε περισσότερα!

5.2 Τύποι εγγεγραμμένων χρηστών

Εφόσον έχετε εγγραφεί, θα πρέπει να *συνδεθείτε* προκειμένου να οδηγηθείτε στο αποθετήριο σας. Αυτό θα σας επιτρέψει να συμμετάσχετε στον *κύκλο ζωής του πόρου* αναλαμβάνοντας έναν από τους ακόλουθους ρόλους.

1. **Επιμελητής (Curator):** είναι υπεύθυνος για τη δημιουργία πόρων, τη μεταφόρτωση των αρχείων περιεχομένου, καθώς επίσης για τη διαχείρισή των πόρων (επεξεργασία, ενημέρωση κλπ.) και την υποβολή τους προς δημοσίευση.

Σημείωση: Με την *σύνδεση*, αποκτάτε **αυτόματα** την ιδιότητα του **επιμελητή** στο αποθετήριο σας.

2. **Επικυρωτής (Validator):** του ανατίθενται πόροι από τον υπεύθυνο του αποθετηρίου προκειμένου να ελέγξει αν τα μεταδεδομένα που περιγράφονται (και τα αρχεία περιεχομένου που έχουν μεταφορτωθεί) είναι σε αντιστοιχία. Αν όχι, ο πόρος επιστρέφει στον επιμελητή για να κάνει διορθώσεις σύμφωνα με τα σχόλια του επικυρωτή.

Σημείωση: Δείτε [εδώ](#) πως ένας χρήστης γίνεται επικυρωτής.

3. **Υπεύθυνος αποθετηρίου (Supervisor):** έχει τον τελευταίο λόγο πριν δημοσιευθεί ένας πόρος και είναι επίσης ο μόνος που μπορεί να άρει την δημοσίευση, αν χρειαστεί.

Σημείωση: Δείτε [εδώ](#) πως ένας χρήστης γίνεται υπεύθυνος αποθετηρίου.

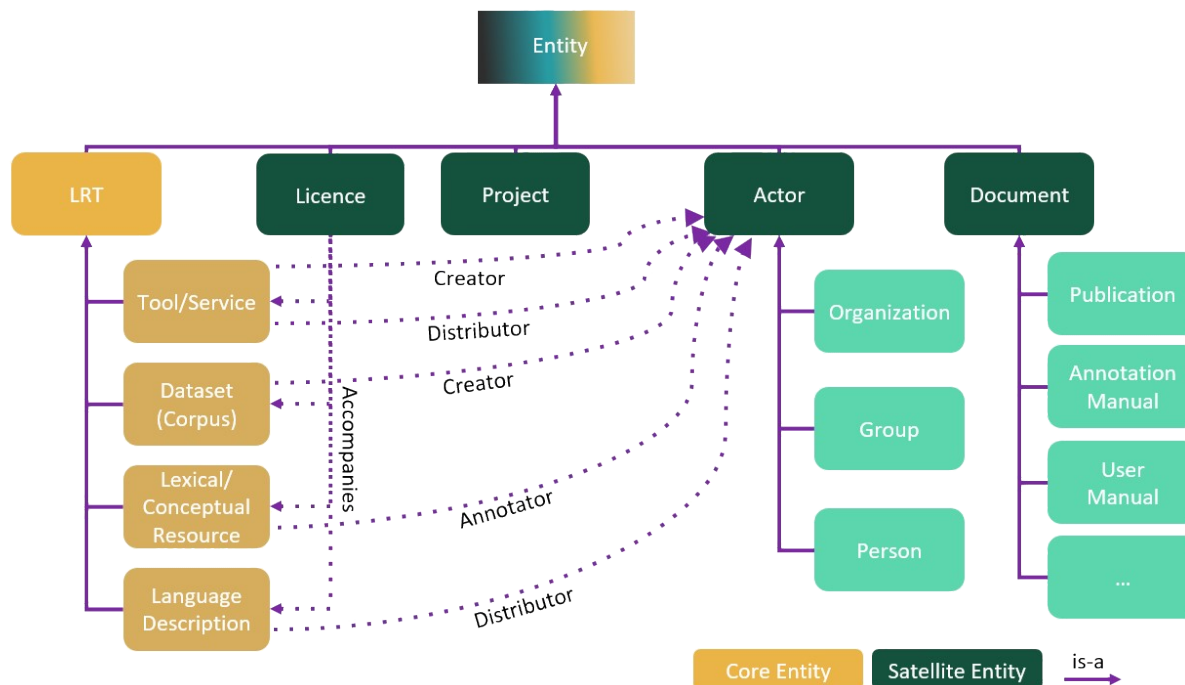
Κάθε ένας από αυτούς τους χρήστες διαχειρίζεται πόρους στους οποίους μπορεί να κάνει κάποιες [ενέργειες](#). Οι ενέργειες αυτές είναι διαθέσιμες από το [dashboard](#) (πίνακας ελέγχου) ή τη [σελίδα θέαςής του πόρου](#).

Τι μπορώ να βρω στον κεντρικό κατάλογο;

Η υποδομή **CLARIN:EL** περιλαμβάνει γλωσσικούς πόρους και τεχνολογίες (Language Resources and Technologies/LRTs), οι οποίοι μπορούν να ταξινομηθούν περαιτέρω ανάλογα με το περιεχόμενό τους σε

- **corpora** (σώματα κειμένων): σύνολα δεδομένων όπως συλλογές κειμένων, ηχητικών απομαγνητοφωνήσεων, ηχογραφήσεων και βιντεοσκοπήσεων κλπ. (δείτε [εδώ](#) ποια είναι τα σώματα κειμένων που μπορούν να χρησιμοποιηθούν για επεξεργασία),
- **lexical/conceptual resources** (λεξικοί/εννοιολογικοί πόροι): υπολογιστικά λεξικά, καταλόγοι λέξεων, οντολογίες, λίστες όρων κλπ.,
- **tools & services** (εργαλεία & υπηρεσίες): κάθε είδος λογισμικού που χρησιμοποιείται για επεξεργασία φυσικής γλώσσας (δείτε [εδώ](#) ποιες είναι οι λειτουργίες επεξεργασίας φυσικής γλώσσας που πραγματοποιούνται στην υποδομή), και
- μοντέλα & υπολογιστικές γραμματικές, που αναφέρονται συνολικά ως **language descriptions** (γλωσσικές περιγραφές).

Στην ακόλουθη εικόνα φαίνεται η ταξινομία των πόρων σε σχέση με άλλες οντότητες, όπως είναι ο **actor** (δράστης), δηλαδή η οντότητα (ένα άτομο, μια ομάδα ατόμων ή ένας οργανισμός) που έχει δημιουργήσει, τροποποιήσει ή επισημειώσει κάποιον πόρο.



Συνήθως ένας πόρος αποτελείται από μια περιγραφή (την εγγραφή μεταδεδομένων) και αρχεία περιεχομένου (π.χ. το σύνολο δεδομένων για ένα σώμα κειμένων, το λογισμικό για ένα εργαλείο κλπ.) Η **περιγραφή** είναι απαραίτητη προϋπόθεση για την εμφάνιση του πόρου στον κεντρικό κατάλογο ενώ τα **αρχεία περιεχομένου** όχι. Ως εκ τούτου, οι πόροι που μπορούν να εμφανιστούν στον κεντρικό κατάλογο εμπίπτουν σε μία από τις ακόλουθες κατηγορίες:

1. Περιγραφές πόρων **με** αρχεία περιεχομένου, τα οποία είναι διαθέσιμα
 - 1.1 μέσω του CLARIN:EL, ή
 - 1.2 μέσω εξωτερικών συνδέσμων στους οποίους ο χρήστης ανακατευθύνεται μέσα από την υποδομή (τέτοια παραδείγματα είναι εξωτερικές τοποθεσίες λήψης ή πρόσβασης αρχείων).
2. Περιγραφές πόρων **χωρίς** αρχεία περιεχομένου διότι οι πόροι
 - 2.1 έχουν πληροφοριακό σκοπό, **for info**, (τα αρχεία μπορεί να μεταφορτωθούν αργότερα), ή
 - 2.2 είναι **metaresources** (πόροι, δηλαδή, που δεν έχουν καθόλου αρχεία περιεχομένου, όπως βιβλιογραφίες, πρακτικά συνεδρίων κλπ.).

Δείτε [εδώ](#) πότε τα αρχεία περιεχομένου ενός πόρου είναι προσβάσιμα.

Πρακτική συμβουλή: Δείτε [εδώ](#) σημαντικές πληροφορίες που αφορούν στην δημιουργία και τον διαμοιρασμό ενός πόρου μέσω της υποδομής CLARIN:EL.

Πότε είναι προσβάσιμο το περιεχόμενο ενός πόρου;

Προσοχή: Αυτή η ενότητα αναφέρεται σε περιγραφές πόρων που **συνοδεύονται** από αρχεία περιεχομένου. Για τους διαφορετικούς τύπους πόρων που βρίσκονται στον κεντρικό κατάλογο του CLARIN:EL δείτε [εδώ](#).

Κατά την πλοήγησή σας στην υποδομή θα δείτε περιγραφές πόρων πολλές από τις οποίες, όπως εξηγήθηκε προηγουμένως, συνοδεύονται από αρχεία περιεχομένου. Για να είναι προσβάσιμα από τον χρήστη τα αρχεία περιεχομένου ενός πόρου πρέπει να πληρούνται δύο προϋποθέσεις:

1. ο πόρος πρέπει να διατίθεται με **άδεια ανοικτής πρόσβασης** (δείτε [εδώ](#) το Προτεινόμενο σχήμα αδειοδότησης Γλωσσικών Πόρων), και
2. τα αρχεία περιεχομένου του πόρου πρέπει να **έχουν μεταφορτωθεί** ή να **έχουν αποθηκευτεί** σε κάποιο σημείο πρόσβασης.

Αυτό ισχύει τόσο για τους πόρους που **διατίθενται** μέσω της υποδομής όσο και για αυτούς για τους οποίους **γίνεται μέσω της υποδομής ανακατεύθυνση σε εξωτερικούς συνδέσμους (URL)**. Στην δεύτερη περίπτωση, πρέπει να ικανοποιούνται πρόσθετα κριτήρια:

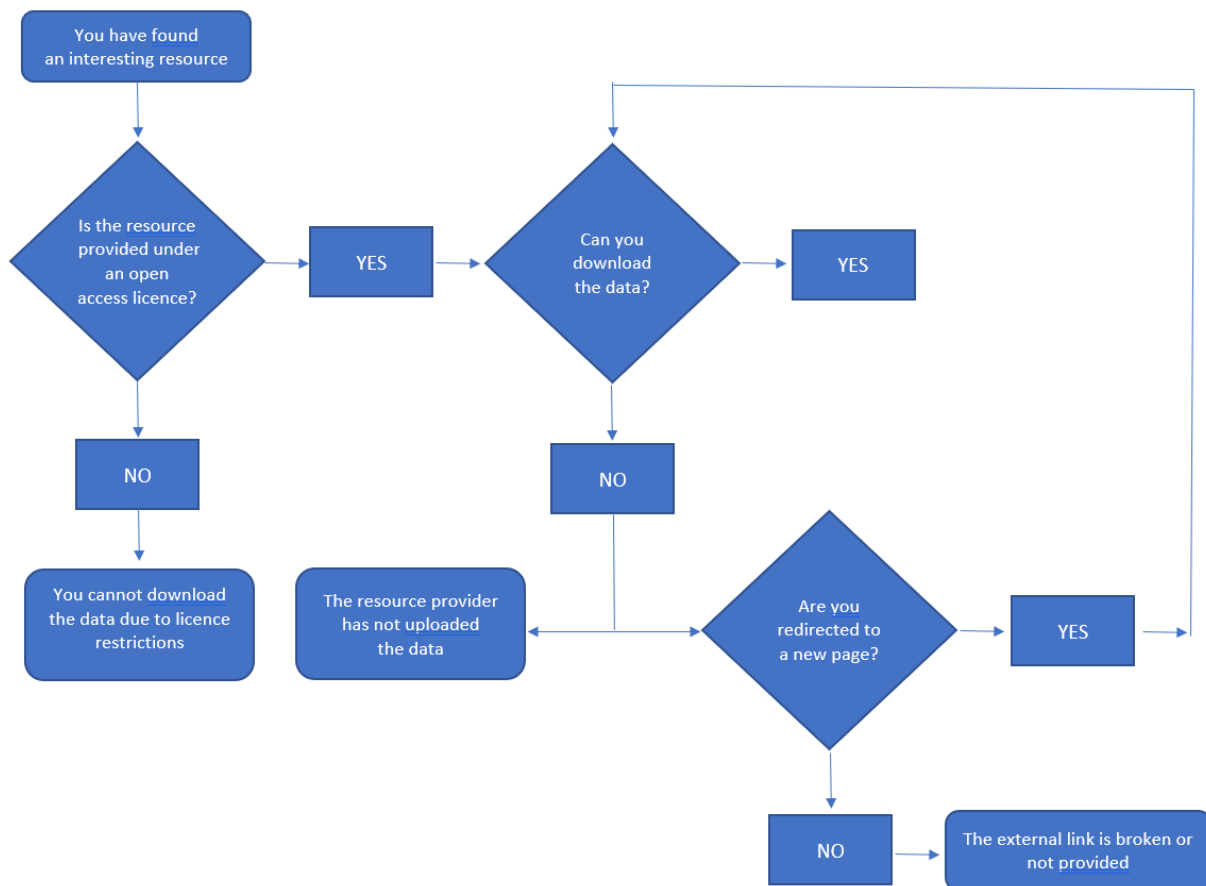
- the link must be provided in the appropriate metadata field in the metadata editor¹/xml file,
- ο σύνδεσμος πρέπει να λειτουργεί (να μην είναι σπασμένος), και
- το περιεχόμενο του συνδέσμου πρέπει να είναι καλά συντηρημένο.

Ανεξάρτητα από το που βρίσκεται το σημείο πρόσβασης (στην υποδομή ή έξω από αυτήν) των αρχείων περιεχομένου, ο ακόλουθος πίνακας δείχνει τους πιθανούς συνδυασμούς ενεργειών που επιτρέπουν ή όχι τη λήψη των αρχείων από τον χρήστη.

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

Διατίθεται ο πόρος με άδεια ανοικτής πρόσβασης;	Έχουν μεταφορτωθεί τα αρχεία περιεχομένου (σύνολα δεδομένων, εργαλεία, λεξικά, κτλ.);	Μπορούν τα αρχεία να καταφορτωθούν;
Ναι	Ναι	Ναι
Ναι	Όχι	Όχι
Όχι	Ναι	Όχι
Όχι	Όχι	Όχι

Το ακόλουθο διάγραμμα ροής απεικονίζει το περιεχόμενο του πίνακα:



Αρχική σελίδα κεντρικού καταλόγου

Η αρχική σελίδα της υποδομής CLARIN:EL προσφέρει στον χρήστη που **δεν έχει ακόμη συνδεθεί ή εγγραφεί** διάφορες επιλογές.



Στην επάνω δεξιά πλευρά της σελίδας, υπάρχουν τρία κουμπιά:

Το *CLARIN:EL portal*: οδηγεί στη **διαδικτυακή πύλη** του CLARIN:EL, όπου μπορείτε να βρείτε πληροφορίες σχετικά με το δίκτυο, νέα και ανακοινώσεις, σχετικές παρουσιάσεις και βιβλιογραφία,

Το *Help*: οδηγεί στις οδηγίες χρήσης της υποδομής (το παρόν εγχειρίδιο), όπου μπορείτε να μάθετε πώς να πλοηγήσετε, να δημιουργείτε και να διαχειρίζεστε πόρους κλπ.,

Το *Sign in*: οδηγεί στην σελίδα όπου μπορείτε να κάνετε **εγγραφή** ή **σύνδεση**.

Στη συνέχεια, υπάρχει ένα πλαίσιο αναζήτησης με την επιλογή **browse** (περιήγηση) στα αριστερά και την επιλογή **search** (αναζήτηση) στα δεξιά. Πρόκειται για σημεία εισόδου στον κεντρικό κατάλογο. Μπορείτε να **περιηγηθείτε** στο σύνολο των πόρων του κεντρικού καταλόγου ή να χρησιμοποιήσετε **φίλτρα και λέξεις-κλειδιά** για να ανακτήσετε ένα υποσύνολο των πόρων που ικανοποιούν τα κριτήρια που θέσατε.

Στη μέση της σελίδας μπορείτε να δείτε δύο στήλες όπου παρουσιάζονται συγκεκριμένες ομάδες πόρων.

Language Technology tools and services

Language processing tools, in the form of web applications or downloadable tools.

- > **CLARIN:EL Workflow registry**
Language processing web services integrated in the CLARIN:EL infrastructure, that operate at various levels of analysis.
- > **Downloadable tools**
Language processing tools which are downloaded to be run locally.
- > **Additional online services**
Language processing services available on the web or forthcoming services currently under construction.

Language resources

Corpora (i.e. structured collections of data such as text documents, audio transcripts, audio and video recordings); lexical and conceptual resources; treebanks etc.

- > **Processable corpora**
Corpora that have the appropriate format in order to be processed by the CLARIN:EL Workflows.
- > **Downloadable resources**
Resources whose license permits downloading.
- > **Metaresources**
Ancillary resources that provide information on various aspects of language and linguistics (such as bibliographical information, studies etc.).

Η πρώτη στήλη περιέχει υποσύνολα εργαλείων ή/και υπηρεσιών ομαδοποιημένα ανάλογα με το αν:

- παρέχονται ως υπηρεσίες στην υποδομή,
- μπορούν να καταφορτωθούν, ή
- είναι προσβάσιμα διαδικτυακά μέσω ανακατεύθυνσης σε εξωτερικούς συνδέσμους.

Με τον ίδιο τρόπο, τα σύνολα δεδομένων ομαδοποιούνται στην επόμενη στήλη ως αυτά που μπορούν να:

- υποστούν επεξεργασία,
- καταφορτωθούν, ή
- έχουν μόνο πληροφοριακό σκοπό.

Στο κάτω μέρος της σελίδας υπάρχει ένας [σύνδεσμος](#) για τους χρήστες που επιθυμούν να μάθουν περισσότερα για την κοινότητα CLARIN:EL και, εφόσον το επιθυμούν, να γίνουν μέλη της.

Join the CLARIN:EL community!

By joining the CLARIN:EL community, you can share your own language data and language processing technologies and have access to resources and technologies from other repositories at the National and European level.

[Learn more](#)



© CLARIN:EL 2021 [Terms of service](#) | [Privacy Policy](#)

Need Help ?

[CLARIN:EL Helpdesks](#)
[FAQs](#)
[Bibliography](#)

Μενού για εγγεγραμμένους χρήστες

Εφόσον έχετε συνδεθεί, βλέπετε πέντε επιλογές στο μενού:

Το *CLARIN:EL portal*: οδηγεί στη [διαδικτυακή πύλη](#) του **CLARIN:EL**, όπου μπορείτε να βρείτε πληροφορίες σχετικά με το δίκτυο, νέα και ανακοινώσεις, σχετικές παρουσιάσεις και βιβλιογραφία,

Το *Dashboard*: οδηγεί στον προσωπικό σας πίνακα ελέγχου, που χρησιμεύει ως σημείο πρόσβασης για το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων¹, τους πόρους που έχετε δημιουργήσει, τις εργασίες σας και τις εργασίες επεξεργασίας που έχετε ζητήσει.

Το *Help*: οδηγεί στις οδηγίες χρήσης της υποδομής (το παρόν εγχειρίδιο), όπου μπορείτε να μάθετε πώς να πλοηγείστε, να δημιουργείτε και να διαχειρίζεστε πόρους κλπ.,

Το *Όνομα σας*: οδηγεί στο προφίλ σας, το οποίο μπορείτε να επεξεργαστείτε.

Το *Exit*: οδηγεί σε αποσύνδεση από την υποδομή.

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.



9.1 Dashboard (Πίνακας ελέγχου)

Το Dashboard λειτουργεί τόσο ως σελίδα επισκόπησης των δραστηριοτήτων σας (μπορείτε να βρείτε πληροφορίες σχετικά με τους πόρους σας, τις εργασίες, τις εργασίες επεξεργασίας που έχετε ζητήσει) όσο και ως σημείο εισόδου στο *περιβάλλον τεκμηρίωσης και ανάρτησης πόρων* και στο *μητρώο ροών εργασίας* για την επεξεργασία πόρων. Όπως φαίνεται παρακάτω, ο πίνακας ελέγχου χωρίζεται σε εννέα διαφορετικά τμήματα. Τα τμήματα 2 και 7-9 είναι ελαφρώς διαφορετικά ανάλογα με τον ρόλο του κάθε χρήστη.

Supervisor H.
supervisor_hosted@gmail.com

2

1 Welcome to your dashboard !

Here you can:

- view and update your profile
- view your resources and your tasks
- access creation forms
- upload resources
- access the workflow registry
- browse your recent activity

6 Processing tasks

Input
A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)

Output
A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) (from MOSES plain text format to Part of Speech XML, XML)

Submission date
19 May 2021, 11:31

Finished

Filename
archive.zip

3 Total resources
Number of resources you have created.
8

4 Upload resources
Upload single or multiple resources in XML format.

5 Workflow registry
Select a workflow to process a corpus.

7 Recent Activity

8 **9**

My Resources		Validation tasks		My repository	
Date created	Title				
19 May 2021	Single DE (raw corpus)				
19 May 2021	Single DE				
19 May 2021	Threeparia				
06 May 2021	Single text EL (txt)				

View all processing tasks

View my resources View my validation tasks View my supervision tasks

Πρακτική συμβουλή: Δείτε εδώ πώς φαίνεται ο πίνακας ελέγχου για τον επιμελητή (curator), τον επικυρωτή (validator) και τον υπεύθυνο αποθετηρίου (supervisor).

9.1.1 1. Welcome (Καλώς Ήρθατε)

Αυτό το τμήμα περιέχει πληροφορίες για το τι μπορείτε να κάνετε στον πίνακα ελέγχου.

Welcome to your dashboard !

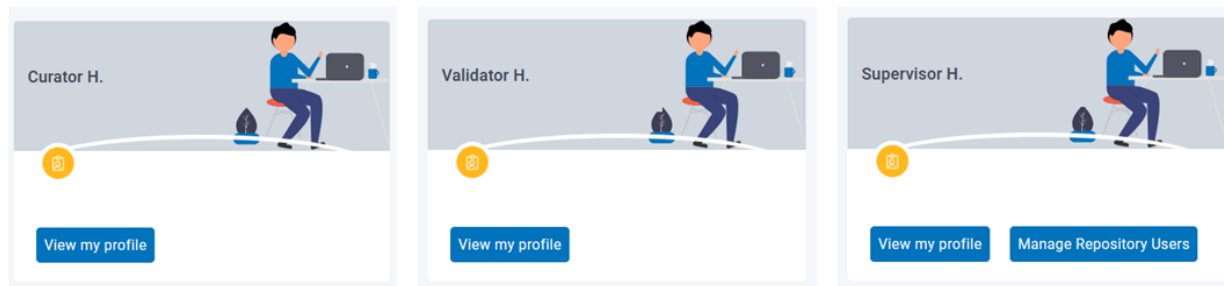
Here you can:

- view and update your profile
- view your resources and your tasks
- access creation forms
- upload resources
- access the workflow registry
- browse your recent activity

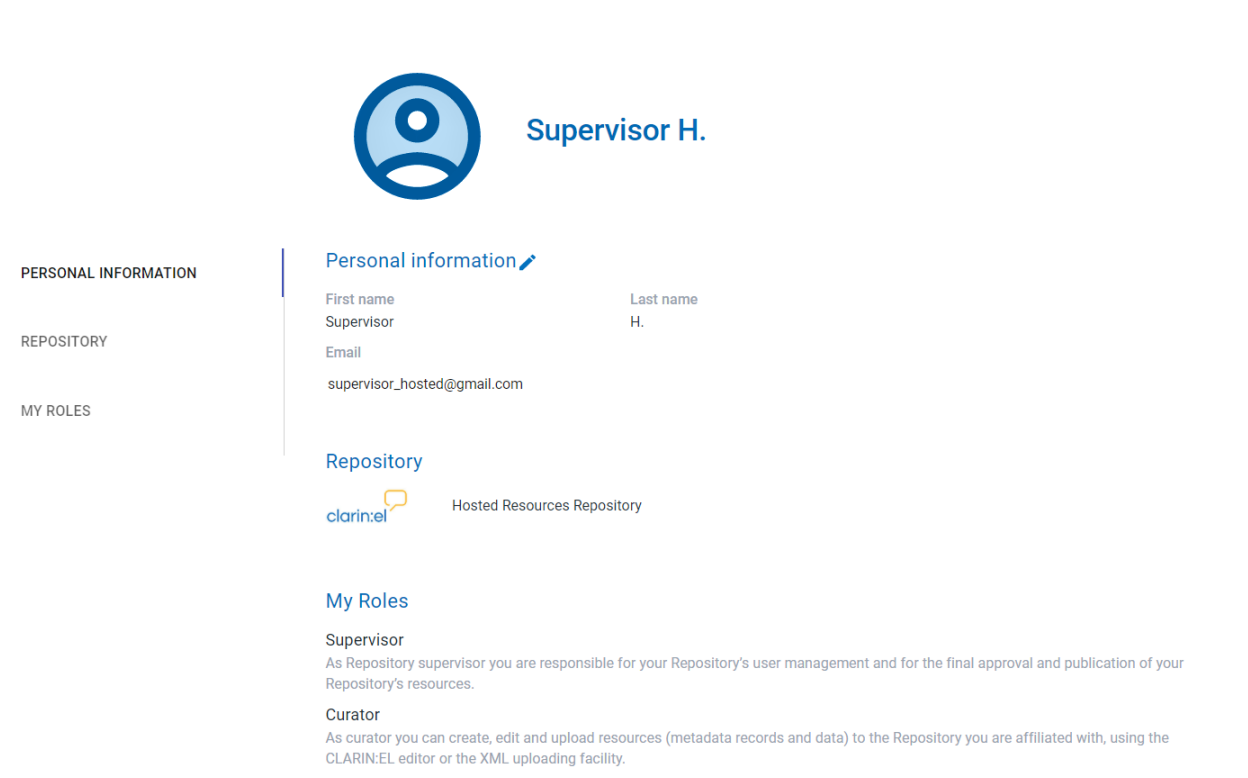


9.1.2 2. Profile (Προφίλ)

Το δεύτερο τμήμα είναι αφιερωμένο στο προφίλ σας. Για τους υπεύθυνους αποθετηρίων υπάρχει μια επιπλέον λειτουργικότητα, η *Manage Repository Users* (Διαχείριση Χρηστών Αποθετηρίου), όπως φαίνεται παρακάτω.



Αν κάνετε κλικ στο *View my profile* (Προβολή του προφίλ μου), θα μεταφερθείτε σε μια νέα σελίδα.



Ορισμένα από τα προσωπικά σας στοιχεία είναι επεξεργάσιμα². Κάντε κλικ στο μολύβι δίπλα στην επικεφαλίδα **Personal information** (Προσωπικά στοιχεία) και θα μεταφερθείτε σε μια νέα σελίδα όπου μπορείτε να επιμεληθείτε τα στοιχεία σας. Στη συνέχεια αποθηκεύστε τις αλλαγές σας.

² Δεν μπορείτε να αλλάξετε το αποθετήριο το οποίο εμφανίζεται καθώς η σύνδεση σας με αυτό γίνεται αυτόματα κατά την εγγραφή σας. Δεν μπορείτε, επίσης, να μεταβάλλετε τους ρόλους που έχετε καθώς αυτοί σας έχουν ανατεθεί από τον υπεύθυνο του αποθετηρίου.

Edit Account * Required fields

Email *

First name *

Last name *

Affiliation

Position

Personal website
Enter an https:// URL



User management (Διαχείριση χρηστών)

Προσοχή: Η λειτουργικότητα αυτή εμφανίζεται μόνο για τους υπεύθυνους αποθετηρίων.

Κάνοντας κλικ στο *Manage Repository Users* (Διαχείριση Χρηστών Αποθετηρίου), θα μεταφερθείτε σε μια σελίδα όπου εμφανίζονται όλοι οι χρήστες του αποθετηρίου σας. Εκεί μπορείτε να εντοπίσετε οποιονδήποτε χρήστη χρησιμοποιώντας το πλαίσιο αναζήτησης στο επάνω μέρος. Μόλις τον/την βρείτε, επιλέξτε το πλαίσιο στα αριστερά του ονόματος.

USERS

Filters

Action

Select an action to perform on your selected items

	User name	Repository	Role	Actions
<input type="checkbox"/>		Hosted Resources Repository		Actions ▾
<input checked="" type="checkbox"/>	K P	Hosted Resources Repository		Actions ▾
<input type="checkbox"/>	Supervisor H.	Hosted Resources Repository	Supervisors	<input checked="" type="checkbox"/> Make metadata validator
<input type="checkbox"/>	Validator H.	Hosted Resources Repository	Metadata Legal	<input checked="" type="checkbox"/> Make legal validator

4 search results

Στη συνέχεια, μπορείτε να διαλέξετε εάν θέλετε να δώσετε στον χρήστη δικαιώματα **Legal Validator** (Επικυρωτής νομικών θεμάτων) ή δικαιώματα **Metadata Validator** (Επικυρωτής μεταδεδομένων). Μπορείτε να κάνετε τις ίδιες ενέργειες και από το πλαίσιο (Action) που βρίσκεται πάνω από τη λίστα των χρηστών. Όποια και αν είναι η απόφασή σας, θα εμφανιστεί ένα νέο παράθυρο που θα σας ζητήσει να την επιβεβαιώσετε.

You are about to Make legal validator the following users.

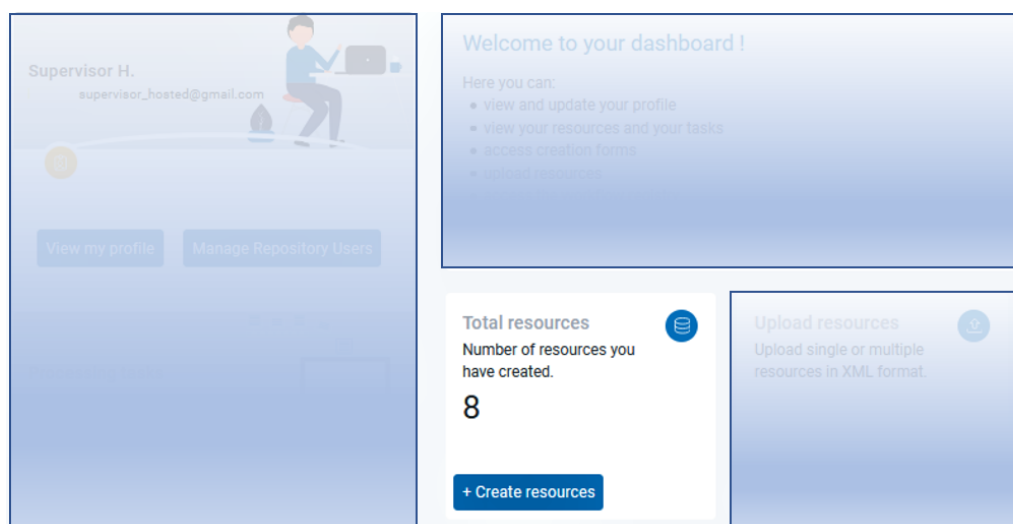
K P

Close

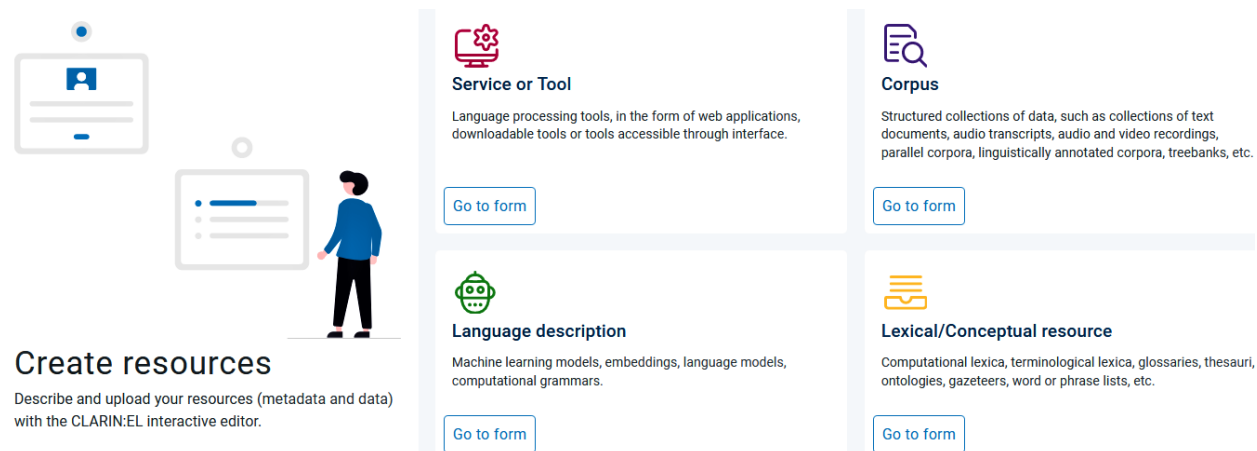
Make legal validator

Κάνοντας κλικ στο **Make legal validator** δίνετε στον επιλεγμένο χρήστη τον συγκεκριμένο ρόλο.

9.1.3 3. Create resources (Δημιουργία πόρων)



Σε αυτό το τμήμα βλέπετε το σύνολο των πόρων που έχετε δημιουργήσει, **αλλά όχι τους ίδιους τους πόρους**³. Κάνοντας κλικ στο **+ Create resources** (Δημιουργία πόρων), μεταφέρεστε σε μια νέα σελίδα όπου θα επιλέξετε τον τύπο του πόρου που θέλετε να δημιουργήσετε.



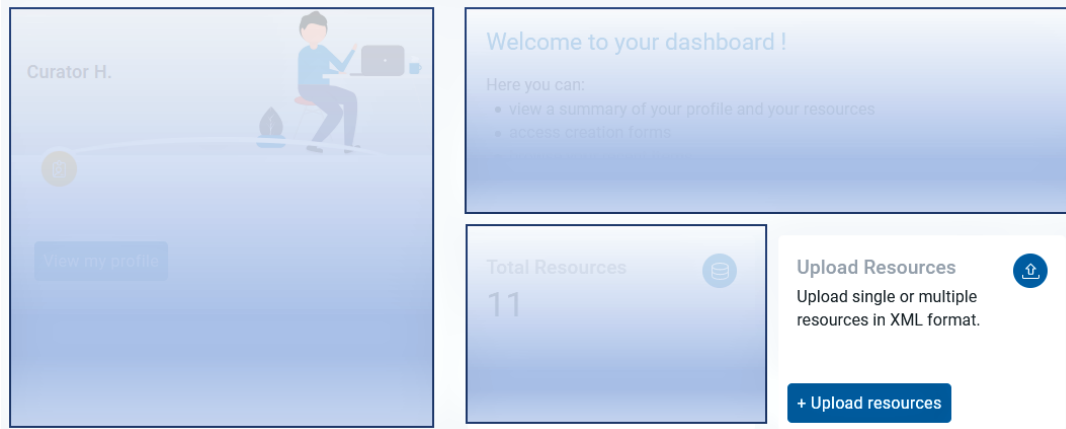
Προτού μεταβείτε στο *περιβάλλον τεκμηρίωσης και ανάρτησης πόρων* κάνοντας κλικ στο **Go to Form (Μετάβαση στη φόρμα)**, δείτε για:

³ Το τμήμα που τιτλοφορείται *My resources* (οι πόροι μου) περιέχει όλους τους πόρους που έχετε δημιουργήσει.

- ένα σώμα κειμένων,
- ένα εργαλείο,
- έναν λεξικό/εννοιολογικό πόρο, ή
- μία γλωσσική περιγραφή,

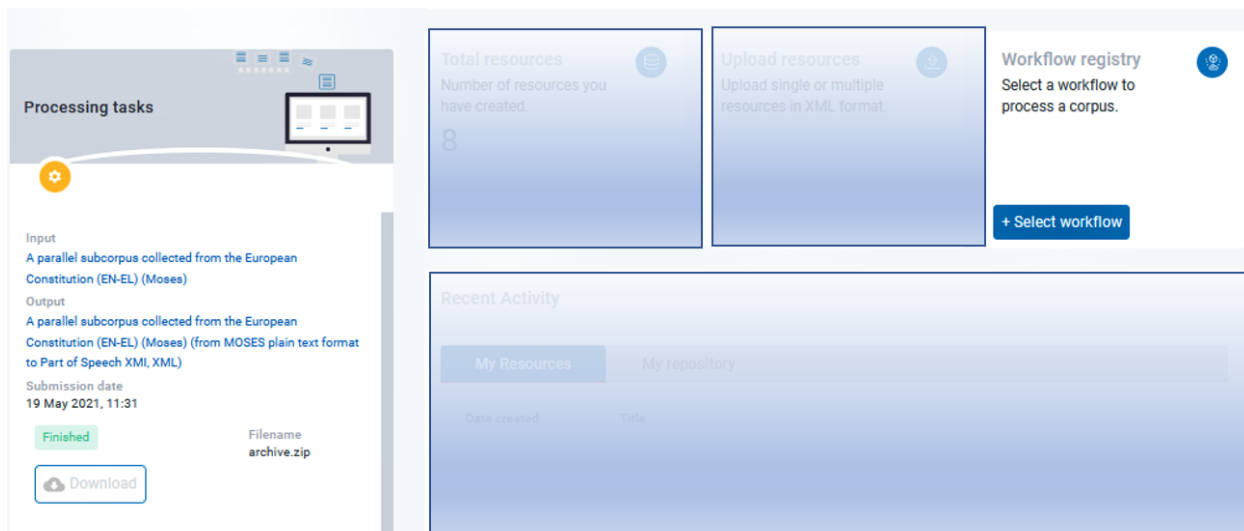
ποια είναι τα *υποχρεωτικά* μεταδεδομένα του σχήματος που θα κληθείτε να συμπληρώσετε.

9.1.4 4. Upload resources (Μεταφόρτωση πόρων)



Μπορείτε να κάνετε κλικ στο + *Upload resources* (Μεταφόρτωση πόρων) αν έχετε ήδη μια (ή περισσότερες) περιγραφές μεταδεδομένων σε μορφή XML. Θα εμφανιστεί ένα νέο παράθυρο με διάφορες επιλογές. Διαβάστε περισσότερα [εδώ](#).

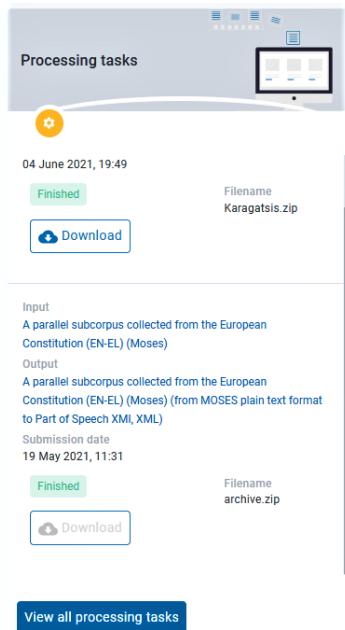
9.1.5 5. Select workflow (Επιλογή ροής εργασίας)



Κάνοντας κλικ στο + *Select workflow* (Επιλογή ροής εργασίας) μεταφέρεστε στο μητρώο όπου υπάρχουν όλες οι **ροές εργασίας** (workflows) που υλοποιούν **λειτουργίες** (functions) επεξεργασίας φυσικής γλώσσας. Δείτε [εδώ](#) ποιες είναι αυτές.

9.1.6 6. Processing Tasks (Εργασίες επεξεργασίας)

Σε αυτό το τμήμα μπορείτε να δείτε τόσο τις εργασίες επεξεργασίας που έχετε ζητήσει να γίνουν (πάνω σε σώματα κειμένων διαθέσιμα από την υποδομή ή δικά σας) όσο και τα αποτελέσματα των εργασιών αυτών, τα οποία, εάν έχει ολοκληρωθεί επιτυχώς η επεξεργασία, μπορείτε να τα καταφορτώσετε.



Πατώντας το *View all processing tasks* (Προβολή όλων των εργασιών επεξεργασίας) μεταφέρεστε σε μία σελίδα που περιέχει το σύνολο των εργασιών που κατά καιρούς έχετε ζητήσει.

MY SUBMITTED PROCESSING TASKS					
Input record	Filename	Output record	Submission date	Status	Download
	Karagatsis.zip		04 June 2021, 19:49	Finished	
A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)	archive.zip	A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) (from MOSES plain text format to Part of Speech XML, XML)	19 May 2021, 11:31	Finished	

Οι εργασίες σας παρουσιάζονται σε έναν πίνακα που χωρίζεται νοητά σε έξι τμήματα:

- στο *πρώτο*, εμφανίζεται το όνομα του πόρου **μόνο εφόσον** ο πόρος αυτός ανήκει στα *επεξεργάσιμα σώματα κειμένων* που προσφέρονται από το **CLARIN:EL** (αν έχετε μεταφορτώσει δικό σας σύνολο δεδομένων για επεξεργασία, η στήλη παραμένει κενή),
- στο *δεύτερο*, είναι το όνομα του συμπίεσμένου αρχείου (κάθε ένα από τα επεξεργάσιμα σώματα κειμένων της υποδομής περιέχεται σε ένα αρχείο που ονομάζεται archive.zip),
- στο *τρίτο*, παρουσιάζεται το όνομα του πόρου όπως προκύπτει μετά την επεξεργασία (και πάλι αν μεταφορτώσατε δικό σας σύνολο δεδομένων, η στήλη αυτή παραμένει κενή),
- στο *τέταρτο*, είναι η ημερομηνία κατά την οποία υποβάλατε την εργασία,
- στο *πέμπτο*, φαίνεται το στάδιο στο οποίο βρίσκεται η εργασία σας, και
- στο *έκτο*, βρίσκεται το κουμπί που μπορείτε να πατήσετε για να καταφορτώσετε τα επεξεργασμένα αρχεία.

9.1.7 7. My resources (Οι πόροι μου)

Σε αυτό το τμήμα εμφανίζονται όλοι οι πόροι που έχετε δημιουργήσει είτε χρησιμοποιώντας το *περιβάλλον τεκμηρίωσης και ανάρτησης πόρων* της υποδομής είτε *μεταφορτώνοντας αρχεία μεταδεδομένων XML*.

Recent Activity

My Resources	
Date created	Title
31 May 2021	Parla (12.1)
30 May 2021	Mandatory Grammar
30 May 2021	Mandatory LCR
29 May 2021	Tool Mandatory
29 May 2021	Mandatory Corpus

[View my resources](#)

Μπορείτε να επιλέξετε έναν πόρο πατώντας πάνω στο όνομά του ή, αν θέλετε να έχετε μια επισκόπηση όλων των πόρων, να κάνετε κλικ στο *View my resources* (Προβολή των πόρων μου).

MY RESOURCES

Search for services, tools, datasets...

Search

21 search results

Resources

+ Corpus (13)

+ Lexical/Conceptual resource (4)

+ Tool/Service (3)

+ Language description (1)

Status

+ draft (12)

+ syntactically valid (6)

+ submitted (2)

+ published (1)

Curator

Username

Search

Has data

+ no (13)

+ yes (8)

Processable

+ no (21)

	Resource name		Actions	Status
<input type="checkbox"/>	Parla (12.1) 1.0.0 (automatically assigned) Corpus Hosted Resources Repository Created: 29 April 2021 Updated: 31 May 2021 has data	legal validator validator_hosted@gmail.com metadata validator validator_hosted@gmail.com curator curator_hosted@gmail.com supervisor supervisor_hosted@gmail.com	Actions ▾	requested for unpublish
<input type="checkbox"/>	Mandatory Grammar 1.0.0 (automatically assigned) Language description Hosted Resources Repository Created: 30 May 2021 Updated: 30 May 2021	curator curator_hosted@gmail.com supervisor supervisor_hosted@gmail.com	Actions ▾	syntactically valid
<input type="checkbox"/>	Mandatory LCR 1.0.0 (automatically assigned) Lexical/Conceptual resource Hosted Resources Repository Created: 30 May 2021 Updated: 30 May 2021	curator curator_hosted@gmail.com supervisor supervisor_hosted@gmail.com	Actions ▾	syntactically valid

9.1. Dashboard (Πίνακας ελέγχου)

31

Σε αυτή τη σελίδα, στα αριστερά, υπάρχουν *φίλτρα* για να σας βοηθήσουν να ταξινομήσετε τους πόρους ανάλογα με τον τύπο (**resource**), το στάδιο στο οποίο βρίσκονται (**status**), το αν συνοδεύονται από μεταφορτωμένα αρχεία (**has data**) ή αν είναι επεξεργάσιμοι (**processable**). Μπορείτε να εφαρμόσετε όσα φίλτρα θέλετε και στη συνέχεια να κάνετε εκκαθάριση (**clear all filters**) πατώντας το αντίστοιχο κουμπί στο πάνω μέρος της σελίδας (δείτε την αμέσως επόμενη εικόνα).

4 search results

Clear all filters

Resources

– Lexical/Conceptual resource (4)

Status

+ draft (2)

+ submitted (1)

	Resource name	Actions	Status
<input type="checkbox"/>	Mandatory LCR 1.0.0 (automatically assigned) Lexical/Conceptual resource Hosted Resources Repository Created: 30 May 2021 Updated: 30 May 2021	curator curator_hosted@gmail.com supervisor supervisor_hosted@gmail.com	Actions syntactically valid

Σημείωση: Ως υπεύθυνος αποθετηρίου θα δείτε επίσης ένα **πλαίσιο αναζήτησης για επιμελητές**. Χρησιμοποιήστε το email ενός επιμελητή για να βρείτε μόνο τους πόρους που έχει δημιουργήσει.

4 search results

Clear all filters

Curator

Username

Search

Search using curator's email

	Resource name	Actions	Status
<input type="checkbox"/>	Mandatory LCR 1.0.0 (automatically assigned) Lexical/Conceptual resource Hosted Resources Repository Created: 30 May 2021 Updated: 30 May 2021	curator curator_hosted@gmail.com supervisor supervisor_hosted@gmail.com	Actions syntactically valid

Όπως μπορείτε να δείτε, κάθε πόρος καταλαμβάνει μια σειρά που χωρίζεται νοητά σε τέσσερα τμήματα:

- στο *πρώτο*, παρέχονται κάποιες βασικές πληροφορίες για τον πόρο,
- στο *δεύτερο*, εμφανίζεται το όνομα του επιμελητή, του υπεύθυνου αποθετηρίου και του επικυρωτή (αν ο πόρος έχει επικυρωθεί),
- στο *τρίτο*, υπάρχει ένα κουμπί για τις διαθέσιμες ενέργειες, και
- στο *τέταρτο* φαίνεται το στάδιο στο οποίο ο πόρος βρίσκεται.

Εάν επιθυμείτε να μάθετε περισσότερα σχετικά με τις ενέργειες που γίνονται σε έναν πόρο, ανατρέξτε στην ενότητα [Διαχείριση πόρων](#).

9.1.8 8. Validation tasks (Εργασίες επικύρωσης)

Προσοχή: Το τμήμα αυτό εμφανίζεται μόνο για τους **επικυρωτές**.

Εδώ μπορείτε να δείτε όλους τους πόρους που σας έχουν ανατεθεί για επικύρωση. Μπορείτε να επιλέξετε έναν πόρο πατώντας πάνω στο όνομά του και θα μεταφερθείτε στη σελίδα του. Σε περίπτωση που έχετε ήδη επικυρώσει αυτόν τον πόρο, ένα αναδυόμενο μήνυμα θα σας ενημερώσει ότι δεν έχετε πλέον δικαιώματα σε αυτόν.

Recent Activity

My Resources	
Validation tasks	
Date created	Title
05 May 2021	Parla (12.1)
29 Apr 2021	Single text EL (txt)
28 Apr 2021	tool1_Maria

inaccessible to you at this status

[View my validation tasks](#)

Εάν επιθυμείτε να έχετε μια επισκόπηση όλων των πόρων που σας έχουν ανατεθεί για επικύρωση (είτε η επικύρωση έχει γίνει είτε όχι), μπορείτε να κάνετε κλικ στο *View my validation tasks* (Προβολή των εργασιών επικύρωσής).

Resources

- + [Corpus \(2\)](#)
- + [Tool/Service \(1\)](#)

Status

- + [published \(2\)](#)
- + [syntactically valid \(1\)](#)

Metadata valid

- + [yes \(2\)](#)

Legally valid

- + [yes \(2\)](#)

Validator

Username

Has data

- + [yes \(2\)](#)
- + [no \(1\)](#)

Processable

- + [no \(2\)](#)
- + [yes \(1\)](#)

CLARIN:EL compatible service

- + [no \(3\)](#)

3 search results

Resource name	Status
<p>Parla (12.1)</p> <p>1.0.0 (automatically assigned)</p> <p>Corpus</p> <p>Hosted Resources Repository</p> <p>submitted: 05 May 2021</p> <p>has data</p>	<p>legal validator</p> <p>validator_hosted@gmail.com</p> <p>metadata validator</p> <p>validator_hosted@gmail.com</p> <p>curator</p> <p>curator_hosted@gmail.com</p> <p>supervisor</p> <p>supervisor_hosted@gmail.com</p> <p>published</p> <p>legally valid</p> <p>yes</p> <p>metadata valid</p> <p>yes</p>
<p>Single text EL (txt)</p> <p>1.0.0 (automatically assigned)</p> <p>Corpus</p> <p>Hosted Resources Repository</p> <p>submitted: 29 April 2021</p> <p>processable has data</p> <p>syntactically valid</p> <p>Review comments</p> <p>The licence chosen is inappropriate for corpora. Please check and replace it.</p>	<p>legal validator</p> <p>legalvalidator_hosted@gmail.com</p> <p>metadata validator</p> <p>validator_hosted@gmail.com</p> <p>curator</p> <p>supervisor_hosted@gmail.com</p> <p>supervisor</p> <p>supervisor_hosted@gmail.com</p>
<p>tool1</p> <p>1.0</p>	<p>metadata validator</p> <p>validator_hosted@gmail.com</p> <p>published</p>

Σε αυτή τη σελίδα, στα αριστερά, υπάρχουν *φίλτρα* για να σας βοηθήσουν να ταξινομήσετε τους πόρους. Μπορείτε να εφαρμόσετε όσα φίλτρα θέλετε και στη συνέχεια να κάνετε εκκαθάριση (clear all filters) πατώντας το κουμπί πάνω από αυτά.

1 search results

Clear all filters

Resources

+ [Corpus \(1\)](#)

Status

+ [syntactically valid \(1\)](#)

Validator

Username

Search

Has data

- [yes \(1\)](#)

Processable

- [yes \(1\)](#)

CLARIN:EL compatible service

+ [no \(1\)](#)

Resource name	Status
<p>Single text EL (txt)</p> <p>1.0.0 (automatically assigned)</p> <p>Corpus</p> <p>Hosted Resources Repository</p> <p>submitted: 29 April 2021</p> <div style="display: flex; gap: 5px;"> processable has data </div>	<p>legal validator</p> <p>legalvalidator_hosted@gmail.com</p> <p>metadata validator</p> <p>validator_hosted@gmail.com</p> <p>curator</p> <p>supervisor_hosted@gmail.com</p> <p>supervisor</p> <p>supervisor_hosted@gmail.com</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p>Review comments</p> <p>The licence chosen is inappropriate for corpora. Please check and replace it.</p> </div>

syntactically valid

Όπως μπορείτε να δείτε, κάθε πόρος καταλαμβάνει μια γραμμή που χωρίζεται νοητά σε τρία τμήματα:

- στο *πρώτο*, παρέχονται κάποιες βασικές πληροφορίες για τον πόρο,
- στο *δεύτερο*, εμφανίζονται τα ονόματα του επιμελητή, του υπεύθυνου αποθετηρίου και των επικυρωτών, και
- στο *τρίτο*, φαίνεται το στάδιο στο οποίο ο πόρος βρίσκεται. Εδώ εμφανίζεται επίσης η πληροφορία για το αν ο πόρος έχει επικυρωθεί (από τον Επικυρωτή νομικών θεμάτων ή/και τον Επικυρωτή μεταδεδομένων).

Επιπλέον, υπάρχει ένα πλαίσιο με τα σχόλια του κάθε επικυρωτή, αν έχουν γίνει. Μπορείτε να δείτε *εδώ* πώς γίνεται η επικύρωση ενός πόρου.

9.1.9 9. My repository (Το αποθετήριό μου)

Προσοχή: Το τμήμα αυτό εμφανίζεται μόνο για τους **υπεύθυνους αποθετηρίων**.

Αυτό το τμήμα περιέχει όλους τους πόρους του αποθετηρίου σας, ανεξάρτητα από το ποιος τους έχει δημιουργήσει. Αν θέλετε να δείτε τους πόρους που έχετε δημιουργήσει **εσείς**, πρέπει να κάνετε κλικ στο [My resources](#) (Οι πόροι μου).

Recent Activity

My Resources	
My repository	
Date created	Title
11 Jun 2021	corpus test upload 2
09 Jun 2021	Language Description (Mandatory elements)
09 Jun 2021	Single text EL (txt)
09 Jun 2021	Lexical/Conceptual Resource (Mandatory elements)
08 Jun 2021	draft Testrecognizer

[View my supervision tasks](#)

Μπορείτε είτε να επιλέξετε έναν πόρο πατώντας στο όνομά του είτε, αν θέλετε να έχετε μια επισκόπηση όλων των πόρων στο αποθετήριο σας, να κάνετε κλικ στο *View my supervision tasks* (Προβολή των εργασιών που επιβλέπω). Σε αυτή τη σελίδα, στα αριστερά, υπάρχουν *φίλτρα* για να σας βοηθήσουν να ταξινομήσετε τους πόρους ανάλογα με τον τύπο (**resource**), το στάδιο στο οποίο βρίσκονται (**status**), το αν συνοδεύονται από μεταφορτωμένα αρχεία (**has data**) κλπ. τους. Μπορείτε να εφαρμόσετε όσα φίλτρα θέλετε και στη συνέχεια να κάνετε εκκαθάριση (**clear all filters**) πατώντας το κουμπί πάνω από αυτά. Στην παρακάτω εικόνα εμφανίζονται μόνο οι πόροι για τους οποίους πρέπει να οριστεί επικυρωτής (με βάση το σχετικό φίλτρο).

Resources	<input type="checkbox"/>	Created: 24 April 2021 Updated: 11 June 2021	yes	Actions	legally valid not validated metadata valid not validated
+ Corpus (5)		has data			
+ Lexical/Conceptual resource (1)					
Action required		Review comments for testing, for testing purposes2, for testing purposes			
+ yes (6)					
Status		test_corpus_1			submitted
+ submitted (6)		1.0.0 (automatically assigned) Corpus	curator admin@email.com		
Metadata valid	<input type="checkbox"/>	Hosted Resources Repository Created: 05 May 2021 Updated: 06 May 2021	supervisor supervisor_hosted@gmail.com	Actions	legally valid yes metadata valid not validated
+ not validated (6)			validator assignment required yes		
Legally valid		TestK_1.1.1			submitted
+ yes (4)		1.0.0	curator curator_hosted@gmail.com		
+ not validated (2)		Lexical/Conceptual resource Hosted Resources Repository Created: 27 April 2021 Updated: 27 April 2021	supervisor supervisor_hosted@gmail.com	Actions	legally valid yes metadata valid not validated
Has data	<input type="checkbox"/>		validator assignment required yes		
+ no (4)					
+ yes (2)					
Processable		TestK_1			submitted
+ no (6)		1.0.0	curator curator_hosted@gmail.com		
CLARIN:EL compatible service	<input type="checkbox"/>	Corpus Hosted Resources Repository Created: 27 April 2021 Updated: 27 April 2021	supervisor supervisor_hosted@gmail.com	Actions	legally valid not validated metadata valid not validated
+ no (6)		has data	validator assignment required yes		
Requested for unpublish					
+ no (6)					
Validator assignment required					
- yes (6)					

Όπως μπορείτε να δείτε, κάθε πόρος καταλαμβάνει μια σειρά που χωρίζεται νοητά σε τέσσερα τμήματα:

- στο *πρώτο*, παρέχονται κάποιες βασικές πληροφορίες για τον πόρο,
- στο *δεύτερο*, παρουσιάζονται τα ονόματα του επιμελητή, του υπεύθυνου αποθετηρίου και του επικυρωτή (αν ο πόρος έχει επικυρωθεί),

- στο *τρίτο*, υπάρχει ένα κουμπί για τις διαθέσιμες ενέργειες, και
- στο *τέταρτο* φαίνεται το στάδιο στο οποίο ο πόρος βρίσκεται και αν έχει επικυρωθεί ή όχι.

Εάν επιθυμείτε να μάθετε περισσότερα σχετικά με τις ενέργειες που γίνονται σε έναν πόρο, ανατρέξτε στην ενότητα [Διαχείριση πόρων](#).

9.2 Help (Βοήθεια)

Αυτός ο σύνδεσμος σας οδηγεί στο *παρόν εγχειρίδιο* όπου δίνονται πληροφορίες σχετικά με τους πόρους, τους χρήστες, τα δικαιώματα και τις διαδικασίες. Μπορείτε να αναζητήσετε ένα συγκεκριμένο θέμα που σας ενδιαφέρει χρησιμοποιώντας το πλαίσιο αναζήτησης (που βρίσκεται στο πάνω δεξιά μέρος της σελίδας) ή να περιηγηθείτε στα διάφορα κεφάλαια και τις ενότητες.

The screenshot shows the CLARIN:EL manual page. On the left is a sidebar with the CLARIN logo and a search bar. Below the search bar are navigation links categorized into 'INTRODUCTION', 'BASIC CONCEPTS', and 'NAVIGATING THE INFRASTRUCTURE'. The main content area is titled 'How to use this manual' and contains the following text:

This manual ¹ aims to help you explore and/or use the CLARIN:EL infrastructure to make your resources available to the **Humanities and Social Sciences** community. It is not meant to be read in sequence (although it can be) but to help you find specific information depending on your needs. There are chapters with general information on [basic concepts](#); others describing the [process](#) ² through which a resource comes to life and chapters which will specifically help you:

- to [browse](#) and [search](#) through the central inventory so as to find resources to [download](#) and [process](#),
- to create resources via the [editor](#) or by [uploading XML files](#), and
- to perform [actions on resources](#) depending on your responsibilities.

If you are looking for something specific, please, use the search box on the top left side of the navigation bar, below the CLARIN:EL logo.

[1] : The current version documents the third official release of the CLARIN:EL infrastructure, launched on May 31st, 2021. More functionalities are continuously added, and this manual keeps on being updated following the evolution of the CLARIN:EL platform.

[2] : Before you start, please see [the lifecycle of a resource](#) to find out what each type of user needs to do throughout the procedure. To assume a role you must have [registered](#) first.

At the bottom of the main content area, there are 'Previous' and 'Next' navigation buttons. The footer contains the copyright information: '© Copyright 2020, CLARIN Technical Team. Built with Sphinx using a theme provided by Read the Docs.'

9.3 Your name (Το όνομά σας)




Κάνοντας κλικ στο όνομά σας μεταφέρεστε στη σελίδα του *προφίλ* σας.

9.4 Exit (Έξοδος)

Εάν θέλετε να αποσυνδεθείτε, κάντε κλικ σε αυτό το εικονίδιο.

Περιήγηση

Από την *αρχική σελίδα του κεντρικού καταλόγου* κάντε κλικ στο κουμπί **browse** (περιήγηση) ή **search** (αναζήτηση) και θα μεταφερθείτε στον κεντρικό κατάλογο. Στην αριστερή πλευρά της σελίδας μπορείτε να δείτε τα διαθέσιμα φίλτρα τα οποία επεξηγούνται *εδώ*. Στον κατάλογο κάθε πόρος εμφανίζεται μέσα σε ένα πλαίσιο.

Repository	▼	 CGL Humanities Metalanguage Corpus: A descriptive directory of Text Corpora for Modern Greek <div>processable</div> <div>accessible through interface</div> <p>Corpus</p> <p>The resource consists of reviews of three Modern Greek-language text corpora: - The corpus of the National Thesaurus of the Greek Language (ΕΘΕΓ) (The review of the corpus is 3500 words) - The newspaper corpus by the Cen ...</p> <p>Language: Modern Greek (1453-)</p> <p>Media Type: text</p> <p>Keywords: monolingual text/plain</p>
Resource type	▼	
Media type	▼	
Languages	▼	
Linguality type	▼	
Domains	▼	 Hellenic National Corpus <div>accessible through interface</div> <p>Corpus</p> <p>The Hellenic National Corpus (HNC) currently contains about 47.000.000 words, and is constantly being updated. It consists of samples of written language exclusively. Texts in the HNC represent modern Greek language use. ...</p> <p>Language: Modern Greek (1453-)</p> <p>Media Type: text</p> <p>Keywords: monolingual text/plain non-fiction segmentation</p>
Time coverage	▼	
Processable	▼	 Golden Part of Speech Tagged Corpus <div>downloadable</div> <p>Corpus</p> <p>The GoldenPart-of-Speech Tagged Corpus is a subset of the Hellenic National Corpus (HNC), the size of which is 100.000 words; it consists of selected texts from a variety of sources covering various domains. These texts ...</p>
Annotation type	▼	
Processing service	▼	
Licences	▼	

Μέσα στο πλαίσιο, στα αριστερά υπάρχει το λογότυπο του αποθετηρίου στο οποίο δημιουργήθηκε ο πόρος. Το όνομα του πόρου, ακριβώς δίπλα, είναι σε μορφή υπερσυνδέσμου για να οδηγεί στη *σελίδα θέασης*. Ακριβώς

κάτω από το όνομα, υπάρχει ο τύπος του πόρου, ακολουθούμενος από τις πρώτες γραμμές της περιγραφής. Στη συνέχεια, εμφανίζονται πληροφορίες σχετικά με τη γλώσσα που καλύπτει (ή μπορεί να επεξεργαστεί) ο πόρος, τις λέξεις-κλειδιά με τις οποίες έχει περιγραφεί και το μέσο στο οποίο διατίθεται (κειμένο, ήχος, εικόνα ή/και βίντεο). Στη δεξιά πλευρά της σελίδας υπάρχουν ετικέτες που παρέχουν πληροφορίες σχετικά με το αν ο πόρος είναι επεξεργάσιμος (**processable**)¹ και αν, και με ποιο τρόπο είναι, προσβάσιμος (**accessible**)². Ορισμένες από τις πιο συχνές ετικέτες είναι οι ακόλουθες:

- *processable* (επεξεργάσιμος): ένας πόρος (σώμα κειμένων) που έχει τεχνικά χαρακτηριστικά συμβατά με τις υπηρεσίες της υποδομής και μπορεί να υποστεί επεξεργασία [για να γίνει αυτό, πρέπει ο χρήστης από την σελίδα θέασης του πόρου να επιλέξει το κουμπί [Process](#) που βρίσκεται στην καρτέλα [Access](#)].
- *processing service* (υπηρεσία επεξεργασίας): ένα εργαλείο το οποίο έχει ενσωματωθεί στην υποδομή ως υπηρεσία και μπορεί να χρησιμοποιηθεί για την επεξεργασία σωμάτων κειμένων [για να γίνει αυτό, πρέπει ο χρήστης από την σελίδα θέασης του πόρου να επιλέξει το κουμπί [Use](#) που βρίσκεται στην καρτέλα [Access](#)].
- *downloadable* (καταφορτώσιμος): ένας πόρος του οποίου τα αρχεία μπορεί να αποκτήσει ο χρήστης με καταφόρτωση, είτε απευθείας από το **CLARIN:EL** είτε από εξωτερικό σύνδεσμο μέσω ανακατεύθυνσης [για να γίνει αυτό, πρέπει ο χρήστης από την σελίδα θέασης του πόρου να επιλέξει το κουμπί [Download](#) που βρίσκεται στην καρτέλα [Access](#)].
- *accessible through interface*: a resource which is accessed by an interface available through an external link [to do so, you must visit the resource view page and click the [Access](#) button found in the [Access](#) tab].





Ακριβώς κάτω από το πλαίσιο αναζήτησης στο πάνω μέρος της σελίδας υπάρχουν δύο εικονίδια όπως φαίνεται στην ακόλουθη εικόνα.



Τα εικονίδια προσφέρουν δύο επιλογές για την διάταξη των πόρων του κεντρικού καταλόγου. Εξ ορισμού οι πόροι εμφανίζονται ως λίστα, αλλά αν κάνετε κλικ στο εικονίδιο στα δεξιά, μπορείτε να τους δείτε ως μεγάλα εικονίδια διατεταγμένα σε πίνακα.

¹ Οι ακόλουθοι είναι οι τρόποι με τους οποίους ένα σώμα κειμένων ή ένας λεξικός/εννοιολογικός πόρος είναι προσβάσιμος (accessible) από τον χρήστη: **downloadable** (καταφορτώσιμος), **CD-ROM**, **DVD-R**, **accessible through interface** (προσβάσιμος μέσω διεπαφής), **accessible through query** (προσβάσιμος μέσω ερωτήματος), **bluRay**, **hard disk**, **other** (προσβάσιμος μέσω οποιουδήποτε άλλου τρόπου), **unspecified** (προσβάσιμος μέσω τρόπου που δεν μπορεί να καθοριστεί).

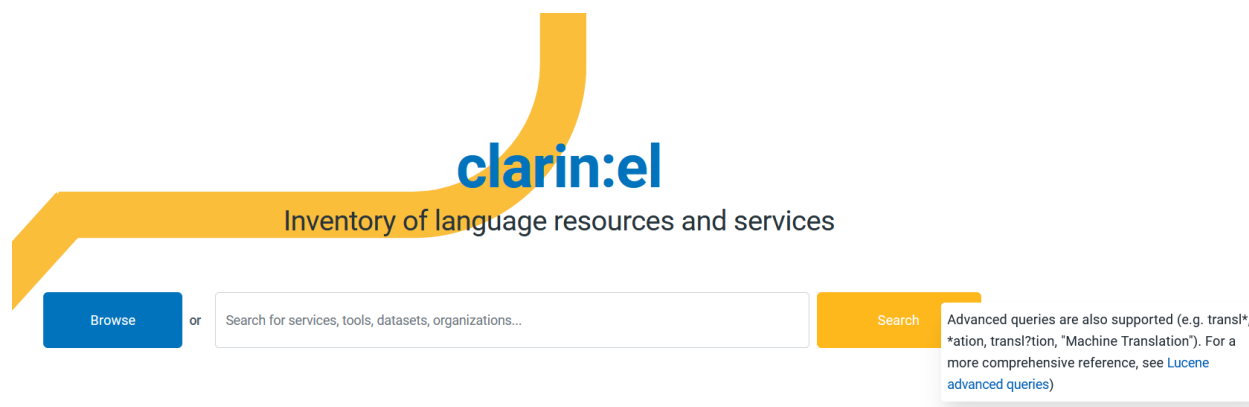
² Οι ακόλουθοι είναι οι τρόποι με τους οποίους ένα εργαλείο ή μία υπηρεσία είναι προσβάσιμος (accessible) από τον χρήστη: **library** (βιβλιοθήκη), **plugin** (πρόσθετο), **source code** (πηγαίος κώδικας), **source and executable code** (πηγαίος και εκτελέσιμος κώδικας), **web service** (διαδικτυακή υπηρεσία), **workflow file** (αρχείο ροής εργασιών), **docker image** ή **other** (οποιοσδήποτε άλλος τρόπος δεν ανήκει στους προαναφερθέντες).

Repository	▼	 A parallel corpus collected from the European Constitution Corpus	 A parallel corpus of KDE4 localization files (v.2) Corpus
Resource type	▼	A parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 total number of tokens: 3.01M total number of sentence fragments: 0.22M	92 languages, 4,101 bitexts total number of files: 75,535 total number of tokens: 60.75M total number of sentence fragments: 8.89M
Media type	▼	Keywords: multilingual parallel application/xml alignment	Keywords: multilingual parallel application/x-xces+xml technicalTexts
Languages	▼	Languages: Latvian Estonian Czech ...	Languages: Galician Icelandic Kashubian ...
Language variety	▼	Media Type: text	Media Type: text
Linguality type	▼	downloadable	downloadable
Multilinguality type	▼		
Domains	▼	 A parallel subcorpus collected from the European Constitution (DE-EL) (Moses) Corpus	 A parallel subcorpus collected from the European Constitution (DE-EL) (Moses) (from MOSES plain text format to Part of Speech XML, XMI) Corpus
Time coverage	▼	The EUconst subcorpus (DE-EL) (Moses) is a parallel subcorpus for German and Greek, subset of the EUconst, a parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 ...	A parallel subcorpus collected from the European Constitution (DE-EL) (Moses) (MOSES plain text format) annotated by the ILSP Feature-based multi-tiered POS Tagger (ILSP workflow v1.0.0) and the OpenNLP Part-of-Speech Ta ...
Geographic coverage	▼	Keywords: bilingual parallel text/plain administrativeTexts	Keywords: bilingual parallel text/plain administrativeTexts
Processable	▼	Languages: Modern Greek (1453-) German	Languages: Modern Greek (1453-) German
Annotation type	▼	Media Type: text	Media Type: text
Processing service	▼	processable	downloadable
		downloadable	downloadable

Κάνοντας κλικ πάνω στο όνομα ενός πόρου, μπορείτε να μεταβείτε στη *σελίδα θέαςής* του.

Αναζήτηση

Στην αρχική σελίδα του κεντρικού καταλόγου του CLARIN:EL υπάρχει ένα πλαίσιο αναζήτησης. Για την αναζήτηση μπορείτε να χρησιμοποιήσετε λέξεις-κλειδιά ή, όπως φαίνεται στην εικόνα, ειδικούς χαρακτήρες για προηγμένα ερωτήματα τύπου Lucene¹.



Για την αναζήτηση, έχετε επίσης τη δυνατότητα να χρησιμοποιήσετε **φίλτρα**. Για να το κάνετε αυτό, κάντε κλικ στο κουμπί **browse** (περιήγηση) ή **search** (αναζήτηση) προκειμένου να μεταφερθείτε στον κεντρικό κατάλογο. Στην αριστερή πλευρά της σελίδας, μπορείτε να δείτε όλα τα διαθέσιμα φίλτρα.

¹ Προς το παρόν, τα προηγμένα ερωτήματα μπορούν να χρησιμοποιηθούν μόνο για αναζήτηση στον κεντρικό κατάλογο, όχι στα πλαίσια αναζήτησης που βρίσκονται στα *My resources* (οι πόροι μου), *Validation tasks* (εργασίες επικύρωσης), και *My repository* (το αποθετήριο μου).

clarin:el

CLARIN:EL portal >

Help

Sign in

Search for services, tools, datasets...

Search

?

Repository

Resource type

Media type

Languages

Language variety

Linguality type

Multilinguality type

Domains

Time coverage

Geographic coverage

Processable

Annotation type

Processing service

Service functions

Licences

ACQUA

A parallel corpus collected from the European Constitution

downloadable

Corpus

A parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 total number of tokens: 3.01M total number of sentence fragments: 0.22M

Languages: Finnish Irish Modern Greek (1453-) Slovak ...

Media Type: text

Keywords: multilingual parallel application/xml alignment

ACQUA

A parallel corpus of KDE4 localization files (v.2)

downloadable

Corpus

92 languages, 4,101 bitexts total number of files: 75,535 total number of tokens: 60.75M total number of sentence fragments: 8.89M

Languages: Macedonian Hebrew Slovak Norwegian Nynorsk ...

Media Type: text

Keywords: multilingual parallel application/x-aces+xml technicalTexts

ACQUA

A parallel subcorpus collected from the European Constitution (DE-EL) (Moses)

processable

downloadable

Corpus

The EUconst subcorpus (DE-EL) (Moses) is a parallel subcorpus for German and Greek, subset of the EUconst, a parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 ...

Languages: Modern Greek (1453-) German

Media Type: text

Keywords: bilingual parallel text/plain administrativeTexts

ACQUA

A parallel subcorpus collected from the European Constitution (DE-EL) (Moses) (from MOSES plain text format to Part of Speech XML, XMI)

downloadable

Corpus

A parallel subcorpus collected from the European Constitution (DE-EL) (Moses) (MOSES plain text format) annotated by the ILSP Feature-based multi-tiered POS Tagger (ILSP workflow v1.0.0) and the OpenNLP Part-of-Speech Ta ...

Languages: Modern Greek (1453-) German

Τα φίλτρα είναι κλειστά, αλλά μόλις επιλέξετε το βέλος που υπάρχει δίπλα στο κάθε φίλτρο εμφανίζεται μία αναδυόμενη λίστα τιμών.

Repository

- Athena RC Repository (341)
- University of Crete Repository (41)
- Centre For The Greek Language Repository (30)
- University of Thessaloniki Repository (28)
- Hosted Resources Repository (24)
- Show more

Resource type

- Corpus (412)
- Lexical/Conceptual resource (78)
- Tool/Service (41)
- Language description (2)

Media type

A parallel corpus collected from the European Constitution downloadable

Corpus

A parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 total number of tokens: 3.01M total number of sentence fragments: 0.22M

Languages: Finnish Irish Modern Greek (1453-) Slovak ...

Media Type: text

Keywords: multilingual parallel application/xml alignment

A parallel corpus of KDE4 localization files (v.2) downloadable

Corpus

92 languages, 4,101 bitexts total number of files: 75,535 total number of tokens: 60.75M total number of sentence fragments: 8.89M

Languages: Macedonian Hebrew Slovak Norwegian Nynorsk ...

Media Type: text

Keywords: multilingual parallel application/x-xces+xml technicalTexts

Αν θέλετε να ανατρέξετε την εφαρμογή όλων των φίλτρων, κάντε κλικ στο κουμπί **clear all filters** (καθαρισμός όλων των φίλτρων). Εναλλακτικά, διαγράψτε με το **x** που υπάρχει δίπλα στο όνομα του φίλτρου όσα από τα φίλτρα δεν επιθυμείτε να εφαρμόζονται πλέον.

Search for services, tools, datasets... Search ?

Clear all filters 52 search results

English ATHENA RC Repository Corpus

Στον κεντρικό κατάλογο υπάρχουν τα φίλτρα που αναλύονται στην συνέχεια. Η εφαρμογή τους οδηγεί στην ανάκτηση των πόρων με βάση:

- το **αποθετήριο** (*Repository*) στο οποίο ανήκουν,
- τον τύπο του πόρου (*Resource type*), αν δηλαδή είναι **σώματα κειμένων** ή **εργαλεία/υπηρεσίες** ή **λεξικοί/εννοιολογικοί πόροι** ή **γλωσσικές περιγραφές**,
- τον τύπο του μέσου (*Media type*), αν δηλαδή το περιεχόμενό τους είναι -ή μπορούν να επεξεργαστούν ως εργαλεία/υπηρεσίες υλικό- **κειμενικό** ή **ηχητικό** ή **βίντεο** ή **εικόνα** (επίσης, ένας πόρος μπορεί να **απαρτίζεται από/στοχεύει σε** συνδυασμούς αυτών των μέσων),
- τις **γλώσσες** (*Languages*), που καλύπτουν ή που μπορούν να επεξεργαστούν ως εργαλεία/υπηρεσίες,
- τη **γλωσσική ποικιλία** (*Language variety*), που καλύπτουν ή που μπορούν να επεξεργαστούν ως εργαλεία/υπηρεσίες,
- τη διάθεσή τους σε μία ή περισσότερες γλώσσες (*Linguality type*), αν δηλαδή οι πόροι είναι **μονόγλωσσοι** ή **δύγλωσσοι** ή **πολύγλωσσοι**
- τη στοίχιση ή μη των γλωσσών στις οποίες διατίθενται (*Multilinguality type*), αν δηλαδή οι πόροι είναι **παράλληλοι** ή **συγκρίσιμοι**,
- τις **θεματικές περιοχές** (*Domains*), δηλαδή τα διάφορα πεδία γνώσης που έχουν χρησιμοποιηθεί για την ταξινόμησή τους ή το περιεχόμενό τους,
- τη **χρονική κάλυψη** (*Time coverage*), τη χρονική δηλαδή περίοδο που καλύπτει το περιεχόμενό τους, όπως για παράδειγμα είναι οι **γλωσσικοί πόροι του 16ου αιώνα**,
- τη **γεωγραφική κάλυψη** (*Geographic coverage*), τη γεωγραφική δηλαδή περιοχή που σχετίζεται με το περιεχόμενό τους, όπως για παράδειγμα είναι οι **γλωσσικοί πόροι που σχετίζονται με τη Γαλλία**,

- την **επεξεργασιμότητα** (*Processable*), τα τεχνικά χαρακτηριστικά που κάποια σώματα κειμένων έχουν και τα καθιστούν συμβατά με τις υπηρεσίες του *μητρώου ροής εργασιών*,
- τον **τύπο επισημείωσης/σχολιασμού** (*Annotation type*), για παράδειγμα, την στοίχιση (alignment) σωμάτων κειμένων, όπως συμβαίνει με τους δίγλωσσους ή πολύγλωσσους γλωσσικούς πόρους που είναι *παράλληλοι*,
- τις **υπηρεσίες επεξεργασίας** (*Processing Services*), τα εργαλεία δηλαδή που έχουν ενσωματωθεί ως *υπηρεσίες* στο CLARIN:EL,
- τη **λειτουργία** (*Service functions*) που επιτελεί ένα εργαλείο/υπηρεσία, όπως για παράδειγμα η *δημιουργία λεξικού*, και
- την **άδεια χρήσης** (*Licences*), όπως για παράδειγμα πόροι που διατίθενται με άδεια *CC-BY 4.0*.

Προσοχή: Προκειμένου να μπορεί ένας πόρος να ανακτηθεί με τη χρήση φίλτρων, θα πρέπει τα αντίστοιχα μεταδεδομένα να έχουν συμπληρωθεί από τον επιμελητή.

Θέαση & Καταφόρτωση

Κάθε πόρος έχει μία **σελίδα θέασης**. Σε αυτήν ο χρήστης μπορεί να δει μία σειρά από **μεταδεδομένα** (όπως η περιγραφή και τα στοιχεία ταυτοποίησης του πόρου, τα τεχνικά χαρακτηριστικά του, σχετικές δημοσιεύσεις κ.α.), να κάνει **χρήση** του πόρου (όπως καταφόρτωση, επεξεργασία κτλ.) αλλά και να επιλέξει κάποιες **ενέργειες** ανάλογα με τον ρόλο που έχει (όπως να επιμεληθεί τα μεταδεδομένα, να επικυρώσει τον πορο, κτλ.).

Προσοχή: Παρόλο που όλοι οι χρήστες μπορούν να δουν τη σελίδα θέασης ενός πόρου, οι χρήσεις του πόρου *Process* και *Use* καθώς και οι ενέργειες (actions) προσφέρονται μόνο στους **συνδεδεμένους** χρήστες.

Η σελίδα θέασης χωρίζεται σε δύο τμήματα (το ανώτερο και το κατώτερο), καθένα από τα οποία υποδιαιρείται επιπλέον. Αν και η διάταξη της σελίδας θέασης είναι **κοινή** για όλους τους τύπους πόρων, τα μεταδεδομένα που παρουσιάζονται είναι **διαφορετικά**, ανάλογα με τον τύπο του κάθε πόρου. Το παράδειγμα που ακολουθεί παρουσιάζει τη σελίδα θέασης ενός **σώματος κειμένων**.

12.1 Το ανώτερο τμήμα

Το τμήμα αυτό περιέχει το **όνομα** και τη **συντομογραφία** του πόρου (εάν υπάρχει), τον **τύπο**, την **έκδοση** και το **PID** ακολουθούμενα από την **περιγραφή**. Κάτω από την περιγραφή, υπάρχει μια ετικέτα που δείχνει ότι ο πόρος αυτός είναι *processable* (επεξεργάσιμος). Στο πλαϊνό μέρος της σελίδας υπάρχει ένα κουμπί **επιλογής γλώσσας**. Εξ ορισμού, όλες οι εγγραφές μεταδεδομένων παρουσιάζονται στα αγγλικά (**en**) και αυτό διότι η γλώσσα του *περιβάλλοντος δημιουργίας και τεκμηρίωσης πόρων* αλλά και του *σχήματος μεταδεδομένων* είναι τα αγγλικά. Ωστόσο, υπάρχει η δυνατότητα προσθήκης μεταδεδομένων και σε άλλες γλώσσες, και εφόσον κάτι τέτοιο έχει γίνει από τον επιμελητή, ο χρήστης έχει τη δυνατότητα να διαλέξει τη γλώσσα (για παράδειγμα ελληνικά ή γαλλικά) στην οποία θα ήθελε να δει τα μεταδεδομένα. Η αλλαγή της γλώσσας αφορά **μόνο** τις τιμές που έχουν λάβει τα διάφορα πεδία μεταδεδομένων (και αυτό όπου ήταν δυνατόν) και όχι το σχήμα μεταδεδομένων το οποίο αυτή τη στιγμή διατίθεται μόνο στα αγγλικά. Ακολουθεί το λογότυπο του αποθετηρίου και ακριβώς από κάτω η βιβλιογραφική παραπομπή (citation text) και το PID ξανά. Για να **αντιγράψετε** την βιβλιογραφική παραπομπή κάντε κλικ στο εικονίδιο που φαίνεται στην εικόνα.

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)

Corpus

EUconst subcorpus EN-EL (Moses)

Version: 1

<http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A>

To EUconst subcorpus EN-EL (Moses) είναι ένα παράλληλο σώμα κειμένων για τα αγγλικά και ελληνικά που αποτελεί υποσύνολο του EUconst, a parallel corpus collected from the European Constitution (ένα παράλληλο σώμα κειμένων με υλικό από το Ευρωπαϊκό Σύνταγμα). 21 γλώσσες, 210 bitexts συνολικός αριθμός αρχείων: 986 συνολικός αριθμός μονάδων (tokens): 3.01M συνολικός αριθμός αποσπασμάτων προτάσεων: 0.22M

processable

Select Language

el

en

el



Ερευνα & Καινοτομία
Τεχνολογία Πληροφοριών

Cite

Citation text

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) (2015). Version 1. [Dataset (Text corpus)]. CLARIN:EL. <http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A>



Copy

Η σελίδα θέασης πόρου, παρέχει πρόσθετες λειτουργίες όταν είστε *συνδεδεμένοι* (εφόσον έχετε αναλάβει κάποιον *ρόλο* στο αποθετήριο σας). Έτσι, ο συνδεδεμένος χρήστης βλέπει το πλαίσιο *actions* (ενέργειες) που φαίνεται στην ακόλουθη εικόνα, το οποίο για τον μη συνδεδεμένο δεν είναι ορατό.

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)

Corpus

EUconst subcorpus EN-EL (Moses)

Version: 1

<http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A>

The EUconst subcorpus EN-EL (Moses) is a parallel subcorpus for English and Greek, subset of the EUconst, a parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 total number of tokens: 3.01M total number of sentence fragments: 0.22M

processable

Select Language

en



Cite

Citation text



A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) (2015). Version 1. [Dataset (Text corpus)]. CLARIN:EL. <http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A>

Actions

Actions

Πρακτική συμβουλή: Δείτε *εδώ* ποιες είναι οι ενέργειες που μπορείτε να κάνετε σε έναν πόρο.

Κάτω από την περιγραφή, υπάρχουν πλαίσια με πληροφορίες για ορισμένα από τα **βασικά χαρακτηριστικά** του πόρου. Για τα **σώματα κειμένων** και τους **λεξικούς/επικοινωνιακούς πόρους** αυτά είναι τα ακόλουθα: **Language** (γλώσσα), **Keywords** (λέξεις-κλειδιά που περιγράφουν το περιεχόμενο τους), **Domain** (θεματική περιοχή) και **Subclass** (υποκατηγορία του πόρου). Εφόσον τα αντίστοιχα μεταδεδομένα έχουν συμπληρωθεί, θα μπορούσαν επίσης να εμφανίζονται και τα **Time coverage** (χρονική κάλυψη) και **Geographic coverage** (γεωγραφική κάλυψη) του πόρου, όπως φαίνεται στην ακόλουθη εικόνα.

Language Modern Greek (1453-)	Keywords monolingual other	Coverage Time coverage  1989-1994 1997-2018 Geographic coverage  Greece
Domain KKE2516 Political Science	Corpus subclass annotated corpus	

Για τα **εργαλεία/υπηρεσίες**, αντιστοίχως, τα πλαίσια που παρουσιάζονται είναι τα **Function** (λειτουργία) και **Language dependent**, δηλαδή αν έχουν δημιουργηθεί για επεξεργασία υλικού σε συγκεκριμένη γλώσσα.

Function Discourse analysis	Language dependent no	Keywords text analysis concordances word frequencies n-grams
---------------------------------------	---------------------------------	-------------------------------------------------------------------------------







12.2 Το κατώτερο τμήμα

Το κατώτερο τμήμα έχει τις καρτέλες: *Overview* (Επισκόπηση), *Technical* (Τεχνικά χαρακτηριστικά), *Relations* (Σχέσεις) και *Access* (Πρόσβαση). Οι καρτέλες που εμφανίζονται σε μία σελίδα θέασης καθορίζονται από τον τύπο του πόρου αλλά και τα μεταδεδομένα που έχουν (ή δεν έχουν) συμπληρωθεί. Κάθε φορά που μία καρτέλα επιλέγεται από τον χρήστη, αλλάζει χρώμα από μπλε σε πορτοκαλί.

Overview	Technical	Relations	Access
----------	-----------	-----------	--------

12.2.1 1. Overview (Επισκόπηση)

Σε αυτήν την καρτέλα εμφανίζονται ορισμένα από τα *υποχρεωτικά μεταδεδομένα* για τα σώματα κειμένων, όπως τα γλωσσικά μεταδεδομένα: **Linguality** (ο αριθμός γλωσσών), **Multilinguality** (η σχέση των γλωσσών μεταξύ τους) και **Languages** (οι ίδιες οι γλώσσες). Όπως φαίνεται από τη σήμανση **TEXT** στα αριστερά, κάθε τμήμα από το οποίο (ενδεχομένως) απαρτίζεται ένα σώμα κειμένων παρουσιάζεται διακριτά με τις πληροφορίες που το αφορούν. Στην ίδια καρτέλα δίνονται πληροφορίες σχετικά με τις επισημειώσεις (**annotations**), αν έχουν γίνει, καθώς επίσης και για το είδος (**genre**) και την κατηγορία (**category**) του τμήματος υπό εξέταση. Η επόμενη στήλη σας δίνει την δυνατότητα να κάνετε εξαγωγή (**export**) της περιγραφής του πόρου σε ένα XML αρχείο¹. Από κάτω μπορείτε να μάθετε ποιος είναι ο πάροχος (**provider**) του πόρου και να βρείτε στοιχεία επικοινωνίας (**contact**), όπως το email ή/και μια ιστοσελίδα όπου μπορείτε να μάθετε περισσότερα για τον πόρο.

Information for Corpus part  <div> Language info Linguality type bilingual Multilinguality type parallel Language English, Modern Greek (1453-) </div> <div> Categories Text genre official Text type Category label administrative texts </div>	Export XML Resource provider  OPUS  Website Contact  Jörg Tiedemann  Email  Landing page
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Σε περίπτωση που το σώμα κειμένων απαρτίζεται από περισσότερα του ενός τμήματα, όπως προαναφέρθηκε, αυτά εμφανίζονται κάθετα στην αριστερή στήλη. Μπορείτε να δείτε τα μεταδεδομένα του κάθε τμήματος κάνο-

¹ Προς το παρόν η εγγραφή μεταδεδομένων μπορεί να εξαχθεί μόνο σε XML αρχείο αλλά σύντομα θα υποστηρίζονται και άλλες τύποι αρχείων.

ντας κλικ στο **μέσο** (εδώ text ή video). Το επιλεγμένο τμήμα επισημαίνεται με πορτοκαλί χρώμα, όπως φαίνεται στην επόμενη εικόνα.

Information for Corpus part

TEXT

VIDEO

Language info

Linguality type
multilingual

Multilinguality type
parallel

Language
Greek Sign Language, British Sign Language, German Sign Language, French Sign Language

Export

[XML](#)

Resource creator



School of Computing Sciences
[Website](#)



Research Institute of Computer Science
[Website](#)

Η καρτέλα Overview (Επισκόπηση) για τα εργαλεία/υπηρεσίες περιέχει πληροφορίες σχετικά με τα χαρακτηριστικά εισόδου (**input**) και εξόδου (**output**) του υλικού που επεξεργάζονται, ανάμεσα στα οποία είναι η γλώσσα (**language**), ο τύπος του πόρου (**processing resource type**), το μορφότυπο των δεδομένων (**data format**) κλπ.

Input

Input item

Language
Modern Greek (1453-)

Processing resource type
[user input text](#)

Media type
[text](#)

Output

Output item

Language
Modern Greek (1453-)


Processing resource type
[output text](#)

Media type
[text](#)


Export

[XML](#)

Resource creator

 Institute for Language and Speech
Processing
[Website](#)

Contact

 Athanassios Protopapas
[Email](#)

[Landing page](#)

Στο κάτω μέρος της σελίδας, εμφανίζονται, για όλους τους τύπους πόρων, τα σχετικά **άρθρα** και **έγγραφα τεκμηρίωσης**. Ειδικά στην περίπτωση των *σωμάτων κειμένων* και των *λεξικών/εννοιολογικών πόρων*, σε αυτό το σημείο υπάρχουν πληροφορίες σχετικά με το αν ο πόρος περιλαμβάνει προσωπικά (**personal**) ή/και ευαίσθητα προσωπικά (**sensitive**) δεδομένα.

Information for the resource

Documentation

Preferred citation
Jörg Tiedemann, 2009, News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov and K. Bontcheva and G. Angelova and R. Mitkov (eds.) Recent Advances in Natural Language Processing (vol V), pages 237-248, John Benjamins, Amsterdam/Philadelphia

Documented in
News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces

Ethics

Personal data included
no

Sensitive data included
no

12.2.2 2. Technical (Τεχνικά χαρακτηριστικά)

Προσοχή: Η καρτέλα αυτή εμφανίζεται **μόνο** στις περιγραφές των **εργαλείων/υπηρεσιών**!


Παρέχει πληροφορίες σχετικά με τη χρήση του του εργαλείου/υπηρεσίας όπως αυτή προβλέπεται από τον πάροχο και όπως γίνεται στην πραγματικότητα (**intended** vs **actual** use). Σε αυτήν την καρτέλα εμφανίζεται η πληροφορία για το αν το εργαλείο/υπηρεσία έχει αξιολογηθεί ή όχι (**evaluated**).

Actual use Used in application: Editing Language Technology	Evaluated: no	Intended application Language Technology Editing
-----------------------------------------------------------------------------	---------------	--------------------------------------------------------------


12.2.3 3. Relations (Σχέσεις)

Αυτή η καρτέλα περιέχει υπερσυνδέσμους για τους **πόρους που σχετίζονται** με τον υπό εξέταση πόρο (για παράδειγμα, εδώ εμφανίζονται σχέσεις μέρους-όλου, στοιχισμένες-μη στοιχισμένες εκδόσεις σωμάτων κειμένων κλπ.).

Relations to other resources

Part of
 A parallel corpus collected from the European Constitution (1)
<http://hdl.handle.net/11500/ATHENA-0000-0000-258B-A> (Handle)

Relations to other entities

Aligned versions
 A parallel subcorpus collected from the European Constitution (EN-EL) (TMX) 1


Προσοχή: Αυτή η καρτέλα δεν εμφανίζεται όταν ο πόρος δεν σχετίζεται με άλλους πόρους.

12.2.4 4. Access (Πρόσβαση)

Η καρτέλα **Access** (Πρόσβαση) περιέχει όλες τις πληροφορίες σχετικά με αυτό που αποκαλείται **Modes of distribution** (τρόποι διάθεσης), τις μορφές δηλαδή με τις οποίες ένας πόρος μπορεί να αποκτηθεί από τον χρήστη, π.χ. ως CD-ROM, ως υλικό που καταφορτώνεται από έναν εξωτερικό σύνδεσμο, κλπ. Κάθε μορφή διάθεσης συνοδεύεται από συγκεκριμένη αδειοδότηση, **licence terms** (όροι χρήσης), και, αν χρειάζεται, από άλλες πληροφορίες.



Modes of distribution

Download 

Sign in to process

Access info

Dataset distribution form
downloadable

Licensing Info

Licence

Creative Commons Attribution 4.0 International
<https://creativecommons.org/licenses/by/4.0/legalcode>
<https://creativecommons.org/licenses/by/4.0/>

Attribution text

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses). Άδεια: Creative Commons Attribution 4.0 International
(<https://creativecommons.org/licenses/by/4.0/legalcode>, <https://creativecommons.org/licenses/by/4.0/>). Πηγή: <http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A> (CLARIN:EL)

Features



Προσοχή: Αυτή η καρτέλα εμφανίζεται μόνο όταν ο πόρος συνοδεύεται από αρχεία περιεχομένου. Στην αντίθετη περίπτωση, όπως για παράδειγμα στα *meta-resources* παραλείπεται.


4.1 Download (Καταφόρτωση)

Εάν ο πόρος διατίθεται με άδεια χρήσης που επιτρέπει την καταφόρτωσή του, θα δείτε το κουμπί *Download* (Καταφόρτωση). Κάνοντας κλικ πάνω του θα λάβετε τα αρχεία περιεχομένου του πόρου.

Προσοχή: Παρακαλούμε ελέγξτε ποιες είναι οι προϋποθέσεις για να είναι ένας πόρος *προσβάσιμος* (και κατά συνέπεια καταφορτώσιμος) στην υποδομή.

4.2 Process (Επεξεργασία)

Εάν ο πόρος διαθέτει τα χαρακτηριστικά που τον καθιστούν συμβατό με τις ροές εργασίας της υποδομής και συνεπώς είναι *processable* (επεξεργάσιμος), εμφανίζεται το κουμπί *Process* (Επεξεργασία)². Δείτε, για παράδειγμα, *εδώ* πώς γίνεται επεξεργασία ενός δίγλωσσου σώματος κειμένων³.

[Download](#) 
[Process](#)

Access info	Licencing Info
Dataset distribution form downloadable	Licence Creative Commons Attribution 4.0 International https://creativecommons.org/licenses/by/4.0/legalcode https://creativecommons.org/licenses/by/4.0/

4.3 Use (Χρήση)

Προσοχή: Το κουμπί *Use* (Χρήση) εμφανίζεται μόνο σε εγγραφές μεταδεδομένων για εργαλεία/υπηρεσίες!

Εάν ο πόρος διατίθεται με άδεια χρήσης που σας επιτρέπει να τον χρησιμοποιήσετε, εμφανίζεται το κουμπί *Use*. Δείτε *εδώ* ένα παράδειγμα χρήσης εργαλείου/υπηρεσίας⁴.

[Use](#)

Access info	Licencing Info
Software distribution form docker image Operating system OS-independent Private yes	Licence CLARIN-EL Terms of Service http://inventory.clarin.ilsp.gr/catalogue_backend/static/licence/clarinelTOS.pdf Attribution text Annotator of Named Entities GrNE-Tagger used under CLARIN-EL Terms of Service (http://inventory.clarin.ilsp.gr/catalogue_backend/static/licence/clarinelTOS.pdf). Source: None (CLARIN:EL)

Features (Χαρακτηριστικά)

Προσοχή: Η ακόλουθη ενότητα εμφανίζεται **μόνο** σε περιγραφές **σωμάτων κειμένων και λεξικών/εννοιολογικών πόρων**!

Στο κάτω μέρος της σελίδας **Access** (Πρόσβαση), υπάρχει μια κατηγορία μεταδεδομένων που ονομάζεται **Features** (Χαρακτηριστικά). Για να εμφανιστούν αυτά τα χαρακτηριστικά κάντε κλικ στο βέλος. Θα δείτε για κάθε τμήμα (κειμενικό, ηχητικό, βίντεο ή/και εικόνα) του σώματος κειμένων ή του λεξικού/εννοιολογικού πόρου στοιχεία όπως τα **size** (μέγεθος), **data format** (μορφότυπος δεδομένων) και **encoding** (κωδικοποίηση).

² Εάν δεν έχετε *συνδεθεί*, το κουμπί σας προτρέπει να το κάνετε (*Sign in to process*). Μόλις το κάνετε, ανακατευθύνετε αυτόματα στη σελίδα θέασης του πόρου όπου πλέον έχει ενεργοποιηθεί και εμφανίζεται το κουμπί *Process* (Επεξεργασία).

³ Το σώμα κειμένων που χρησιμοποιήθηκε για αυτό το παράδειγμα είναι η Παράλληλη συλλογή κειμένων υποσύνολο του Ευρωπαϊκού Συντάγματος (EN-EL) (Moses).

⁴ Ο πόρος που χρησιμοποιήθηκε για αυτό το παράδειγμα είναι ο Επισημειωτής Ονοματικών Οντοτήτων GrNE-Tagger.

Features

Text feature

size

280000 word

534 kilobyte

Data format

MOSES plain text format

Character encoding

UTF-8

Επεξεργασία

Με τον όρο **επεξεργασία** (process) στο **CLARIN:EL** νοείται η εφαρμογή μίας **λειτουργίας** (*function*) σε ένα σώμα κειμένων (corpus). Οι λειτουργίες αναλύουν το σώμα κειμένων με διαφορετικό τρόπο και αποτέλεσμα κάθε φορά. Στην υποδομή οι λειτουργίες που μπορούν να γίνουν είναι ο **διαχωρισμός σε προτάσεις** (Sentence Splitting), η **αναγνώριση και επισημείωση λεκτικών μονάδων** (Tokenization), η **αναγνώριση και επισημείωση μερών του λόγου** (Below POS Tagging), η **αναγνώριση και επισημείωση λήμματος** (Lemmatization), η **επιφανειακή συντακτική ανάλυση και επισημείωση** (Chunking), η **συντακτική ανάλυση και επισημείωση** (Dependency Parsing), η **αναγνώριση και επισημείωση ονοματικών οντοτήτων** (Named Entity Recognition), η **κατηγοριοποίηση κειμένου** (Text Categorization) και η **ανάλυση λεκτικής επιθετικότητας** (Verbal aggression analysis). Κάθε μία από τις λειτουργίες αυτές για να γίνει χρειάζεται εργαλεία ή/και υπηρεσίες που υπάρχουν στον κεντρικό κατάλογο και συνολικά ονομάζονται **υπηρεσίες επεξεργασίας** (*processing services*). Κάθε φορά που ένας χρήστης θέλει να επεξεργαστεί ένα σώμα κειμένων ανακατευθύνεται στο **μητρώο ροών εργασίας** (*Workflow Registry*) και αυτό διότι τόσο οι λειτουργίες (functions) όσο και οι υπηρεσίες επεξεργασίας (processing services) υλοποιούνται με συγκεκριμένα βήματα που ονομάζονται **ροές εργασίας** (workflows). Αυτές οι ροές εργασίας συνδυάζονται τόσο με τα επεξεργάσιμα σώματα κειμένων¹ που διαθέτει η υποδομή όσο και με υλικό συγκεκριμένων προδιαγραφών που μπορεί να μεταφορτώσει ο ίδιος ο χρήστης. Η συγκεκριμένη αρχιτεκτονική έχει υιοθετηθεί για να διευκολύνει τον χρήστη που μπορεί από διαφορετικά σημεία στην υποδομή να ξεκινήσει μία επεξεργασία. Οι διαφορετικές διαδρομές περιγράφονται στη συνέχεια.

Προσοχή: Μόνο οι **συνδεδεμένοι χρήστες** μπορούν να επεξεργαστούν σώματα κειμένων. Αν δεν έχετε λογαριασμό, δείτε πρώτα πώς να κάνετε *εγγραφή/σύνδεση*

¹ Όλα τα σώματα κειμένων που πληρούν αυτά τα κριτήρια επισημαίνονται ως *processable* (επεξεργάσιμα) και παρουσιάζονται ως προεπιλογή στην αρχική σελίδα του κεντρικού καταλόγου.

13.1 1. Ξεκινώντας από τα σώματα κειμένων

Στην υποδομή υπάρχουν σώματα κειμένων που μπορείτε να επεξεργαστείτε. Αυτά επισημαίνονται με την ετικέτα *processable*² και μπορείτε να τα βρείτε συγκεντρωμένα ως *προεπιλογή* στην αρχική σελίδα του κεντρικού καταλόγου. Πρόκειται για σώματα κειμένων που είναι είτε *μονόγλωσσα* (στα *ελληνικά*, *αγγλικά*, *γερμανικά* ή *πορτογαλικά*) είτε *δύγλωσσα* (με ζεύγη γλωσσών που περιλαμβάνουν τα *ελληνικά* ως τη μία γλώσσα και τα *αγγλικά*, *γερμανικά* ή *πορτογαλικά* ως την άλλη).

Για να επεξεργαστείτε ένα από αυτά τα σώματα κειμένων, ακολουθήστε τα επόμενα **βήματα**:

13.1.1 Βήμα 1. Επιλέξτε ένα σώμα κειμένων

Επιλέξτε ένα σώμα κειμένων κάνοντας χρήση των *φίλτρων*, με λέξεις κλειδιά στο *πεδίο αναζήτησης* ή συνδυαστικά. Ο πόρος που επιλέχθηκε για το σενάριο που θα εξετάσουμε είναι μία παράλληλη συλλογή κειμένων σε ελληνικά και αγγλικά, υποσύνολο του Ευρωπαϊκού Συντάγματος και σε μορφότυπο txt (moses): *A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)*.

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses)

 Corpus

EUconst subcorpus EN-EL (Moses)

Version: 1

<http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A>

The EUconst subcorpus EN-EL (Moses) is a parallel subcorpus for English and Greek, subset of the EUconst, a parallel corpus collected from the European Constitution. 21 languages, 210 bitexts total number of files: 986 total number of tokens: 3.01M total number of sentence fragments: 0.22M

 processable


Select Language

en



Cite

Citation text

A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) (2015). Version 1. [Dataset (Text corpus)]. CLARIN-EL. <http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A> 

Μετακινηθείτε στο *κατώτερο τμήμα* της σελίδας του πόρου και επιλέξτε την καρτέλα **Access** (Πρόσβαση). Στη συνέχεια, κάντε κλικ στην επιλογή *Process* (Επεξεργασία)³.

² Η ετικέτα αυτή εμφανίζεται τόσο στο απόσπασμα περιγραφής ενός πόρου στον *κεντρικό κατάλογο* όσο και στη *σελίδα θέασης του πόρου*.

³ Εάν δεν έχετε *συνδεθεί*, το κουμπί σας προτρέπει να το κάνετε (*Sign in to process*). Μόλις το κάνετε, ανακατευθύνετε αυτόματα στη σελίδα θέασης του πόρου όπου πλέον έχει ενεργοποιηθεί και εμφανίζεται το κουμπί *Process* (Επεξεργασία).

Overview Relations **Access**

Modes of distribution

Download Process

Access info

Dataset distribution form
downloadable

Licencing Info

Licence
Creative Commons Attribution 4.0 International
<https://creativecommons.org/licenses/by/4.0/legalcode>
<https://creativecommons.org/licenses/by/4.0/>Attribution text
A parallel subcorpus collected from the European Constitution (EN-EL) (Moses) used under Creative Commons Attribution 4.0 International
(<https://creativecommons.org/licenses/by/4.0/legalcode>, <https://creativecommons.org/licenses/by/4.0/>). Source: <http://hdl.handle.net/11500/ATHENA-0000-0000-23FF-A> (CLARIN:EL)

Features

13.1.2 Βήμα 2. Επιλέξτε μία λειτουργία

Μόλις κάνετε κλικ στο κουμπί *Process* (Επεξεργασία), θα οδηγηθείτε στο *μητρώο ροών εργασίας*. Εδώ θα πρέπει πρώτα να διαλέξετε *λειτουργία* (function). Ανάλογα με τα χαρακτηριστικά που έχει το σώμα κειμένων που επιλέξατε θα δείτε τις διαθέσιμες (όλες ή κάποιες) λειτουργίες στην οριζόντια μπάρα. Στη συνέχεια για τη συγκεκριμένη λειτουργία θα πρέπει να επιλέξετε την ροή εργασίας που θέλετε. Για κάθε λειτουργία υπάρχουν μία ή περισσότερες ροές εργασίας (ο αριθμός τους εξαρτάται από τα διαθέσιμα και ενσωματωμένα στην υποδομή εργαλεία/υπηρεσίες και τη γλώσσα). Και πάλι θα δείτε **μόνο τις διαθέσιμες** ροές εργασίας. Εφόσον το σώμα κειμένων που θα επεξεργαστείτε είναι δίγλωσσο, σας παρουσιάζονται οι ροές εργασίας για την κάθε γλώσσα χωριστά με πρώτες τις επιλογές για το **ελληνικό τμήμα** (ειδοποίηση εμφανίζεται στο πάνω μέρος της σελίδας).



Available services and functions for Modern Greek (1453-)

BelowPosTagging NamedEntityRecognition Chunking

ILSP Feature-based multi-tiered POS Tagger

The ILSP Feature-based multi-tiered POS Tagger is a language processing workflow that annotates every word of a text with the corresponding part of speech (POS) tag, e.g. noun, verb, adjective, adverb, etc., based on its context. This tool performs POS tagging for texts in Modern Greek and it is the last step of this workflow, which consists of the ILSP Sentence Splitter and Tokenizer and the POS Tagger. It accepts as input plain text (txt) in UTF8 encoding and generates as output an XMI document with POS tags assigned to each token, besides labelled sentences and tokens.

Input

input_language
Modern Greek (1453-)
input_character_encoding
UTF-8
input_data_format
MOSESFormat

Output

output_language
Modern Greek (1453-)
output_character_encoding
UTF-8
output_data_format
Xmi

Use this workflow

HBrill

HBrill is used for annotating a word in Greek texts with the corresponding part of speech (POS) tag (e.g. noun, verb, adjective, adverb, etc.). It accepts as input XML documents in UTF8 encoding, and generates as output a standoff XML document in UTF8 encoding, with part of speech annotations. It belongs to the language processing workflows of NCSR Demokritos.

Input

input_language
Modern Greek (1453-)
input_character_encoding
UTF-8
input_data_format
MOSESFormat

Output

output_language
Modern Greek (1453-)
output_character_encoding
UTF-8
output_data_format
Xmi

Use this workflow

cancel Select service for Modern Greek (1453-)

Διαλέξτε μία ροή εργασίας και πατήστε το **use this workflow**. Το χρώμα του κουμπιού αλλάζει αυτόματα από γαλάζιο σε πράσινο. Στη συνέχεια κάντε κλικ στο *Select service for Modern Greek (1453-)* για να προχωρήσετε.

Available services and functions for Modern Greek (1453-)

[BelowPosTagging](#) [NamedEntityRecognition](#) [Chunking](#)

ILSP Feature-based multi-tiered POS Tagger

Input	Output
input_language Modern Greek (1453-)	output_language Modern Greek (1453-)
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format MOSESFormat	output_data_format Xmi

Use this workflow

HB Brill

Input	Output
input_language Modern Greek (1453-)	output_language Modern Greek (1453-)
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format MOSESFormat	output_data_format Xmi

Use this workflow

cancel [Select service for Modern Greek \(1453-\)](#)

Επαναλάβετε την ίδια διαδικασία για το **αγγλικό τμήμα**.

Available services for BelowPosTagging English

[BelowPosTagging](#)

OpenNLP Part-of-Speech Tagger (English)

Input	Output
input_language English	output_language English
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format MOSESFormat	output_data_format Xmi

Use this workflow

Back cancel [Select service for English](#)

Εφόσον έχει ολοκληρωθεί η επιλογή και για τις δύο γλώσσες, εμφανίζεται ένα νέο παράθυρο που σας ζητά να ελέγξετε τις ροές εργασίας πριν την τελική υποβολή.

[Review your selections and press submit to start the processing or back to choose again](#)

BelowPosTagging - ILSP Feature-based multi-tiered POS Tagger

Input	Output
input_language Modern Greek (1453-)	output_language Modern Greek (1453-)
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format MOSESFormat	output_data_format Xmi

BelowPosTagging - OpenNLP Part-of-Speech Tagger (English)

Input	Output
input_language English	output_language English
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format MOSESFormat	output_data_format Xmi

Back cancel [Submit for process](#)

Μόλις πατήσετε το κουμπί **Submit for process** (Υποβολή για επεξεργασία) εμφανίζεται το ακόλουθο μήνυμα: *θα ειδοποιηθείτε με email όταν ολοκληρωθεί η επεξεργασία*.

The dataset is being processed. You will be notified by email when it's finished.

[close](#)

13.1.3 Βήμα 3. Καταφορτώστε τα επεξεργασμένα αρχεία

Θα λάβετε ειδοποίηση με email μόλις η επεξεργασία ολοκληρωθεί. Για να δείτε τα αποτελέσματα μεταβείτε στο **dashboard** (πίνακας ελέγχου) και ελέγξτε τις εργασίες επεξεργασίας (*Processing tasks*).

Προσοχή: Δημιουργείται **αυτόματα** μία περιγραφή μεταδεδομένων με τα επισημειωμένα δεδομένα και ο πόρος δημοσιεύεται στον κεντρικό κατάλογο.

13.2 2. Ξεκινώντας από τη λειτουργία

13.2.1 2.1 Μητρώο ροής εργασιών

Για να δείτε τις διαθέσιμες λειτουργίες επεξεργασίας, πρέπει να μεταβείτε στο μητρώο ροής εργασιών. Πρόσβαση έχετε από την *αρχική σελίδα* του κεντρικού καταλόγου ή από το *dashboard* (πίνακας ελέγχου).

1 ————— 2 ————— 3
 Select function Upload data Process

SentenceSplitting
Tokenization
MorphosyntacticTagging
Lemmatization
DependencyParsing
NamedEntityRecognition
Chunking

TextCategorization
Verbal aggression analysis

ILSP Sentence splitter and tokenizer for Greek text (ILSP workflow v1.0.0)

The ILSP Sentence splitter and tokenizer for Greek texts splits the input text into smaller units, such as sentences, words, punctuation marks, dates, numbers or symbols. It is the first step of all ILSP language processing workflows for Greek texts. It accepts as input plain text (txt) in UTF8 encoding and generates as output a standoff XML document in UTF8 encoding, with labelled sentences and tokens.

Aa...λ Input	Output
input_language el	output_language el
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format Text	output_data_format Xmi
input_typesystem	output_typesystem ILSP_NLP typesystem

Use this workflow

OpenNLP Sentence Detector (English) (OpenNLP workflow v1.0.0)

The OpenNLP Sentence Detector for English splits the input English text into smaller units, such as sentences, words, punctuation marks, dates, numbers or symbols. It accepts as input plain text (txt) in UTF8 encoding and generates as output XML document in UTF8 encoding. It is the first step of OpenNLP language processing workflows for English.

Aa...λ Input	Output
input_language en	output_language en
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format Text	output_data_format Xmi
input_typesystem	output_typesystem ILSP_OpenNLP typesystem

Use this workflow

OpenNLP Sentence Detector (German) (OpenNLP workflow v1.0.0)

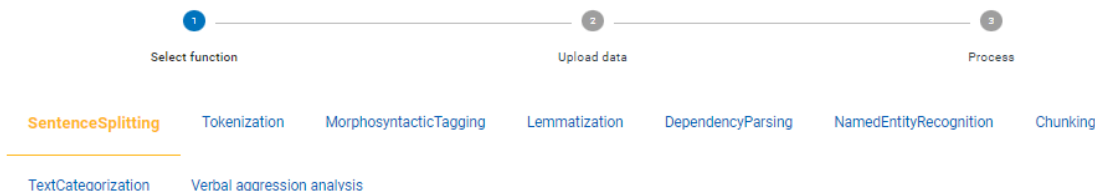
The OpenNLP Sentence Detector for German splits the input German text into smaller units, such as sentences, words, punctuation marks, dates, numbers or symbols. It accepts as input plain text (txt) in UTF8 encoding and generates as output XML document in UTF8 encoding. It is the first step of OpenNLP language processing workflows for German.

OpenNLP Sentence Detector (Portuguese) (OpenNLP workflow v1.0.0)

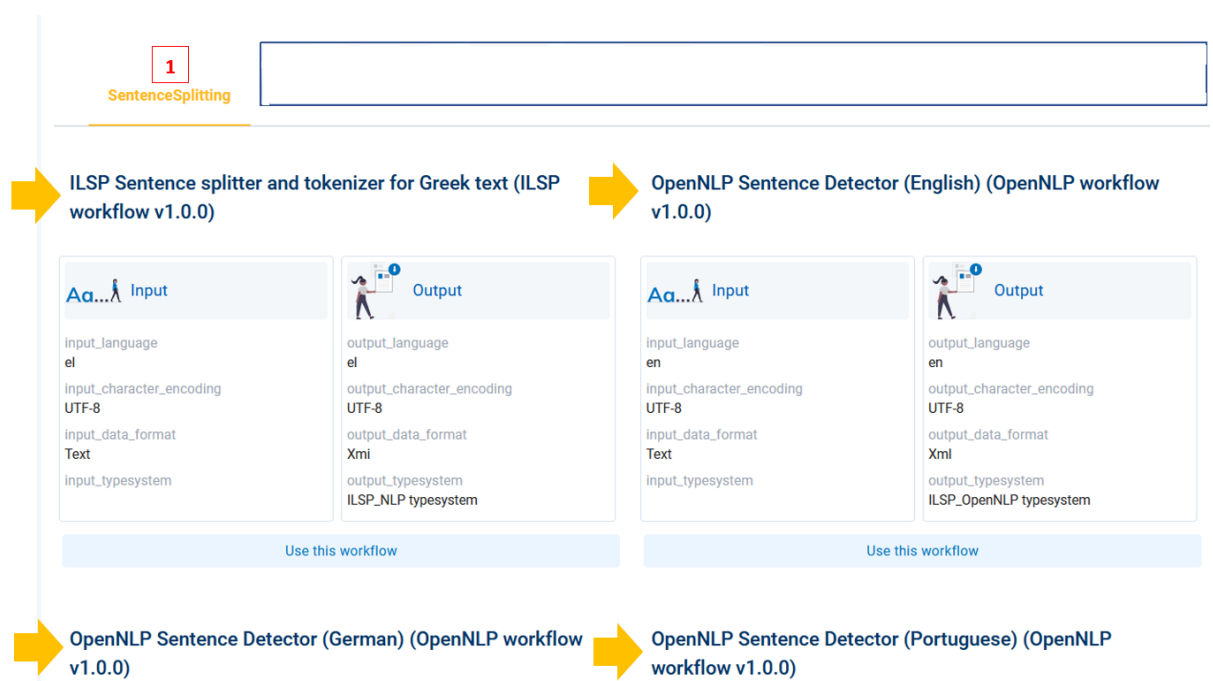
The OpenNLP Sentence Detector for Portuguese splits the input Portuguese text into smaller units, such as sentences, words, punctuation marks, dates, numbers or symbols. It accepts as input plain text (txt) in UTF8 encoding and generates as output XML document in UTF8 encoding. It is the first step of OpenNLP language processing workflows for Portuguese.

Αυτήν τη στιγμή η υποδομή υποστηρίζει τις ακόλουθες **εννέα** λειτουργίες: **Sentence Splitting** (διαχωρισμός σε προτάσεις), **Tokenization** (αναγνώριση και επισημείωση λεκτικών μονάδων), **Below POS Tagging** (αναγνώριση και επισημείωση μερών του λόγου), **Lemmatization** (αναγνώριση και επισημείωση λήμματος), **Dependency**

Parsing (συντακτική ανάλυση και επισημείωση), **Named Entity Recognition** (αναγνώριση και επισημείωση ονοματικών οντοτήτων), **Chunking** (επιφανειακή συντακτική ανάλυση και επισημείωση), **Text Categorization** (κατηγοριοποίηση κειμένου) και **Verbal aggression analysis** (ανάλυση λεκτικής επιθετικότητας).



Για κάθε λειτουργία, π.χ. **sentence splitting** (διαχωρισμός προτάσεων), το CLARIN:EL προσφέρει μία ή περισσότερες ροές εργασίας, όπως φαίνεται στην ακόλουθη εικόνα. Ως ροή εργασίας νοούνται τα αλληπάλλληλα στάδια που απαιτούνται προκειμένου να υλοποιηθεί η λειτουργία. Ο αριθμός των ροών εργασίας ανά λειτουργία εξαρτάται από τα διαθέσιμα και ενσωματωμένα στην υποδομή εργαλεία και τη γλώσσα.



Τα **βήματα** για τη χρήση του μητρώου ροών εργασίας είναι τα εξής:

Βήμα 1. Επιλέξτε μία λειτουργία

Επιλέξτε μια λειτουργία, ανάλογα με το είδος της επεξεργασίας που θέλετε να εκτελέσετε, κάνοντας κλικ στο όνομά της. Η επιλεγμένη λειτουργία, π.χ. **tokenization** (αναγνώριση λεκτικών μονάδων) αλλάζει χρώμα, από μπλε σε πορτοκαλί.

Βήμα 2. Επιλέξτε μία ροή εργασίας

Για τη συγκεκριμένη λειτουργία διατίθενται πολλαπλές ροές εργασίας: κάποιες για ελληνικά σώματα κειμένων, μία για αγγλικά, μία για γερμανικά και μία για πορτογαλικά. Επιλέξτε τη ροή εργασίας που χρειάζεστε κάνοντας κλικ στο *Use this workflow* (Χρήση της επιλεγμένης ροής εργασίας).

ILSP Sentence splitter and tokenizer for Greek text (ILSP workflow v1.0.0)

Input	Output
input_language el	output_language el
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format Text	output_data_format Xmi
input_typesystem	output_typesystem ILSP_NLP typesystem

Use this workflow

OpenNLP Tokenizer (English) (OpenNLP workflow v1.0.0)

Input	Output
input_language en	output_language en
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format Text	output_data_format Xmi
input_typesystem ILSP_OpenNLP typesystem	output_typesystem ILSP_OpenNLP typesystem

Use this workflow

Βήμα 3. Μεταφορτώστε το δικό σας σύνολο δεδομένων

Στο νέο παράθυρο βλέπετε ποια είναι τα τεχνικά χαρακτηριστικά που θα πρέπει να έχει το δικό σας σύνολο δεδομένων ώστε να επιτραπεί η μεταφόρτωση και να είναι εφικτή η επεξεργασία.

Προσοχή: Μπορείτε να μεταφορτώσετε **μόνο μονόγλωσσα σώματα δεδομένων στα ελληνικά, αγγλικά, γερμανικά ή πορτογαλικά**. Οι ροές εργασίας μπορούν επίσης να χρησιμοποιηθούν στα δίγλωσσα σώματα κειμένων της υποδομής τα οποία έχουν επισημανθεί ως **επεξεργάσιμα**.

You can now upload your zip file, which should

1. have a size of max 2MB
2. have a filename in Latin characters, with no spaces in it
3. contain only plain txt files in UTF-8 encoding
4. not contain other zip files

Drag & Drop your files or Browse

ILSP Sentence splitter and tokenizer for Greek text (ILSP workflow v1.0.0)

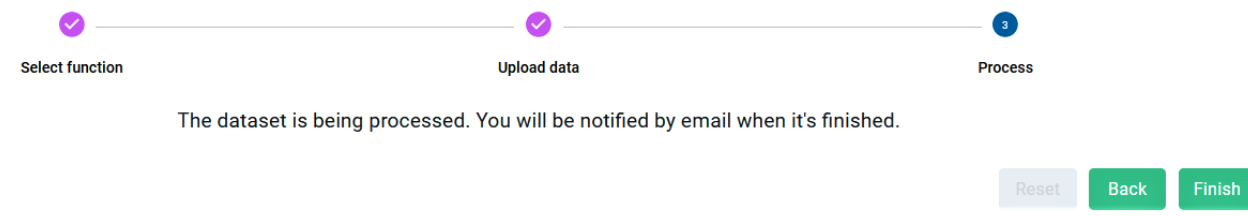
Input	Output
input_language el	output_language el
input_character_encoding UTF-8	output_character_encoding UTF-8
input_data_format Text	output_data_format Xmi
input_typesystem	output_typesystem ILSP_NLP typesystem

Reset Back Next

File has been successfully uploaded.

Need Help ?

Μετά την επιτυχή μεταφόρτωση, ενεργοποιείται το κουμπί **Next** (Επόμενο) το οποίο και πρέπει να πατήσετε.



Βήμα 4. Καταφορτώστε τα επεξεργασμένα αρχεία

Θα λάβετε ειδοποίηση με email μόλις η επεξεργασία ολοκληρωθεί. Για να δείτε τα αποτελέσματα μεταβείτε στο **dashboard** (πίνακας ελέγχου) και ελέγξτε τις εργασίες επεξεργασίας (*Processing tasks*).

Προσοχή: Τα δεδομένα που μεταφορτώνονται και τα δεδομένα που παράγονται από την επεξεργασία **δεν αποθηκεύονται** μόνιμα στην υποδομή. Σύμφωνα με την πολιτική του CLARIN:EL τα αρχεία που προέκυψαν από εργασίες επεξεργασίας **διαγράφονται αυτόματα 48 ώρες** μετά την ολοκλήρωση της επεξεργασίας. Σε περίπτωση που επιθυμείτε να τα καταφορτώσετε πρέπει να το κάνετε πριν παρέλθει το διάστημα αυτό.

13.2.2 2.2 Υπηρεσίες επεξεργασίας

Με τον όρο **processing services** (υπηρεσίες επεξεργασίας) χαρακτηρίζονται τα εργαλεία/υπηρεσίες που φιλοξενοούνται στο CLARIN:EL και χρησιμοποιούνται για να γίνουν οι διάφορες λειτουργίες (functions) επεξεργασίας. Οι διαθέσιμες υπηρεσίες επεξεργασίας εμφανίζονται με την εφαρμογή του αντίστοιχου φίλτρου στον κεντρικό κατάλογο. Για να χρησιμοποιήσετε μία εξ αυτών, ακολουθήστε τα επόμενα **βήματα**:

Clear all filters 19 search results

yes

Repository

Resource type

Languages

Processable

Processing service (19)

Service functions

Licences

Annotator of Named Entities GrNE-Tagger (processing service, docker image)

Ο Επισημειωτής Ονοματικών Οντοτήτων GrNE-Tagger επεξεργάζεται κείμενα αναγνωρίζοντας αυτόματα βάσει κανόνων και ταξινομώντας τις Ονοματικές Οντότητες (κύρια ονόματα) που αυτά περιέχουν στις ακόλουθες πέντε κατηγορίες: Ά...

Language: Modern Greek (1453-)

Keywords: named entities text

HBrill (processing service, docker image)

HBrill

Language: Modern Greek (1453-)

Keywords: segmentation writtenLanguage morphosyntactic analysis text

HNPPChunker (processing service, docker image)

HNPPChunker

Language: Modern Greek (1453-)

Keywords: segmentation writtenLanguage morphosyntactic analysis text

Βήμα 1: Επιλέξτε μία υπηρεσία

Κάντε κλικ στο όνομα της υπηρεσίας που θέλετε να χρησιμοποιήσετε. Θα μεταφερθείτε στην αντίστοιχη σελίδα θέασης. Μετακινηθείτε στο κατώτερο τμήμα της σελίδας και επιλέξτε την καρτέλα **Access** (Πρόσβαση).

Annotator of Named Entities GrNE-Tagger

 ToolService

GrNE-Tagger

Version: 0.4

<http://hdl.handle.net/11500/ATHENA-0000-0000-23F2-7>

Ο Επισημειωτής Ονοματικών Οντοτήτων GrNE-Tagger επεξεργάζεται κείμενα αναγνωρίζοντας αυτόματα βάσει κανόνων και ταξινομώντας τις Ονοματικές Οντότητες (κύρια ονόματα) που αυτά περιέχουν στις ακόλουθες πέντε κατηγορίες: Άτομο/Πρόσωπο (PERSON), Τοποθεσία (LOCATION), Οργανισμός (ORGANIZATION), Ονόματα κτιρίων ή/και άλλων ανθρώπινων κατασκευών (FACILITY), Γεωπολιτική Οντότητα (Geopolitical entity - GPE). Ο GrNE-Tagger χρησιμοποιείται για την επεξεργασία κειμένων της νέας Ελληνικής και δέχεται ως είσοδο κείμενα σε μορ ... [Read More](#)

Select Language

en



Cite

Citation text

Annotator of Named Entities GrNE-Tagger
(2015). Version 0.4. [Software (Tool/Service)]
CLARIN:EL: None

Function

Named Entity Recognition

Language dependent

yes

Keywords

named entities text

[Overview](#)

[Technical](#)

[Relations](#)

[Access](#)



Modes of distribution

[Use](#)

Access info

Software distribution form
docker image
Operating system

Licencing Info

Licence
CLARIN-EL Terms of Service
http://inventory.clarin.llsp.gr/catalogue_backend/static/licence/clarinelTOS.pdf

Κάντε κλικ στο κουμπί *Use* (Χρήση). Στο επόμενο παράθυρο, θα σας παρουσιαστεί η ροή εργασίας (workflow) που αντιστοιχεί στο εργαλείο/υπηρεσία που επιλέξατε. Πρέπει να κάνετε κλικ στο **Use this workflow** (Χρήση της επιλεγμένης ροής εργασίας).

1

2

3

Select function

Upload data

Process

NamedEntityRecognition

Aa... Input

input_language
el

input_character_encoding
UTF-8

input_data_format
Xml

input_typesystem

Output

output_language
el

output_character_encoding
UTF-8

output_data_format
Xml

output_typesystem

Use this workflow

Βήμα 2. Μεταφορτώστε το δικό σας σύνολο δεδομένων

Στο νέο παράθυρο βλέπετε ποια είναι τα τεχνικά χαρακτηριστικά που θα πρέπει να έχει το δικό σας σύνολο δεδομένων ώστε να επιτραπεί η μεταφόρτωση και να είναι εφικτή η επεξεργασία.

✓

2

3

Select function

Upload data

Process

Aa... Input

input_language
el

input_character_encoding
UTF-8

input_data_format
Xml

input_typesystem

Output

output_language
el

output_character_encoding
UTF-8

output_data_format
Xml

output_typesystem

Reset

Back

Next

You can now upload your zip file, which should

1. have a size of max 2MB
2. have a filename in Latin characters, with no spaces in it
3. contain only plain txt files in UTF-8 encoding
4. not contain other zip files

Drag & Drop your files or Browse

File has been successfully uploaded.

Μετά την επιτυχή μεταφόρτωση, ενεργοποιείται το κουμπί **Next** (Επόμενο) το οποίο και πρέπει να πατήσετε.

✓

✓

3

Select function

Upload data

Process

Aa... Input

input_language
el

input_character_encoding
UTF-8

input_data_format
Xml

input_typesystem

Output

output_language
el

output_character_encoding
UTF-8

output_data_format
Xml

output_typesystem

Reset

Back

Next

The dataset is being processed. You will be notified by email when it's finished.

Reset

Back

Finish

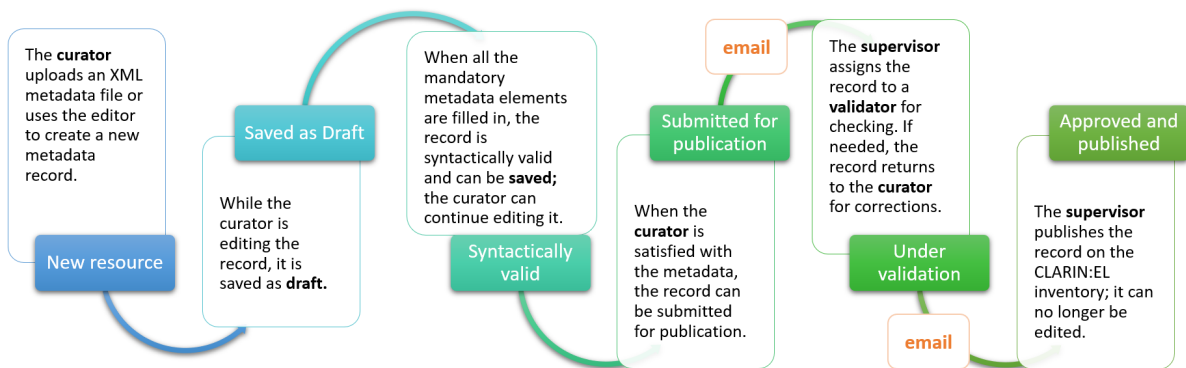
Βήμα 3. Καταφορτώστε τα επεξεργασμένα αρχεία

Θα λάβετε ειδοποίηση με email μόλις η επεξεργασία ολοκληρωθεί. Για να δείτε τα αποτελέσματα μεταβείτε στο **dashboard** (πίνακας ελέγχου) και ελέγξτε τις εργασίες επεξεργασίας (*Processing tasks*).

Προσοχή: Τα δεδομένα που μεταφορτώνονται και τα δεδομένα που παράγονται από την επεξεργασία **δεν αποθηκεύονται** μόνιμα στην υποδομή. Σύμφωνα με την πολιτική του CLARIN:EL τα αρχεία που προέκυψαν από εργασίες επεξεργασίας **διαγράφονται αυτόματα 48 ώρες** μετά την ολοκλήρωση της επεξεργασίας. Σε περίπτωση που επιθυμείτε να τα καταφορτώσετε πρέπει να το κάνετε πριν παρέλθει το διάστημα αυτό.

Ο κύκλος ζωής του πόρου

Το CLARIN:EL είναι μια υποδομή για γλωσσικούς πόρους και εργαλεία/υπηρεσίες που δημιουργούνται στα αποθετήρια των μελών του δικτύου CLARIN:EL και εν συνεχεία συλλέγονται και παρουσιάζονται στον κεντρικό κατάλογο. Ο πόρος (δηλ. η εγγραφή μεταδεδομένων και, ενίοτε, τα αρχεία περιεχομένου που αντιστοιχούν στο φυσικό αντικείμενο) περνάει από μία σειρά αλληπάλλων σταδίων κατά τη διαδικασία προετοιμασίας για δημοσίευση στον κατάλογο, όπως απεικονίζεται στο ακόλουθο σχήμα:



Αναλυτικά, κατά τα στάδια αυτά συμβαίνουν τα ακόλουθα:

- Ο πόρος γεννιέται στο αποθετήριο ως **νέα εγγραφή**: ο *επιμελητής* δημιουργεί έναν πόρο, φτιάχνοντας μια

εγγραφή μεταδεδομένων μέσω του *περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων*¹ ή μεταφορτώνοντας ένα *αρχείο μεταδεδομένων XML*. Σε περίπτωση που για την ολοκλήρωση της περιγραφής του πόρου χρειάζεται να γίνει μεταφόρτωση αρχείων περιεχομένου, ο επιμελητής μπορεί να επιλέξει αν θα το κάνει κατά την *έναρξη* της διαδικασίας δημιουργίας ή *μεταγενέστερα*.

- Ο πόρος αποθηκεύεται ως **προσχέδιο** (στάδιο 1): ο επιμελητής μπορεί να αποθηκεύσει² την εγγραφή μεταδεδομένων ως προσχέδιο (save draft), χωρίς να έχει συμπληρώσει όλα τα υποχρεωτικά μεταδεδομένα. Κατά την αποθήκευση, η υποδομή ελέγχει μόνο αν οι τιμές που έχουν δοθεί στα διάφορα μεταδεδομένα είναι σύμφωνες με τις προδιαγραφές του *σχήματος* (π.χ. στις περιπτώσεις που πρέπει να συμπληρωθεί ένα πεδίο με URL η υποδομή ελέγχει αν περιέχει μόνο λατινικούς χαρακτήρες και έχει τη σωστή σύνταξη).
- Ο πόρος είναι **συντακτικά έγκυρος** (στάδιο 2): η υποδομή ελέγχει αν η εγγραφή συμφωνεί με το σχήμα μεταδεδομένων *CLARIN:EL metadata schema* και όλα τα *υποχρεωτικά* μεταδεδομένα έχουν συμπληρωθεί. Ο επιμελητής μπορεί να συνεχίσει να *επεξεργάζεται* τον πόρο μέχρι τη στιγμή που θα τον υποβάλει για δημοσίευση (**submit for publication**). Μόλις ο επιμελητής υποβάλει τον πόρο για δημοσίευση (submitted), παύει η δυνατότητα επεξεργασίας των μεταδεδομένων του (στάδιο 3).
- ο πόρος **ανατίθεται σε επικυρωτή**: ο *υπεύθυνος αποθετηρίου*³ λαμβάνει ένα email που τον ενημερώνει για την υποβολή και στην συνέχεια καλείται να αναθέσει τον πόρο σε *επικυρωτές*. Ανάλογα με τον τύπο του πόρου (και τα μεταφορτωμένα αρχεία περιεχομένου, εάν υπάρχουν), ο πόρος ελέγχεται από τον επικυρωτή μεταδεδομένων και τον επικυρωτή νομικών θεμάτων⁴. Η επικύρωση αποσκοπεί στο να ελεγχθεί η συνέπεια της περιγραφής και δεν περιλαμβάνει καμία ποιοτική αξιολόγηση. Σε περίπτωση που οι επικυρωτές εντοπίσουν κάποιο πρόβλημα, επικοινωνούν με τον επιμελητή για να λάβουν περισσότερες πληροφορίες και μπορεί να ζητήσουν να γίνει επιπλέον επεξεργασία των μεταδεδομένων. Σε αυτές τις περιπτώσεις, η κατάσταση του πόρου από το **στάδιο 3 (submitted)** επανέρχεται στο **στάδιο 2 (syntactically valid)** και ο επιμελητής λαμβάνει ειδοποίηση προκειμένου να προβεί στις κατάλληλες αλλαγές/διορθώσεις.
- ο πόρος **δημοσιεύεται** (στάδιο 5): όταν οι επικυρωτές έχουν **εγκρίνει** (approved) έναν πόρο (στάδιο 4), ο υπεύθυνος αποθετηρίου ενημερώνεται με email και μπορεί να προβεί στην δημοσίευσή του. Μόλις ο πόρος εμφανιστεί στον κεντρικό κατάλογο CLARIN:EL, περνά στο επόμενο στάδιο και η εγγραφή μεταδεδομένων παύει να είναι τροποποιήσιμη. Η επιστροφή του πόρου στην κατάσταση *συντακτικά έγκυρος* (syntactically valid) γίνεται μόνο εάν υπάρξει αίτημα άρσης δημοσίευσης (**requested for unpublishing**), (στάδιο 6). Εάν ο υπεύθυνος αποθετηρίου κρίνει πως πρέπει να γίνει δεκτό το αίτημα, ο πόρος **αποδημοσιεύεται**, (στάδιο 7).

Προσοχή: Μόνο οι εγγραφές μεταδεδομένων που δημιουργούνται μέσω του περιβάλλοντος τεκμηρίωσης και ανάρτησης μπορούν να αποθηκευτούν ως **προσχέδιο**. Τα αρχεία μεταδεδομένων XML εισάγονται στην υποδομή μετά από έλεγχο κατά τον οποίο διαπιστώνεται αν είναι **συντακτικά έγκυρα** και κατά συνέπεια αποκτούν αυτομάτως αυτόν τον χαρακτηρισμό.

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

² Δείτε τις διαφορές μεταξύ της *αποθήκευσης ως προσχέδιο* και της *αποθήκευσης εδώ*.

³ Όταν υπάρχουν περισσότεροι του ενός υπεύθυνοι σε ένα αποθετήριο, θα πρέπει ο πόρος πρώτα να ανατεθεί σε κάποιον από αυτούς. Δείτε *εδώ* πώς να το κάνετε αυτό. Στη συνέχεια, αυτός ο υπεύθυνος πρέπει να επιλέξει και να *αναθέσει* επικυρωτές στον πόρο.

⁴ Όταν δεν έχουν μεταφορτωθεί αρχεία περιεχομένου, ο πόρος θεωρείται αυτόματα νομικά έγκυρος.

Σημαντικές πληροφορίες

Πριν ξεκινήσετε να δημιουργείτε γλωσσικούς πόρους και εργαλεία στην υποδομή CLARIN:EL θα πρέπει να λάβετε υπόψη σας τα ακόλουθα:

- Οι Όροι Χρήσης των Υπηρεσιών CLARIN:EL¹ είναι **δεσμευτικοί** και για το λόγο αυτό παρακαλείστε να τους διαβάσετε προσεκτικά.
- Το νομικό καθεστώς σύμφωνα με το οποίο διατίθεται κάθε πόρος πρέπει να είναι σαφές και να υπάρχει **άδεια χρήσης**.
- Ένας πόρος μπορεί να δημιουργηθεί στην υποδομή είτε μέσω του *περιβάλλοντος τεκμηρίωσης και ανάρτησης* είτε με *μεταφόρτωση ενός αρχείου μεταδεδομένων XML*. Σε κάθε περίπτωση θα πρέπει να συμπληρωθούν τα **υποχρεωτικά μεταδεδομένα** προκειμένου να δημιουργηθεί μία εγγραφή.
- Για την πληρότητα της περιγραφής ενός πόρου, θα πρέπει το φυσικό αντικείμενο (εφεξής **αρχεία περιεχομένου ή σύνολο δεδομένων**) να είναι προσβάσιμο είτε *άμεσα από την υποδομή* είτε έμμεσα μέσω εξωτερικού συνδέσμου.

Ανάλογα με τον τύπο του πόρου που θέλετε να δημιουργήσετε, ανατρέξτε στα αντίστοιχα κεφάλαια όπου περιγράφονται τα υποχρεωτικά μεταδεδομένα που απαιτούνται εάν θέλετε να δημιουργήσετε:

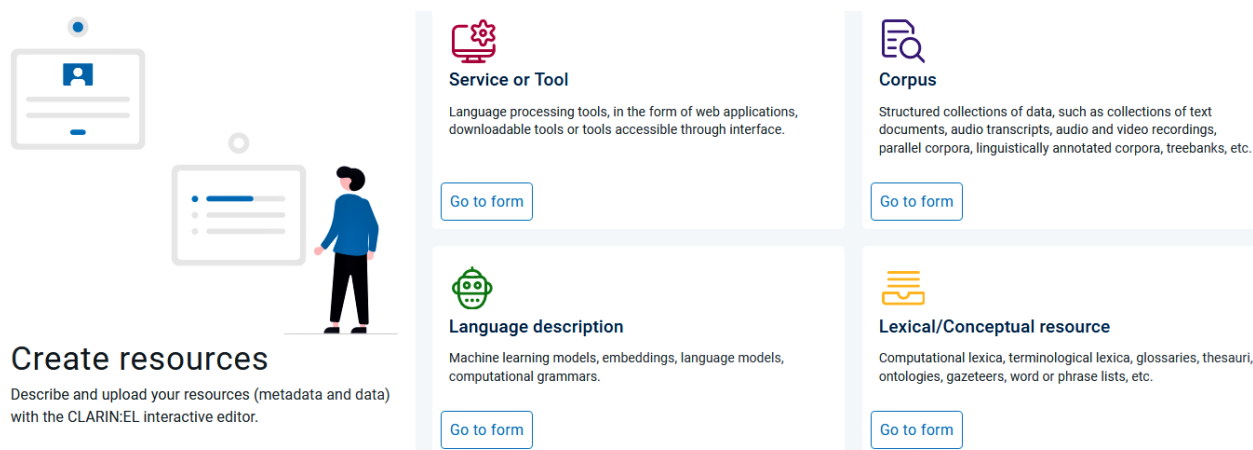
1. ένα *σώμα κειμένων*,
2. ένα *εργαλείο*,
3. έναν *λεξικό/εννοιολογικό πόρο*, ή
4. μία *γλωσσική περιγραφή*,

¹ Δείτε επίσης ποια είναι η Πολιτική απορρήτου





Πώς να δημιουργήσετε έναν πόρο χρησιμοποιώντας το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων

Προσοχή: Αυτή η ενότητα είναι αφιερωμένη στα **βήματα** που πρέπει να κάνετε για να δημιουργήσετε έναν πόρο χρησιμοποιώντας το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor**¹). Παρόλο που παρουσιάζεται ο τρόπος δημιουργίας ενός σώματος κειμένων, τα βήματα (εκτός αν αναφέρεται διαφορετικά) είναι κοινά για όλους τους τύπους πόρων.

Προϋπόθεση για να χρησιμοποιήσετε το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων είναι να έχετε συνδεθεί. Στην συνέχεια, επισκεφθείτε το *Dashboard* (πίνακας ελέγχου) και κάντε κλικ στο *create resources* (δημιουργία πόρων). Στο νέο παράθυρο, επιλέξτε τον τύπο του πόρου που θέλετε να δημιουργήσετε κάνοντας κλικ στο **Go to Form** (μετάβαση στη φόρμα συμπλήρωσης μεταδεδομένων).



Create resources
Describe and upload your resources (metadata and data) with the CLARIN:EL interactive editor.

 <p>Service or Tool Language processing tools, in the form of web applications, downloadable tools or tools accessible through interface.</p> <p>Go to form</p>	 <p>Corpus Structured collections of data, such as collections of text documents, audio transcripts, audio and video recordings, parallel corpora, linguistically annotated corpora, treebanks, etc.</p> <p>Go to form</p>
 <p>Language description Machine learning models, embeddings, language models, computational grammars.</p> <p>Go to form</p>	 <p>Lexical/Conceptual resource Computational lexica, terminological lexica, glossaries, thesauri, ontologies, gazeteers, word or phrase lists, etc.</p> <p>Go to form</p>

Πριν ξεκινήσετε, δείτε ποια είναι τα *υποχρεωτικά* μεταδεδομένα που θα πρέπει να συμπληρώσετε προκειμένου να δημιουργήσετε μια εγγραφή για:

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

- ένα σώμα κειμένων,
- ένα εργαλείο,
- έναν λεξικό/εννοιολογικό πόρο, ή
- μία γλωσσική περιγραφή,

16.1 Βήμα 1: Δώστε ένα όνομα στον πόρο σας

Το πρώτο βήμα είναι να δώσετε στον πόρο σας ένα όνομα. Η υποδομή ελέγχει αν αυτό το όνομα είναι ήδη σε χρήση και, εάν αυτό συμβαίνει, εμφανίζεται ένας κατάλογος με τους πόρους οι οποίοι έχουν (ολόκληρο ή τμηματικά) το όνομα που θέλετε να χρησιμοποιήσετε. Εάν επιθυμείτε να δημιουργήσετε μια νέα εγγραφή, πρέπει να χρησιμοποιήσετε άλλο όνομα².

corpus name

test corpus

Check if the corpus is already listed in the catalogue

ACCURAT balanced test corpus for under resourced languages

new test corpus m

test corpus	🔒
Test corpus1	🔒
Test corpus1	🔒
Test corpus1	🔒
Test corpus1	🔒
test corpus for validation	🔒
test corpus	🔒

Create test corpus Cancel

Εάν δεν βρεθούν εγγραφές με το όνομα που αρχικώς επιλέξατε, μπορείτε να προχωρήσετε στην επόμενη φάση κάνοντας κλικ στο κουμπί **create** (δημιουργία)³.

corpus name

Demo resource

Check if the corpus is already listed in the catalogue

No matches

Create Demo resource Cancel

Προσοχή: Κατά την δημιουργία των εργαλείων, και αφότου έχει ελεγχθεί το όνομα, εμφανίζεται μια διαφορετική οθόνη που σας ρωτά αν θέλετε να **ενσωματωθεί το εργαλείο στην υποδομή CLARIN:EL ως υπηρεσία**.

² Εναλλακτικά, μπορείτε να επιλέξετε έναν από τους πόρους της λίστας για να επιμεληθείτε τα μεταδεδομένα του, εφόσον έχετε το δικαίωμα.

³ Το κουμπί απενεργοποιείται αυτόματα όταν το όνομα που έχετε χρησιμοποιήσει εμφανίζεται και σε άλλες εγγραφές και ενεργοποιείται μόλις χρησιμοποιήσετε ένα νέο μοναδικό όνομα.

Do you want to contribute a tool that will be integrated in CLARIN:EL as a functional service (i.e., available through the CLARIN:EL APIs)?

- ☐ Yes
- ☐ No

Ανάλογα με την απάντησή σας, θα δείτε στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων το πλαίσιο που φαίνεται στην ακόλουθη εικόνα επιλεγμένο ή μη. Η απόφασή σας αυτή δεν είναι δεσμευτική και μπορείτε να την αλλάξετε όποτε θέλετε απο-επιλέγοντας (ή επιλέγοντας) αυτό το πλαίσιο.

CREATE A NEW SERVICE OR TOOL

Info

- At any point you are able to save your progress as draft and continue editing at a later time
- Once you submit your record you will not be able to save it as draft any more, but you will still be able to make changes and submit them.
- Visit all tabs in order to fill in as much information as possible for better visibility of your record.
- If the metadata record is for a resource that you plan to deliver later, please check the "work in progress" box.

LANGUAGE
RESOURCE/TECHNOLOGY

TOOL/SERVICE

DISTRIBUTION

DATA

☐ For information
☒ CLARIN:EL compatible service

Save draft

Save

IDENTITY

CATEGORIES

CONTACT

DOCUMENTATION

RELATED LRTS

LRT name *

Demo tool

The official name or title of the language resource/technology

LRT identifier

A string used to uniquely identify the language resource/technology

LRT short name

An abbreviation, acronym, etc. used for the language resource/technology

Description *

language

English

select language

language

English

select language

language

English

select language

16.2 Βήμα 2: Μεταφορτώστε τα αρχεία περιεχομένου του πόρου

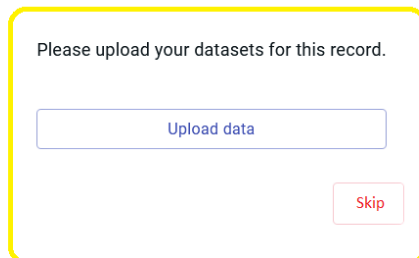
Προσοχή: Αν και η λειτουργικότητα αυτή διατίθεται για όλους τους τύπους πόρων, συνιστάται τα εργαλεία να **ενσωματώνονται ως υπηρεσίες**, σε συνεργασία με την τεχνική ομάδα του CLARIN:EL, αντί να μεταφορτώνεται το λογισμικό τους.

Τα αρχεία περιεχομένου (ή *σύνολο δεδομένων*) του πόρου πρέπει για να μεταφορτωθούν να βρίσκονται σε έναν **συμπίεσμένο φάκελο**. Οι επιτρεπτές επεκτάσεις συμπίεσης είναι οι ακόλουθες: **.zip, .tgz, .gz, .tar**⁴. Για την ονομασία του φακέλου χρησιμοποιήστε το λατινικό αλφάβητο χωρίς κενά μεταξύ των λέξεων. Δείτε [εδώ](#) πως να προετοιμάσετε τα δεδομένα σας πριν την μεταφόρτωση.

⁴ Εάν τα αρχεία περιεχομένου είναι διαθέσιμα σε διάφορα μορφώματα, (π.χ. σε XML, TXT και PDF), συνιστάται να τα οργανώσετε σε διαφορετικούς φακέλους ανά μορφότυπο. Στην συνέχεια κάθε φάκελος θα πρέπει να συμπίεστεί και να συνδεθεί με διαφορετικό τρόπο διάθεσης. Με αυτόν τον τρόπο η υποδομή μπορεί να διαχειριστεί με τον βέλτιστο τρόπο κάθε μορφότυπο, συμβατό ή μη με τις υλοποιημένες ροές εργασίας.

16.2.1 Skip and upload later

Μόλις κάνετε κλικ στο **creation** (δημιουργία), εμφανίζεται ένα νέο παράθυρο που σας ζητά να μεταφορτώσετε τα αρχεία περιεχομένου του πόρου. Έχετε τη δυνατότητα να **παραλείψετε** (skip) αυτό το βήμα και να κάνετε την μεταφόρτωση αργότερα.

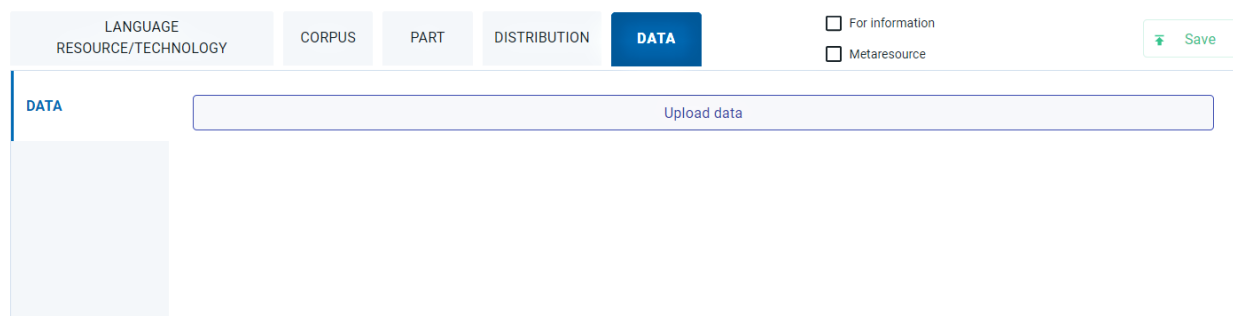


Please upload your datasets for this record.

Upload data

Skip

Εάν αποφασίσετε να μεταφορτώσετε τα αρχεία σε μεταγενέστερο στάδιο, θα πρέπει να μεταβείτε στο περιβάλλον τεκμηρίωσης και ανάρτησης. Η μετάβαση γίνεται από τις διαθέσιμες για τον πόρο **ενέργειες** διαλέγοντας την επεξεργασία μεταδεδομένων (*edit*). Εφόσον έχετε μεταφερθεί στο περιβάλλον, επιλέξτε την ενότητα *Data* (Δεδομένα) όπου βρίσκεται το κουμπί **upload** (μεταφόρτωση).



LANGUAGE RESOURCE/TECHNOLOGY CORPUS PART DISTRIBUTION **DATA**

☐ For information
☐ Metaresource

Save

DATA

Upload data

16.2.2 Upload immediately

Εναλλακτικά, μπορείτε να μεταφορτώσετε απευθείας τα αρχεία περιεχομένου. Θα παρουσιαστούν μια σειρά από οθόνες (όπως φαίνεται στην παρακάτω εικόνα) που θα σας καθοδηγήσουν:

1. να επιλέξετε το σύνολο δεδομένων,
2. να το μεταφορτώσετε,
3. να δείτε τις λεπτομέρειες της μεταφόρτωσής όταν αυτή ολοκληρωθεί, και
4. να κάνετε κλικ στο κουμπί **finish** (ολοκλήρωση διαδικασίας) για να επιστρέψετε στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων.

Please select a .zip file in order to upload a dataset for this record

Drag & Drop your files or [Browse](#)

[cancel](#) [upload dataset](#)

Please select a .zip file in order to upload a dataset for this record

[x](#) singleTXT.zip 8 KB [↑](#)

[cancel](#) [upload dataset](#)

Please upload your datasets for this record.

[Upload data](#)

Name	Upload date	Assigned to distribution	Actions
singleTXT.zip	27 May 2021	No	x

*In order to delete a dataset you should first unlink it from the corresponding distribution and submit/save your record.

[Finish](#)

16.2.3 Assign to distribution

Ένας πόρος ως φυσικό αντικείμενο μπορεί να διατίθεται με διάφορους τρόπους (για παράδειγμα, ως CD-ROM, ως καταφορτώσιμο υλικό από υπερσύνδεσμο κλπ.) καθένας εκ των οποίων έχει τους δικούς του όρους χρήσης (**licence terms**). Αυτές οι μορφές ονομάζονται **modes of distribution**. Για την πληρότητα της περιγραφής του πόρου πρέπει να δηλώσετε όλους τους τρόπους διάθεσης και να συνδέσετε κάθε έναν από αυτούς με το σύνολο δεδομένων όταν χρειάζεται. Όπως φαίνεται στην προηγούμενη εικόνα, η σύνδεση αυτή (**assigned to distribution**) δεν έχει γίνει.

Προσοχή: Η σύνδεση δεν χρειάζεται να γίνει κατά τη μεταφόρτωση των αρχείων περιεχομένου. Όταν αποφασίσετε να την κάνετε, θα πρέπει να μεταβείτε στο περιβάλλον τεκμηρίωσης και ανάρτησης. Η μετάβαση γίνεται από τις διαθέσιμες για τον πόρο *ενέργειες* διαλέγοντας την επεξεργασία μεταδεδομένων (*edit*). Εφό-

σον έχετε μεταφερθεί στο περιβάλλον, επιλέξτε την ενότητα *Distribution* (Διάθεση).

Dataset distribution form *
downloadable

Select the form or delivery channel through which the corpus is distributed

Private
☐ Yes
☒ No

Specifies whether the resource is private so that its access/download location remains hidden

Associate a dataset with this distribution

singleTXT .zip

Αρχικά επιλέξτε τον τρόπο διάθεσης (**data distribution form**) από την αναπτυσσόμενη λίστα και στη συνέχεια κάντε κλικ στο όνομα του αρχείου zip για να συσχετιστεί αυτό το σύνολο δεδομένων με αυτή τη μορφή διάθεσης. Αποθηκεύστε τις αλλαγές που κάνετε. Την επόμενη φορά που θα ελέγξετε την εγγραφή μεταδεδομένων θα δείτε ότι η σύνδεση έχει καταχωρηθεί όπως φαίνεται στην επόμενη εικόνα.

LANGUAGE RESOURCE/TECHNOLOGY CORPUS PART DISTRIBUTION **DATA** ☐ For information ☐ Metaresource [Save](#)

DATA Upload data

Name	Upload date	Assigned to distribution	Actions
singleTXT.zip	27 May 2021	Yes	

*In order to delete a dataset you should first unlink it from the corresponding distribution and save your record.

Η υποδομή δεν θα σας επιτρέψει να **υποβάλετε έναν πόρο για δημοσίευση** (submit for publication), προτού συσχετίσετε κάθε σύνολο δεδομένων που έχετε μεταφορτώσει με έναν ή περισσότερους τρόπους διάθεσης. Στην περίπτωση που επιχειρήσετε να υποβάλετε έναν πόρο για δημοσίευση χωρίς να το έχετε κάνει, θα μεταφερθείτε αυτόματα στο περιβάλλον τεκμηρίωσης και ανάρτησης για να ολοκληρώσετε την διαδικασία.

LANGUAGE RESOURCE/TECHNOLOGY CORPUS PART DISTRIBUTION **DATA** ☐ For information ☐ Metaresource [Save](#)

IDENTITY LRT name * demo corpus language English [+](#)
The official name or title of the language resource/technology select language

LRT identifier [Fill in](#)
A string used to uniquely identify the language resource/technology

LRT short name language English [+](#)
An abbreviation, acronym, etc. used for the language resource/technology select language

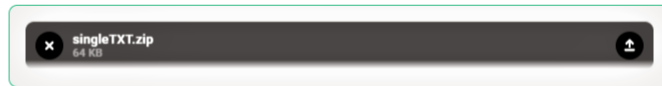
Description * language

Invalid record. The record has data not related with any Dataset Distribution

16.2.4 Αντικατάσταση αρχείων

Εάν πρέπει να αντικαταστήσετε τα αρχεία περιεχομένου που έχετε ανεβάσει, δημιουργήστε ένα νέο συμπιεσμένο αρχείο **με το σωστό σύνολο δεδομένων**. Στο νέο σύνολο δεδομένων δώστε το **ίδιο όνομα** με το παλιό και συμπιέστε το με τον ίδιο τρόπο. Από την ενότητα *Data* επιλέξτε να κάνετε μεταφόρτωση. Θα εμφανιστεί το ακόλουθο προειδοποιητικό μήνυμα.

Please select a .zip file in order to upload a dataset for this record



Warning

There is already a file with the same name. By proceeding you are going to delete the already uploaded file and replace it with this one. Are you sure?

cancel replace dataset

Κάντε κλικ στο **replace dataset** (αντικατάσταση συνόλου δεδομένων) και αν η διαδικασία ολοκληρωθεί με επιτυχία, θα ενημερωθείτε με ένα μήνυμα στο κάτω δεξιό μέρος της σελίδας. Στην καρτέλα πληροφοριών, το νέο σύνολο δεδομένων έχει την καινούργια ημερομηνία μεταφόρτωσης, όπως φαίνεται στην παρακάτω εικόνα.

LANGUAGE
RESOURCE/TECHNOLOGY
CORPUS
PART
DISTRIBUTION
DATA
☐ For information
☐ Metaresource
Save

DATA

Upload data

Name	Upload date	Assigned to distribution	Actions
singleTXT.zip	03 July 2021	Yes	

*In order to delete a dataset you should first unlink it from the corresponding distribution and save your record.

File has been replaced successfully.

Αποθηκεύστε την εγγραφή μεταδεδομένων. Θα μεταφερθείτε αυτόματα στη σελίδα θέασης του πόρου λαμβάνοντας παράλληλα ενημέρωση ότι τα μεταδεδομένα του πόρου έχουν επικαιροποιηθεί επιτυχώς.

Single text EL (txt)

Corpus

Version: 1.0.0 (automatically assigned)

<http://hdl.handle.net/11239/CLARIN-EL-0000-0000-60FD-7-TEST>

Singletxt RbNh6VucQs7yYdVUhsCR6A (text/plain) annotated by the ILSP Sentence splitter and tokenizer for Greek text (ILSP workflow v1.0.0).

Automatically generated metadata. Please edit them.

clarin:el

Cite

Citation text

Single text EL (txt) (2021). Version 1.0.0 (automatically assigned). [Dataset (Text corpus)]. CLARIN-EL.
<http://hdl.handle.net/11239/CLARIN-EL-0000-0000-60FD-7-TEST>

Actions

Language

Modern Greek (1453-)

Keywords

corpus annotated corpus

Corpus subclass

annotated corpus

Record updated successfully.

16.2.5 Διαγραφή αρχείων

Εφόσον έχετε συσχετίσει τα αρχεία περιεχομένου με έναν τρόπο διάθεσης, η δυνατότητα διαγραφής στην ενότητα *Data* απενεργοποιείται και εμφανίζεται το ακόλουθο προειδοποιητικό μήνυμα.

LOGY CORPUS PART DISTRIBUTION **DATA** ☐ For information ☐ Metaresource Save

Upload data

Name	Upload date	Assigned to distribution	Actions
singleTXT.zip	27 May 2021	Yes	

*In order to delete a dataset you should first unlink it from the corresponding distribution and save your record.

In order to delete a dataset you should first unlink it from the corresponding distribution and submit/save your record.

Για κάνετε διαγραφή, ακολουθήστε τα επόμενα βήματα:

1. Μεταβείτε στην ενότητα *Distribution* (Διάθεση).
2. Επιλέξτε την κενή επιλογή στο αντίστοιχο πεδίο μεταδεδομένων (*Associate a dataset with this distribution*). Αυτή η ενέργεια **αποσυνδέει** το σύνολο δεδομένων.

Dataset distribution form *
downloadable

Select the form or delivery channel through which the corpus is distributed

Private
☐ Yes
☒ No

Specifies whether the resource is private so that its access/download location remains hidden

Associate a dataset with this distribution

singleTXT .zip

3. Αποθηκεύστε την εγγραφή μεταδεδομένων.
4. Επιλέξτε **και πάλι** να επεξεργαστείτε την εγγραφή μεταδεδομένων.
5. Μεταβείτε στην ενότητα *Data*. Το κουμπί διαγραφής έχει ενεργοποιηθεί.
6. Διαγράψτε τα δεδομένα.

Θα ενημερωθείτε με μήνυμα ότι η ενέργεια ολοκληρώθηκε επιτυχώς.

LANGUAGE RESOURCE/TECHNOLOGY CORPUS PART DISTRIBUTION **DATA** ☐ For information ☐ Metaresource Save

DATA Upload data

File deleted.

7. Αποθηκεύστε την εγγραφή μεταδεδομένων.

16.3 Βήμα 3: Συμπληρώστε τα υποχρεωτικά μεταδεδομένα

Όποιον τύπο πόρου και αν επιλέξετε να δημιουργήσετε, πρέπει να δώσετε πληροφορίες για ορισμένα *υποχρεωτικά* μεταδεδομένα (διαφορετικά ανά τύπο πόρου), τα οποία κατανέμονται σε διάφορες **ενότητες** (οργανωμένες οριζόντια) και διάφορες **καρτέλες** (που παρουσιάζονται κάθετα) στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων.

Κατά την τεκμηρίωση του πόρου, μπορείτε να αποθηκεύετε την εγγραφή μεταδεδομένων ως προσχέδιο (**save draft**)⁵. Δεν θα μπορέσετε να προχωρήσετε σε τελική αποθήκευση (**save**) αν προηγουμένως δεν έχετε συμπληρώσει *όλα τα υποχρεωτικά μεταδεδομένα*. Κάθε φορά που κάνετε κλικ στο κουμπί **save** (αποθήκευση), τα μεταδεδομένα που μέχρι στιγμής έχετε συμπληρώσει ελέγχονται. Εάν κάποια από τα υποχρεωτικά μεταδεδομένα λείπουν ή έχουν λανθασμένες τιμές, εμφανίζεται ένα μήνυμα που σας προτρέπει να κάνετε προσθήκες/διορθώσεις.

❗ Correct the following errors in order to proceed

Language Resource/Technology > Identity > Description is required
 Language Resource/Technology > Identity > Description language is required
 Language Resource/Technology > Categories > Keyword is a required field
 Language Resource/Technology > Contact > Additional information is a required field
 Corpus > Technical > Corpus subclass is a required field
 Corpus > Technical > Personal data included is a required field
 Corpus > Technical > Sensitive data included is a required field
 Part > Media part > Corpus part is a required field
 Distribution > Technical > Dataset distribution is a required field

[Close](#) [Proceed](#)

Μόλις κλείσετε αυτό το μήνυμα εμφανίζεται ένα πλαίσιο με χρωματική επισήμανση. Σε αυτό το πλαίσιο υπάρχουν όλα τα ελλιπή/λανθασμένα μεταδεδομένα με υπερσυνδέσμους. Έτσι, πατώντας πάνω σε κάποιο, μεταφέρεστε στο σημείο (δηλαδή στην καρτέλα της ενότητας) όπου αυτό εντοπίζεται μέσα στο περιβάλλον τεκμηρίωσης. Για διευκόλυνση του χρήστη τα μεταδεδομένα που λείπουν επισημαίνονται με μια κάθετη κόκκινη γραμμή.

⁵ Για την αποθήκευση ως προσχέδιο αρκεί μόνο το όνομα του πόρου.

CREATE A NEW CORPUS

Correct the following errors in order to proceed

1. Language Resource/Technology > Identity > Description is required
2. Language Resource/Technology > Identity > Description language is required
3. Language Resource/Technology > Categories > Keyword is a required field
4. Language Resource/Technology > Contact > Additional information is a required field
5. Corpus > Technical > Corpus subclass is a required field
6. Corpus > Technical > Personal data included is a required field
7. Corpus > Technical > Sensitive data included is a required field
8. Part > Media part > Corpus part is a required field
9. Distribution > Technical > Dataset distribution is a required field

LANGUAGE RESOURCE/TECHNOLOGY CORPUS PART DISTRIBUTION DATA

☐ For information
☐ Metaresource

[Save draft](#) [Save](#)

IDENTITY

LRT name * Demo resource
The official name or title of the language resource/technology

language English
select language

CATEGORIES

CONTACT

DOCUMENTATION

RELATED LRTS

LRT identifier
A string used to uniquely identify the language resource/technology [Fill in](#)

LRT short name
An abbreviation, acronym, etc. used for the language resource/technology

language English
select language

Description *
General information on the language resource/technology

language English
select language

Μετά την ολοκλήρωση της διαδικασίας, εάν επιχειρήσετε εκ νέου να κάνετε αποθήκευση (save), ενδέχεται να ενημερωθείτε με μήνυμα ότι έχουν προκύψει **καινούργια υποχρεωτικά μεταδεδομένα**. Αυτό συμβαίνει επειδή ορισμένες από τις τιμές που έχετε εισαγάγει έχουν δημιουργήσει νέες απαιτήσεις (πρόκειται για τα *υπό συνθήκη υποχρεωτικά* μεταδεδομένα). Και πάλι, τα νέα μεταδεδομένα που πρέπει να συμπληρώσετε εμφανίζονται στο επάνω μέρος της σελίδας μέσα στο πλαίσιο με χρωματική επισήμανση.

Correct the following errors in order to proceed

- Part > Media part > Corpus text part > Language is a required field
- Distribution > Technical > Text features is a required field
- Distribution > Technical > Licence is a required field
- Distribution > Technical > Distribution location is a required field

[Close](#) [Proceed](#)

Το ενημερωτικό αυτό μήνυμα μπορεί να εμφανιστεί αρκετές φορές πριν συμπληρωθούν **όλα** τα απαραίτητα μεταδεδομένα, υποχρεωτικά και υπό συνθήκη υποχρεωτικά. Αφού ολοκληρωθεί η διαδικασία, θα σας επιτραπεί τελικά να αποθηκεύσετε την νέα εγγραφή μεταδεδομένων.

☒ Save

You are about to update Demo resource. Are you sure?

[Close](#) [Proceed](#)

16.4 Βήμα 4: Δείτε πως φαίνεται ο πόρος που δημιουργήσατε

Αφού επιλέξετε να προχωρήσετε κάνοντας κλικ στο **proceed**, λαμβάνετε μήνυμα ότι η εγγραφή μεταδεδομένων δημιουργήθηκε επιτυχώς και μεταφέρεστε στη σελίδα θέασης του πόρου.

The screenshot displays the CLARIN:EL portal interface. At the top, there is a navigation bar with the CLARIN:EL logo, a 'CLARIN:EL portal' button, and links for 'Dashboard', 'Help', and 'Curator H.'. A 'Go to catalogue' link is also present. Below the navigation bar, a 'STATUS' section provides instructions on the record's lifecycle: draft, syntactically valid, submitted, approved, and published. A progress bar indicates the current status is 'syntactically valid'. The main content area shows details for a 'corpus test' record, including its version (1.0.0), a link to the PID URL, and the text 'Hello'. A 'clarin:el' logo and an 'Actions' dropdown menu are also visible. At the bottom, there are three input fields for 'Language' (Kalaallisut; Greenlandic), 'Keyword' (Hello), and 'Corpus subclass' (raw corpus). A green notification bar at the bottom right states 'Record updated successfully.'

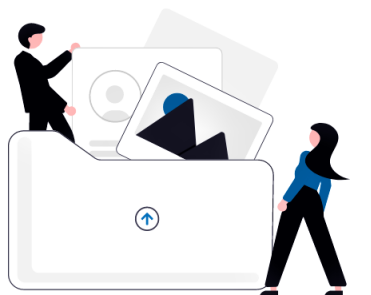
Μπορείτε να συνεχίσετε να *επεξεργάζεστε* τα μεταδεδομένα μέχρι να αποφασίσετε να υποβάλετε τον πόρο για δημοσίευση.

Πώς να δημιουργήσετε έναν πόρο μεταφορτώνοντας ένα αρχείο μεταδεδομένων XML

Μπορείτε να δημιουργήσετε οποιονδήποτε τύπο πόρου μεταφορτώνοντας ένα αρχείο μεταδεδομένων XML που περιέχει τουλάχιστον τα *υποχρεωτικά μεταδεδομένα*.

Προσοχή: Αυτή η ενότητα είναι αφιερωμένη στα **βήματα** που πρέπει να κάνετε για να δημιουργήσετε έναν πόρο με μεταφόρτωση αρχείου μεταδεδομένων XML. Παρόλο που παρουσιάζεται ο τρόπος δημιουργίας ενός σώματος κειμένων, τα βήματα (εκτός αν αναφέρεται διαφορετικά) είναι κοινά για όλους τους τύπους πόρων. Δείτε *εδώ* συγκεκριμένα παραδείγματα XML αρχείων που περιγράφονται διαφορετικοί τύποι πόρων.

Για να μεταφορτώσετε ένα αρχείο μεταδεδομένων XML πρέπει να είστε *συνδεδεμένοι*. Στην συνέχεια επισκεφθείτε το *Dashboard* (πίνακας ελέγχου) και κάντε κλικ στο *upload resources* (μεταφόρτωση πόρων). Στο νέο παράθυρο, εμφανίζονται τρεις καρτέλες, όπως φαίνεται στην παρακάτω εικόνα.



Upload records

Upload your records (metadata and data files). The metadata files must be validated for compliance with the CLARIN:EL schema.

VALIDATE XML FILES

UPLOAD SINGLE ITEM

UPLOAD MULTIPLE ITEMS

Before uploading your metadata XML file(s) you need to validate them against the CLARIN:EL schema.

If you want to validate multiple files, please add them all directly in the zip file (without subfolders).

Drag & Drop your file or Browse

Πρώτα πρέπει να επικυρώσετε (validate) το αρχείο μεταδεδομένων XML, πρέπει δηλαδή να το μεταφορτώσετε προκειμένου να ελεγχθεί. Αν και μπορείτε να παραλείψετε αυτό το βήμα, συνιστάται να το κάνετε. Θα εξοικονομήσετε χρόνο διότι ο έλεγχος που γίνεται από την υποδομή δείχνει ασυνέπειες ή μεταδεδομένα που λείπουν, όπως στην παρακάτω εικόνα.

VALIDATE XML FILES

UPLOAD SINGLE ITEM

UPLOAD MULTIPLE ITEMS

Before uploading your metadata XML file(s) you need to validate them against the CLARIN:EL schema.

If you want to validate multiple files, please add them all directly in the zip file (without subfolders).

Corpus.xml

Upload complete

tap to undo

Validation report

```
{
  "info": "file 'Corpus.xml' not valid",
  "errors": [
    {
      "line": 213,
      "error": "Element 'LicenceIdentifier': This element is not expected. Expected is one of (
    }
  ]
}
```

Αφού κάνετε τις απαραίτητες διορθώσεις/αλλαγές, προσπαθήστε να επικυρώσετε ξανά το αρχείο. Αν όλα έχουν διορθωθεί, θα ενημερωθείτε ότι το αρχείο xml είναι έγκυρο.

VALIDATE XML FILES

UPLOAD SINGLE ITEM

UPLOAD MULTIPLE ITEMS

Before uploading your metadata XML file(s) you need to validate them against the CLARIN:EL schema.

If you want to validate multiple files, please add them all directly in the zip file (without subfolders).



Validation report

```
{
  "info": "file 'Corpus.xml' valid"
}
```

Στη συνέχεια, μπορείτε να προχωρήσετε στην επόμενη καρτέλα για να μεταφορτώσετε το αρχείο με επιτυχία. Εάν ο πόρος που θέλετε να δημιουργήσετε εμπίπτει σε μία από τις κατηγορίες¹ που φαίνονται στην ακόλουθη εικόνα πρέπει να επιλέξετε το αντίστοιχο πλαίσιο πριν μεταφορτώσετε το αρχείο μεταδεδομένων XML.

VALIDATE XML FILES

UPLOAD SINGLE ITEM

UPLOAD MULTIPLE ITEMS

You can upload one XML file at a time.

It is highly recommended that you [validate](#) your XML file against the CLARIN:EL schema before you proceed.

☒ For information

If the metadata record is for a resource that you plan to upload at a later stage, please check this box.

☐ CLARIN:EL compatible service

If the metadata records are for services to be integrated in CLARIN:EL, please check the box "CLARIN:EL compatible service".


☐ Metaresource

If this is a metaresource, please check the box "metaresource".


Drag & Drop your file or [Browse](#)

Θα ενημερωθείτε μόλις η διαδικασία ολοκληρωθεί επιτυχώς και θα μεταφερθείτε αυτόματα στη σελίδα θέασης του πόρου.

¹ Δείτε [εδώ](#) τις διάφορες κατηγορίες πόρων.



CLARIN:EL portal >

DashboardHelpCurator H. 

Go to inventory

STATUS

You can continue editing your metadata record while the status is draft (syntactically valid); when you are satisfied with it, you can submit it for publication; the resource will be validated and published by the supervisor of the repository, or, if required, you will be contacted for further information

draft


syntactically valid

submitted

approved

published

Demo corpus

Corpus

short name

Version: 1.0.0

clarin:el

Actions ▾

Short description of the resource for human readers with the basic features (e.g., language, size, format, domain or genre, etc); the example for this template is for an video of lectures in English, the audio recordings, their transcriptions, the subtitles in English and their translations into German

for info

Language

EnglishGerman

Keywords

multimedia corpusvideo lecturessubtitles

Corpus subclass

raw corpus

File has been successfully uploaded. ✕

Πώς να προετοιμάσετε τα δεδομένα πριν τη μεταφόρτωση

Η υποδομή CLARIN:EL προσφέρει στον χρήστη τη δυνατότητα να μεταφορτώσει δεδομένα σε δύο περιπτώσεις:

- όταν θέλει να κατατεθεί το **περιεχόμενο** ενός πόρου (βλ. *κεφάλαιο 1*), ή
- όταν θέλει να τροφοδοτήσει με δεδομένα (**input**) μια υπηρεσία επεξεργασίας (βλ. *κεφάλαιο 2*).

Προσοχή: Σε κάθε περίπτωση τόσο ο εμπλεκόμενος χρήστης όσο και οι απαιτήσεις για τα δεδομένα είναι διαφορετικές. Και στις δύο περιπτώσεις ο χρήστης θα πρέπει να έχει *συνδεθεί*. Εάν επιθυμείτε να μεταφορτώσετε δεδομένα και δεν είστε χρήστης του CLARIN:EL θα πρέπει πρώτα να κάνετε *εγγραφή*.

18.1 I. Τα δεδομένα ως περιεχόμενο του πόρου

18.1.1 1. ΤΥΠΟΙ

Παρόλο που όλοι οι *πόροι* στην υποδομή CLARIN:EL έχουν πάντα περιγραφή μεταδεδομένων, δεν συνοδεύονται κάθε φορά και από αρχεία περιεχομένου (βλ. για παράδειγμα τους πόρους που χαρακτηρίζονται ως *metaresources* ή «for info»). Όταν, ωστόσο, οι πόροι έχουν περιεχόμενο, αυτό ποικίλλει ανάλογα με τον τύπο *resource* τους.

- Τα **Σώματα κειμένων** είναι συλλογές από:
 - *πρωτογενή δεδομένα* διάφορων μέσων:
 - * ψηφιακά/ψηφιοποιημένα γραπτά κείμενα (π.χ. ψηφιοποιημένα βιβλία, κείμενα του διαδικτύου, εφημερίδες, σώματα κειμένων κλπ.), ηχογραφήσεις προφορικού λόγου (π.χ. συνεντεύξεις, ραδιοφωνικές εκπομπές κλπ.)
 - * βιντεοσκοπήσεις (π.χ. τηλεοπτικές εκπομπές, συλλογές από εκφράσεις προσώπου, χειρονομίες κλπ.)
 - * εικόνες (π.χ. ψηφιακές/ψηφιοποιημένες φωτογραφίες με τις λεζάντες τους κλπ.)

ή

- επεξεργασμένα δεδομένα
 - * διάφοροι τύποι επισημειώσεων κειμένων,
 - * δεδομένα ήχου και πολυμέσων, που δημιουργούνται αυτόματα ή με το χέρι (π.χ. μορφολογικά επισημειωμένα κείμενα, μεταγραφές προφορικών δεδομένων, επισημειωμένα βίντεο κλπ.)
- Οι **λεξικοί/εννοιολογικοί πόροι** και **γλωσσικές περιγραφές** είναι:
 - δομημένα γλωσσικά δεδομένα (π.χ. κατάλογοι λέξεων, λεξικά, θησαυροί, γραμματικές κλπ.) που χρησιμοποιούνται για επεξεργασία και μελέτη πρωτογενών και επεξεργασμένων δεδομένων.
- Τα **εργαλεία** είναι:
 - *πηγαίος κώδικας*, ή
 - *λογισμικό*

προγραμμάτων/εφαρμογών που εκτελούν διάφορους τύπους γλωσσικής επεξεργασίας (π.χ. αντιστοίχιση πολύ-γλωσσου κειμένου, μορφολογικός επισημείωση, λημματοποίηση, επιφανειακή συντακτική ανάλυση, εξαγωγή πληροφορίας κλπ.).

18.1.2 2. ΧΡΗΣΗ

Η μεταφόρτωση δεδομένων στο CLARIN:EL δεν συνεπάγεται αυτόματα ότι τα δεδομένα αυτά είναι άμεσα προσβάσιμα από τους χρήστες του CLARIN:EL ή ότι είναι επεξεργάσιμα από τα εργαλεία και τις υπηρεσίες του CLARIN:EL. Πρέπει, για να συμβεί αυτό, να ληφθούν υπόψη οι παράμετροι της **προσβασιμότητας** (accessibility) και της **επεξεργασιμότητας** (processability), με την δεύτερη να αφορά αποκλειστικά τα σώματα δεδομένων (*processable*) και τα εργαλεία (*processing services*). Προκειμένου να γίνει σωστά η προετοιμασία των δεδομένων, πρέπει να έχετε, πριν ξεκινήσετε, δώσει απάντηση στα ακόλουθα ερωτήματα.

Σημαντικό:

- Θέλω τα δεδομένα μου να είναι **προσβάσιμα** στους χρήστες του CLARIN:EL;
 - Εάν η απάντηση είναι **ναι**, παρακαλούμε δείτε την *αντίστοιχη ενότητα* πριν διαβάσετε τις ακόλουθες οδηγίες. ddd
- Θέλω τα δεδομένα που θα μεταφορτώσω για ένα σώμα κειμένων να είναι **επεξεργάσιμα** από τις υπηρεσίες του CLARIN:EL;
 - If the answer is **yes**, please check the necessary metadata values along with all the other instructions.
- Θέλω τα δεδομένα που θα μεταφορτώσω για ένα εργαλείο να μετατραπούν σε **υπηρεσία επεξεργασίας** στο CLARIN:EL;
 - Εάν η απάντηση είναι **ναι**, παρακαλούμε ελέγξτε ποιες είναι οι *απαραίτητες τιμές μεταδεδομένων* που πρέπει να έχει η περιγραφή ενός εργαλείου μαζί με όλες τις άλλες οδηγίες.

Οι ακόλουθες οδηγίες χωρίζονται σε δύο τμήματα: οι γενικές οδηγίες ισχύουν για όλους τους τύπους πόρων, ενώ οι ειδικές οδηγίες ισχύουν μόνο για τα σώματα κειμένων και τα εργαλεία, όπως δηλώνεται.

18.1.3 3. Βήματα που πρέπει να ακολουθήσετε

3.1. Γενικές οδηγίες

Υπάρχουν διάφορα νομικά έγγραφα τα οποία πρέπει να συμβουλευτείτε πριν προχωρήσετε. Βεβαιωθείτε ότι έχετε διαβάσει προσεκτικά

- την πολιτική απορρήτου και
- τους όρους παροχής υπηρεσιών

του CLARIN:EL καθώς η μεταφόρτωση δεδομένων στην υποδομή συνεπάγεται ότι **έχετε αποδεχθεί** τα όσα αναφέρονται στα προαναφερθέντα νομικά έγγραφα.

Εάν **ανήκετε** σε έναν φορέα-μέλος του CLARIN:EL, πριν τη μεταφόρτωση των δεδομένων, πρέπει να επικοινωνήσετε με τον **Επιστημονικό Υπεύθυνο** του φορέα σας.

Εάν **δεν ανήκετε** σε έναν φορέα-μέλος του CLARIN:EL, πριν τη μεταφόρτωση των δεδομένων, πρέπει να υπογράψετε το **Συμφωνητικό Κατάθεσης Γλωσσικών Πόρων**.

Βεβαιωθείτε, επίσης, ότι τα δεδομένα διατίθενται με **σαφείς όρους χρήσης** και ότι έχει ληφθεί **άδεια από όλους τους εμπλεκόμενους κατόχους δικαιωμάτων**. Εάν τα δεδομένα έχουν περισσότερους από έναν **τρόπους διάθεσης** θα πρέπει να αναφέρετε τους όρους χρήσης για κάθε έναν από αυτούς. Επιπροσθέτως, τα ίδια τα δεδομένα μπορεί να διατίθενται με διάφορους **όρους χρήσης** ανάλογα με τη φύση του χρήστη ή την προβλεπόμενη χρήση (ακαδημαϊκή έναντι εμπορικής).

Στη συνέχεια, μπορείτε να προχωρήσετε με τα τρία στάδια προετοιμασίας των δεδομένων: **συλλογή, συστηματοποίηση/ κατηγοριοποίηση και συμπίεση**.



Βήμα 1: Συλλογή

Συλλέξτε δεδομένα με άξονα μια συγκεκριμένη ιδέα (π.χ. ένα γλωσσάρι φεμινιστικής θεωρίας). Συλλέξτε **όλα** και **μόνο** τα απαραίτητα δεδομένα. Εάν περιλαμβάνονται **προσωπικά, ευαίσθητα** ή δεδομένα **εμπιστευτικού** χαρακτήρα, παρακαλούμε να τα ανωνυμοποιήσετε ή να τα αφαιρέσετε πριν τη μεταφόρτωση.

Βήμα 2: Συστηματοποίηση/ Κατηγοριοποίηση

Τα δεδομένα που συλλέγονται μπορεί να είναι αποτέλεσμα διαφόρων σταδίων επεξεργασίας: βιντεοσκοπήσεις που έχουν απομαγνητοφωνηθεί, αρχεία PDF που έχουν καθαριστεί (από εικόνες και διευθύνσεις URL) και έχουν μετατραπεί σε αρχεία TXT. Σε αυτές τις περιπτώσεις, τα **πρωτογενή** και τα δεδομένα **που είναι αποτέλεσμα μετατροπής** αποτελούν μια ενότητα που περιλαμβάνει πολλούς και διαφορετικούς **μορφότυπους** (formats), **μέσα** (media) και **γλώσσες** (languages), που ίσως δεν θέλετε να διαχωρίσετε. Για να το κάνετε αυτό και να τα παρουσιάσετε όλα σε μια ενιαία εγγραφή μεταδεδομένων, πρέπει να οργανώσετε τα δεδομένα σας με σαφή δομή και κατανοητό τρόπο. Με την ομαδοποίησή τους με τρόπο που έχει συνοχή και συνεκτικότητα, όχι μόνο θα διευκολύνετε τους άλλους χρήστες αλλά και θα καταστήσετε τα δεδομένα συμβατά με τις υπηρεσίες και τις ροές εργασίας της υποδομής. Οι ακόλουθες κατευθυντήριες γραμμές έχουν ως στόχο να σας βοηθήσουν με τέτοιο τρόπο ώστε **να μην χάνεται καμία πληροφορία** και το κειμενικό τμήμα των δεδομένων που απαρτίζουν ένα σώμα κειμένων να είναι επεξεργάσιμο.

Προσοχή: Οι παρούσες κατευθυντήριες γραμμές δεν αφορούν την κατηγοριοποίηση με βάση τη θεματική περιοχή (domain), τη χρονική/γεωγραφική κάλυψη (time/geographic coverage) κλπ.

Πολλαπλά μορφότυπα (formats)

Εάν τα δεδομένα είναι σε διάφορους μορφότυπους (π.χ. XML, TXT, PDF κλπ.), οργανώστε τα αρχεία βάσει του μορφότυπου. Ομαδοποιήστε όλα τα αρχεία του ίδιου μορφότυπου σε έναν φάκελο (π.χ. όλα τα αρχεία XML μαζί). Μπορείτε να μεταφορτώσετε δύο διαφορετικά σύνολα δεδομένων (π.χ. XML έναντι TXT) στην ίδια εγγραφή μεταδεδομένων, συσχετίζοντας το καθένα από αυτά με διαφορετική *διανομή*.

Πρακτική συμβουλή: Δείτε τον κατάλογο των *συνιστώμενων μορφότυπων αρχείων* για την υποδομή CLARIN:EL.

Πολλαπλά μέσα (media)

Εάν τα δεδομένα, ή τμήματά τους, διατίθενται σε διάφορα μέσα (π.χ. κείμενο, ήχος κλπ.), οργανώστε τα αρχεία βάσει του μέσου. Ομαδοποιήστε όλα τα αρχεία του ίδιου μέσου σε έναν φάκελο (π.χ. όλα τα αρχεία κειμένου μαζί σε έναν φάκελο και όλα τα αρχεία ήχου σε έναν άλλο). Μπορείτε να μεταφορτώσετε δύο διαφορετικά σύνολα δεδομένων (π.χ. ηχογραφήσεις και απομαγνητοφωνήσεις) στην ίδια εγγραφή μεταδεδομένων, συσχετίζοντας το καθένα από αυτά με διαφορετική *διανομή*.

Ονοματοδοσία αρχείων και φακέλων

Ονομάστε τόσο τα αρχεία όσο και τους φακέλους με τρόπο που να αντικατοπτρίζει **με νόημα και με συνέπεια** το περιεχόμενό τους. Χρησιμοποιήστε το λατινικό αλφάβητο και μην αφήνετε κενά μεταξύ των λέξεων. Εάν έχετε αρχεία σε διαφορετικούς μορφότυπους, μέσα ή/και γλώσσες, ονομάστε τα ανάλογα (π.χ. news1_el.txt, news1_en.txt).

Σημαντικό: Οποιαδήποτε σχετική τεκμηρίωση (π.χ. εγχειρίδια, ερωτηματολόγια, codebooks, εκθέσεις έργου κλπ.) θα πρέπει να περιγραφεί στα αντίστοιχα πεδία μεταδεδομένων στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων¹ όπου πρέπει να μεταφορτωθεί και το αντίστοιχο αρχείο. Παρ' όλα αυτά, εάν επιθυμείτε να συμπεριλάβετε οποιαδήποτε τεκμηρίωση στο φάκελο δεδομένων, δημιουργήστε ένα ξεχωριστό αρχείο και ονομάστε το «README» (σε μορφή TXT ή PDF). Το αρχείο αυτό θα πρέπει να περιέχει όλες τις απαραίτητες πληροφορίες σχετικά με τις μεθόδους που χρησιμοποιήθηκαν για τη συλλογή/παραγωγή των δεδομένων και επεξηγήσεις σχετικά με τη δομή, την ονομασία των αρχείων ή κάθε άλλου είδους πληροφορία που μπορεί να βοηθήσει τον

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

χρήστη.

Συνέπεια

Τα μεταδεδομένα που χρησιμοποιούνται για την περιγραφή των δεδομένων σας θα πρέπει να τα αντικατοπτρίζουν με σαφήνεια. Βεβαιωθείτε ότι δεν υπάρχουν ασυνέπειες (π.χ. ελέγξτε ότι τα αρχεία σας είναι όντως σε μορφή PDF και όχι απλώς σαρωμένες εικόνες - αν παρέχετε πληροφορίες για ένα σχολιασμένο σώμα δεδομένων, αναφέρετε το εργαλείο σχολιασμού κλπ.) για να αποφύγετε τυχόν προβλήματα. Διαβάστε [εδώ](#) ποια είναι τα υποχρεωτικά μεταδεδομένα για όλους τους τύπους πόρων, αλλά λάβετε επίσης υπόψη σας ότι μια περιγραφή πόρου είναι πληρέστερη εάν παρέχονται και τα συνιστώμενα μεταδεδομένα.

Βήμα 3: Συμπίεση

Τα αρχεία περιεχομένου πρέπει να βρίσκονται σε έναν **συμπιεσμένο φάκελο** σε έναν από τους ακόλουθους μορφότυπους: **.zip, .tgz, .gz, .tar**. Κατά την ονομασία του φακέλου πρέπει να χρησιμοποιείτε το λατινικό αλφάβητο και να μην αφήνετε κενά μεταξύ των λέξεων.

Προσοχή: Μην συμπιέζετε τα εμφωλευμένα αρχεία/φακέλους καθώς κάτι τέτοιο καθιστά τα δεδομένα μη επεξεργάσιμα από τις υπηρεσίες CLARIN:EL (δηλ. μην δημιουργείτε .zip αρχεία μέσα σε ένα .zip αρχείο).

3.2. Ειδικές οδηγίες

Corpora

Για να γίνει επεξεργάσιμο, ένα σώμα κειμένων πρέπει να έχει τα χαρακτηριστικά που περιγράφονται παρακάτω:

- multilinguality:
 - για τα **μονόγλωσσα** σώματα κειμένων, η γλώσσα πρέπει να είναι τα *ελληνικά*, τα *αγγλικά*, τα *γερμανικά* ή τα *πορτογαλικά* (μέχρι στιγμής αυτές είναι οι γλώσσες που υποστηρίζονται από τις υπηρεσίες),
 - για τα **δίγλωσσα** σώματα κειμένων, τα *ελληνικά* πρέπει να είναι η μία γλώσσα του γλωσσικού ζεύγους όπου τα *αγγλικά*, τα *γερμανικά* ή τα *πορτογαλικά* είναι η άλλη.
- medium: *Text*
- format:
 - για τα **μονόγλωσσα** σώματα κειμένων τα μορφότυπα πρέπει να είναι είτε *Plain Text* είτε *XCES*,
 - για τα **δίγλωσσα** σώματα κειμένων τα μορφότυπα πρέπει να είναι είτε *TMX* είτε *MOSES*.
- encoding: *UTF-8*
- size: *< 60Mb*
- licence: Άδειες Creative Commons (CC, ξεκινώντας από την Creative Commons Zero (CC-0) και προχωρώντας σε όλους τους δυνατούς συνδυασμούς βάσει των διαφοροποιημένων δικαιωμάτων χρήσης κατά CC). Δείτε επίσης το [Προτεινόμενο σχήμα αδειοδότησης Γλωσσικών Πόρων](#).

Τα σώματα κειμένων με αυτά τα χαρακτηριστικά είναι συμβατά με τις ροές εργασίας της υποδομής και επισημαίνονται αυτόματα ως *processable* (επεξεργάσιμα) και παρουσιάζονται ως **προεπιλογή** στην αρχική σελίδα του κεντρικού καταλόγου.

Εργαλεία

Εάν θέλετε να ενσωματώσετε ένα εργαλείο στην υποδομή του CLARIN:EL ως συμβατή υπηρεσία, δηλώστε την επιλογή σας κατά τη δημιουργία του πόρου και επικοινωνήστε με την τεχνική ομάδα του [CLARIN:EL](#).

Do you want to contribute a tool that will be integrated in CLARIN:EL as a functional service (i.e., available through the CLARIN:EL APIs)?

☐ Yes

☐ No

18.1.4 4. ΜΕΤΑΦΟΡΤΩΣΗ

Όταν τελειώσετε, μπορείτε να μεταφορτώσετε τα δεδομένα.

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο στους *συνδεδεμένους επιμελητές*.

Ως επιμελητής έχετε δύο επιλογές για τη μεταφόρτωση:

- *κατά τη δημιουργία της εγγραφής μεταδεδομένων του πόρου,*
- *σε μεταγενέστερο στάδιο.*

Μόλις ολοκληρώσετε τη μεταφόρτωση, πρέπει να συσχετίσετε τα δεδομένα με μια **μορφή διανομής** (distribution), τον τρόπο ή το κανάλι διανομής μέσω του οποίου διανέμονται τα δεδομένα, που περιγράφεται *εδώ*.

Μπορείτε να επαναλάβετε τη διαδικασία (μεταφόρτωση δεδομένων → συσχέτιση με διανομή) όσες φορές χρειάζεται, έχοντας διαφορετικά σύνολα δεδομένων συσχετισμένα με διάφορες διανομές. Αυτή η λειτουργικότητα εξυπηρετεί όχι μόνο τους διάφορους τρόπους με τους οποίους **τα ίδια δεδομένα** διανέμονται (π.χ. ένα CD-ROM, ένας σύνδεσμος από όπου μπορεί να γίνει καταφόρτωση ενός συνόλου δεδομένων, κλπ.), αλλά και τους **διαφορετικούς μορφότυπους ή μέσα δεδομένων** (π.χ. PDF vs TXT, αρχεία ήχου vs μεταγραφές, κλπ.), που μπορούν να αντιμετωπιστούν ανεξάρτητα.

Πρακτική συμβουλή: Εάν αντιμετωπίσετε οποιοδήποτε πρόβλημα κατά τη διάρκεια της μεταφόρτωσης, παρακαλούμε επικοινωνήστε με το *γραφείο τεχνικής υποστήριξης*

18.2 II. Τα δεδομένα ως τροφοδοσία μίας υπηρεσίας

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη σε όλους τους *συνδεδεμένους χρήστες*.

Τόσο τα δεδομένα που μεταφορτώνονται προς επεξεργασία όσο και τα δεδομένα που προκύπτουν από την επεξεργασία **δεν αποθηκεύονται** μόνιμα στην υποδομή - η πολιτική του CLARIN:EL είναι να διαγράφονται τα επεξεργασμένα δεδομένα 48 ώρες μετά την ολοκλήρωση της επεξεργασίας.

Πρακτική συμβουλή: Αν θέλετε, μπορείτε να δημιουργήσετε μια εγγραφή μεταδεδομένων, στην οποία μπορείτε να μεταφορτώσετε τα δεδομένα, είτε χρησιμοποιώντας το *περιβάλλον τεκμηρίωσης και ανάρτησης πόρων* είτε

με τη μεταφόρτωση ενός αρχείου XML. Λάβετε υπόψη ότι για να το κάνετε αυτό θα πρέπει πρώτα να έχετε συνδεθεί.

Οι υπηρεσίες CLARIN:EL δέχονται στην είσοδό τους μικρά σύνολα δεδομένων με τα ακόλουθα χαρακτηριστικά:

- multilinguality: **μονόγλωσσα** σώματα κειμένων στα ελληνικά, αγγλικά, γερμανικά ή πορτογαλικά,
- medium: *Text*
- format: *Plain Text*
- encoding: *UTF-8*
- size: *< 2Mb*

Επιπλέον, τα αρχεία περιεχομένου πρέπει να βρίσκονται σε έναν **συμπιεσμένο φάκελο** σε έναν από τους ακόλουθους μορφότευπους: **.zip, .tgz, .gz, .tar**. Κατά την ονομασία του φακέλου πρέπει να χρησιμοποιείτε το λατινικό αλφάβητο και να μην αφήνετε κενά μεταξύ των λέξεων.

Προσοχή: Μην συμπιέζετε τα εμφωλευμένα αρχεία/φακέλους καθώς κάτι τέτοιο καθιστά τα δεδομένα μη επεξεργάσιμα από τις υπηρεσίες CLARIN:EL (δηλ. μην δημιουργείτε .zip αρχεία μέσα σε ένα .zip αρχείο).

Για να μάθετε περισσότερα σχετικά με την επεξεργασία, δείτε πώς να μπορείτε να αποκτήσετε πρόσβαση:

1. στις *υπηρεσίες*,
2. στις *ροές εργασίας*.

Συνιστώμενα μορφότυπα (formats) αρχείων

19.1 Συστάσεις για την επιλογή μορφότυπων αρχείων που διασφαλίζουν μακροπρόθεσμα την προσβασιμότητα και διαλειτουργικότητα των δεδομένων

Στην παρούσα ενότητα¹ παρουσιάζονται τα μορφότυπα αρχείων που συνιστώνται για την μεταφόρτωση δεδομένων στο CLARIN:EL.

19.1.1 Μορφότυπα (formats) αρχείων που υπηρετούν την Πολιτική Ψηφιακής Διατήρησης

Κατά τον καθορισμό των συνιστώμενων μορφότυπων αρχείων, η ομάδα του CLARIN:EL συνυπολόγισε τους ακόλουθους παράγοντες που επιτρέπουν στα δεδομένα να παραμένουν προσβάσιμα και να μπορούν να επαναχρησιμοποιηθούν σε βάθος χρόνου.

- Για την **επεξεργασιμότητα** (processability) λήφθηκαν υπόψιν:
 - η καταλληλότητα του μορφότυπου για τον τύπο του πόρου ή/και τον τύπο της επεξεργασίας, και
 - η συμβατότητά του με τις **ροές εργασίας CLARIN:EL** (τα κειμενικά δεδομένα, προκειμένου να είναι επεξεργάσιμα, πρέπει να είναι σε έναν από τους μορφότυπους που επεξεργάζονται οι ροές εργασίας).
- Για την **διατήρηση** (preservation) λήφθηκαν υπόψιν:
 - η καταλληλότητα του μορφότυπου για έρευνα, όπως έχει οριστεί από τις σχετικές κοινότητες,
 - ο βαθμός διάδοσής του (προτιμήθηκαν ευρέως χρησιμοποιούμενα μορφότυπα, που δεν έχουν καταργηθεί, γνωστά στις καθορισμένες κοινότητες),
 - ο τρόπος διάθεσης (επιλέχθηκαν μορφότυπα ανοικτού έναντι κλειστού, ιδιόκτητου, κώδικα), και
 - το αν ο μορφότυπος εφαρμόζει συμπίεση με ή χωρίς απώλειες.

¹ Οι παρούσες οδηγίες έχουν συνταχθεί από την **τεχνική ομάδα του CLARIN:EL** στην οποία μπορείτε να απευθυνθείτε για οποιαδήποτε ερώτηση ή παρατήρηση.

Η πολιτική, που βασίζεται στα όσα προαναφέρθηκαν, ανταποκρίνεται στην αποστολή του CLARIN:EL να συλλέγει, να διατηρεί και να διανέμει ψηφιακούς γλωσσικούς πόρους και υπηρεσίες γλωσσικής επεξεργασίας για την υποστήριξη ερευνητών, ακαδημαϊκών, φοιτητών, επαγγελματιών του γλωσσικού τομέα, επιστημόνων καθώς και του ευρύτερου κοινού. Προκειμένου να δημιουργηθούν οι κατάλληλες συστάσεις για τους επιμέρους μορφότυπους αρχείων ή να αποφασιστεί η καταλληλότητά τους για συγκεκριμένα είδη ερευνητικών δραστηριοτήτων/τύπων, πρέπει να εξεταστεί ο σκοπός για τον οποίο προορίζονται. Για παράδειγμα, ενώ ο μορφότυπος PDF/A έχει αναπτυχθεί για την χωρίς προβλήματα μακροπρόθεσμη αρχειοθέτηση και αποτελεί εξαιρετική επιλογή μορφότυπου για τεκμηρίωση, είναι αναμφίβολα *ακατάλληλος* για κειμενικά δεδομένα που προορίζονται για γλωσσική επεξεργασία. Επομένως, με βάση τους τύπους πόρων που εμπίπτουν στο πεδίο εφαρμογής των χρηστών του CLARIN:EL και των υπηρεσιών που προσφέρονται/υποστηρίζονται, η ομάδα του CLARIN:EL διακρίνει τις ακόλουθες κατηγορίες, ως προς τους ψηφιακούς γλωσσικούς πόρους, για τις οποίες παρέχονται συγκεκριμένες συστάσεις:

- **CLARIN:EL processable data** (επεξεργάσιμα δεδομένα στην υποδομή CLARIN:EL): πρόκειται για κειμενικά δεδομένα² που μπορούν να τροφοδοτήσουν τις *ροές εργασίας* του CLARIN:EL,
- **Textual Data** (κειμενικά δεδομένα): γραπτό μη δομημένο/απλό κείμενο (plain text) ή δομημένο (π.χ. HTML) χωρίς γλωσσική ή άλλη σήμανση που προστίθεται για ερευνητικούς σκοπούς (*μη επεξεργάσιμα* από τις ροές εργασίας CLARIN:EL),
- **Text Annotation** (κειμενική επισημείωση): επισημειώσεις κειμενικών δεδομένων της γλώσσας προέλευσης, με το αρχικό κείμενο να συμπεριλαμβάνεται ή να υφίσταται ως ξεχωριστό έγγραφο (stand-off document),
- **Language Description** (γλωσσική περιγραφή): δεδομένα που περιγράφουν μια γλώσσα ή κάποια πτυχή/πτυχές μιας γλώσσας μέσω συστηματικής τεκμηρίωσης γλωσσικών δομών (γραμματικές, μοντέλα μηχανικής μάθησης (ML), μοντέλα N-γραμμάτων),
- **Lexical/Conceptual Resource** (λεξικός/εννοιολογικός πόρος): πόρος οργανωμένος με βάση λεξιλογικές ή εννοιολογικές καταχωρίσεις (λεξιλογικά στοιχεία, όρους, έννοιες κ.λπ.) με συμπληρωματικές πληροφορίες (π.χ. μορφολογικές, σημασιολογικές, στατιστικές πληροφορίες κλπ.),
- **Image data** (δεδομένα εικόνας): ψηφιοποιημένες εικόνες αναλογικών πηγών δεδομένων γραπτού λόγου για ερευνητικούς σκοπούς (π.χ. σαρώσεις χειρογράφων, φωτογραφίες επιγραφών) ή δισδιάστατες εικόνες ή σχήματα που διανέμονται με τα σχετικά κειμενικά δεδομένα για ανάλυση επεξεργασίας φυσικής γλώσσας, NLP, (π.χ. ιατρικές εικόνες, *δεδομένα εικόνας*, συνοδευόμενες από ακτινολογικές γνωματεύσεις, *κειμενικά δεδομένα*),
- **Audio data** (ηχητικά δεδομένα): ηχογραφήσεις που προσφέρουν δεδομένα προφορικού λόγου για ερευνητικούς σκοπούς (π.χ. αρχεία ήχου με την προφορά λέξεων για ένα λεξικό, ηχογραφημένες συνεντεύξεις, ραδιοφωνικές εκπομπές κλπ.),
- **Video data** (δεδομένα βίντεο): βιντεοσκοπήσεις που προσφέρουν πολυτροπικά ή δεδομένα νοηματικής γλώσσας για ερευνητικούς σκοπούς.

19.1.2 Συστάσεις μορφότυπων

Προτιμώνται τα μορφότυπα που πληρούν τα κριτήρια της Πολιτικής Ψηφιακής Διατήρησης που προαναφέρθηκαν. Ωστόσο, γίνονται δεκτά και άλλα μορφότυπα, ως *επίπεδο πρώτης εισόδου*, που προτείνεται να μετατραπούν στους συνιστώμενους.

Ως εκ τούτου, τα μορφότυπα αρχείων κατηγοριοποιούνται σε δύο επίπεδα διατήρησης (**recommended**/συνιστώμενα, **acceptable**/αποδεκτά) πάντα στο πλαίσιο της κάθε περίπτωσης. Ο κατάλογος των αποδεκτών μορφότυπων δεν είναι εξαντλητικός, ιδίως στην περίπτωση της κειμενικής επισημείωσης (text annotation), αλλά μάλλον ενδεικτικός, ενώ προτείνεται και η μετατροπή των αποδεκτών μορφότυπων σε συνιστώμενα.

² Δείτε *εδώ* τις οδηγίες για τα επεξεργάσιμα σώματα κειμένων (*processable*).

	Recommended	Acceptable
CLARIN:EL processable data	Monolingual textual data: plain text Monolingual encoded data: XCES-ILSP variant (XML based format compliant with the XCES model for corpora) Bi-/Multilingual encoded data: TMX (XML based format for aligned data), MOSES (text-based format for parallel data)	
Textual Data	File Formats: plain text Formatted/Encoded: ODT, DOCX, PDF/A, HTML, Latex, TeX, MOSES	PDF, SGML, Rich Text Format (.rtf), Microsoft Word (.doc, .docx), PostScript
Text Annotation	File Formats: XML, XMI, CSV, TSV, RDF (all serialisation formats RDF/XML, Turtle, Notation3, N-Triples, TriG, N-Quads, JSON-LD, HDT), JSON Models: XCES for corpora and structural annotation, TEI for structural and linguistic annotation, GrAF linguistic annotation, TMX for aligned, GATE linguistic annotation, CoNLL family (CoNLL-U, CoNLL-2000, CoNLL-2002, CoNLL-2003, CoNLL-2006, CoNLL-2008, CoNLL-2009, CoNLL-2012) for linguistic annotation, NIF linguistic annotation for RDF data, WARC for web crawled data	SGML, Plain Text, Microsoft Excel (.xlsx, .xls), ELLOGON
Language Description	ML Model: H5, ProtoBuf, ONNX, PMML, Pickle, MLeap, YAML, JSON N-gram model: ARPA	
Lexical/Conceptual Resource	File Formats: XML, CSV, TSV, RDF (RDF/XML, Turtle, Notation3, N-Triples, TriG, N-Quads, JSON-LD, HDT), OWL Models: LMF for lexica, OWL for ontologies, SKOS for thesauri, OntoLex-Lemon for lexica, TBX for terminological data	Microsoft Excel (.xlsx, .xls), Plain Text, SQL
Image data	All images: TIFF, SVG, JPEG 2000, PNG, GIF Scanned images: PDF/A	JPEG, BMP, Photoshop, NifTi, FlashPix, PDF
Audio data	WAV, AIFF, FLAC	MP3, MPEG, Windows Media Audio
Video data	AVI	MPEG-4, RealNetworks 'Real Video', Windows Media Video, Flash Video, QuickTime Video

Τι είναι τα μεταδεδομένα και γιατί είναι τόσο σημαντικά;

ορισμός

Τα μεταδεδομένα είναι «δεδομένα που παρέχουν πληροφορίες για άλλα δεδομένα»

Τα δεδομένα για τα οποία επιθυμούμε να έχουμε πληροφορίες είναι **γλωσσικά δεδομένα** και **εργαλεία/υπηρεσίες** που τα επεξεργάζονται. Τα βασικά **μεταδεδομένα** που χρησιμοποιούνται για την περιγραφή των προαναφερθέντων είναι:

- corpora (σώματα κειμένων - δηλαδή συλλογές κειμένων ή άλλων μέσων),
- lexical/conceptual resources (λεξικοί/εννοιολογικοί πόροι - δηλ. συλλογές όρων),
- language descriptions (γλωσσικές περιγραφές - π.χ. γραμματικές), και
- tools or services (εργαλεία ή υπηρεσίες - π.χ. λογισμικό για την επεξεργασία φυσικής γλώσσας).

Αυτά τα μεταδεδομένα έχουν πολλά και διαφορετικά χαρακτηριστικά και ιδιότητες. Για παράδειγμα, το μεταδεδομένο corpus έχει πολλά παιδιά (*children*), ιεραρχικά εξαρτώμενα στοιχεία, όπως φαίνεται στην εικόνα, τα οποία είναι με τη σειρά τους τα ίδια μεταδεδομένα:

Element **ms:Corpus**

Namespace	http://w3id.org/meta-share/meta-share/
Annotations	<input checked="" type="checkbox"/> A structured collection of pieces of data (textual, audio, video, multimodal/multimedia, etc.) typically of considerable size and selected according to criteria external to the data (e.g. size, type of language, type of text producers or expected audience, etc.) to represent as comprehensively as possible the object of study
Diagram	<input checked="" type="checkbox"/>
Properties	<input checked="" type="checkbox"/> Content complex
Used by	<input checked="" type="checkbox"/> Element ms:LRSubclass
Model	<input checked="" type="checkbox"/> ms:lrType , ms:corpusSubclass , ms:CorpusMediaPart+ , ms:DatasetDistribution+ , ms:personalDataIncluded , ms:personalDataDetails+ , ms:sensitiveDataIncluded , ms:sensitiveDataDetails+ , ms:anonymized(0,1) , ms:anonymizationDetails+ , ms:isAnalysedBy+ , ms:isEditedBy+ , ms:isElicitedBy+ , ms:isAnnotatedVersionOf(0,1) , ms:isAlignedVersionOf(0,1) , ms:isConvertedVersionOf+ , ms:timeCoverage+ , ms:geographicCoverage+ , ms:register+ , ms:userQuery(0,1)
Children	<input checked="" type="checkbox"/> ms:CorpusMediaPart , ms:DatasetDistribution , ms:anonymizationDetails , ms:anonymized , ms:corpusSubclass , ms:geographicCoverage , ms:isAlignedVersionOf , ms:isAnalysedBy , ms:isAnnotatedVersionOf , ms:isConvertedVersionOf , ms:isEditedBy , ms:isElicitedBy , ms:lrType , ms:personalDataDetails , ms:personalDataIncluded , ms:register , ms:sensitiveDataDetails , ms:sensitiveDataIncluded , ms:timeCoverage , ms:userQuery
Instance	<input checked="" type="checkbox"/> <pre> <ms:Corpus xmlns:ms="http://w3id.org/meta-share/meta-share/"> <ms:lrType>{1,1}</ms:lrType> <ms:corpusSubclass>{1,1}</ms:corpusSubclass> <ms:CorpusMediaPart>{1,unbounded}</ms:CorpusMediaPart> <ms:DatasetDistribution>{1,unbounded}</ms:DatasetDistribution> <ms:personalDataIncluded>{1,1}</ms:personalDataIncluded> <ms:personalDataDetails xml:lang="">{0,unbounded}</ms:personalDataDetails> <ms:sensitiveDataIncluded>{1,1}</ms:sensitiveDataIncluded> <ms:sensitiveDataDetails xml:lang="">{0,unbounded}</ms:sensitiveDataDetails> <ms:anonymized>{0,1}</ms:anonymized> <ms:anonymizationDetails xml:lang="">{0,unbounded}</ms:anonymizationDetails> <ms:isAnalysedBy>{0,unbounded}</ms:isAnalysedBy> <ms:isEditedBy>{0,unbounded}</ms:isEditedBy> <ms:isElicitedBy>{0,unbounded}</ms:isElicitedBy> <ms:isAnnotatedVersionOf>{0,1}</ms:isAnnotatedVersionOf> <ms:isAlignedVersionOf>{0,1}</ms:isAlignedVersionOf> <ms:isConvertedVersionOf>{0,unbounded}</ms:isConvertedVersionOf> <ms:timeCoverage xml:lang="">{0,unbounded}</ms:timeCoverage> <ms:geographicCoverage xml:lang="">{0,unbounded}</ms:geographicCoverage> <ms:register xml:lang="">{0,unbounded}</ms:register> <ms:userQuery>{0,1}</ms:userQuery> </ms:Corpus> </pre>
Source	<input checked="" type="checkbox"/> <pre> <xs:element name="Corpus"> <xs:annotation> <xs:documentation xml:lang="en">A structured collection of pieces of data (textual, audio, video, multimodal/multimedia, etc.) typically of considerable size and selected according to criteria external to the data (e.g. size, type of language, type of text producers or expected audience, etc.) to represent as comprehensively as possible the object of study</xs:documentation> </xs:annotation> <xs:appinfo> <identifier>http://w3id.org/meta-share/meta-share/Corpus</identifier> <label xml:lang="en">Corpus</label> <subclassOf>http://w3id.org/meta-share/meta-share/DataLanguageResource</subclassOf> </xs:appinfo> </xs:element> <xs:complexType> <xs:sequence> <xs:element name="lrType" fixed="Corpus"> <xs:annotation> <xs:documentation xml:lang="en">Classifies the language resource described by a metadata record to one of the major classes</xs:documentation> </xs:annotation> <xs:appinfo> <indexed>true</indexed> </xs:appinfo> </xs:element> </xs:sequence> </xs:complexType> </pre>

Αυτό που φαίνεται στην εικόνα είναι ένα μέρος του σχήματος μεταδεδομένων *CLARIN:EL metadata schema* που είναι αφιερωμένο στο στοιχείο `corpus`. Ένα **σχήμα** (schema) είναι ένας περίπλοκος λεπτομερής *χάρτης* όπου όλα τα στοιχεία εντοπίζονται, ορίζονται, περιγράφονται και συσχετίζονται μεταξύ τους ιεραρχικά. Όλες αυτές οι πληροφορίες αποθηκεύονται σε ένα εξωτερικό έγγραφο που ονομάζεται **XSD: XML Schema Documentation**.

XML σημαίνει **eXtensible Markup Language**. Είναι μια γλώσσα που έχει σχεδιαστεί για την επισήμανση δεδομένων με τη χρήση **ετικετών** `<>` (tags)¹. Οι ετικέτες αντιπροσωπεύουν τη δομή των δεδομένων και περιέχουν **μεταδεδομένα**. Το XSD περιέχει, επίσης, κανόνες με τους οποίους πρέπει να συμμορφώνεται ένα έγγραφο XML προκειμένου να είναι *έγκυρο* (valid) σύμφωνα με το συγκεκριμένο σχήμα.

Ένα σχήμα δημιουργείται για να βοηθήσει διαφορετικούς τύπους χρηστών να **περιγράψουν**, να **οργανώσουν**, να **ανακτήσουν** και να **επαναχρησιμοποιήσουν** τους πόρους (για περισσότερες πληροφορίες δείτε την ενότητα *Fair Principles*). Όσον αφορά στους πόρους που βρίσκονται στο **CLARIN:EL**, το σχήμα που δημιουργήθηκε παρέχει πληροφορίες σε ερωτήματα, όπως τα ακόλουθα:

- **Ποια** είναι η φύση των πόρων;
- **Πώς** δημιουργήθηκαν οι πόροι;
- **Γιατί** δημιουργήθηκαν;
- **Πότε** δημιουργήθηκαν;
- **Ποιος** τους δημιούργησε;
- **Ποιά** ήταν τα **πρότυπα/εργαλεία/τεχνικές** που χρησιμοποιήθηκαν (αν χρησιμοποιήθηκαν);
- Ποιο είναι το **μέγεθός τους** (σε διάφορες μονάδες μέτρησης);

¹ Μπορείτε να εξαγάγετε την περιγραφή μεταδεδομένων ενός πόρου σε αρχείο XML, επισκεπτόμενοι τη [σελίδα θέασης](#) του πόρου.

- Ποια ήταν η **πηγή** τους;

Το σχήμα μεταδεδομένων CLARIN:EL έχει επίσης προβλέψει για τα διάφορα μέσα (media), τις γλώσσες και άλλες χρήσιμες πληροφορίες που αφορούν όλους τους τύπους πόρων, που εκφράζονται από τα αντίστοιχα μεταδεδομένα.

Κάθε πληροφορία που κωδικοποιείται ως μεταδεδομένο είναι *περισσότερο ή λιγότερο αναγκαία* για την περιγραφή ενός πόρου. Αυτό εκφράζεται από τους διάφορους βαθμούς **προαιρετικότητας** (optionality) όπως φαίνεται στον ακόλουθο πίνακα:

Εάν ένα μεταδεδομένο είναι	Τότε
υποχρεωτικό (mandatory)	πρέπει να συμπληρώνεται πάντα
συνιστώμενο (recommended)	και αυτό είναι σημαντικό, επομένως θα πρέπει να συμπληρώνεται
υπό συνθήκη υποχρεωτικό (mandatory upon condition)	γίνεται υποχρεωτικό μετά τη συμπλήρωση μιας συγκεκριμένης τιμής <i>ενός άλλου μεταδεδομένου</i>
υπό συνθήκη συνιστώμενο (recommended upon condition)	γίνεται συνιστώμενο μετά τη συμπλήρωση μιας συγκεκριμένης τιμής <i>ενός άλλου μεταδεδομένου</i>
προαιρετικό (optional)	«μη λέτε ποτέ αυτό το μεταδεδομένο δεν είναι χρήσιμο· δείξτε γενναιοδωρία και συμπληρώστε το!» ²

Πρακτική συμβουλή: Δείτε *εδώ* τα υποχρεωτικά μεταδεδομένα για το CLARIN:EL.

Κάθε μεταδεδομένο λαμβάνει μια συγκεκριμένη τιμή (value). Αυτή η τιμή είναι το αποδεκτό περιεχόμενο που περιλαμβάνεται μεταξύ των ετικετών και ποικίλλει, από αλφαριθμητικές συμβολοσειρές (alphanumeric strings) έως δεκαδικούς αριθμούς (float numbers), διευθύνσεις URL κλπ. Στα ακόλουθα παραδείγματα εμφανίζονται κάποιες από αυτές τις τιμές (*κάντε κλικ στο βέλος για να εμφανιστεί το παράδειγμα*).

`<ms:keyword xml:lang=»en»> alignment </ms:keyword>`

`<ms:categoryLabel xml:lang=»en»> Political Science </ms:categoryLabel>`

`<ms:description xml:lang=»en»> This is a collection of the raw minutes of the Greek Parliament plenary sessions of the last 30 years (more than 1.000.000 speeches). The existing corpus has all raw data in txt format. In order to make the resource more processable, we have also split it into smaller subcorpora, with a maximum compressed folder size of 40 Mb per subcorpus. The created subcorpora are thematically organized per Greek parliamentary terms. </ms:description>`

`<ms:creationStartDate> 2005-10-01 </ms:creationStartDate>`

`<ms:amount> 100000.0 </ms:amount>`

`<ms:website> http://www.ilsp.gr/ </ms:website>`

`<ms:email> name@athenarc.gr </ms:email>`

`<ms:additionalInfo> <ms:email> name@athenarc.gr </ms:email> </ms:additionalInfo>`

Μπορείτε να δείτε περισσότερα παραδείγματα *εδώ*.

² FAIR Principles > F2: Data are described with rich metadata.

Η υποδομή CLARIN:EL και το *σχήμα μεταδεδομένων* υποστηρίζουν τις αρχές FAIR : **Ευρεσιμότητα** (Findability), **Προσβασιμότητα** (Accessibility), **Διαλειτουργικότητα** (Interoperability) και **Επαναχρησιμοποίηση** (Reuse) των ψηφιακών αντικειμένων. Αυτή η ενότητα παρέχει μια επισκόπηση των αρχών FAIR. Για περισσότερες πληροφορίες, παρακαλούμε, επισκεφθείτε τον δικτυακό τόπο **GoFair** όπου κάθε μία από τις αρχές αναλύεται περαιτέρω.

21.1 Ευρεσιμότητα (Findability)

Πρώτα τα δεδομένα πρέπει να βρεθούν. Ένας από τους τρόπους εντοπισμού ενός ψηφιακού αντικειμένου ανεξάρτητα από τις αλλαγές της θέσης του στο διαδίκτυο είναι με το **μόνιμο αναγνωριστικό PID** (persistent identifier). Το PID είναι μια συμβολοσειρά που προσδιορίζει μοναδικά ένα ψηφιακό αντικείμενο. Στο **CLARIN:EL** κάθε πόρος λαμβάνει ένα PID κατά τη δημοσίευσή του και μπορεί να ανακτηθεί με αυτό ακόμη και αν ο πόρος έχει αφαιρεθεί από τον κεντρικό κατάλογο. Για παράδειγμα, το *Sentiment Analysis Tool* δεν βρίσκεται πλέον δημοσιευμένο στον κεντρικό κατάλογο, αλλά χρησιμοποιώντας το PID του (<http://hdl.handle.net/11500/DEMOKRITOS-0000-0000-24A2-0>) ο χρήστης οδηγείται στη σελίδα προβολής του πόρου όπου μια ετικέτα ενημερώνει ότι **αυτός ο πόρος δεν είναι προσωρινά διαθέσιμος**.

STATUS

this resource is temporarily unavailable

Sentiment Analysis Tool

 ToolService

Sent

Version: 1.0

<http://hdl.handle.net/11500/DEMOKRITOS-0000-0000-24A2-0>

The sentiment analysis tool is a text classification and sentiment extraction tool based on n-gram graph text representations. It may be paired with various machine learning algorithms for the generation of the language model. It can be accessed by a URL endpoint as a REST service. It has been used as is, or as a part


[Read more](#)

Select Language

en



Cite current version

Giannakopoulos, George (2015). Sentiment Analysis Tool. Version 1.0. [Software (Tool/Service)]. CLARIN:EL. <http://hdl.handle.net/11500/DEMOKRITOS-0000-0000-24A2-0> 

Actions

Ένας άλλος τρόπος εύρεσης δεδομένων είναι μέσω των **περιγραφών μεταδεδομένων τους**. Όσο περισσότερα (και ακριβή) μεταδεδομένα παρέχονται, τόσο το καλύτερο. Στον ιστότοπο του GoFair επισημαίνεται η σημασία των μεταδεδομένων με έναν απλό κανόνα: «*μη λέτε ποτέ αυτό το μεταδεδομένο δεν είναι χρήσιμο*»· δείξτε γενναιοδωρία και συμπληρώστε το!»

21.2 Προσβασιμότητα (Accessibility)

Μόλις βρεθούν τα δεδομένα, η πρόσβαση σε αυτά θα πρέπει να είναι κάτι γνωστό στον χρήστη: «*οποιοσδήποτε με υπολογιστή και σύνδεση στο διαδίκτυο μπορεί να έχει πρόσβαση τουλάχιστον στα μεταδεδομένα*»¹. Προσβασιμότητα σε αυτό το πλαίσιο είναι η δυνατότητα ανάκτησης δεδομένων και μεταδεδομένων χωρίς εξειδικευμένα ή ιδιόκτητα εργαλεία ή μεθόδους επικοινωνίας. Ωστόσο, η προσβασιμότητα υπόκειται σε περιορισμούς. Όπου είναι απαραίτητο θα μπορούσε να χρειαστεί πιστοποίηση ταυτότητας (authentication) ή/και εξουσιοδότηση (authorisation). Στο CLARIN:EL η πιστοποίηση ταυτότητας και η εξουσιοδότηση απαιτούνται όταν ένας χρήστης ζητά πρόσβαση σε συγκεκριμένα δικαιώματα (ως *επιμελητής*, *επικυρωτής* ή *υπεύθυνος αποθετηρίου*) ή όταν κάποιος θέλει πρόσβαση στις υπηρεσίες επεξεργασίας. Σε αυτές τις περιπτώσεις ο χρήστης πρέπει πρώτα να *εγγραφεί/συνδεθεί* - η περιήγηση στον κεντρικό κατάλογο, η θέαση πόρων, η εξαγωγή περιγραφών μεταδεδομένων καθώς και η καταφόρτωση πόρων επιτρέπονται και σε μη εγγεγραμμένους χρήστες.

Η προσβασιμότητα εξασφαλίζεται επίσης όταν τα μεταδεδομένα είναι διαθέσιμα ακόμη και όταν τα δεδομένα δεν είναι. Εκτός από την προαναφερθείσα ετικέτα για πόρους των οποίων η δημοσίευση έχει αρθεί, το CLARIN:EL έχει επίσης μόνο **ενημερωτικούς πόρους** (for info), είτε επειδή τα δεδομένα τους είναι υπό επεξεργασία και δεν είναι έτοιμα να δημοσιευτούν είτε επειδή εκκρεμεί η εκκαθάριση των νομικών δικαιωμάτων που διέπουν τη χρήση τους. Αυτές οι εγγραφές μεταδεδομένων εξακολουθούν να προσφέρουν όλες τις απαραίτητες πληροφορίες σχετικά με τα επερχόμενα δεδομένα και δίνουν και στοιχεία επικοινωνίας.

clarin:el

CLARIN:EL portal >

Help Sign in

Go to inventory

Greek Short Stories in Greek Literary Periodicals (1880-1930)
[Bibliographic Database]

Corpus

GSSLP

Version: 1.17 (2020-09-08)

<http://hdl.handle.net/11500/UOC-0000-0000-5ED0-C>

The language resource "Short stories in the Greek Periodicals (1880-1930)" is the modified form of the research project "From classic to modern: Charting the formal evolution of Modern Greek short stories in literary periodicals (1880 -1930)", which was prepared in the framework of the "Research laboratory of literary g

Read more

for info

Select Language
en

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Cite current version

Laboratory of Literary Genres and Literary History - Department of Philology (2020, September 08). Greek Short Stories in Greek Literary Periodicals (1880-1930) [Bibliographic Database]. Version 1.17, [Dataset (Text corpus)]. CLARIN:EL.
<http://hdl.handle.net/11500/UOC-0000-0000-5ED0-C>

¹ FAIR Principles > A1.1: The protocol is open, free and universally implementable.

21.3 Διαλειτουργικότητα (Interoperability)

Η διαλειτουργικότητα αφορά τόσο τα δεδομένα και τα μεταδεδομένα όσο και την αντίληψή τους από ανθρώπους και υπολογιστές. Με απλά λόγια, η ανταλλαγή και η ερμηνεία των δεδομένων θα πρέπει να είναι μια απρόσκοπτη προσπάθεια μεταξύ ανθρώπων ή μηχανών. Για να καταστεί δυνατή η αναγνωσιμότητα χωρίς την ανάγκη πρόσθετου λογισμικού (αλγόριθμοι, μεταφραστές, αντιστοιχίσεις) θα πρέπει να χρησιμοποιούνται κοινά αποδεκτά «ελεγχόμενα λεξιλόγια, οντολογίες, θησαυροί και ένα καλό μοντέλο δεδομένων (ένα καλά καθορισμένο πλαίσιο για την περιγραφή και τη δόμηση των (μετα)δεδομένων)»².

Για τον σκοπό αυτό, χρειάζονται και συγκεκριμένοι τρόποι αναφοράς μεταξύ των πόρων (διασταυρούμενες αναφορές που δηλώνουν με ρητό τρόπο πως συνδέονται οι πόροι). Στο CLARIN:EL το σχήμα μεταδεδομένων έχει προβλέψει τέτοιους τρόπους σύνδεσης, ένα παράδειγμα των οποίων παρουσιάζεται στην ακόλουθη εικόνα. Το OROSSIMO Corpus - Economics είναι, όπως φαίνεται στο τμήμα των σχέσεων (relations) της σελίδας θέασης του πόρου μέρος (part of) του OROSSIMO Corpus και είχε ως αποτέλεσμα (has outcome) τον λεξικό/εννοιολογικό πόρο Orossimo Terminological Resource - Economics.

Relations to other resources

Part of
[OROSSIMO Corpus \(1.0.0 \(automatically assigned\)\)](http://hdl.handle.net/11500/ATHENA-0000-0000-2410-5)
[http://hdl.handle.net/11500/ATHENA-0000-0000-2410-5 \(Handle\)](http://hdl.handle.net/11500/ATHENA-0000-0000-2410-5)

Relations to other entities

Has outcome
[Orossimo Terminological Resource - Economics 1.0.0 \(automatically assigned\)](#)

Επιπλέον, για όλους τους πόρους η παραπομπή (citation) περιλαμβάνει το PID τους.

OROSSIMO Corpus - Economics



Version: 1.0.0 (automatically assigned)

<http://hdl.handle.net/11500/ATHENA-0000-0000-240D-A>

A corpus of academic discourse texts belonging to the Economics domain (according to the Dewey Decimal classification, DDC33 - Economics), annotated at structural level conformant to the XCES standard. The terms contained in the texts are included in a bilingual terminological glossary (see Orossimo terminological reso

[Read more](#)

processable

Select Language

en



Cite current version

Institute for Language and Speech Processing
- Athena Research Center (2015). OROSSIMO
Corpus - Economics. Version 1.0.0
(automatically assigned). [Dataset (Text
corpus)]. CLARIN:EL.
<http://hdl.handle.net/11500/ATHENA-0000-0000-240D-A>

Cite all versions

Institute for Language and Speech Processing
- Athena Research Center (2015). OROSSIMO
Corpus - Economics. [Dataset (Text corpus)].
CLARIN:EL.
<http://hdl.handle.net/11500/CLARIN-EL-0000-0000-6948-A>

² FAIR Principles > I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

21.4 Επαναχρησιμοποίηση (Reuse)

Για να είναι δυνατή η επαναχρησιμοποίηση των δεδομένων, τα μεταδεδομένα με τα οποία θα περιγραφούν θα πρέπει να είναι πλούσια, καλύπτοντας όλες τις πλευρές που σχετίζονται με τον τρόπο παραγωγής των δεδομένων. Ο όρος **πληθώρα** (plurality) χρησιμοποιείται «για να υποδηλώσει ότι ο επιμελητής των μεταδεδομένων θα πρέπει να είναι όσο το δυνατόν πιο γενναιόδωρος στην παροχή μεταδεδομένων, ακόμη και στην περίπτωση πληροφοριών που μπορεί να μη φαίνονται σχετικές»³.

Για τον ίδιο λόγο, το καθεστώς αδειοδότησης των δεδομένων θα πρέπει να είναι σαφές. Το *σχήμα μεταδεδομένων* CLARIN:EL διαθέτει πολλά μεταδεδομένα που καλύπτουν όλες τις πτυχές νομικής αδειοδότησης (όροι αδειοδότησης, URL, περιορισμοί χρήσης κλπ.) ενός πόρου. Αυτά εμφανίζονται ως πεδία μεταδεδομένων στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων⁴, όπως δείχνει η ακόλουθη εικόνα:

Licence name *

Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International

The full title of the licence (or terms of use, terms of service)

language

English

select language

remove

+

Licence/terms of use/service URL *

The URL where the text of the licence is found

+

Condition of use *

Select the condition(s) of use for accessing the corpus; the list includes the most frequently used conditions and aims to provide brief human readable information, while the proper exposition of all conditions and possible exceptions is found inside the licence text

Licence identifier

A string used to uniquely identify the licence

Fill in

Add

Cost

The cost for using the corpus

Fill in

Access rights

The rights for accessing the distributable form(s) of the corpus (preferably in accordance to a formalised vocabulary)

Fill in

³ FAIR Principles > R1: (Meta)data are richly described with a plurality of accurate and relevant attributes

⁴ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

Υποχρεωτικά μεταδεδομένα

Το *σχήμα μεταδεδομένων* έχει στοιχεία των οποίων η αναγκαιότητα, δηλαδή κατά πόσο θεωρούνται απαραίτητα για την περιγραφή ενός πόρου, ποικίλλει. Ανεξάρτητα από τον τρόπο που επιλέγεται για τη δημιουργία ενός πόρου (μέσω του *περιβάλλοντος τεκμηρίωσης και ανάρτησης*¹ ή μέσω *μεταφόρτωσης αρχείου μεταδεδομένων XML*), τα **υποχρεωτικά** μεταδεδομένα πρέπει να συμπληρωθούν. Διαφορετικά, η εγγραφή δεν θα αποθηκευτεί ή το αρχείο XML δεν θα μπορέσει να εισαχθεί στην υποδομή για την δημιουργία εγγραφής. Οι ακόλουθες ενότητες περιέχουν εικόνες που προσφέρουν μια επισκόπηση αυτών των μεταδεδομένων ανά τύπο πόρου και εξηγούνται εν συντομία. Τα μεταδεδομένα που επισημαίνονται με **αστερίσκο (*)** είναι **υπό συνθήκη υποχρεωτικά**. Αυτό σημαίνει ότι η αναγκαιότητά τους πηγάζει από τις τιμές άλλων μεταδεδομένων που επιλέγει ο χρήστης. Οι εικόνες αναπαράγουν το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (με τις ενότητες να βρίσκονται οριζόντια και τις επιμέρους καρτέλες κάθε ενότητας κάθετα) ώστε να μπορείτε εύκολα να εντοπίσετε κάθε μεταδεδομένο. Μέσα στο περιβάλλον τεκμηρίωσης, όλα τα μεταδεδομένα, υποχρεωτικά και μη, επεξηγούνται με ορισμούς και παραδείγματα.

22.1 Κοινά υποχρεωτικά μεταδεδομένα

Πρακτική συμβουλή: Κάποια μεταδεδομένα είναι υποχρεωτικά για όλους τους πόρους!

Για να **περιγράψετε** έναν πόρο αποτελεσματικά, πρέπει να του δώσετε ένα **όνομα** (resource name) και να δημιουργήσετε μια μικρή **περιγραφή** (description). Για να διευκολύνετε τον αναγνώστη και να κάνετε πιο ελκυστική την περιγραφή σας, χρησιμοποιήστε τις λειτουργικότητες που προσφέρονται για διαμόρφωση του κειμένου (στυλ, μορφοποίηση, υπερσύνδεση, στοίχιση κλπ.) όπως φαίνονται στην ακόλουθη εικόνα.

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

LANGUAGE
RESOURCE/TECHNOLOGY

TOOL/SERVICE

DISTRIBUTION

DATA

☐ For information
☐ CLARIN:EL compatible service

Save draft

Save

IDENTITY

CATEGORIES

CONTACT

DOCUMENTATION

RELATED LRTS

LRT name *

X tagger for French, Y speech recognizer, Z corpus

The official name or title of the language resource/technology

language

English

select language

LRT identifier

A string used to uniquely identify the language resource/technology

Fill in

LRT short name

An abbreviation, acronym, etc. used for the language resource/technology

language

English

select language

Description

Formatting

Hyperlinking

Alignment

Styling

Numbering/Bullets

language

English

select language

Στη συνέχεια δηλώστε τον αριθμό της **έκδοσης** (version) του πόρου. Αν δεν δώσετε εσείς αριθμό έκδοσης, *το σύστημα θεωρεί αυτομάτως ότι η έκδοση είναι η πρώτη και την προσθέτει (v1.0.0.)*. Θα χρειαστεί να βάλετε, ακόμη, μία ή περισσότερες **λέξεις κλειδιά** (keywords) και ένα **email** ή μια **ιστοσελίδα** (landing page) για όποιον επιθυμεί να έχει **περισσότερες πληροφορίες** σχετικά με τον πόρο.

Αυτά τα μεταδεδομένα βρίσκονται στην ενότητα *Language Resource/Technology* (Γλωσσικός πόρος/Γλωσσική τεχνολογία) στην καρτέλα *Identity* (Ταυτότητα).

Πρέπει επίσης να περιγράψετε ανεξάρτητα κάθε **τρόπο διανομής** του πόρου. Όλους, δηλαδή, τους τρόπους με τους οποίους ο χρήστης μπορεί να τον αποκτήσει, είτε σε συμπαγή μορφή, ως CD-ROM για παράδειγμα, είτε από διάφορα URL, όπως είναι η **τοποθεσία πρόσβασης** (access location), η **τοποθεσία λήψης/καταφόρτωσης** (download location) και η **τοποθεσία διανομής** (distribution location). Ταυτόχρονα, θα πρέπει να δηλώσετε τους **όρους χρήσης** (licence terms) κάθε μορφής διάθεσης του πόρου (δείτε [εδώ](#) το Προτεινόμενο Σχήμα Αδειοδότησης Γλωσσικών Πόρων).

Αυτά τα μεταδεδομένα βρίσκονται στην ενότητα *Distribution* (Διανομή) στην καρτέλα *Technical* (Τεχνικά).

22.2 Υποχρεωτικά μεταδεδομένα ανά τύπο πόρου

22.2.1 Μία γρήγορη ματιά σε ένα σώμα κειμένων

LANGUAGE RESOURCE/ TECHNOLOGY	CORPUS	PART	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none"> Resource Name Description Version CATEGORIES <ul style="list-style-type: none"> Keyword CONTACT <ul style="list-style-type: none"> Additional Information DOCUMENTATION	TECHNICAL <ul style="list-style-type: none"> Corpus subclass Personal Data Sensitive Data Anonymized (*) 	MEDIA PART <ul style="list-style-type: none"> Corpus Part Linguality type (text, audio, video, image) Multilinguality type (text, audio, video) Language (text, audio, video, image) Type of content (video, image, textNumerical) 	TECHNICAL <ul style="list-style-type: none"> Dataset Distribution Dataset Distribution Form Distribution Location (*) Download Location (*) Access Location (*) Distribution Medium Features (*) Data Format Size Licence Terms 	DATA
RELATED LRTs				

Πέρα από τα *κοινά υποχρεωτικά μεταδεδομένα*, για να περιγραφεί ένα σώμα κειμένων απαιτούνται επίσης πληροφορίες σχετικά με την **υποκατηγορία** (subclass) του, αν είναι δηλαδή τα δεδομένα πρωτογενή (raw) ή επισημειωμένα (annotated), και αν περιλαμβάνονται **προσωπικά** (personal) ή **ευαίσθητα προσωπικά** (sensitive) δεδομένα. Εάν συμβαίνει αυτό, πρέπει να αναφέρετε εάν έχουν **ανωνυμοποιηθεί** (anonymized). Το σώμα κειμένων μπορεί επίσης να απαρτίζεται από διάφορα **είδη μέσων** (media types), όπως είναι το κείμενο (text), το βίντεο (video), ο ήχος (audio), η εικόνα (image) και το textNumerical, μία υποκατηγορία κειμένου με αριθμητικά δεδομένα. Κάθε ένα από αυτά τα μέρη πρέπει να περιγραφεί ανεξάρτητα. Για παράδειγμα, αν έχετε ένα σώμα κειμένου με απομαγνητοφωνημένες συζητήσεις μαζί με τα αρχεία ήχου, πρέπει να δώσετε πληροφορίες και για τα δύο μέσα, αναφέροντας το **μέγεθος** (size) και το **μορφότυπο** (data format) του κάθε μέσου.

Πρακτική συμβουλή: Δείτε [εδώ](#) παραδείγματα περιγραφών μεταδεδομένων σε XML για σώματα κειμένων.

22.2.2 Μία γρήγορη ματιά σε ένα εργαλείο

LANGUAGE RESOURCE/ TECHNOLOGY	TOOL SERVICE	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none"> Resource Name Description Version CATEGORIES <ul style="list-style-type: none"> Keyword CONTACT <ul style="list-style-type: none"> Additional Information DOCUMENTATION	CATEGORIES <ul style="list-style-type: none"> Function TECHNICAL <ul style="list-style-type: none"> Language Dependent Input Content Resource Processing Resource Type Language (*) EVALUATION <ul style="list-style-type: none"> Evaluated 	TECHNICAL <ul style="list-style-type: none"> Software Distribution Software Distribution Form Web Service Type (*) (Download location, Access location, Execution location) Licence Terms 	DATA
RELATED LRTs			

Πέρα από τα *κοινά υποχρεωτικά μεταδεδομένα*, για να περιγραφεί ένα εργαλείο πρέπει να αναφέρετε ποια **λειτουργία** (function) επιτελεί, καθώς επίσης και αν έχει δημιουργηθεί για επεξεργασία υλικού σε **συγκεκριμένη**

γλώσσα (language dependent), ή όχι. Εάν ισχύει κάτι τέτοιο, πρέπει να δηλώσετε μία ή περισσότερες **γλώσσες** (languages) που μπορεί να χειριστεί το εργαλείο. Πρέπει επίσης να οριστεί ο **τύπος πόρου** (resource type) τον οποίο δέχεται στην **είσοδό** (input) του (π.χ. σώμα κειμένων).

Πρακτική συμβουλή: Δείτε [εδώ](#) παραδείγματα περιγραφών μεταδεδομένων σε XML για εργαλεία.

22.2.3 Μία γρήγορη ματιά σε έναν λεξικό/εννοιολογικό πόρο

LANGUAGE RESOURCE/ TECHNOLOGY	CORPUS	PART	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none"> Resource Name Description Version CATEGORIES <ul style="list-style-type: none"> Keyword CONTACT <ul style="list-style-type: none"> Additional Information DOCUMENTATION	TECHNICAL <ul style="list-style-type: none"> Encoding Level Personal Data Sensitive Data Anonymized (*) 	MEDIA PART <ul style="list-style-type: none"> Lexical/Conceptual Resource Part Linguality type (text, audio, video, image) Language (text, audio, video, image) Type of content (video, image) 	TECHNICAL <ul style="list-style-type: none"> Dataset Distribution Dataset Distribution Form Distribution Location (*) Download Location (*) Access Location (*) Distribution Medium Features (*) Data Format Size Licence Terms 	DATA
RELATED LRTs				

Πέρα από τα **κοινά υποχρεωτικά μεταδεδομένα**, για να περιγραφεί ένας λεξικός/εννοιολογικός πόρος, απαιτούνται επίσης πληροφορίες σχετικά με το **επίπεδο κωδικοποίησης** (encoding level), δηλαδή το επίπεδο γλωσσικής ανάλυσης (π.χ. μορφολογία, φωνολογία, σημασιολογία κλπ.) και αν περιλαμβάνονται **προσωπικά** (personal) ή **ευαίσθητα προσωπικά** (sensitive) δεδομένα. Εάν συμβαίνει αυτό, πρέπει να αναφέρετε εάν έχουν **ανωνυμοποιηθεί** (anonymized). Ο λεξικός/εννοιολογικός πόρος μπορεί επίσης να απαρτίζεται από διάφορα **είδη μέσων** (media types), όπως είναι το κείμενο (text), το βίντεο (video), ο ήχος (audio) και η εικόνα (image). Κάθε ένα από αυτά τα μέρη πρέπει να περιγραφεί ανεξάρτητα. Για παράδειγμα, εάν έχετε ένα ψηφιακό λεξικό με ορισμούς σε κειμενική μορφή και ηχητικά αρχεία για την προφορά, πρέπει να δώσετε πληροφορίες και για τα δύο μέσα, αναφέροντας το **μέγεθος** (size) και το **μορφότυπο** (data format) του κάθε μέσου.

Πρακτική συμβουλή: Δείτε [εδώ](#) παραδείγματα περιγραφών μεταδεδομένων σε XML για λεξικούς/εννοιολογικούς πόρους.

22.2.4 Μία γρήγορη ματιά σε μία γλωσσική περιγραφή

LANGUAGE RESOURCE/ TECHNOLOGY	CORPUS	PART	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none"> Resource Name Description Version CATEGORIES <ul style="list-style-type: none"> Keyword CONTACT <ul style="list-style-type: none"> Additional Information DOCUMENTATION	TECHNICAL <ul style="list-style-type: none"> Language Description Subclass ML Model (*) Grammar (*) Ngram Model (*) Encoding Level Personal Data Sensitive Data Anonymized (*) 	MEDIA PART <ul style="list-style-type: none"> Lexical/Conceptual Resource Part Linguality type (text, audio, video, image) Language (text, audio, video, image) Type of content (video, image) 	TECHNICAL <ul style="list-style-type: none"> Dataset Distribution Dataset Distribution Form Distribution Location (*) Download Location (*) Access Location (*) Distribution Medium Features (*) Data Format Size Licence Terms 	DATA
RELATED LRTs				

Πέρα από τα *κοινά υποχρεωτικά μεταδεδομένα*, για να περιγραφεί μια γλωσσική περιγραφή απαιτούνται επίσης πληροφορίες σχετικά με την **υποκατηγορία** (subclass) της, αν είναι δηλαδή **γραμματική** (grammar) ή ένα **μοντέλο ML** (ML model) ή ένα **μοντέλο n-gram** (n-gram model) και αν περιλαμβάνονται **προσωπικά** (personal) ή **ευαίσθητα προσωπικά** (sensitive) δεδομένα. Εάν συμβαίνει αυτό, πρέπει να αναφέρετε εάν έχουν **ανωνυμοποιηθεί** (anonymized). Η γλωσσική περιγραφή μπορεί επίσης να απαρτίζεται από διάφορα μέρη με ποικίλα **είδη μέσων** (media types), όπως είναι το κείμενο (text), το βίντεο (video), ο ήχος (audio) και η εικόνα (image). Κάθε ένα από αυτά τα μέρη πρέπει να περιγραφεί ανεξάρτητα. Για παράδειγμα, αν έχετε μια γραμματική σε κειμενική μορφή μαζί με τα αρχεία βίντεο όπου παρουσιάζονται παραδείγματα, πρέπει να δώσετε πληροφορίες και για τα δύο μέσα, αναφέροντας το **μέγεθος** (size) και τον **μορφότυπο** (data format) του κάθε μέσου.

Πρακτική συμβουλή: Δείτε *εδώ* παραδείγματα περιγραφών μεταδεδομένων σε XML για γλωσσικές περιγραφές.

Προσοχή: Παρόλο που μόνο τα υποχρεωτικά μεταδεδομένα απαιτούνται για τη δημιουργία μιας εγγραφής, η περιγραφή ενός πόρου είναι πληρέστερη εάν έχουν συμπληρωθεί και τα **συνιστώμενα** (recommended) μεταδεδομένα.

Γενικές οδηγίες για τα μεταδεδομένα

23.1 Γλώσσα

Όλα τα μεταδεδομένα εμφανίζονται και πρέπει να συμπληρώνονται, κατά κύριο λόγο, στα αγγλικά. Παρόλα αυτά, μπορείτε να συμπληρώσετε πληροφορίες σε **οποιαδήποτε άλλη γλώσσα**. Για να το κάνετε αυτό, κάντε κλικ στο **σύμβολο προσθήκης** [+] δίπλα στο μεταδεδομένο που θέλετε να περιγράψετε σε διαφορετική γλώσσα (π.χ. LRT name).

The screenshot shows a form with two input fields. The first field is labeled 'LRT name *' and contains the text 'Demo corpus'. Below it is the subtitle 'The official name or title of the language resource/technology'. To the right of this field is a 'language' dropdown menu currently set to 'English', with the subtitle 'select language'. A yellow box highlights a '+' button next to the language dropdown, with a tooltip that says 'Add in another language'.

Δημιουργείται μία νέα εγγραφή για το συγκεκριμένο πεδίο μεταδεδομένων και η νέα γλώσσα στην οποία μπορείτε να δώσετε πληροφορίες είναι, από προεπιλογή, η **νέα ελληνική**.

This screenshot shows the form after a new language entry has been added. The 'LRT name *' field still contains 'Demo corpus'. The 'language' dropdown menu now shows 'English' with a '+' button next to it. Below this, a new entry has been added: the 'LRT name *' field is empty, and the 'language' dropdown menu is set to 'Modern Greek (1453-)' with an 'x' button next to it. The subtitle for the new entry is 'select language'.

Προσοχή: Εάν δεν θέλετε να προσθέσετε πληροφορίες στα νέα ελληνικά, **επιλέξτε μία άλλη γλώσσα** από την αναδυόμενη λίστα.

LRT name *

The official name or title of the language resource/technology

language
Bulgarian
x
Remove

English
Modern Greek (1453-)
Bulgarian
Croatian
Czech
Danish
Dutch; Flemish

LRT identifier
A string used to uniquely identify the language resource/technology

LR Identifier scheme *
other
Select the scheme or the authority that issued the identifier

Identifier value *
UPTAP-id
The value of the identifier; if applicable, use the

Εάν αποφασίσετε να **αφαιρέσετε το πρόσθετο πεδίο**, κάντε κλικ στο **σύμβολο αφαίρεσης [x]**. Αυτή η ενέργεια θα αφαιρέσει ολόκληρο το πεδίο, δηλαδή το όνομα του λεξικού/εννοιολογικού πόρου (LRT name), όχι μόνο τη γλώσσα.

23.2 Συνέπεια

Τα μεταδεδομένα που χρησιμοποιούνται για την περιγραφή των δεδομένων σας θα πρέπει να **τα αντικατοπτρίζουν με σαφήνεια**. Βεβαιωθείτε ότι δεν υπάρχουν ασυνέπειες (π.χ. ελέγξτε ότι τα αρχεία σας είναι πράγματι σε μορφή PDF και όχι απλώς σκαναρισμένες εικόνες- αν παρέχετε πληροφορίες για ένα επισημειωμένο σώμα δεδομένων, αναφέρετε το εργαλείο επισημείωσης κλπ.). Ελέγξτε πρώτα ποια είναι τα **υποχρεωτικά** μεταδεδομένα ανά τύπο πόρου και στη συνέχεια δείτε τις **συγκεκριμένες οδηγίες** για το πώς να τα συμπληρώσετε.

23.3 Πληρότητα

Βεβαιωθείτε ότι παρέχετε όλες τις απαραίτητες πληροφορίες σε όλες τις επιλεγμένες γλώσσες. Συνιστάται ιδιαίτερα να παρέχετε τον τίτλο και την περιγραφή στα **αγγλικά** και στα **ελληνικά** (καθώς και σε οποιαδήποτε άλλη γλώσσα, εάν χρειάζεται), καθώς έτσι διευκολύνεται η αναζήτηση και η ανάκτηση των πόρων.

23.4 Επεξεργασία

Η επεξεργασία των μεταδεδομένων είναι απαραίτητη όταν υπάρχουν τυπογραφικά λάθη, ασυνέπειες, ελλείψεις πληροφοριών κλπ. Αν είστε **επιμελητής**, μπορείτε να **επεξεργαστείτε** τα μεταδεδομένα του πόρου όσες φορές θέλετε πριν τον υποβάλετε για δημοσίευση. Μόλις ο πόρος υποβληθεί, η επεξεργασία επιτρέπεται μόνο στον **υπεύθυνο αποθετηρίου** (δείτε **εδώ** τις διαφορετικές ενέργειες που μπορεί να κάνει ο κάθε χρήστης ανάλογα με τον ρόλο του).

Προσοχή: Λάβετε υπόψη ότι η **επεξεργασία** αφορά μόνο **μικρές αλλαγές** στα μεταδεδομένα του εν λόγω πόρου. Η επεξεργασία λαμβάνει χώρα κατά τη διάρκεια της δημιουργίας ενός πόρου μέχρι την υποβολή του για δημοσίευση. Διαφέρει από την δημιουργία νέας έκδοσης (versioning) στο ότι **οι αλλαγές δεν συνιστούν μια νέα οντότητα**. Για παράδειγμα, αν έχει γίνει **λάθος στο μέγεθος του πόρου**, απαιτείται **επεξεργασία**, ενώ αν το μέγεθος του πόρου έχει αλλάξει (όπως συμβαίνει με τα γλωσσάρια που συμπληρώνονται) απαιτείται **νέα έκδοση**.

23.5 Δημιουργία νέας έκδοσης (versioning)

Εάν υπάρχουν σημαντικές αλλαγές στον πόρο που τον διαφοροποιούν από την υπάρχουσα οντότητα που περιγράφεται, τότε πρέπει να δημιουργηθεί μια νέα έκδοση. Τέτοιες αλλαγές είναι νέες εκδόσεις γλωσσαρίων με περισσότερους όρους, νέες βελτιωμένες καθαρισμένες εκδόσεις σωμάτων κειμένων, νέα δεδομένα (π.χ. παράλληλες προτάσεις) με τα οποία έχει εμπλουτιστεί ένα δίγλωσσο σώμα κειμένων ή ενημερώσεις εργαλείων/λογισμικού. Τόσο ο *επιμελητής* όσο και ο *υπεύθυνος αποθετηρίου* μπορούν να δημιουργήσουν μια νέα έκδοση μετά τη δημοσίευση του πόρου.

Προσοχή: Λάβετε υπόψη ότι μόνο η **τελευταία έκδοση** βρίσκεται στον κεντρικό κατάλογο, ενώ όλες οι εκδόσεις είναι προσβάσιμες από τη σελίδα **θέασης του πόρου**, όπως φαίνεται στην παρακάτω εικόνα.

Overview
Technical
Relations
Access

Information for Lexical/ Conceptual resource part

TEXT

Language info

Linguality type
bilingual

Multilinguality type
parallel

Language
Modern Greek (1453-), English

Export

XML

All versions

Terms on Fora! – English-Greek Glossary of Terms
(20.0.0)
<http://hdl.handle.net/11500/CLARIN-EL-0000-0000-6A02-7>
(Handle)

Terms on Fora! – English-Greek Glossary of Terms (16)
<http://hdl.handle.net/11500/CLARIN-0000-0000-5CF4-6>
(Handle)

Συγκεκριμένες οδηγίες για τα υποχρεωτικά μεταδεδομένα

Αυτή η ενότητα παρέχει οδηγίες σχετικά με τον τρόπο συμπλήρωσης συγκεκριμένων μεταδεδομένων τα οποία είναι υποχρεωτικά για *σώματα κειμένων*, *λεξικούς/εννοιολογικούς πόρους*, *εργαλεία/υπηρεσίες* και *γλωσσικές περιγραφές*. Δεδομένου ότι ορισμένα μεταδεδομένα είναι *κοινά* για όλους τους πόρους, αυτά παρουσιάζονται πρώτα και ακολουθούν τα υπόλοιπα μεταδεδομένα που είναι διαφορετικά για τον κάθε τύπο πόρου.

Κάθε στοιχείο μεταδεδομένων επεξηγείται εν συντομία και παρέχονται παραδείγματα όπου είναι δυνατόν. Τα παραδείγματα καλύπτουν τόσο τις βέλτιστες πρακτικές όσο και τα συνηθισμένα λάθη που πρέπει να αποφεύγονται (σημειώνονται με αστερίσκο *). Επιπλέον, για κάθε στοιχείο μεταδεδομένων υπάρχει σύνδεσμος προς το *XSD* με την πλήρη αναπαράστασή του.

24.1 1. resourceName

~Το επίσημο όνομα ή ο τίτλος του γλωσσικού πόρου ή της γλωσσικής τεχνολογίας~

Το όνομα πρέπει να αντικατοπτρίζει το περιεχόμενο (και τον τύπο) του πόρου- πρέπει να παρουσιάζει όλες τις απαραίτητες πληροφορίες για τον πόρο, αλλά δεν πρέπει να είναι υπερβολικά περιγραφικό- λεπτομερείς πληροφορίες πρέπει να παρέχονται στην περιγραφή. Μην χρησιμοποιείτε πλήρεις φράσεις, σημεία στίξης (εκτός αν είναι απαραίτητο) ή συντομογραφίες στον τίτλο του πόρου. Δώστε την πλήρη ονομασία του πόρου και χρησιμοποιήστε τη συντομογραφία/ακρωνύμιο (εάν υπάρχει) στο αντίστοιχο πεδίο μεταδεδομένων.

Παραδείγματα

Ναι: Glossary of medical terms; Old and New Testament; Ellogon annotation tool

Όχι: *This is a glossary of medical terms; *Old and New Testament!; *Ellogon ann. tool

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου resourceName στο *XSD*
-

24.2 2. description

~Μια σύντομη παρουσίαση του γλωσσικού πόρου ή της γλωσσικής τεχνολογίας~

Η περιγραφή πρέπει να περιέχει όλες τις σημαντικές πληροφορίες για τον πόρο. Μην επαναλαμβάνετε (ή επαναδιατυπώνετε) απλώς τον τίτλο του πόρου χωρίς να προσθέτετε άλλες πληροφορίες. Μόλις διαβάσει κανείς την περιγραφή και χωρίς να δει τα υπόλοιπα μεταδεδομένα, θα πρέπει να είναι σε θέση να καταλάβει τι είναι και τι περιλαμβάνει ο πόρος. Ορίστε τον τύπο του πόρου και δώστε κάθε χρήσιμη πληροφορία σχετικά με το πώς, πότε και από ποιον δημιουργήθηκε, ποια είναι η γλώσσα και το μέγεθός του και ποιος είναι ο σκοπός που εξυπηρετεί, αν υπάρχει. Αναφέρετε τυχόν ιδιαιτερότητες ή περιορισμούς σχετικά με τα δεδομένα ή το εργαλείο/υπηρεσία που πρέπει να γνωρίζουν οι χρήστες. Η περιγραφή πρέπει να είναι ελεύθερο κείμενο μιας παραγράφου τουλάχιστον. Μπορείτε επίσης να χρησιμοποιήσετε τις λειτουργικότητες (μορφοποίηση, υπερσύνδεση, λίστες κλπ.) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων (editor¹) για να κάνετε την περιγραφή ευανάγνωστη.

Παραδείγματα

Do: 1) Bilingual glossary (German / Greek) made in 2019/2020 by students of DFLTI (Ionian University) under the supervision of Mr. Olaf Immanuel Seel in the framework of the department's cooperation with the EU TermCord.

- 2) Texts corpus from the transcription of recorded children's speech focused on narration. The corpus was collected from interviews conducted by undergraduate and postgraduate students of the Department of Mediterranean Studies of the University of the Aegean with children with whom they are related either by friendship or kinship. Files with both the questions and answers are provided, where K=girl and A=boy, as well as cleaned files containing only the children's answers (clean).

Όχι: *Symposium Proceedings; *Bilingual lexicon on the Greek economy

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου description στο [XSD](#)
-

24.3 3. version

~Μια συγκεκριμένη μορφή ενός πόρου που διαφέρει σε ορισμένα σημεία από μια προηγούμενη μορφή~

Το συνιστώμενο μορφότυπο (format) για μια έκδοση είναι: μεγάλη_έκδοση.μικρή_έκδοση.επιδιόρθωση (patch)².

Παραδείγματα

Ναι: 1.0.0-alpha; 2.1.1

Όχι: *1.0.1-alpha; *0.0.2

Η υποδομή αναθέτει αυτόματα την έκδοση **1.0.0** σε όλους τους πόρους. Αν αυτό δεν ισχύει για τον πόρο σας, γράψτε τον αριθμό έκδοσης στο πλαίσιο (π.χ. 2.0.0) και στη συνέχεια κάντε κλικ στην ημερομηνία έκδοσης για να εμφανιστεί το ημερολόγιο. Επιλέξτε την ημερομηνία κατά την οποία κυκλοφόρησε αυτή η έκδοση και κάντε κλικ στο OK.

¹ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

² Δείτε τις συγκεκριμένες οδηγίες για τη δημιουργία έκδοσης.

Το περιβάλλον τεκμηρίωσης και ανάρτησης πόρων παρέχει επίσης τη δυνατότητα να *δημιουργηθεί αυτόματα μια νέα έκδοση* ενός υπάρχοντος πόρου. Δείτε τις *οδηγίες για νέα έκδοση* πριν προχωρήσετε σε μία τέτοια ενέργεια.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου version στο XSD

24.4 4. keyword

~Μια λέξη ή φράση χαρακτηριστική του γλωσσικού πόρου ή της γλωσσικής τεχνολογίας που μπορεί να χρησιμοποιηθεί κατά την αναζήτηση~

Οι λέξεις-κλειδιά είναι λέξεις ή μικρές φράσεις που χρησιμοποιούνται για την αναζήτηση ενός πόρου. Όσο περισσότερες λέξεις-κλειδιά χρησιμοποιούνται, τόσο το καλύτερο για την ανάκτηση του πόρου. Ωστόσο, οι λέξεις-κλειδιά πρέπει να αναδεικνύουν πτυχές του πόρου που δεν καλύπτονται ήδη από τα **υποχρεωτικά** μεταδεδομένα. Εάν, για παράδειγμα, περιγράφετε ένα *μονόγλωσσο σχολιασμένο σώμα κειμένων που δημιουργήθηκε για να βελτιώσει τη διαδικασία εκμάθησης μη φυσικών ομιλητών*, οι λέξεις-κλειδιά δεν πρέπει να είναι **πρωτίστως ή αποκλειστικά** τα «corpus», «annotated» ή «monolingual», καθώς αυτές είναι οι τιμές των μεταδεδομένων resourceType, corpusSubclass και linguality αντίστοιχα και μπορούν να ανακτηθούν από αυτά κατά την αναζήτηση. Αντ' αυτού, χρησιμοποιήστε ως λέξεις-κλειδιά τις φράσεις «non native speaker» και «learning process» που τονίζουν την προβλεπόμενη χρήση του πόρου. Επικουρικά μπορείτε να προσθέσετε ως λέξεις-κλειδιά τα «corpus», «annotated» και «monolingual».

Παραδείγματα

Ναι: non native speaker; learning process (corpus; annotated; monolingual)

Όχι: *corpus; *annotated; *monolingual

Αφού πληκτρολογήσετε τη λέξη-κλειδί που θέλετε, κάντε **κλικ στην προτροπή** που εμφανίζεται κάτω από το πλαίσιο: **Add** «non native speaker». Μόνο τότε η τιμή θα αποθηκευτεί. Εάν παραλείψετε αυτό το βήμα, η λέξη-κλειδί **δεν θα εμφανίζεται** όταν επισκεφθείτε ξανά αυτό το τμήμα του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

Keyword

A word or phrase characteristic of the language resource/technology that can be used at search (multiple values possible)

Keyword *

non native speaker

×

language

English

▼

Remove

+

Missing **non native speaker**? Add "non native speaker"

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου keyword στο [XSD](#)

24.5 5. additionalInformation

~Ένα URL (σελίδα προσγείωσης/landing page) ή μία ηλεκτρονική διεύθυνση (π.χ. μία ηλεκτρονική διεύθυνση υποστήριξης/support email) όπου ο χρήστης μπορεί να βρει ή να ζητήσει περισσότερες πληροφορίες~

Αυτό το μεταδεδομένο είναι είτε μια ιστοσελίδα με πρόσθετες πληροφορίες σχετικά με τον γλωσσικό πόρο ή την γλωσσική τεχνολογία (π.χ. μία διεύθυνση όπου βρίσκεται το περιεχόμενό του, ένας σύνδεσμος για την τοποθεσία πρόσβασης στον πόρο κλπ.) είτε η ηλεκτρονική διεύθυνση του υπεύθυνου για την παροχή πληροφοριών. Βεβαιωθείτε ότι έχετε εισάγει μία έγκυρη ηλεκτρονική διεύθυνση ή ένα URL.

Παραδείγματα

Ναι: person@athenarc.gr; <http://www.clarin.gr>

Όχι: *person@athenarc.g; <http://clarin.gr>

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου additionalInformation στο [XSD](#)

24.6 6. distribution και σχετικά μεταδεδομένα

~Η μορφή (ή οι μορφές) που είναι διαθέσιμος ένας πόρος~

Ένας πόρος μπορεί να είναι διαθέσιμος με περισσότερους από έναν τρόπους, σε συμπαγή μορφή (όπως ένα CD-ROM, ένα DVD-R, ένας σκληρός δίσκος κλπ.) ή μέσω ενός σημείου πρόσβασης. Εάν υπάρχουν περισσότεροι του ενός τρόποι διάθεσης (distribution) για έναν πόρο, πρέπει να περιγραφούν ο καθένας **ανεξάρτητα**. Μια αναπτυσσόμενη λίστα προσφέρει μια ποικιλία μορφών διάθεσης για να επιλέξετε.

Dataset distribution 1

Describe separately each distributable form of the corpus (e.g., downloadable form in CSV, XML formats, form accessible via an i/f)

Dataset distribution form *

- CD-ROM
- DVD-R
- accessible through interface
- accessible through query
- bluRay
- downloadable
- hard disk
- other
- unspecified

Μόλις μια τιμή επιλεγεί, δημιουργούνται τα αντίστοιχα πεδία μεταδεδομένων, τα οποία πρέπει και αυτά να συμπληρωθούν. Εάν, για παράδειγμα, ένας πόρος είναι *προσβάσιμος μέσω διεπαφής* (accessible through interface), δημιουργείται το μεταδεδομένο `access location` (σημείο πρόσβασης) όπου πρέπει να συμπληρώσετε το URL στο οποίο ο πόρος είναι προσβάσιμος.

Dataset distribution 1

Describe separately each distributable form of the corpus (e.g., downloadable form in CSV, XML formats, form accessible via an i/f)

Dataset distribution form *

accessible through interface

Select the form or delivery channel through which the corpus is distributed

Private

☐ Yes

☐ No

Specifies whether the resource is private so that its access/download location remains hidden

Access location *

The URL where the corpus distribution can be accessed from (e.g. a URL with the access i/f, or a landing page where the user needs to follow some links, provide some information and check some boxes before accessing the corpus)

Προσοχή: Η υποδομή CLARIN:EL φιλοξενεί κυρίως πόρους που συνοδεύονται από δεδομένα. Ανεξάρτητα από το αν τα δεδομένα έχουν *μεταφορτωθεί* κατά τη δημιουργία του πόρου ή σε μεταγενέστερο στάδιο, πρέπει να συσχετιστούν με μια μορφή διάθεσης (distribution). Ο κατάλληλος τρόπος διάθεσης έχει την τιμή *downloadable* (καταφορτώσιμο) και παρόλο που παράγει το πεδίο μεταδεδομένων `download location` (σημείο καταφόρτωσης), αυτό δεν χρειάζεται να συμπληρωθεί (αφού τα δεδομένα καταφορτώνονται από την υποδομή CLARIN:EL). Ωστόσο, αυτό που πρέπει να γίνει είναι να δημιουργηθεί η *συσχέτιση* (*association*) μεταξύ του τρόπου διάθεσης και των δεδομένων, όπως φαίνεται στην ακόλουθη εικόνα. Κάντε κλικ στο όνομα του συμπίεσμένου αρχείου zip για να δημιουργηθεί η συσχέτιση. Τέλος, για να ολοκληρωθεί η διαδικασία, πρέπει να **αποθηκεύσετε** (ή να **αποθηκεύσετε ως προσχέδιο**) την εγγραφή μεταδεδομένων.

Dataset distribution 1

Describe separately each distributable form of the corpus (e.g., downloadable form in CSV, XML formats, form accessible via an i/f)

Dataset distribution form *
downloadable

Select the form or delivery channel through which the corpus is distributed

Private
☐ Yes
☐ No

Specifies whether the resource is private so that its access/download location remains hidden

Associate a dataset with this distribution

Parla_12_1.zip

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου distribution στο XSD

24.7 7. licenceTerms και σχετικά μεταδεδομένα

~Οι όροι σύμφωνα με τους οποίους διατίθεται ένας πόρος~

Τα μεταδεδομένα που σχετίζονται με τα licenceTerms είναι το όνομα της licence (άδειας), οι licence terms (όροι χρήσης της άδειας) και οι πιο συχνά χρησιμοποιούμενοι conditions of use (περιορισμοί χρήσης). Το όνομα της licence αποκαλύπτεται μόλις αρχίσετε να πληκτρολογείτε στο αντίστοιχο πεδίο. Εάν έχει ήδη χρησιμοποιηθεί από άλλον χρήστη, θα σας παρουσιαστεί το πλήρες επίσημο όνομά. Κάντε κλικ στο όνομα και τα σχετικά μεταδεδομένα θα συμπληρωθούν αυτόματα. Για παράδειγμα, αν η εν λόγω άδεια είναι η **cc-by-nc-sa**, τότε όταν αρχίσετε να πληκτρολογείτε, θα παρουσιαστούν οι αντίστοιχες επιλογές, όπως φαίνεται στην παρακάτω εικόνα.

Licence

Start typing to select the licence of the corpus distribution, or add a new value

Licence name *
cc-by-nc-sa

Missing **cc-by-nc-sa**? Add "cc-by-nc-sa"

Creative Commons Attribution Non Commercial Share Alike 3.0 Unported
<https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode>
 CC-BY-NC-SA-3.0 - SPDX

Creative Commons Attribution Non Commercial Share Alike 4.0 International
<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>
 CC-BY-NC-SA-4.0 - SPDX

Κάντε κλικ στο όνομα της άδειας που επιθυμείτε από την προτεινόμενη λίστα. Για τη συγκεκριμένη άδεια, η πλήρης ονομασία είναι **Creative Commons Attribution Non Commercial Share Alike 4.0 International**, η διεύθυνση URL όπου μπορείτε να βρείτε τους όρους της άδειας είναι <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>, και οι περιορισμοί είναι η **attribution** (αναφορά δημιουργού), η **non commercial use** (μη εμπορική χρήση) και η **share-alike** (παρόμοια διανομή), που συμπληρώθηκαν αυτόματα στα αντίστοιχα πεδία μεταδεδομένων όπως φαίνεται στην ακόλουθη εικόνα.

Licence

Start typing to select the licence of the corpus distribution, or add a new value

Licence name *
Creative Commons Attribution Non Commercial Share Alike 4.0 International

language
English

Remove

+

Licence/terms of use/service URL *
https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

+

Licence/terms of use/service URL *
https://creativecommons.org/licenses/by-nc-sa/4.0/

×

Condition of use *
attribution × non-commercial use × share alike ×

Select the condition(s) of use for accessing the corpus; the list includes the most frequently used conditions and aims to provide brief human readable information, while the proper exposition of all conditions and possible exceptions is found inside the licence text

Εάν η άδεια που επιθυμείτε να χρησιμοποιήσετε δεν έχει εφαρμοστεί ποτέ πριν, θα πρέπει να συμπληρώσετε με το χέρι τα προαναφερθέντα μεταδεδομένα. Ανατρέξτε επίσης στο προτεινόμενο σχήμα αδειότησης γλωσσικών πόρων αν χρειάζεστε βοήθεια στην επιλογή άδειας για τους πόρους σας.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου licenceTerms στο XSD

24.8 8. data

~Τα αρχεία περιεχομένου ενός πόρου~

Δεν έχουν όλοι οι πόροι αρχεία περιεχομένου. Μια περιγραφή μεταδεδομένων μπορεί να συνοδεύεται ή να μην συνοδεύεται από αρχεία περιεχομένου (δείτε [εδώ](#) για περισσότερες πληροφορίες). Δείτε επίσης τις λεπτομερείς οδηγίες για τον τρόπο *προετοιμασίας των δεδομένων*, τα *συνιστώμενα μορφότυπα (format)* και τον τρόπο *μεταφόρτωσής τους*.

24.9 9. personalData, sensitiveData & anonymized

~Πληροφορίες σχετικά με το αν ο πόρος περιέχει προσωπικά ή/και ευαίσθητα προσωπικά δεδομένα~

Προσοχή: Αυτό το μεταδεδομένο είναι υποχρεωτικό για **σώματα κειμένων, λεξικούς/εννοιολογικούς πόρους και γλωσσικές περιγραφές**.

Πρέπει να προσδιορίσετε εάν ο πόρος περιέχει προσωπικά δεδομένα (π.χ. ονόματα) ή/και ευαίσθητα προσωπικά δεδομένα (π.χ. ιατρικά/σχετικά με την υγεία κλπ.) και συνεπώς απαιτεί ειδικό χειρισμό. Εάν αυτό ισχύει, παρουσιάζονται νέα πεδία μεταδεδομένων στα οποία μπορείτε να δώσετε πρόσθετες πληροφορίες σχετικά με τις όποιες ειδικές απαιτήσεις.

Personal data included *
☒ Yes
☐ No
Specify if personal data are included

Personal data details

language
English
select language
+

Provide additional information on special requirements, if needed

Sensitive data included *
☒ Yes
☐ No
Specify whether the language resource contains sensitive data (e.g., medical/health-related, etc.) and thus requires special handling

Sensitive data details

language
English
select language
+

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `personalData` στο [XSD](#)
- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `sensitiveData` στο [XSD](#)

Η ύπαρξη προσωπικών ή/και ευαίσθητων προσωπικών δεδομένων παράγει³ ένα νέο μεταδεδομένο, αυτό της **ανωνυμοποίησης**. Εδώ μπορείτε να δώσετε όλες τις πληροφορίες σχετικά με την ανωνυμοποίηση/ ψευδο-ανωνυμοποίηση, το εργαλείο που χρησιμοποιήθηκε, αν γράφτηκε κώδικας, τυχόν συμβάσεις που υιοθετήθηκαν κλπ.

Anonymized *
☒ Yes
☐ No
Specify if the corpus has been anonymized

Anonymization details

language
English
select language
+

More information on the anonymization

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `anonymized` στο [XSD](#)

³ Το μεταδεδομένο `anonymized` (ανωνυμοποιημένο) ανήκει στα μεταδεδομένα που είναι **υπό συνθήκη υποχρεωτικά**, η αναγκαιότητα των οποίων εξαρτάται από τις τιμές άλλων μεταδεδομένων που παρέχει ο χρήστης, όπως η απάντηση «ναι» στην ερώτηση σχετικά με την ύπαρξη προσωπικών ή/και ευαίσθητων προσωπικών δεδομένων σε έναν πόρο.

24.10 10. Subclass και σχετικά μεταδεδομένα

~Οι κατηγορίες στις οποίες ένας γλωσσικός πόρος μπορεί να κατηγοριοποιηθεί περαιτέρω ανάλογα με τον τύπο του~

Προσοχή: Αυτό το μεταδεδομένο είναι υποχρεωτικό για σώματα κειμένων, λεξικούς/εννοιολογικούς πόρους και γλωσσικές περιγραφές.

24.10.1 10.1 corpusSubclass

Για τα σώματα κειμένων οι κατηγορίες corpusSubclass είναι:

- **raw** (πρωτογενή δεδομένα), για μη επεξεργασμένα σώματα κειμένων,
- **annotated** (επισημειωμένα), για σώματα κειμένων που περιλαμβάνουν τόσο τα πρωτογενή δεδομένα όσο και το επεξεργασμένο αποτέλεσμα,
- **annotations** (επισημειώσεις), για σώματα κειμένων που αποτελούνται μόνο από το επεξεργασμένο αποτέλεσμα, και
- **unspecified** (απροσδιόριστο), για σώματα κειμένων που δεν μπορούν να περιγραφούν με μία από τις προαναφερθείσες κατηγορίες.

Το μεταδεδομένο αυτό βρίσκεται στην ενότητα *Corpus* (καρτέλα *Technical*) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου corpusSubclass στο [XSD](#)

24.10.2 10.2 lcrSubclass

Ένας λεξικός/εννοιολογικός πόρος μπορεί να κατηγοριοποιηθεί περαιτέρω με το στοιχείο lcrSubclass ως εξής:

- **annotation scheme** (σχήμα επισημείωσης): Ένα σύνολο στοιχείων και τιμών που έχουν σχεδιαστεί για την επισημείωση δεδομένων. Συνήθως αποτελείται από μια τυπική αναπαράσταση. Αποσκοπεί στην αναπαράσταση ενός συγκεκριμένου επιπέδου πληροφορίας, όπως είναι τα μορφολογικά χαρακτηριστικά λέξεων, οι συντακτικές σχέσεις εξάρτησης μεταξύ φράσεων, πληροφορίες σε επίπεδο λόγου κλπ. Μπορεί να αποτελείται από μια επίπεδη δομή στοιχείων και τιμών (π.χ. ετικέτες με τα μέρη του λόγου) ή μπορεί να είναι πιο σύνθετη με αλληλένδετα στοιχεία (π.χ. συγκεκριμένα μορφολογικά χαρακτηριστικά που χρησιμοποιούνται για κάθε μέρος του λόγου).⁴

⁴ Η διαφορά μεταξύ του συστήματος τύπων (**typesystem**) και του σχήματος επισημείωσης (**annotation scheme**) βασίζεται στο αν χρησιμοποιούνται από εργαλεία ή ορίζονται από τους χρήστες: το **σχήμα επισημείωσης** περιέχει **custom types** (τροποποιημένους τύπους) ενώ το **σύστημα τύπων** χρησιμοποιείται κυρίως για ενσωματωμένους τύπους.

- **computational lexicon** (υπολογιστικό λεξικό): ένα λεξικό που προορίζεται για υπολογιστικούς σκοπούς και συνεπώς περιέχει λέξεις που σχετίζονται με πληροφορίες σχετικές με τους συγκεκριμένους σκοπούς.
- **dictionary** (λεξικό): βιβλίο ή ηλεκτρονική πηγή που περιέχει κατάλογο λέξεων (συνήθως με αλφαβητική σειρά) και εξηγεί τη σημασία τους ή δίνει τη λέξη που αντιστοιχεί σε άλλη γλώσσα και άλλες πληροφορίες (π.χ. ορθογραφία, προφορά κλπ.).
- **FrameNet**: μια λεξική βάση δεδομένων που βασίζεται σε επισημειωμένα παραδείγματα για τον τρόπο χρήσης των λέξεων σε πραγματικά κείμενα σε συνάρτηση με την έννοια του «semantic frame» (σημασιολογικού πλαισίου), που είναι μία σχηματική αναπαράσταση μιας κατάστασης με διάφορους συμμετέχοντες και άλλους εννοιολογικούς ρόλους) - αρχικά κατασκευάστηκε για την αγγλική γλώσσα και επεκτάθηκε σε άλλες γλώσσες σύμφωνα με τις ίδιες αρχές σχεδιασμού.
- **lexicon** (λεξικό): όλες οι λέξεις (σε έναν κατάλογο) που χρησιμοποιούνται σε μια συγκεκριμένη γλώσσα ή σε ένα συγκεκριμένο θέμα.
- **Machine Readable Dictionary** (λεξικό αναγνώσιμο από μηχανή): ένα λεξικό που προορίζεται συνήθως για ανθρώπους σε μορφή που μπορεί να επεξεργαστεί ένας υπολογιστής.
- **mapping of resources** (αντιστοίχιση πόρων): ένας πόρος που αποτελείται από την αντιστοίχιση τιμών ή/και κανόνων μεταξύ δύο πόρων.
- **morphological lexicon** (μορφολογικό λεξικό): ένα λεξικό που παράλληλα με τα λήμματά του έχει και μορφολογική πληροφορία.
- **ontology** (οντολογία): ένα σύνολο εννοιών και κατηγοριών σε ένα θεματικό πεδίο ή τομέα που δείχνει τις ιδιότητές τους και τις σχέσεις μεταξύ τους.
- **other** (άλλο): τιμή που χρησιμοποιείται όταν καμία από τις συνιστώμενες τιμές δεν είναι κατάλληλη για ένα στοιχείο.
- **tagset** (σύνολο ετικετών): ένας σταθερός κατάλογος έγκυρων τιμών (ετικέτες) που έχουν σχεδιαστεί για να επισημειώνονται δεδομένα. Συνήθως αντιστοιχεί σε συγκεκριμένο τύπο επισημείωσης ή σε σύνολο τύπων επισημείωσης.⁵
- **terminological resource** (ορολογικός πόρος): ένας λεξικός πόρος που απαριθμεί έννοιες που αφορούν ένα συγκεκριμένο πεδίο.
- **thesaurus** (θησαυρός): ένα έργο αναφοράς που απαριθμεί λέξεις ομαδοποιημένες ανάλογα με την ομοιότητα μίας έννοιας (που περιέχει συνώνυμα και μερικές φορές αντώνυμα).
- **typesystem** (σύστημα τύπων): ένα σύνολο στοιχείων σχεδιασμένων για την επισημείωση δεδομένων. Συνήθως περιέχει μόνο μία λίστα τύπων επισημείωσης, δηλαδή συγκεκριμένες ετικέτες που χρησιμοποιούνται για την επισημείωση (π.χ. μέρος του λόγου, πρόσωπο, οργανισμός κλπ.), και είναι συνήθως ενσωματωμένο στο λογισμικό επισημείωσης.⁶
- **unspecified** (απροσδιόριστο): τιμή που χρησιμοποιείται για υποχρεωτικά στοιχεία των οποίων η τιμή είναι άγνωστη ή δεν μπορεί να προσδιοριστεί.
- **WordNet**: μια λεξική βάση δεδομένων που δημιουργήθηκε αρχικά για την αγγλική γλώσσα και επεκτάθηκε σε άλλες γλώσσες, η οποία ομαδοποιεί τις λέξεις σε σύνολα συνωνύμων που ονομάζονται synsets, παρέχει σύντομους ορισμούς και παραδείγματα χρήσης και καταγράφει έναν αριθμό σχέσεων μεταξύ αυτών των συνόλων συνωνύμων ή των μελών τους.
- **wordlist** (λίστα λέξεων): μια γραπτή συλλογή όλων των λέξεων που προέρχονται από μια συγκεκριμένη πηγή ή μοιράζονται κάποιο άλλο χαρακτηριστικό.

⁵ Η διαφορά μεταξύ ενός συστήματος τύπων (**typesystem**) και ενός συνόλου ετικετών (**tagset**) είναι ότι το **σύστημα τύπων** περιλαμβάνει μόνο τύπους επισημείωσης (π.χ. έναν τύπο επισημείωσης POS για την αναπαράσταση επισημειώσεων μερών του λόγου) ενώ το **σύνολο ετικετών** περιέχει έναν κατάλογο των έγκυρων τιμών ετικετών (π.χ. το Penn Treebank Tagset).

⁶ Η διαφορά μεταξύ του συστήματος τύπων (**typesystem**) και του σχήματος επισημείωσης (**annotation scheme**) βασίζεται στο αν χρησιμοποιούνται από εργαλεία ή ορίζονται από τους χρήστες: το **σχήμα επισημείωσης** περιέχει **custom types** (τροποποιημένους τύπους) ενώ το **σύστημα τύπων** χρησιμοποιείται κυρίως για ενσωματωμένους τύπους.

Οι κατηγορίες `lcrSubclass` παρουσιάζονται επίσης αλφαβητικά ως αναδυόμενη λίστα στην ενότητα *LCR* (καρτέλα *Technical*) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

The screenshot shows the 'LCR' tab in the CLARIN interface. Under the 'TECHNICAL' section, there is a dropdown menu for 'LCR subclass'. The dropdown is open, showing a list of options: 'annotation scheme', 'computational lexicon', 'dictionary', and 'FrameNet'. The 'LCR subclass' label is highlighted with a yellow box.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `lcrSubclass` στο [XSD](#)

24.10.3 10.3 LanguageDescriptionSubclass

Μια **γλωσσική περιγραφή** έχει τρεις κατηγορίες από τις οποίες μπορεί κανείς να επιλέξει για να περιγράψει με πιο λεπτομερή τρόπο τον πόρο:

- **grammar** (γραμματική): ένα σύνολο κανόνων που ρυθμίζουν ποιες ακολουθίες είναι έγκυρες ή επιτρεπτές σε μια γλώσσα ή ένα κείμενο.
- **ML model** (μοντέλο Μηχανικής Μάθησης/Machine Learning): το μοντέλο ML που πρέπει να χρησιμοποιηθεί μαζί με το εργαλείο/υπηρεσία για την εκτέλεση της επιθυμητής εργασίας.
- **n-gram model** (μοντέλο n-γραμμάτων): ένα γλωσσικό μοντέλο που αποτελείται από n-grams (n-γράμματα), δηλαδή συγκεκριμένες ακολουθίες ενός αριθμού λέξεων.

Αυτές οι κατηγορίες παρουσιάζονται στην ενότητα *Language Description* (καρτέλα *Technical*) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων ως αναπτυσσόμενη λίστα.

The screenshot shows the 'LANGUAGE DESCRIPTION' tab in the CLARIN interface. Under the 'TECHNICAL' section, there is a dropdown menu for 'Language description type'. The dropdown is open, showing a list of options: 'grammar', 'ML model', and 'n-gram model'. The 'Language description type' label is highlighted with a yellow box.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `LanguageDescriptionSubclass` στο [XSD](#)

24.11 11. encodingLevel

~Πληροφορίες σχετικά με το περιεχόμενο ενός πόρου όσον αφορά το γλωσσικό επίπεδο ανάλυσης που καλύπτει~

Προσοχή: Αυτό το μεταδεδομένο είναι υποχρεωτικό για **λεξικούς/εννοιολογικούς πόρους** και **γλωσσικές περιγραφές**.

Οι τιμές για την κωδικοποίηση αναφέρονται σε επίπεδα γλωσσικής ανάλυσης. Τα επίπεδα αυτά παρουσιάζονται με αλφαβητική σειρά στην συνέχεια και δίνεται το αντικείμενο μελέτης τους:

- **morphology** (μορφολογία): σχηματισμός λέξεων (όπως κλίση, παράγωγη και σύνθεση),
- **other** (άλλο): τιμή που χρησιμοποιείται όταν καμία από τις συνιστώμενες τιμές δεν είναι κατάλληλη για ένα στοιχείο.
- **phonetics** (φωνητική): ήχοι ομιλίας,
- **phonology** (φωνολογία): ήχοι ομιλίας που συνιστούν τα θεμελιώδη στοιχεία μιας γλώσσας,
- **pragmatics** (πραγματολογία): η σχέση των προτάσεων με το περιβάλλον στο οποίο εμφανίζονται,
- **semantics** (σημασιολογία): το νόημα μιας λέξης, φράσης κλπ,
- **syntax** (σύνταξη): η δομή των γλωσσικών μονάδων (φράσεων, προτάσεων),
- **unspecified** (απροσδιόριστο): τιμή που χρησιμοποιείται για υποχρεωτικά στοιχεία των οποίων η τιμή είναι άγνωστη ή δεν μπορεί να προσδιοριστεί.

Το μεταδεδομένο βρίσκεται στην ενότητα *LRC* (καρτέλα *Technical*) για τους λεξικούς/εννοιολογικούς πόρους πάνω από την *lcrSubclass*, όπως φαίνεται στην παρακάτω εικόνα.

Για τις γλωσσικές περιγραφές το μεταδεδομένο βρίσκεται στην ενότητα *Language Description* (καρτέλα *Technical*) κάτω από την επιλεγμένη *LanguageDescriptionSubclass*.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `encodingLevel` στο [XSD](#)

24.12 12. function

~Η λειτουργία/εργασία που εκτελεί ένα λογισμικό~

Προσοχή: Αυτό το μεταδεδομένο είναι υποχρεωτικό μόνο για **εργαλεία/υπηρεσίες**.

Η αναδυόμενη λίστα στο αντίστοιχο πεδίο μεταδεδομένων περιλαμβάνει πολλές τιμές που δεν μπορούν να παρουσιαστούν όλες εδώ. Αν αρχίσετε να πληκτρολογείτε, ωστόσο, ο κατάλογος θα μειωθεί μόνο στις τιμές που ταιριάζουν με τα κριτήριά σας. Εάν η λειτουργία του εργαλείου/υπηρεσίας σας ταιριάζει με μία από τις προτεινόμενες τιμές, **κάντε κλικ σε αυτήν** και θα προστεθεί. Εάν η λειτουργία του εργαλείου/υπηρεσίας σας **δεν** ταιριάζει με μια από τις προτεινόμενες τιμές, πρέπει να **κάνετε κλικ στην προτροπή** (*missing...? add*). Μόνο τότε η τιμή θα αποθηκευτεί. Εάν παραλείψετε αυτό το βήμα, η λειτουργία **δεν θα εμφανιστεί** όταν επισκεφθείτε ξανά αυτό το τμήμα του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

The screenshot shows the CLARIN-EL interface with the 'TOOL/SERVICE' tab selected. On the left, under 'CATEGORIES', the 'EVALUATION' checkbox is checked. The main form has a 'Function' field with the text 'event' entered. A dropdown menu is open below the field, showing suggestions: 'Event detection', 'Event labelling', and 'Sound event annotation'. Above the suggestions, it says 'Missing event? Add "event"'. At the top right, there are checkboxes for 'For information' and 'CLARIN-EL compatible service', and buttons for 'Save draft' and 'Save'.

Το μεταδεδομένο βρίσκεται στην ενότητα *Tool/Service* (καρτέλα *Categories*) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `function` στο [XSD](#)

24.13 13. inputContentResource

~Οι απαιτήσεις που θέτει ένα εργαλείο/υπηρεσία για τον πόρο (περιεχόμενο) που επεξεργάζεται~

Προσοχή: Αυτό το μεταδεδομένο είναι υποχρεωτικό μόνο για **εργαλεία/υπηρεσίες**.

Πρόκειται για ένα σύνθετο μεταδεδομένο που συνεπάγεται την περιγραφή τεσσάρων άλλων μεταδεδομένων: **input resource type** (τύπος πόρου στην είσοδο του εργαλείου/υπηρεσίας), **media type** (τύπος μέσου), **data format** (μορφότυπο δεδομένων) and **annotation type** (τύπος επισημείωσης). Όλα αυτά τα στοιχεία παρέχουν τις απαραίτητες πληροφορίες σχετικά με τον πόρο που επεξεργάζεται ένα εργαλείο ή μία υπηρεσία.

LANGUAGE
RESOURCE/TECHNOLOGY

TOOL/SERVICE

DISTRIBUTION

DATA

☐ For information
☐ CLARIN:EL compatible service

Save draft
Save

CATEGORIES

TECHNICAL

EVALUATION

Specify if the tool is language dependent *
☐ Language dependent

Input content resource
Describe the requirements that a data resource must fulfill in order to be processed

Input resource type *
Select the type of the resource that can be processed by the tool (e.g., corpus, single file, lexicon, etc.)

Media type
Select the media type of the data resource that can be processed by the tool

Data format
Indicates the format(s) of a data resource

Annotation type
If the tool takes as input an annotated resource (e.g. a parser that takes as input a tagged corpus), select the annotation type (multiple values possible)

Για τον πόρο που χρησιμοποιείται στην είσοδο του εργαλείου ή της υπηρεσίας, μια αναδυόμενη λίστα παρέχει τις τιμές που φαίνονται στην ακόλουθη εικόνα. Για να επιλέξετε μία από αυτές, κάντε κλικ στην τιμή.

Remove Input content resource

Input resource type *
corpus

corpus
file
language description
lexical/conceptual resource
other
output text
unspecified
user input text

Το επόμενο πεδίο που πρέπει να συμπληρωθεί, απαιτεί πληροφορίες σχετικά με το μέσο του πόρου που χρησιμοποιείται. Και πάλι, κάντε κλικ σε μια τιμή για να την προσθέσετε.

Media type

audio
image
text
numerical text
video

Για το μορφότυπο (format) δεδομένων που ακολουθεί, πρέπει να πληκτρολογήσετε στο πλαίσιο για να εμφανιστούν οι τιμές που ταιριάζουν με τα κριτήριά σας και να παραλειφθούν όλες οι άλλες από την αναδυόμενη λίστα. Μόλις εντοπίσετε την κατάλληλη τιμή, κάντε κλικ πάνω της.

Data format

mp

Wikipedia template filtered article

BMP

audio **mp3**

MPEG-4

audio **mpg**

Τέλος, εάν ο πόρος που χρησιμοποιείται ως είσοδος στο εργαλείο ή την υπηρεσία είναι επισημειωμένος, πρέπει να δηλώσετε τον τύπο επισημείωσης. Για άλλη μια φορά, αρχίστε να πληκτρολογείτε στο πλαίσιο για να εμφανιστούν οι πιθανές αντίστοιχες τιμές. Επιλέξτε μια από αυτές κάνοντας κλικ πάνω της.

Annotation type

mov|

Body **movement**

Gaze eye **movement**

Head **movement**

Lip **movement**

Το μεταδεδομένο `inputContentResource` βρίσκεται στην ενότητα *Tool/Service* (καρτέλα *Technical*) του περιβάλλοντος τεκμηρίωσης και ανάρτησης πόρων.

LANGUAGE RESOURCE/TECHNOLOGY

TOOL/SERVICE

DISTRIBUTION

DATA

☐ For information

☐ CLARIN:EL compatible service

Save draft

Save

CATEGORIES

Specify if the tool is language dependent *

☐ Language dependent

TECHNICAL

Input content resource

Describe the requirements that a data resource must fulfill in order to be processed

- Δείτε την λεπτομερή αναπαράσταση του στοιχείου `inputContentResource` στο [XSD](#)

Παραδείγματα μεταδεδομένων

Στόχος αυτής της ενότητας είναι να εξοικειωθούν οι χρήστες με τη χρήση των **μεταδεδομένων**. Για το σκοπό αυτό, έχουν εξαχθεί¹ περιγραφές πόρων από την υποδομή **CLARIN:EL** και από αυτές έχουν αντιγραφεί αποσπάσματα που έχουν ενδιαφέρον. Κάθε μεταδεδομένο παρουσιάζεται αυτοτελώς και στη συνέχεια επεξηγείται εν συντομία. Υπάρχουν επίσης σύνδεσμοι προς το πλήρες αρχείο μεταδεδομένων XML για όποιον επιθυμεί να δει πού εντάσσεται το συγκεκριμένο μεταδεδομένο στο σύνολο της περιγραφής του πόρου και προς το **XSD**, όπου υπάρχει λεπτομερής αναπαράσταση του υπό εξέταση στοιχείου.

25.1 resourceName

Το πρώτο μεταδεδομένο είναι το **resourceName** από το **Greek Parliament Plenary Sessions (1989-2019)**, μια συλλογή των πρακτικών της Ολομέλειας του Ελληνικού Κοινοβουλίου των τελευταίων 30 ετών (περισσότερες από 1.000.000 ομιλίες).

XML

```
<ms:resourceName xml:lang="en">Greek Parliament Plenary Sessions (1989-2019)</
↪ms:resourceName>
<ms:resourceName xml:lang="el">Πρακτικά της Ολομέλειας του Ελληνικού Κοινοβουλίου
↪ (1989-2019)</ms:resourceName>
```

Όπως φαίνεται, είναι δυνατό να δοθεί το όνομα του πόρου σε περισσότερες από μία γλώσσες - η πρώτη γλώσσα, εξ ορισμού, είναι η αγγλική (xml:lang=»el»), ενώ η δεύτερη είναι ελεύθερη επιλογής. Εδώ η γλώσσα που επιλέγεται είναι η νέα ελληνική (xml:lang=»en»).

- *Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML*
- *Λεπτομερής αναπαράσταση του στοιχείου στο XSD*

¹ Αυτές οι ετικέτες έρχονται σε ζεύγη - οι ετικέτες έναρξης και λήξης είναι πανομοιότυπες εκτός από το σύμβολο /.

25.2 resourceCreator

Το δεύτερο απόσπασμα προέρχεται από την **KELLY word-list**, έναν μονόγλωσσο λεξικό/εννοιολογικό πόρο. Οι λίστες λέξεων KELLY δημιουργήθηκαν για να διευκολύνουν την εκμάθηση μιας ξένης/δεύτερης γλώσσας. Το ελληνικό μέρος δημιουργήθηκε από το *Ινστιτούτο Επεξεργασίας Γλώσσας και Λόγου* το οποίο είναι ένας **οργανισμός**.

XML

```
<ms:resourceCreator>
  <ms:Organization>
    <ms:actorType>Organization</ms:actorType>
    <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου</
↪ms:organizationName>
    <ms:organizationName xml:lang="en">Institute for Language and Speech
↪Processing</ms:organizationName>
    <ms:website>http://www.ilsp.gr/</ms:website>
  </ms:Organization>
</ms:resourceCreator>
```

Οι απαραίτητες πληροφορίες σχετικά με τον δημιουργό περικλείονται μεταξύ των ετικετών resourceCreator. Αρχικά, ορίζεται ο τύπος του δημιουργού (actorType)- ένας πόρος μπορεί να έχει ως δημιουργό ένα *άνθρωπο*, μια *ομάδα* ανθρώπων ή έναν *οργανισμό*, όπως συμβαίνει στην περίπτωση του Kelly world-list. Στη συνέχεια παρέχεται το όνομα του οργανισμού (σε δύο γλώσσες, xml:lang=»el» και xml:lang=»en») καθώς και ο ιστότοπός του.

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.3 isPartOf

Το επόμενο παράδειγμα είναι από το **Golden Part of Speech Tagged Corpus**, ένα μονόγλωσσο επισημειωμένο σώμα κειμένων στα ελληνικά με 100.000 λέξεις. Το σώμα αυτό είναι **υποσύνολο** του **Hellenic National Corpus** που περιέχει περισσότερες από 97 εκατομμύρια λέξεις από διάφορες πηγές και διάφορους τομείς. Η σχέση υποσυνόλου εκφράζεται μέσω του μεταδεδομένου isPartOf στο *σχήμα μεταδεδομένων CLARIN:EL*.

XML

```
<ms:isPartOf>
  <ms:resourceName xml:lang="el">Ελληνικός Θησαυρός της Ελληνικής Γλώσσας</
↪ms:resourceName>
  <ms:resourceName xml:lang="en">Hellenic National Corpus</ms:resourceName>
  <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
    >http://hdl.handle.net/11500/ATHENA-0000-0000-23E2-9</ms:LRIdentifier>
  <ms:version>3.0</ms:version>
</ms:isPartOf>
```

Το στοιχείο `isPartOf` περιλαμβάνει το όνομα του πόρου (`resourceName`) από τον οποίο έχει προκύψει ο υπό εξέταση πόρος, δηλαδή το *Hellenic National Corpus*, σε δύο γλώσσες (`xml:lang=»el»` και `xml:lang=»en»`) μαζί με το αναγνωριστικό του (`LRIdentifier`) και την έκδοσή του (`version`).

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.4 annotationType

Η στοίχιση (`alignment`) είναι η διαδικασία που καθορίζει μεταφραστικές ισοδυναμίες μεταξύ δομικών μονάδων (λέξεις, προτάσεις κλπ.) ενός κειμένου σε μια δεδομένη γλώσσα και ενός κειμένου με παρόμοιο νόημα σε άλλη(-ες) γλώσσα(-ες). Το *Greek-Bulgarian Bul-TM parallel corpus* είναι ένα διγλωσσικό σώμα κειμένων και όπως υποδηλώνει το επίθετο *παράλληλο*, τα ελληνικά και τα βουλγαρικά έχουν **στοιχηθεί**.

XML

```
<ms:annotation>
  <ms:annotationType>http://w3id.org/meta-share/omtd-share/Alignment1</
↪ms:annotationType>
  <ms:segmentationLevel>http://w3id.org/meta-share/meta-share/sentence</
↪ms:segmentationLevel>
  <ms:annotationStandoff>false</ms:annotationStandoff>
  <ms:annotationMode>http://w3id.org/meta-share/meta-share/automatic</
↪ms:annotationMode>
  <ms:isAnnotatedBy>
    <ms:resourceName xml:lang="en">TrAid</ms:resourceName>
    <ms:version>unspecified</ms:version>
  </ms:isAnnotatedBy>
</ms:annotation>
```

Η στοίχιση θεωρείται ένας τύπος `annotation` (επισημείωσης). Οι δύο γλώσσες έχουν παραλληλοποιηθεί σε *επίπεδο πρότασης* (`segmentationLevel`) και δεν υπάρχουν ξεχωριστά έγγραφα (`annotationStandoff`) με περιεχόμενο στην κάθε γλώσσα. Η διαδικασία έχει γίνει *αυτόματα* (`annotationMode`)- το εργαλείο που χρησιμοποιήθηκε για την στοίχιση (`isAnnotatedBy`) ονομάζεται *TrAid* αλλά δεν υπάρχει διαθέσιμη έκδοση (*unspecified*).

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.5 multilingualityType

Το **DICTA-SIGN corpus** είναι ένα **πολυμεσικό** σώμα κειμένων, αποτελούμενο από δύο τμήματα (ένα βίντεο και ένα κείμενο), για τέσσερις νοηματικές γλώσσες (αγγλικά, γαλλικά, γερμανικά και ελληνικά).

XML

```
<ms:multilingualityType>http://w3id.org/meta-share/meta-share/parallel</
↪ms:multilingualityType>
  <ms:language>
    <ms:languageTag>gss</ms:languageTag>
    <ms:languageId>gss</ms:languageId>
  </ms:language>
  <ms:language>
    <ms:languageTag>bfi</ms:languageTag>
    <ms:languageId>bfi</ms:languageId>
  </ms:language>
  <ms:language>
    <ms:languageTag>gsg</ms:languageTag>
    <ms:languageId>gsg</ms:languageId>
  </ms:language>
  <ms:language>
    <ms:languageTag>fsl</ms:languageTag>
    <ms:languageId>fsl</ms:languageId>
  </ms:language>
```

Κάθε τμήμα του σώματος κειμένων περιγράφεται ξεχωριστά. Αυτό το απόσπασμα περιγράφει το περιεχόμενο του τμήματος του πόρου που είναι για το **βίντεο**. Οι γλώσσες στο βίντεο είναι νοηματικές γλώσσες και είναι παραλληλοποιημένες όπως υποδεικνύεται από την επιλογή της τιμής *parallel* για το στοιχείο `multilingualityType`. Στη συνέχεια, κάθε γλώσσα (`language`) παρουσιάζεται ξεχωριστά με την ετικέτα της (`languageTag`) και το αναγνωριστικό της (`languageId`): `gss` (Greek Sign Language/Ελληνική Νοηματική Γλώσσα), `bfi` (British Sign Language/Βρετανική Νοηματική Γλώσσα), `gsg` (German Sign Language/Γερμανική Νοηματική Γλώσσα) και `fsl` (French Sign Language/Γαλλική Νοηματική Γλώσσα).

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.6 isDocumentedBy

Μερικές φορές υπάρχουν επιπλέον πληροφορίες για έναν πόρο σε εξωτερικά έγγραφα, όπως άρθρα ή/και ανακοινώσεις συνεδρίων. Τέτοια είναι η περίπτωση του **Orossimo Terminological Resource - History** που τεκμηριώνεται στο *Collection of digital terminological resources: methodology and results*.

XML

```
<ms:isDocumentedBy>
  <ms:title xml:lang="el">Συλλογή ηλεκτρονικών ορολογικών πόρων:
↪μεθοδολογία και αποτελέσματα</ms:title>
```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

    <ms:title xml:lang="en">Collection of digital terminological
↪resources: methodology and results</ms:title>
</ms:isDocumentedBy>

```

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.7 fundingProject

Το ακόλουθο παράδειγμα είναι πιο σύνθετο, καθώς περιλαμβάνει διάφορα μεταδεδομένα. Είναι από το *Trilingual Terminological Dictionary*, έναν λεξικό/εννοιολογικό πόρο με τριπλό στόχο: να βοηθήσει τους μαθητές στην εκμάθηση των μαθημάτων του προγράμματος σπουδών, στη βελτίωση των γλωσσικών τους δεξιοτήτων στην ελληνική γλώσσα και στην εξοικείωση με την τεχνολογία της πληροφορίας.

XML

```

<ms:fundingProject>
  <ms:projectName xml:lang="el">Τρίγλωσσο Ορολογικό Λεξικό</ms:projectName>
  <ms:projectName xml:lang="en">Trilingual Terminological Dictionary</
↪ms:projectName>
  <ms:website>https://bit.ly/2V4hWLe</ms:website>
  <ms:website>https://www.ilsp.gr/projects/tol/</ms:website>
  <ms:fundingType>http://w3id.org/meta-share/meta-share/euFunds</ms:fundingType>
  <ms:fundingType>http://w3id.org/meta-share/meta-share/nationalFunds</
↪ms:fundingType>
  <ms:funder>
    <ms:Organization>
      <ms:actorType>Organization</ms:actorType>
      <ms:organizationName xml:lang="en">Ministry of Education and
↪Religious Affairs</ms:organizationName>
    </ms:Organization>
  </ms:funder>
  <ms:funder>
    <ms:Organization>
      <ms:actorType>Organization</ms:actorType>
      <ms:organizationName xml:lang="el">Ευρωπαϊκή Επιτροπή</
↪ms:organizationName>
      <ms:organizationName xml:lang="en">European Commission</
↪ms:organizationName>
      <ms:website>https://ec.europa.eu/info/index_en</ms:website>
    </ms:Organization>
  </ms:funder>
</ms:fundingProject>

```

Ο πόρος είναι το αποτέλεσμα ενός έργου (fundingProject) με το ίδιο όνομα (projectName), *Trilingual Terminological Dictionary*. Οι πληροφορίες που παρέχονται για το έργο είναι οι διαθέσιμες ιστοσελίδες (websites), ο τύπος χρηματοδότησης (fundingType) και οι χρηματοδότες (funders). Το έργο δημιουργήθηκε με *ευρωπαϊκούς και εθνικούς πόρους*, ενώ το χρηματοδότησαν δύο οργανισμοί, το *Υπουργείο Παιδείας και Θρησκευμάτων* και η *Ευρωπαϊκή Επιτροπή*.

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
 - Λεπτομερής αναπαράσταση του στοιχείου στο XSD
-

25.8 inputContentResource

Το ακόλουθο απόσπασμα XML παρέχει πληροφορίες σχετικά με τα μορφώτυπα δεδομένων που δέχονται στην είσοδό τους τα *Voyant Tools*, ένα διαδικτυακό περιβάλλοντος γλωσσικής ανάλυσης.

XML

```
<ms:inputContentResource>
  <ms:processingResourceType>http://w3id.org/meta-share/meta-share/corpus</
  ↪ms:processingResourceType>
  <ms:mediaType>http://w3id.org/meta-share/meta-share/text</ms:mediaType>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Pdf</ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Rtf</ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Xml</ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/ConllU</ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Html</ms:dataFormat>
</ms:inputContentResource>
```

Τα Voyant tools μπορούν να επεξεργαστούν, λαμβάνοντας στην είσοδό τους (inputContentResource), σώματα κειμένων (processingResourceType) των οποίων τα μορφώτυπα (dataFormat) είναι *plain text*, *PDF*, *RTF*, *XML*, *ConllU* και *HTML*.

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
 - Λεπτομερής αναπαράσταση του στοιχείου στο XSD
-

25.9 outputResource

Στο επόμενο απόσπασμα παρουσιάζεται η έξοδος του εργαλείου *ILSP Language Identification System*.

XML

```
      <ms:outputResource>
      <ms:processingResourceType>http://w3id.org/meta-share/meta-share/corpus</
      ↪ms:processingResourceType>
      <ms:language>
        <ms:languageTag>el-Latn</ms:languageTag>
        <ms:languageId>el</ms:languageId>
        <ms:scriptId>Latn</ms:scriptId>
        <ms:languageVarietyName xml:lang="en">Greeklish</ms:languageVarietyName>
      </ms:language>
      <ms:language>
        <ms:languageTag>el-Grek</ms:languageTag>
        <ms:languageId>el</ms:languageId>
```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

      <ms:scriptId>Grek</ms:scriptId>
    </ms:language>
    <ms:language>
      <ms:languageTag>fr</ms:languageTag>
      <ms:languageId>fr</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>en</ms:languageTag>
      <ms:languageId>en</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>de</ms:languageTag>
      <ms:languageId>de</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>nl</ms:languageTag>
      <ms:languageId>nl</ms:languageId>
    </ms:language>
    <ms:mediaType>http://w3id.org/meta-share/meta-share/text</ms:mediaType>
  </ms:outputResource>

```

Αυτό το εργαλείο αναγνωρίζει γλώσσες όπως τα *greekish*, *ελληνικά*, *αγγλικά*, *γερμανικά*, *ολλανδικά* και *γαλλικά*. Τα *greekish*, όπως φαίνεται στο παραπάνω απόσπασμα, είναι μια ποικιλία (*languageVarietyName*) της ελληνικής γλώσσας: η γλώσσα (*languageId*) ορίζεται ως η *ελληνική* (*el*) αλλά η γραφή (*scriptId*) γίνεται με το *λατινικό* (*Latn*) αλφάβητο.

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

25.10 attributionText

Το τελευταίο παράδειγμα παρουσιάζει το κείμενο αναφοράς δημιουργού (*attributionText*) ενός πόρου γλωσσικής περιγραφής, του *PANACEA Environment Corpus n-grams EL*.

XML

```

<ms:attributionText xml:lang="el">PANACEA σώμα ελληνικών n-γραμμάτων (n-grams)
↳ περιβαλλοντικού τομέα. Δημιουργός:
Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό Κέντρο Αθηνά. Άδεια: Creative Commons
↳ Attribution Share Alike 4.0
International (https://creativecommons.org/licenses/by-sa/4.0/legalcode,
https://creativecommons.org/licenses/by-sa/4.0/). Πηγή: http://hdl.handle.net/11500/
↳ ATHENA-0000-0000-23DA-3
(CLARIN:EL)</ms:attributionText>
<ms:attributionText xml:lang="en">PANACEA Environment Corpus n-grams EL (Greek) by
↳ Institute for Language and Speech
Processing - Athena Research Center used under Creative Commons Attribution Share
↳ Alike 4.0 International
(https://creativecommons.org/licenses/by-sa/4.0/legalcode, https://creativecommons.
↳ org/licenses/by-sa/4.0/). Source:
http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3 (CLARIN:EL)</ms:attributionText>

```

Η άδεια χρήσης του πόρου είναι η *CC-BY-SA 4.0 International*. «Αυτή η άδεια επιτρέπει σε άλλους να επαναχρησιμοποιούν, να προσαρμόζουν και να αξιοποιούν το έργο σας ακόμη και για εμπορικούς σκοπούς, αρκεί να σας αναφέρουν και να αδειοδοτούν τις νέες τους δημιουργίες με τους ίδιους όρους. Αυτή η άδεια συγκρίνεται συχνά με τις άδειες ελεύθερου και ανοικτού λογισμικού «copyleft». Όλα τα νέα έργα που βασίζονται στο δικό σας θα φέρουν την ίδια άδεια, οπότε οποιαδήποτε παράγωγα θα επιτρέπουν επίσης την εμπορική χρήση.»² Η αναφορά δημιουργού εξυπηρετεί αυτόν ακριβώς τον σκοπό, καθώς παρέχει σε κάποιον κείμενο που περιέχει τις πληροφορίες για τον δημιουργό του πόρου, το *Institute for Language and Speech Processing - Athena Research Center* και την άδεια χρήσης με την οποία θα διανέμεται ο πόρος και όλα τα παράγωγά του.

- Πλήρης περιγραφή μεταδεδομένων του πόρου σε XML
- Λεπτομερής αναπαράσταση του στοιχείου στο XSD

² Περισσότερες πληροφορίες στον ιστότοπο του Creative Commons.

Πλήρεις περιγραφές μεταδεδομένων σε XML

Οι περιγραφές μεταδεδομένων σε XML που παρουσιάζονται εδώ έχουν εξαχθεί από την υποδομή **CLARIN:EL** και αντιστοιχούν σε πραγματικούς πόρους. Κάντε κλικ στο κουτάκι με το όνομα του πόρου για να εμφανιστεί η περιγραφή μεταδεδομένων του πόρου σε XML. Σε όλα τα αρχεία τα ονόματα, τα επώνυμα και οι ηλεκτρονικές διευθύνσεις των δημιουργών, των επιμελητών και των λοιπών επαφών έχουν ανωνυμοποιηθεί.

26.1 1. Σώματα κειμένων

26.1.1 Μονόγλωσσο σώμα κειμένων #1

```
<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2019-12-17</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-11-24</ms:metadataLastDateUpdated>
<ms:metadataCurator>
<ms:actorType>Person</ms:actorType>
<ms:surname xml:lang="en">Person_Surname</ms:surname>
<ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
<ms:actorType>Person</ms:actorType>
<ms:surname xml:lang="en">Person_Surname</ms:surname>
<ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
<ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
<ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
<ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">Πρακτικά της Ολομέλειας του Ελληνικού
    ↪ Κοινοβουλίου
      (1989-2019)</ms:resourceName>
    <ms:resourceName xml:lang="en">Greek Parliament Plenary Sessions
      (1989-2019)</ms:resourceName>
    <ms:description xml:lang="el">Σώμα κειμένων με τα πρακτικά των συνεδριάσεων
    ↪ της
      Ολομέλειας του Ελληνικού Κοινοβουλίου τα τελευταία 30 χρόνια
    ↪ (περισσότερες από
      1.000.000 ομιλίες). Το παρόν σώμα κειμένων περιλαμβάνει όλα τα πρακτικά
    ↪ σε μορφή
      txt. Για ευκολότερη επεξεργασία έχουν δημιουργηθεί και μικρότερα
    ↪ υποσύνολα, με
      ανώτατο μέγεθος συμπιεσμένων αρχείων τα 40 Mb ανά υποσύνολο, που
    ↪ ανταποκρίνονται
      στις εκάστοτε βουλευτικές περιόδους.</ms:description>
    <ms:description xml:lang="en">This is a collection of the raw minutes of the
    ↪ Greek
      Parliament plenary sessions of the last 30 years (more than 1.000.000
    ↪ speeches). The
      existing corpus has all raw data in txt format. In order to make the
    ↪ resource more
      processable, we have also split it into smaller subcorpora, with a maximum
      compressed folder size of 40 Mb per subcorpus. The created subcorpora are
      thematically organized per Greek parliamentary terms.</ms:description>
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
      >http://hdl.handle.net/11500/ATHENA-0000-0000-5D62-A</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
    <ms:additionalInfo>
      <ms:email>person@ilsp.gr</ms:email>
    </ms:additionalInfo>
    <ms:contact>
      <ms:Person>
        <ms:actorType>Person</ms:actorType>
        <ms:surname xml:lang="en">Person_Surname</ms:surname>
        <ms:givenName xml:lang="en">Person_Name</ms:givenName>
      </ms:Person>
    </ms:contact>
    <ms:citationText xml:lang="el">Πρακτικά της Ολομέλειας του Ελληνικού
    ↪ Κοινοβουλίου
      (1989-2019) (2019). Version 1.0.0 (automatically assigned). [Dataset
    ↪ (Text corpus)].
      CLARIN:EL. http://hdl.handle.net/11500/ATHENA-0000-0000-5D62-A</
    ↪ ms:citationText>
      <ms:citationText xml:lang="en">Greek Parliament Plenary Sessions (1989-2019)
    ↪ (2019).
      Version 1.0.0 (automatically assigned). [Dataset (Text corpus)].
    ↪ CLARIN:EL.
      http://hdl.handle.net/11500/ATHENA-0000-0000-5D62-A</ms:citationText>
    <ms:keyword xml:lang="en">monolingual, political science</ms:keyword>
    <ms:resourceProvider>
      <ms:Organization>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

    <ms:actorType>Organization</ms:actorType>
    <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
        Λόγου</ms:organizationName>
    <ms:organizationName xml:lang="en">Institute for Language and Speech
        Processing</ms:organizationName>
    <ms:website>http://www.ilsp.gr/</ms:website>
  </ms:Organization>
</ms:resourceProvider>
<ms:LRS subclass>
  <ms:Corpus>
    <ms:lrType>Corpus</ms:lrType>
    <ms:corpusSubclass>http://w3id.org/meta-share/meta-share/rawCorpus</
↪ms:corpusSubclass>
    <ms:CorpusMediaPart>
      <ms:CorpusTextPart>
        <ms:corpusMediaType>CorpusTextPart</ms:corpusMediaType>
        <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
        <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪monolingual</ms:lingualityType>
        <ms:language>
          <ms:languageTag>el</ms:languageTag>
          <ms:languageId>el</ms:languageId>
        </ms:language>
      </ms:CorpusTextPart>
    </ms:CorpusMediaPart>
    <ms:DatasetDistribution>
      <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
      <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
      <ms:distributionTextFeature>
        <ms:size>
          <ms:amount>5079.0</ms:amount>
          <ms:sizeUnit>http://w3id.org/meta-share/meta-share/file</
↪ms:sizeUnit>
        </ms:size>
        <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Text</
↪ms:dataFormat>
        <ms:characterEncoding>http://w3id.org/meta-share/meta-share/
↪UTF-8</ms:characterEncoding>
      </ms:distributionTextFeature>
      <ms:licenceTerms>
        <ms:licenceTermsName xml:lang="en">Creative Commons
↪Attribution 4.0
        International</ms:licenceTermsName>
        <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/legalcode</ms:licenceTermsURL>
        <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/</ms:licenceTermsURL>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
        <ms:LicenceIdentifier
            ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
            >CC-BY-4.0</ms:LicenceIdentifier>
        </ms:licenceTerms>
        <ms:attributionText xml:lang="el">Πρακτικά της Ολομέλειας του
↪Ελληνικού
            Κοινοβουλίου (1989-2019). Άδεια: Creative Commons Attribution
↪4.0
            International (https://creativecommons.org/licenses/by/4.0/
↪legalcode,
            https://creativecommons.org/licenses/by/4.0/). Πηγή:
            http://hdl.handle.net/11500/ATHENA-0000-0000-5D62-A
            (CLARIN:EL) </ms:attributionText>
        <ms:attributionText xml:lang="en">Greek Parliament Plenary
↪Sessions
            (1989-2019) used under Creative Commons Attribution 4.0
↪International
            (https://creativecommons.org/licenses/by/4.0/legalcode,
            https://creativecommons.org/licenses/by/4.0/). Source:
            http://hdl.handle.net/11500/ATHENA-0000-0000-5D62-A
            (CLARIN:EL) </ms:attributionText>
        </ms:DatasetDistribution>
        <ms:personalDataIncluded>>false</ms:personalDataIncluded>
        <ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
    </ms:Corpus>
</ms:LRSubclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```

26.1.2 Μονόγλωσσο σώμα κειμένων #2

```

<?xml version="1.0" encoding="utf-8"?>
    <ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.
↪clarin.gr/metadata-schema/CLARIN-SHARE.xsd">
        <ms:metadataCreationDate>2021-02-25</ms:metadataCreationDate>
        <ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
        <ms:metadataCurator>
            <ms:actorType>Person</ms:actorType>
            <ms:surname xml:lang="en">Person_Surname</ms:surname>
            <ms:givenName xml:lang="en">Person_Name</ms:givenName>
        </ms:metadataCurator>
        <ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</
↪ms:compliesWith>
        <ms:metadataCreator>
            <ms:actorType>Person</ms:actorType>
            <ms:surname xml:lang="en">Person_Surname</ms:surname>
            <ms:givenName xml:lang="en">Person_Name</ms:givenName>
        </ms:metadataCreator>
    </ms:MetadataRecord>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</
↪ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</
↪ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</
↪ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
  <ms:resourceName xml:lang="en">Golden Part of Speech Tagged Corpus</
↪ms:resourceName>
  <ms:description xml:lang="el">Το Golden Part of Speech Tagged Corpus είναι ένα
↪σύνολο
  δεδομένων που αποτελείται από κείμενα επιλεγμένα από το σύνολο των κειμένων
↪του
  ΕΘΕΓ, και το οποίο είναι μεγέθους 100.000 λέξεων. Τα κείμενα συλλέχθηκαν από
↪το
  διαδίκτυο με τεχνικές διάσχισης σημασιολογικού ιστού (web crawling), με τα
↪εξής
  κριτήρια: κείμενα αποκλειστικά με ελεύθερη διάθεση (CC0 4.0) ή με αναφορά
  δημιουργού (CC BY 4.0) και ευρεία ποικιλία πηγών και θεμάτων. Η διαδικασία που
  ακολουθήθηκε για την υλοποίηση αυτού του σώματος δεδομένων περιλαμβάνει τα
↪ακόλουθα
  βήματα: - Καθαρισμό από άχρηστα στοιχεία (boilerplate material) - Διόρθωση
  ορθογραφικών λαθών με το χέρι - Αυτόματη μορφοσυντακτική επισημείωση με χρήση
↪του
  αυτόματου μορφολογικού επισημειωτή του ΙΕΛ/ΕΚ ΑΘΗΝΑ (ILSP Feature-based multi-
↪tiered
  POS Tagger), με την οποία για κάθε λέξη του σώματος δεδομένων αποδόθηκε ο
  γραμματικός της χαρακτηρισμός και το λήμμα στο οποίο ανήκει - Έλεγχος και
↪διόρθωση
  των αποτελεσμάτων της αυτόματης επισημείωσης, επίσης με το χέρι.
  Το Golden Corpus XML διατίθεται σε μορφή XML, ως αρχείο το οποίο περιλαμβάνει
↪όλα τα
  κείμενα. Η δομή κάθε κειμένου περιλαμβάνει αρχικά ορισμένα μεταδεδομένα για το
  κείμενο και μετά το ίδιο το κείμενο, που είναι δομημένο
  σε επισημειωμένες παραγράφους, προτάσεις και λέξεις. Κάθε λέξη συνοδεύεται από
  πληροφορίες για το λήμμα της, τον γραμματικό χαρακτηρισμό της και την θέση
↪της μέσα
  στην πρόταση με βάση τους χαρακτήρες έναρξης και λήξης. Επιλέχθηκε ως μορφή
  διάθεσης η XML ώστε το σύνολο δεδομένων να μπορεί να χρησιμοποιηθεί σε
↪διαφορετικά
  περιβάλλοντα, ανεξαρτήτως λειτουργικού συστήματος.</ms:description>
  <ms:description xml:lang="en">The Golden Part-of-Speech Tagged Corpus is a subset
↪of the
  Hellenic National Corpus (HNC), the size of which is 100.000 words; it
↪consists of
  selected texts from a variety of sources covering various domains. These
↪texts have
  been crawled from the web and are licensed under either CC0 4.0 or CC BY 4.0..
↪The
  corpus underwent the following stages: • cleaning and removal of
  boilerplate material, • manual correction of typos and spelling mistakes, •
  automatic lemmatization and part-of-speech tagging for each word, using the
↪ILSP

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

Feature-based multi-tiered POS Tagger, and • manual correction of the ILSP Feature-based multi-tiered POS Tagger results. The Golden Part of Speech Tagged Corpus is available as a single XML file, containing all texts in the following structure: first, some metadata about the text and then the text itself with annotation at the level of paragraphs, sentences and words. Each word comes with information on its lemma, POS and its boundaries (beginning and end). XML was chosen as the most appropriate format as it can be used in various environments, regardless of the operating system in use.

```

</ms:description>
<ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
>http://hdl.handle.net/11500/ATHENA-0000-0000-5E7D-C</ms:LRIdentifier>
<ms:version>1</ms:version>
<ms:additionalInfo>
  <ms:email>person@ilsp.athena-innovation.gr</ms:email>
</ms:additionalInfo>
<ms:contact>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:contact>
<ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
Κέντρο
Αθηνά (2021). Golden Part of Speech Tagged Corpus. Version 1. [Dataset (Text
corpus)]. CLARIN:EL.
http://hdl.handle.net/11500/ATHENA-0000-0000-5E7D-C</ms:citationText>
<ms:citationText xml:lang="en">Institute for Language and Speech Processing -
Athena
Research Center (2021). Golden Part of Speech Tagged Corpus. Version 1.
[Dataset
(Text corpus)]. CLARIN:EL.
http://hdl.handle.net/11500/ATHENA-0000-0000-5E7D-C</ms:citationText>
<ms:keyword xml:lang="en">monolingual</ms:keyword>
<ms:keyword xml:lang="en">morphosyntacticAnnotation-posTagging</ms:keyword>
<ms:resourceCreator>
  <ms:Organization>
    <ms:actorType>Organization</ms:actorType>
    <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
    Λόγου</ms:organizationName>
    <ms:organizationName xml:lang="en">Institute for Language and Speech
    Processing</ms:organizationName>
    <ms:website>http://www.ilsp.gr</ms:website>
  </ms:Organization>
</ms:resourceCreator>
<ms:isPartOf>
  <ms:resourceName xml:lang="el">Ελληνικός Θησαυρός της Ελληνικής
  Γλώσσας</ms:resourceName>
  <ms:resourceName xml:lang="en">Hellenic National Corpus</ms:resourceName>
  <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
  >http://hdl.handle.net/11500/ATHENA-0000-0000-23E2-9</ms:LRIdentifier>
  <ms:version>3.0</ms:version>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

</ms:isPartOf>
<ms:LRS subclass>
  <ms:Corpus>
    <ms:lrType>Corpus</ms:lrType>
    <ms:corpusSubclass>http://w3id.org/meta-share/meta-share/annotatedCorpus</
↪ms:corpusSubclass>
    <ms:CorpusMediaPart>
      <ms:CorpusTextPart>
        <ms:corpusMediaType>CorpusTextPart</ms:corpusMediaType>
        <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
        <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪monolingual</ms:lingualityType>
        <ms:language>
          <ms:languageTag>el</ms:languageTag>
          <ms:languageId>el</ms:languageId>
        </ms:language>
        <ms:annotation>
          <ms:annotationType>http://w3id.org/meta-share/omtd-share/
↪PartOfSpeech</ms:annotationType>
          <ms:isAnnotatedBy>
            <ms:resourceName xml:lang="en">ILSP Feature-based multi-
↪tiered
              POS Tagger</ms:resourceName>
            <ms:LRIIdentifier>
              ms:LRIIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
              >http://hdl.handle.net/11500/ATHENA-0000-0000-23E8-3</
↪ms:LRIIdentifier>
              <ms:version>1</ms:version>
            </ms:isAnnotatedBy>
          </ms:annotation>
          <ms:annotation>
            <ms:annotationType>http://w3id.org/meta-share/omtd-share/
↪StructuralAnnotationType</ms:annotationType>
            <ms:segmentationLevel>http://w3id.org/meta-share/meta-share/
↪sentence</ms:segmentationLevel>
            <ms:segmentationLevel>http://w3id.org/meta-share/meta-share/
↪token1</ms:segmentationLevel>
            </ms:annotation>
            <ms:annotation>
              <ms:annotationType>http://w3id.org/meta-share/omtd-share/Lemma
↪</ms:annotationType>
            </ms:annotation>
          </ms:CorpusTextPart>
        </ms:CorpusMediaPart>
      <ms:DatasetDistribution>
        <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
        <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
        <ms:accessLocation>http://fixme.com</ms:accessLocation>
        <ms:distributionTextFeature>
          <ms:size>
            <ms:amount>100000.0</ms:amount>
            <ms:sizeUnit>http://w3id.org/meta-share/meta-share/word3</
↪ms:sizeUnit>

```

(συνέχεια στην επόμενη σελίδα)

```

</ms:size>
<ms:dataFormat>http://w3id.org/meta-share/omtd-share/Xml</
↪ms:dataFormat>
<ms:characterEncoding>http://w3id.org/meta-share/meta-share/UTF-8
↪</ms:characterEncoding>
</ms:distributionTextFeature>
<ms:licenceTerms>
  <ms:licenceTermsName xml:lang="en">Creative Commons Attribution 4.
↪0
    International</ms:licenceTermsName>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.0/
↪legalcode</ms:licenceTermsURL>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.0/</
↪ms:licenceTermsURL>
  <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/public</
↪ms:licenceCategory>
  <ms:LicenceIdentifier
    ms:LicenceIdentifierScheme="http://w3id.org/meta-share/meta-
↪share/SPDX"
    >CC-BY-4.0</ms:LicenceIdentifier>
  </ms:licenceTerms>
  <ms:attributionText xml:lang="el">Golden Part of Speech Tagged Corpus.
    Δημιουργός: Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό Κέντρο
↪Αθηνά.
    Άδεια: Creative Commons Attribution 4.0 International
    (https://creativecommons.org/licenses/by/4.0/legalcode,
    https://creativecommons.org/licenses/by/4.0/). Πηγή:
    http://hdl.handle.net/11500/ATHENA-0000-0000-5E7D-C
    (CLARIN:EL) </ms:attributionText>
  <ms:attributionText xml:lang="en">Golden Part of Speech Tagged Corpus
↪by
    Institute for Language and Speech Processing - Athena Research
↪Center
    used under Creative Commons Attribution 4.0 International
    (https://creativecommons.org/licenses/by/4.0/legalcode,
    https://creativecommons.org/licenses/by/4.0/). Source:
    http://hdl.handle.net/11500/ATHENA-0000-0000-5E7D-C
    (CLARIN:EL) </ms:attributionText>
</ms:DatasetDistribution>
<ms:personalDataIncluded>>false</ms:personalDataIncluded>
<ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
  </ms:Corpus>
  </ms:LRSubclass>
  </ms:LanguageResource>
  </ms:DescribedEntity>
</ms:MetadataRecord>

```

26.1.3 Δίγλωσσο σώμα κειμένων

```
<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
  <ms:metadataCreationDate>2015-09-11</ms:metadataCreationDate>
  <ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
  <ms:metadataCurator>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:metadataCurator>
  <ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
  <ms:metadataCreator>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:metadataCreator>
  <ms:sourceOfMetadataRecord>
    <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
    <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
    <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
  </ms:sourceOfMetadataRecord>
  <ms:DescribedEntity>
    <ms:LanguageResource>
      <ms:entityType>LanguageResource</ms:entityType>
      <ms:resourceName xml:lang="el">Ελληνο-Βουλγαρικό κειμενικό σώμα Bul-TM</
ms:resourceName>
      <ms:resourceName xml:lang="en">Greek-Bulgarian Bul-TM parallel corpus</
ms:resourceName>
      <ms:resourceShortName xml:lang="en">Bul-TM</ms:resourceShortName>
      <ms:description xml:lang="el">Δίγλωσσο σώμα παράλληλων προτάσεων (βουλγαρικά -
ελληνικά)
      από κείμενα γενικής γλώσσας που έχουν αντληθεί από το διαδίκτυο. Το σώμα
      <ms:description>
        <ms:description xml:lang="en">Parallel bilingual corpus (web documents,
general domain)
        aligned at sentence level; the corpus is available in TMX format.</
ms:description>
      <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
>http://hdl.handle.net/11500/ATHENA-0000-0000-23E4-7</ms:LRIdentifier>
      <ms:version>1.0.0 (automatically assigned)</ms:version>
      <ms:additionalInfo>
        <ms:email>person@ilsp.gr</ms:email>
      </ms:additionalInfo>
      <ms:additionalInfo>
        <ms:email>person@ilsp.gr</ms:email>
      </ms:additionalInfo>
      <ms:contact>
        <ms:Person>
          <ms:actorType>Person</ms:actorType>
          <ms:surname xml:lang="en">Person_Surname</ms:surname>
          <ms:givenName xml:lang="en">Person_Name</ms:givenName>
        </ms:Person>
      </ms:contact>
    </ms:LanguageResource>
  </ms:DescribedEntity>
</ms:MetadataRecord>
```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

</ms:contact>
<ms:contact>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:contact>
<ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↪Κέντρο
  Αθηνά (2015). Ελληνο-Βουλγαρικό κειμενικό σώμα Bul-TM. Version 1.0.0
↪(automatically
  assigned). [Dataset (Text corpus)]. CLARIN:EL.
  http://hdl.handle.net/11500/ATHENA-0000-0000-23E4-7</ms:citationText>
<ms:citationText xml:lang="en">Institute for Language and Speech Processing -
↪Athena
  Research Center (2015). Greek-Bulgarian Bul-TM parallel corpus. Version 1.
↪0.0
  (automatically assigned). [Dataset (Text corpus)]. CLARIN:EL.
  http://hdl.handle.net/11500/ATHENA-0000-0000-23E4-7</ms:citationText>
<ms:keyword xml:lang="en">bilingual</ms:keyword>
<ms:keyword xml:lang="en">parallel</ms:keyword>
<ms:keyword xml:lang="en">alignment</ms:keyword>
<ms:keyword xml:lang="en">writtenLanguage</ms:keyword>
<ms:domain>
  <ms:categoryLabel xml:lang="en">Political Science</ms:categoryLabel>
  <ms:DomainIdentifier
↪DDC_classification"
    >DDC320</ms:DomainIdentifier>
  </ms:domain>
<ms:domain>
  <ms:categoryLabel xml:lang="en">society</ms:categoryLabel>
</ms:domain>
<ms:resourceCreator>
  <ms:Organization>
    <ms:actorType>Organization</ms:actorType>
    <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
      Λόγου</ms:organizationName>
    <ms:organizationName xml:lang="en">Institute for Language and Speech
      Processing</ms:organizationName>
    <ms:website>http://www.ilsp.gr/</ms:website>
  </ms:Organization>
</ms:resourceCreator>
<ms:creationStartDate>2005-10-01</ms:creationStartDate>
<ms:creationEndDate>2007-09-30</ms:creationEndDate>
<ms:fundingProject>
  <ms:projectName xml:lang="en">Development of a Bulgarian to Greek and
↪Greek to
  Bulgarian Translation Memory workbench</ms:projectName>
  <ms:website>https://www.ilsp.gr/projects/bultm/</ms:website>
  <ms:fundingType>http://w3id.org/meta-share/meta-share/nationalFunds</
↪ms:fundingType>
  <ms:funder>
    <ms:Organization>
      <ms:actorType>Organization</ms:actorType>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:organizationName xml:lang="en">Ministry of Economy and
            Finances</ms:organizationName>
        </ms:Organization>
    </ms:funder>
</ms:fundingProject>
    <ms:intendedApplication>
        <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageTechnology</ms:LTCClassRecommended>
        </ms:intendedApplication>
    <ms:intendedApplication>
        <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪MachineTranslation</ms:LTCClassRecommended>
        </ms:intendedApplication>
    <ms:actualUse>
        <ms:usedInApplication>
            <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪MachineTranslation</ms:LTCClassRecommended>
            </ms:usedInApplication>
        <ms:usedInApplication>
            <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageTechnology</ms:LTCClassRecommended>
            </ms:usedInApplication>
        </ms:actualUse>
    <ms:isDocumentedBy>
        <ms:title xml:lang="el">Π2.1-Σώματα_Κειμένων-v1</ms:title>
        <ms:title xml:lang="en">Deliverable 2.1 - Text corpora-v1</ms:title>
    </ms:isDocumentedBy>
    <ms:LRS subclass>
        <ms:Corpus>
            <ms:lrType>Corpus</ms:lrType>
            <ms:corpusSubclass>http://w3id.org/meta-share/meta-share/
↪annotatedCorpus</ms:corpusSubclass>
            <ms:CorpusMediaPart>
                <ms:CorpusTextPart>
                    <ms:corpusMediaType>CorpusTextPart</ms:corpusMediaType>
                    <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
                    <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪bilingual</ms:lingualityType>
                    <ms:multilingualityType>http://w3id.org/meta-share/meta-share/
↪parallel</ms:multilingualityType>
                    <ms:language>
                        <ms:languageTag>el</ms:languageTag>
                        <ms:languageId>el</ms:languageId>
                    </ms:language>
                    <ms:language>
                        <ms:languageTag>bg</ms:languageTag>
                        <ms:languageId>bg</ms:languageId>
                    </ms:language>
                    <ms:modalityType>http://w3id.org/meta-share/meta-share/
↪writtenLanguage</ms:modalityType>
                    <ms:annotation>
                        <ms:annotationType>http://w3id.org/meta-share/omtd-share/
↪Alignment1</ms:annotationType>
                        <ms:segmentationLevel>http://w3id.org/meta-share/meta-
↪share/sentence</ms:segmentationLevel>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:annotationStandoff>false</ms:annotationStandoff>
<ms:annotationMode>http://w3id.org/meta-share/meta-share/
↪automatic</ms:annotationMode>
  <ms:isAnnotatedBy>
    <ms:resourceName xml:lang="en">TrAid</ms:resourceName>
    <ms:version>unspecified</ms:version>
  </ms:isAnnotatedBy>
</ms:annotation>
<ms:hasOriginalSource>
  <ms:resourceName xml:lang="en">The JRC-Acquis Corpus, ↪
↪version
    3.0</ms:resourceName>
  <ms:LRIdentifier
    ms:LRIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
    >http://hdl.handle.net/11500/ATHENA-0000-0000-25C9-4</
↪ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
  </ms:hasOriginalSource>
<ms:hasOriginalSource>
  <ms:resourceName xml:lang="en">SETIMES - A parallel ↪
↪corpus of the
    Balkan languages</ms:resourceName>
  <ms:LRIdentifier
    ms:LRIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
    >http://hdl.handle.net/11500/ATHENA-0000-0000-2591-2</
↪ms:LRIdentifier>
    <ms:version>1</ms:version>
  </ms:hasOriginalSource>
  <ms:creationDetails xml:lang="en">original source: EU
    texts</ms:creationDetails>
</ms:CorpusTextPart>
</ms:CorpusMediaPart>
<ms:DatasetDistribution>
  <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
  <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
  <ms:accessLocation>http://fixme.com</ms:accessLocation>
  <ms:distributionTextFeature>
    <ms:size>
      <ms:amount>10000000.0</ms:amount>
      <ms:sizeUnit>http://w3id.org/meta-share/meta-share/token</
↪ms:sizeUnit>
    </ms:size>
    <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Tmx</
↪ms:dataFormat>
    <ms:characterEncoding>http://w3id.org/meta-share/meta-share/
↪UTF-8</ms:characterEncoding>
  </ms:distributionTextFeature>
  <ms:licenceTerms>
    <ms:licenceTermsName xml:lang="en">Creative Commons ↪
↪Attribution 4.0
      International</ms:licenceTermsName>
    <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/legalcode</ms:licenceTermsURL>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/</ms:licenceTermsURL>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
        <ms:LicenceIdentifier
            ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
            >CC-BY-4.0</ms:LicenceIdentifier>
        </ms:licenceTerms>
        <ms:attributionText xml:lang="el">Ελληνο-Βουλγαρικό κειμενικό
↪σώμα Bul-TM.
        Δημιουργός: Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↪Κέντρο Αθηνά.
        Άδεια: Creative Commons Attribution 4.0 International
        (https://creativecommons.org/licenses/by/4.0/legalcode,
        https://creativecommons.org/licenses/by/4.0/). Πηγή:
        http://hdl.handle.net/11500/ATHENA-0000-0000-23E4-7
        (CLARIN:EL) </ms:attributionText>
        <ms:attributionText xml:lang="en">Greek-Bulgarian Bul-TM parallel
↪corpus by
        Institute for Language and Speech Processing - Athena
↪Research Center
        used under Creative Commons Attribution 4.0 International
        (https://creativecommons.org/licenses/by/4.0/legalcode,
        https://creativecommons.org/licenses/by/4.0/). Source:
        http://hdl.handle.net/11500/ATHENA-0000-0000-23E4-7
        (CLARIN:EL) </ms:attributionText>
    </ms:DatasetDistribution>
    <ms:personalDataIncluded>>false</ms:personalDataIncluded>
    <ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
</ms:Corpus>
</ms:LRS subclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```

26.1.4 Πολύγλωσσο σώμα κειμένων

```

<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
↪gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2015-12-23</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">Σώμα DICTA-SIGN</ms:resourceName>
    <ms:resourceName xml:lang="en">DICTA-SIGN corpus</ms:resourceName>
    <ms:description xml:lang="el">Σώμα πολυμεσικών δεδομένων (βίντεο) για τέσσερις
      νοηματικές γλώσσες (ελληνική, αγγλική, γαλλική και γερμανική).
    <ms:description xml:lang="en">Multimedia corpus (video) for four sign
      languages
      (english, french, german and greek) of at least 14 informants per
      language and a
      session duration of approx. 2 hours using the same elicitation materials
      and tasks) across languages. The data is partially annotated. The corpus
      is
      available through a dedicated website, created and maintained by the
      University of
      Hamburg accessible
      href="http://www.sign-lang.uni-hamburg.de/dicta-sign/portal/index.html"
      target="_blank"
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
      >http://hdl.handle.net/11500/ATHENA-0000-0000-28C5-5</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
    <ms:additionalInfo>
      <ms:landingPage>http://www.sign-lang.uni-hamburg.de/dicta-sign/portal/
      index.html</ms:landingPage>
    </ms:additionalInfo>
    <ms:contact>
      <ms:Person>
        <ms:actorType>Person</ms:actorType>
        <ms:surname xml:lang="en">Person_Surname</ms:surname>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:givenName xml:lang="en">Person_Name</ms:givenName>
      </ms:Person>
    </ms:contact>
    <ms:contact>
      <ms:Person>
        <ms:actorType>Person</ms:actorType>
        <ms:surname xml:lang="en">Person_Surname</ms:surname>
        <ms:givenName xml:lang="en">Person_Name</ms:givenName>
      </ms:Person>
    </ms:contact>
    <ms:citationText xml:lang="el">School of Computing Sciences - University of
↪East Anglia;
    Research Institute of Computer Science - Paul Sabatier University;
↪Institute of
    German Sign Language and Communication of the Deaf - University of
↪Hamburg (2015).
    Έγγρα DICTA-SIGN. Version 1.0.0 (automatically assigned). [Dataset (Text
↪and Video
    corpus)]. CLARIN:EL.
    http://hdl.handle.net/11500/ATHENA-0000-0000-28C5-5</ms:citationText>
    <ms:citationText xml:lang="en">School of Computing Sciences - University of
↪East Anglia;
    Research Institute of Computer Science - Paul Sabatier University;
↪Institute of
    German Sign Language and Communication of the Deaf - University of
↪Hamburg (2015).
    DICTA-SIGN corpus. Version 1.0.0 (automatically assigned). [Dataset (Text
↪and Video
    corpus)]. CLARIN:EL.
    http://hdl.handle.net/11500/ATHENA-0000-0000-28C5-5</ms:citationText>
    <ms:keyword xml:lang="en">multilingual</ms:keyword>
    <ms:keyword xml:lang="en">parallel</ms:keyword>
    <ms:keyword xml:lang="en">scripts</ms:keyword>
    <ms:keyword xml:lang="en">signLanguage</ms:keyword>
    <ms:domain>
      <ms:categoryLabel xml:lang="en">Geography & Travel</ms:categoryLabel>
      <ms:DomainIdentifier
↪DDC_classification"
        >DDC910</ms:DomainIdentifier>
    </ms:domain>
    <ms:resourceCreator>
      <ms:Organization>
        <ms:actorType>Organization</ms:actorType>
        <ms:organizationName xml:lang="en">School of Computing
        Sciences</ms:organizationName>
        <ms:website>https://www.uea.ac.uk/about/school-of-computing-sciences</
↪ms:website>
      </ms:Organization>
    </ms:resourceCreator>
    <ms:resourceCreator>
      <ms:Organization>
        <ms:actorType>Organization</ms:actorType>
        <ms:organizationName xml:lang="en">Research Institute of Computer
        Science</ms:organizationName>
        <ms:organizationName xml:lang="fr">Institut de Recherche en
↪Informatique de

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        Toulouse</ms:organizationName>
        <ms:website>http://www.irit.fr/</ms:website>
    </ms:Organization>
</ms:resourceCreator>
<ms:resourceCreator>
    <ms:Organization>
        <ms:actorType>Organization</ms:actorType>
        <ms:organizationName xml:lang="en">Institute of German Sign Language_
↪and
        Communication of the Deaf</ms:organizationName>
        <ms:website>https://www.idgs.uni-hamburg.de/en.html</ms:website>
    </ms:Organization>
</ms:resourceCreator>
<ms:creationStartDate>2009-02-01</ms:creationStartDate>
<ms:creationEndDate>2012-01-31</ms:creationEndDate>
<ms:fundingProject>
    <ms:projectName xml:lang="en">Sign Language Recognition, Generation and_
↪Modelling
        with application in Deaf Communication</ms:projectName>
        <ms:website>http://www.dictasign.eu/</ms:website>
        <ms:fundingType>http://w3id.org/meta-share/meta-share/euFunds</
↪ms:fundingType>
        <ms:funder>
            <ms:Organization>
                <ms:actorType>Organization</ms:actorType>
                <ms:organizationName xml:lang="el">Ευρωπαϊκή Επιτροπή</
↪ms:organizationName>
                <ms:organizationName xml:lang="en">European Commission</
↪ms:organizationName>
                <ms:website>https://ec.europa.eu/info/index_en</ms:website>
            </ms:Organization>
        </ms:funder>
    </ms:fundingProject>
    <ms:intendedApplication>
        <ms:LTClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageTechnology</ms:LTClassRecommended>
    </ms:intendedApplication>
    <ms:actualUse>
        <ms:usedInApplication>
            <ms:LTClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageTechnology</ms:LTClassRecommended>
        </ms:usedInApplication>
    </ms:actualUse>
    <ms:LRSubclass>
        <ms:Corpus>
            <ms:lrType>Corpus</ms:lrType>
            <ms:corpusSubclass>http://w3id.org/meta-share/meta-share/rawCorpus</
↪ms:corpusSubclass>
            <ms:CorpusMediaPart>
                <ms:CorpusTextPart>
                    <ms:corpusMediaType>CorpusTextPart</ms:corpusMediaType>
                    <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
                    <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪multilingual</ms:lingualityType>
                    <ms:multilingualityType>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:multilingualityType>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:language>
  <ms:languageTag>fr</ms:languageTag>
  <ms:languageId>fr</ms:languageId>
</ms:language>
<ms:language>
  <ms:languageTag>el</ms:languageTag>
  <ms:languageId>el</ms:languageId>
</ms:language>
<ms:language>
  <ms:languageTag>de</ms:languageTag>
  <ms:languageId>de</ms:languageId>
</ms:language>
<ms:language>
  <ms:languageTag>en</ms:languageTag>
  <ms:languageId>en</ms:languageId>
</ms:language>
<ms:modalityType>http://w3id.org/meta-share/meta-share/
↪signLanguage</ms:modalityType>
  <ms:TextGenre>
    <ms:categoryLabel xml:lang="en">scripts</ms:categoryLabel>
  </ms:TextGenre>
  <ms:linkToOtherMedia>
    <ms:otherMedia>http://w3id.org/meta-share/meta-share/video
↪</ms:otherMedia>
    <ms:mediaTypeDetails xml:lang="en">scripts of the tasks↪
↪for the
      video</ms:mediaTypeDetails>
    <ms:synchronizedWithVideo>false</ms:synchronizedWithVideo>
  </ms:linkToOtherMedia>
</ms:CorpusTextPart>
</ms:CorpusMediaPart>
<ms:CorpusMediaPart>
  <ms:CorpusVideoPart>
    <ms:corpusMediaType>CorpusVideoPart</ms:corpusMediaType>
    <ms:mediaType>http://w3id.org/meta-share/meta-share/video</
↪ms:mediaType>
    <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪multilingual</ms:lingualityType>
    <ms:multilingualityType>http://w3id.org/meta-share/meta-share/
↪parallel</ms:multilingualityType>
    <ms:language>
      <ms:languageTag>gss</ms:languageTag>
      <ms:languageId>gss</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>bfi</ms:languageTag>
      <ms:languageId>bfi</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>gsg</ms:languageTag>
      <ms:languageId>gsg</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>fsl</ms:languageTag>
      <ms:languageId>fsl</ms:languageId>
    </ms:language>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:modalityType>http://w3id.org/meta-share/meta-share/
↪signLanguage</ms:modalityType>
        <ms:typeOfVideoContent xml:lang="en">natural
            signers</ms:typeOfVideoContent>
        <ms:textIncludedInVideo>http://w3id.org/meta-share/meta-share/
↪none1</ms:textIncludedInVideo>
        <ms:naturality>http://w3id.org/meta-share/meta-share/elicited
↪</ms:naturality>
        <ms:conversationalType>http://w3id.org/meta-share/meta-share/
↪dialogue</ms:conversationalType>
        <ms:scenarioType>http://w3id.org/meta-share/meta-share/
↪rolePlay</ms:scenarioType>
        <ms:audience>http://w3id.org/meta-share/meta-share/none3</
↪ms:audience>
        <ms:interactivity>http://w3id.org/meta-share/meta-share/
↪interactive1</ms:interactivity>
        <ms:linkToOtherMedia>
            <ms:otherMedia>http://w3id.org/meta-share/meta-share/text
↪</ms:otherMedia>
            <ms:mediaTypeDetails xml:lang="en">scripts of the tasks↪
↪for the
                video</ms:mediaTypeDetails>
            <ms:synchronizedWithText>>false</ms:synchronizedWithText>
        </ms:linkToOtherMedia>
    </ms:CorpusVideoPart>
</ms:CorpusMediaPart>
<ms:DatasetDistribution>
    <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪accessibleThroughInterface</ms:DatasetDistributionForm>
    <ms:accessLocation>http://www.sign-lang.uni-hamburg.de/dicta-sign/
↪portal/index.html</ms:accessLocation>
    <ms:distributionTextFeature>
        <ms:size>
            <ms:amount>10.0</ms:amount>
            <ms:sizeUnit>http://w3id.org/meta-share/meta-share/file</
↪ms:sizeUnit>
        </ms:size>
        <ms:dataFormat>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:dataFormat>
    </ms:distributionTextFeature>
    <ms:distributionVideoFeature>
        <ms:size>
            <ms:amount>25.0</ms:amount>
            <ms:sizeUnit>http://w3id.org/meta-share/meta-share/hour1</
↪ms:sizeUnit>
        </ms:size>
        <ms:dataFormat>http://w3id.org/meta-share/omtd-share/mp4</
↪ms:dataFormat>
    </ms:distributionVideoFeature>
    <ms:licenceTerms>
        <ms:licenceTermsName xml:lang="en">Creative Commons↪
↪Attribution Non
            Commercial No Derivatives 4.0 International</
↪ms:licenceTermsName>
        <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc-nd/4.0/legalcode</ms:licenceTermsURL>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc-nd/4.0/</ms:licenceTermsURL>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪nonCommercialUse</ms:conditionOfUse>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪noDerivatives</ms:conditionOfUse>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
        <ms:LicenceIdentifier
            ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
            >CC-BY-NC-ND-4.0</ms:LicenceIdentifier>
    </ms:licenceTerms>
    <ms:attributionText xml:lang="el">Σώμα DICTA-SIGN. Δημιουργός:↪
↪School of
        Computing Sciences - University of East Anglia, Research↪
↪Institute of
        Computer Science - Paul Sabatier University and Institute of↪
↪German Sign
        Language and Communication of the Deaf - University of↪
↪Hamburg. Άδεια:
        Creative Commons Attribution Non Commercial No Derivatives 4.0
        International
        (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode,
        https://creativecommons.org/licenses/by-nc-nd/4.0/). Πηγή:
        http://hdl.handle.net/11500/ATHENA-0000-0000-28C5-5
        (CLARIN:EL) </ms:attributionText>
    <ms:attributionText xml:lang="en">DICTA-SIGN corpus by School of↪
↪Computing
        Sciences - University of East Anglia, Research Institute of↪
↪Computer
        Science - Paul Sabatier University and Institute of German↪
↪Sign Language
        and Communication of the Deaf - University of Hamburg used↪
↪under
        Creative Commons Attribution Non Commercial No Derivatives 4.0
        International
        (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode,
        https://creativecommons.org/licenses/by-nc-nd/4.0/). Source:
        http://hdl.handle.net/11500/ATHENA-0000-0000-28C5-5
        (CLARIN:EL) </ms:attributionText>
    </ms:DatasetDistribution>
    <ms:personalDataIncluded>>false</ms:personalDataIncluded>
    <ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
</ms:Corpus>
</ms:LRS subclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```


26.2 2. Λεξικοί/Εννοιολογικοί πόροι

26.2.1 Μονόγλωσσος λεξικός/εννοιολογικός πόρος

```
<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2015-12-08</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="en">KELLY word-list Greek</ms:resourceName>
    <ms:resourceShortName xml:lang="en">KELLY word-list EL</ms:resourceShortName>
    <ms:description xml:lang="el">Ο μονόγλωσσος πόρος KELLY EL αποτελεί μέρος του
    ψηφιακού εκπαιδευτικού υλικού που αναπτύχθηκε για εννέα γλώσσες, από τις οποίες 4
    ευρέως διδασκόμενες (Αγγλικά, Αραβικά, Ρωσικά και Κινέζικα) και 5 λιγότερο
    διδασκόμενες (Ελληνικά, Ιταλικά, Σουηδικά, Πολωνικά και Νορβηγικά) με σκοπό να
    υποβοηθήσει την εκμάθησή τους ως ξένης/δεύτερης γλώσσας. Πιο συγκεκριμένα, πρόκειται για
    το ελληνικό κομμάτι ενός υλικού το οποίο ολοκληρώνεται από μία σειρά μονόγλωσσων και
    δίγλωσσων εγγραφών καλύπτοντας συνολικά 36 γλωσσικά ζεύγη. Η επιλογή των λέξεων
    βασίστηκε σε ψηφιακούς γλωσσικούς πόρους για κάθε γλώσσα, η επιλογή των οποίων
    ακολούθησε κοινές προδιαγραφές σε όλες τις γλώσσες, με στόχο την εξασφάλιση ομοιόμορφου
    γλωσσικού υλικού. Συνδυάστηκε δε με επεξεργασία των αποτελεσμάτων από γλωσσολόγους
    και εκπαιδευτικούς για την ένταξη κάθε λέξης στο κατάλληλο επίπεδο του Κοινού
    Ευρωπαϊκού Πλαισίου Αναφοράς (CEFR, &lt;a
    href="https://www.coe.int/en/web/common-european-framework-reference-
    languages"
```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

target="_blank">Common European Framework of Reference for Languages). Η επιλογή του λεξιλογίου υιοθέτησε διαδικασία εξαγωγής γνώσης από τα κείμενα (εφαρμογή ακολουθιών τεχνολογιών όπως δομική επισημείωση κειμένου, μορφολογική επισημείωση και ληματοποίηση), ακολουθούμενη από εφαρμογή στατιστικών για την εξαγωγή των πιο συχνών (και άρα απαραίτητων στην εκμάθηση της γλώσσας) λέξεων. Η παιδαγωγική διάσταση καθόρισε τις αρχές επεξεργασίας των αποτελεσμάτων της γλωσσικής τεχνολογίας από ειδικούς, οι οποίοι αξιολόγησαν τα αποτελέσματα της αυτοματοποιημένης διαδικασίας εξαγωγής των λέξεων από τα κείμενα, κατέταξαν τις λέξεις στα επίπεδα του CEFR, και αντιστοίχισαν διαγλωσσικά τις λέξεις, καταρτίζοντας λίστες λεξιλογίου για 36 ζεύγη γλωσσών. Στην επίσημη ιστοσελίδα, υπάρχουν περισσότερες πληροφορίες για το έργο Kelly. Από το ίδιο σημείο είναι διαθέσιμες οι λίστες λέξεων για τα Αγγλικά, Αραβικά, Ιταλικά, Κινέζικα, Νορβηγικά και Ρωσικά, ενώ υπάρχει και διεπαφή που προσφέρει πρόσβαση στη βάση δεδομένων, όπου μπορεί κανείς να αναζητήσει μια λέξη και να δει τις μεταφράσεις της στις άλλες γλώσσες.</ms:description>

<ms:description xml:lang="en">The monolingual lexical conceptual resource</ms:description>

KELLY EL is part of digital material created for educational purposes, i.e. to facilitate the learning of a foreign/second language. Nine different languages were involved, four commonly learned (English, Arabic, Russian and Chinese) and five less commonly learned (Greek, Italian, Swedish, Polish and Norwegian). More precisely, KELLY EL is the Greek part of a material which consists of monolingual and bilingual word-lists covering 36 language pairs in total. The choice of words was based for each language on digital language resources. The same standards were applied to all languages for the choice of these digital language resources in order to ensure uniformity. The material was analyzed and edited by linguists and education professionals and each word was mapped to the appropriate language level of the Common European Framework of Reference (CEFR, Common European Framework of Reference for Languages).

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

The vocabularies produced were created by extracting knowledge from texts
 ↳ (processes such as lemmatization, morphological and structural annotation were used)
 ↳ followed by statistical metrics techniques for extracting the most frequent (and
 ↳ therefore necessary for language learning) words. The language technology results
 ↳ were examined by experts according to pedagogical principles, evaluated and
 ↳ finally bilingual word-lists with words mapped to the different levels of CEFR
 ↳ for 36 language pairs were created. More information about Kelly is available
 ↳ from the [<!--a href="https://spraakbanken.gu.se/eng/kelly" target="_blank">](https://spraakbanken.gu.se/eng/kelly)
 ↳ official project site<!--/a>, where the lists in English, Arabic, Italian,
 ↳ Chinese, Norwegian and Russian can be downloaded. There is also a [<!--a href="http://kelly.sketchengine.co.uk/" target="_blank">](http://kelly.sketchengine.co.uk/)
 ↳ interface database<!--/a>, which can be used to explore the links between words selected
 ↳ for each of these languages.</ms:description>
 <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
 >http://hdl.handle.net/11500/ATHENA-0000-0000-25C1-C</ms:LRIdentifier>
 <ms:version>1.0.0 (automatically assigned)</ms:version>
 <ms:additionalInfo>
 <ms:email>person@ilsp.athena-innovation.gr</ms:email>
 </ms:additionalInfo>
 <ms:contact>
 <ms:Person>
 <ms:actorType>Person</ms:actorType>
 <ms:surname xml:lang="en">Person_Surname</ms:surname>
 <ms:givenName xml:lang="en">Person_Name</ms:givenName>
 </ms:Person>
 </ms:contact>
 <ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
 ↳ Κέντρο
 Αθηνά (2015). KELLY word-list Greek. Version 1.0.0 (automatically
 ↳ assigned).
 [Dataset (Lexical/Conceptual Resource)]. CLARIN:EL.
 http://hdl.handle.net/11500/ATHENA-0000-0000-25C1-C</ms:citationText>
 <ms:citationText xml:lang="en">Institute for Language and Speech Processing -
 ↳ Athena
 Research Center (2015). KELLY word-list Greek. Version 1.0.0
 ↳ (automatically
 assigned). [Dataset (Lexical/Conceptual Resource)]. CLARIN:EL.
 http://hdl.handle.net/11500/ATHENA-0000-0000-25C1-C</ms:citationText>
 <ms:keyword xml:lang="en">monolingual</ms:keyword>
 <ms:resourceCreator>
 <ms:Organization>
 <ms:actorType>Organization</ms:actorType>
 <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
 Λόγου</ms:organizationName>
 <ms:organizationName xml:lang="en">Institute for Language and Speech

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        Processing</ms:organizationName>
        <ms:website>http://www.ilsp.gr/</ms:website>
    </ms:Organization>
</ms:resourceCreator>
<ms:isToBeCitedBy>
    <ms:title xml:lang="en">Kilgarriff, Adam, Frieda Charalabopoulou, Maria
↪Gavrilidou,
        Janne Bondi Johannessen, Saussan Khalil, Sofie Johansson Kokkinakis,
↪Robert Lew,
        Serge Sharoff, Ravikiran Vadlapudi and Elena Volodina (2014) Corpus-
↪based
        vocabulary lists for language learners for nine languages. Language
↪Resources
        and Evaluation, 48:121-163</ms:title>
    <ms:DocumentIdentifier
        ms:DocumentIdentifierScheme="http://purl.org/spar/datacite/doi"
        >https://doi.org/10.1007/s10579-013-9251-2</ms:DocumentIdentifier>
</ms:isToBeCitedBy>
<ms:isPartOf>
    <ms:resourceName xml:lang="en">KELLY word-lists</ms:resourceName>
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
        >http://hdl.handle.net/11500/ATHENA-0000-0000-5862-F</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
</ms:isPartOf>
<ms:LRSubclass>
    <ms:LexicalConceptualResource>
        <ms:lrType>LexicalConceptualResource</ms:lrType>
        <ms:lcrSubclass>http://w3id.org/meta-share/meta-share/wordlist</
↪ms:lcrSubclass>
        <ms:encodingLevel>http://w3id.org/meta-share/meta-share/other</
↪ms:encodingLevel>
        <ms:LexicalConceptualResourceMediaPart>
            <ms:LexicalConceptualResourceTextPart>
                <ms:lcrMediaType>LexicalConceptualResourceTextPart</
↪ms:lcrMediaType>
                <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
                <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪monolingual</ms:lingualityType>
                <ms:language>
                    <ms:languageTag>el</ms:languageTag>
                    <ms:languageId>el</ms:languageId>
                </ms:language>
            </ms:LexicalConceptualResourceTextPart>
        </ms:LexicalConceptualResourceMediaPart>
        <ms:DatasetDistribution>
            <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
            <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
            <ms:distributionTextFeature>
                <ms:size>
                    <ms:amount>7385.0</ms:amount>
                    <ms:sizeUnit>http://w3id.org/meta-share/meta-share/entry</
↪ms:sizeUnit>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        </ms:size>
        <ms:dataFormat>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:dataFormat>
        </ms:distributionTextFeature>
        <ms:licenceTerms>
            <ms:licenceTermsName xml:lang="en">Creative Commons.
↪Attribution Non
                Commercial 4.0 International</ms:licenceTermsName>
            <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc/4.0/legalcode</ms:licenceTermsURL>
            <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc/4.0/</ms:licenceTermsURL>
            <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
            <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪nonCommercialUse</ms:conditionOfUse>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
            <ms:LicenceIdentifier
                ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
                >CC-BY-NC-4.0</ms:LicenceIdentifier>
        </ms:licenceTerms>
        <ms:attributionText xml:lang="el">KELLY word-list Greek.
↪Δημιουργός:
                Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό Κέντρο Αθηνά.
↪Άδεια:
                Creative Commons Attribution Non Commercial 4.0 International
                (https://creativecommons.org/licenses/by-nc/4.0/legalcode,
                https://creativecommons.org/licenses/by-nc/4.0/). Πηγή:
                http://hdl.handle.net/11500/ATHENA-0000-0000-25C1-C
                (CLARIN:EL) </ms:attributionText>
        <ms:attributionText xml:lang="en">KELLY word-list Greek by
↪Institute for
                Language and Speech Processing - Athena Research Center used
↪under
                Creative Commons Attribution Non Commercial 4.0 International
                (https://creativecommons.org/licenses/by-nc/4.0/legalcode,
                https://creativecommons.org/licenses/by-nc/4.0/). Source:
                http://hdl.handle.net/11500/ATHENA-0000-0000-25C1-C
                (CLARIN:EL) </ms:attributionText>
    </ms:DatasetDistribution>
    <ms:personalDataIncluded>>false</ms:personalDataIncluded>
    <ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
</ms:LexicalConceptualResource>
</ms:LRS subclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```

26.2.2 Δίγλωσσος λεξικός/εννοιολογικός πόρος

```

<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2016-12-07</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">Ορολογικός Πόρος ΟΡΟΣΗΜΟ - Ιστορία</
    ms:resourceName>
    <ms:resourceName xml:lang="en">Orossimo Terminological Resource -
    History</ms:resourceName>
    <ms:description xml:lang="el">Πρόκειται για δίγλωσσο ορολογικό πόρο που
    προέρχεται από
    αντίστοιχο σώμα κειμένων ακαδημαϊκού λόγου του γνωστικού τομέα της
    Ιστορίας.</ms:description>
    <ms:description xml:lang="en">A bilingual terminological glossary extracted
    from
    academic discourse texts belonging to the History domain.</ms:description>
  <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
  >http://hdl.handle.net/11500/ATHENA-0000-0000-4B4B-9</ms:LRIdentifier>
  <ms:version>1.0.0 (automatically assigned)</ms:version>
  <ms:additionalInfo>
    <ms:email>person@ilsp</ms:email>
  </ms:additionalInfo>
  <ms:contact>
    <ms:Person>
      <ms:actorType>Person</ms:actorType>
      <ms:surname xml:lang="en">Person_Surname</ms:surname>
      <ms:givenName xml:lang="en">Person_Name</ms:givenName>
    </ms:Person>
  </ms:contact>
  <ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
  Κέντρο
  Αθηνά (2016). Ορολογικός Πόρος ΟΡΟΣΗΜΟ - Ιστορία. Version 1.0.0
  (automatically
  assigned). [Dataset (Lexical/Conceptual Resource)]. CLARIN:EL.

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

    http://hdl.handle.net/11500/ATHENA-0000-0000-4B4B-9</ms:citationText>
    <ms:citationText xml:lang="en">Institute for Language and Speech Processing -
↪Athena
    Research Center (2016). Orossimo Terminological Resource - History.
↪Version 1.0.0
    (automatically assigned). [Dataset (Lexical/Conceptual Resource)].
↪CLARIN:EL.
    http://hdl.handle.net/11500/ATHENA-0000-0000-4B4B-9</ms:citationText>
    <ms:keyword xml:lang="en">bilingual</ms:keyword>
    <ms:domain>
        <ms:categoryLabel xml:lang="en">History</ms:categoryLabel>
        <ms:DomainIdentifier
↪DDC_classification"
            ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
            >DDC900</ms:DomainIdentifier>
        </ms:domain>
    <ms:resourceCreator>
        <ms:Organization>
            <ms:actorType>Organization</ms:actorType>
            <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
                Λόγου</ms:organizationName>
            <ms:organizationName xml:lang="en">Institute for Language and Speech
                Processing</ms:organizationName>
            <ms:website>http://www.ilsp.gr/</ms:website>
        </ms:Organization>
    </ms:resourceCreator>
    <ms:creationStartDate>1996-01-01</ms:creationStartDate>
    <ms:creationEndDate>1998-12-31</ms:creationEndDate>
    <ms:fundingProject>
        <ms:projectName xml:lang="el">ΟΡΟΣΗΜΟ</ms:projectName>
        <ms:projectName xml:lang="en">OROSSIMO</ms:projectName>
        <ms:website>https://www.ilsp.gr/projects/orosimo/</ms:website>
        <ms:fundingType>http://w3id.org/meta-share/meta-share/nationalFunds</
↪ms:fundingType>
        <ms:funder>
            <ms:Organization>
                <ms:actorType>Organization</ms:actorType>
                <ms:organizationName xml:lang="en">General Secretariat for
↪Research and
                    Technology</ms:organizationName>
            </ms:Organization>
        </ms:funder>
    </ms:fundingProject>
    <ms:hasOriginalSource>
        <ms:resourceName xml:lang="el">Σώμα κειμένων ΟΡΟΣΗΜΟ - Ιστορία</
↪ms:resourceName>
        <ms:resourceName xml:lang="en">OROSSIMO Corpus - History</ms:resourceName>
        <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
            >http://hdl.handle.net/11500/ATHENA-0000-0000-240F-8</ms:LRIdentifier>
        <ms:version>1.0.0 (automatically assigned)</ms:version>
    </ms:hasOriginalSource>
    <ms:isDocumentedBy>
        <ms:title xml:lang="el">Συλλογή ηλεκτρονικών ορολογικών πόρων:
↪μεθοδολογία και
            αποτελέσματα</ms:title>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:title xml:lang="en">Collection of digital terminological resources:
↪methodology
    and results</ms:title>
</ms:isDocumentedBy>
<ms:isPartOf>
    <ms:resourceName xml:lang="el">Ορολογικός Πόρος ΟΡΟΣΗΜΟ</ms:resourceName>
    <ms:resourceName xml:lang="en">Orossimo Terminological Resource</
↪ms:resourceName>
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/
↪handle"
        >http://hdl.handle.net/11500/ATHENA-0000-0000-4B49-B</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
</ms:isPartOf>
<ms:LRSubclass>
    <ms:LexicalConceptualResource>
        <ms:lrType>LexicalConceptualResource</ms:lrType>
        <ms:lcrSubclass>http://w3id.org/meta-share/meta-share/
↪terminologicalResource</ms:lcrSubclass>
        <ms:encodingLevel>http://w3id.org/meta-share/meta-share/semantics</
↪ms:encodingLevel>
        <ms:ContentType>http://w3id.org/meta-share/meta-share/lemma</
↪ms:ContentType>
        <ms:ContentType>http://w3id.org/meta-share/meta-share/domain1</
↪ms:ContentType>
        <ms:ContentType>http://w3id.org/meta-share/meta-share/derivation</
↪ms:ContentType>
        <ms:ContentType>http://w3id.org/meta-share/meta-share/
↪translationEquivalent</ms:ContentType>
        <ms:ContentType>http://w3id.org/meta-share/meta-share/note1</
↪ms:ContentType>
        <ms:LexicalConceptualResourceMediaPart>
            <ms:LexicalConceptualResourceTextPart>
                <ms:lcrMediaType>LexicalConceptualResourceTextPart</
↪ms:lcrMediaType>
                <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
                <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪bilingual</ms:lingualityType>
                <ms:language>
                    <ms:languageTag>en</ms:languageTag>
                    <ms:languageId>en</ms:languageId>
                </ms:language>
                <ms:language>
                    <ms:languageTag>el</ms:languageTag>
                    <ms:languageId>el</ms:languageId>
                </ms:language>
            </ms:LexicalConceptualResourceTextPart>
        </ms:LexicalConceptualResourceMediaPart>
        <ms:DatasetDistribution>
            <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
            <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
            <ms:accessLocation>http://fixme.com</ms:accessLocation>
            <ms:distributionTextFeature>
                <ms:size>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:amount>2353.0</ms:amount>
<ms:sizeUnit>http://w3id.org/meta-share/meta-share/term</
↪ms:sizeUnit>
</ms:size>
<ms:dataFormat>http://w3id.org/meta-share/omtd-share/MsExcel</
↪ms:dataFormat>
</ms:distributionTextFeature>
<ms:licenceTerms>
  <ms:licenceTermsName xml:lang="en">Creative Commons
↪Attribution 4.0
    International</ms:licenceTermsName>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/legalcode</ms:licenceTermsURL>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/</ms:licenceTermsURL>
  <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
  <ms:LicenceIdentifier
    ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
    >CC-BY-4.0</ms:LicenceIdentifier>
  </ms:licenceTerms>
  <ms:attributionText xml:lang="el">Ορολογικός Πόρος ΟΡΟΣΗΜΟ -
↪Ιστορία.
    Δημιουργός: Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↪Κέντρο Αθηνά.
    Άδεια: Creative Commons Attribution 4.0 International
    (https://creativecommons.org/licenses/by/4.0/legalcode,
    https://creativecommons.org/licenses/by/4.0/). Πηγή:
    http://hdl.handle.net/11500/ATHENA-0000-0000-4B4B-9
    (CLARIN:EL) </ms:attributionText>
  <ms:attributionText xml:lang="en">Orossimo Terminological
↪Resource - History
    by Institute for Language and Speech Processing - Athena
↪Research Center
    used under Creative Commons Attribution 4.0 International
    (https://creativecommons.org/licenses/by/4.0/legalcode,
    https://creativecommons.org/licenses/by/4.0/). Source:
    http://hdl.handle.net/11500/ATHENA-0000-0000-4B4B-9
    (CLARIN:EL) </ms:attributionText>
</ms:DatasetDistribution>
<ms:personalDataIncluded>>false</ms:personalDataIncluded>
<ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
</ms:LexicalConceptualResource>
</ms:LRSubclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```


26.2.3 Πολύγλωσσος λεξικός/εννοιολογικός πόρος

```

<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2019-03-27</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">Τρίγλωσσο Ορολογικό Λεξικό (ΤΟΛ)</
    ms:resourceName>
    <ms:resourceName xml:lang="en">Trilingual Term Dictionary (TTD)</
    ms:resourceName>
    <ms:description xml:lang="el">Το Τρίγλωσσο Ορολογικό Λεξικό (ΤΟΛ) απευθύνεται
    σε
      αλλόφωνους μαθητές Γυμνασίου της μουσουλμανικής μειονότητας στη Θράκη με
    στόχο τη
      διευκόλυνση της μαθησιακής τους πορείας στο ελληνικό Γυμνάσιο.
    Περιλαμβάνει όρους
      που χρησιμοποιούνται στη διδασκαλία δεκατριών γνωστικών αντικειμένων -
    βιολογίας,
      γεωγραφίας, ιστορίας, κοινωνικής και πολιτικής αγωγής, λογοτεχνίας,
    μαθηματικών,
      μουσικής, νεοελληνικής γλώσσας, οικιακής οικονομίας, πληροφορικής,
    τεχνολογίας,
      φυσικής και χημείας- με βάση το αναλυτικό πρόγραμμα του Γυμνασίου. Οι
    όροι του ΤΟΛ
      έχουν συλλεχθεί από τα σχολικά εγχειρίδια που χρησιμοποιούνται στη
    διδασκαλία των
      αντίστοιχων μαθημάτων. Οι όροι είναι ταξινομημένοι ανά γνωστικό
    αντικείμενο και
      συνοδεύονται από εύληπτους και επιστημονικά ελεγμένους ορισμούς και
    μεταφράσεις στα
      Τουρκικά και τα Αγγλικά.</ms:description>
    <ms:description xml:lang="en">The Trilingual Term Dictionary (TTD) is
    targeted to
      foreign students of the secondary school in Thrace, Greece. The aim of
    the TTD is

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

threefold: to assist the student in learning the subject areas of the
↳curriculum, to
    improve their language skills in Greek and to familiarize themselves with
    information technology. TTD contains terms that are used in several
↳subject areas
    (e.g. biology, geography, history, social and political studies etc.)
↳that are
    taught in the secondary school. The terms of TDD (more than 5.000) have
↳been
    collected from the schoolbooks. The terms are categorised within the
↳subject areas,
    accompanied by definitions and translated into English and Turkish.</
↳ms:description>
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
        >http://hdl.handle.net/11500/ATHENA-0000-0000-5837-0</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
    <ms:additionalInfo>
        <ms:landingPage>http://www.ilsp.gr/tol/</ms:landingPage>
    </ms:additionalInfo>
    <ms:contact>
        <ms:Person>
            <ms:actorType>Person</ms:actorType>
            <ms:surname xml:lang="en">Person_Surname</ms:surname>
            <ms:givenName xml:lang="en">Person_Name</ms:givenName>
        </ms:Person>
    </ms:contact>
    <ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↳Κέντρο
    Αθηνά (2019). Τρίγλωσσο Ορολογικό Λεξικό (ΤΟΛ). Version 1.0.0
↳(automatically
    assigned). [Dataset (Lexical/Conceptual Resource)]. CLARIN:EL.
    http://hdl.handle.net/11500/ATHENA-0000-0000-5837-0</ms:citationText>
    <ms:citationText xml:lang="en">Institute for Language and Speech Processing -
↳Athena
    Research Center (2019). Trilingual Term Dictionary (TTD). Version 1.0.0
    (automatically assigned). [Dataset (Lexical/Conceptual Resource)].
↳CLARIN:EL.
    http://hdl.handle.net/11500/ATHENA-0000-0000-5837-0</ms:citationText>
    <ms:keyword xml:lang="en">multilingual</ms:keyword>
    <ms:domain>
        <ms:categoryLabel xml:lang="en">Mathematics</ms:categoryLabel>
        <ms:DomainIdentifier
↳DDC_classification"
            >DDC510</ms:DomainIdentifier>
        </ms:domain>
        <ms:domain>
            <ms:categoryLabel xml:lang="en">Physics</ms:categoryLabel>
            <ms:DomainIdentifier
↳DDC_classification"
                >DDC530</ms:DomainIdentifier>
            </ms:domain>
            <ms:domain>
                <ms:categoryLabel xml:lang="en">Biology</ms:categoryLabel>
                <ms:DomainIdentifier

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC570</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Technology</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC600</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Home & Family Management</
↪ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC640</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Chemistry</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC540</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Language</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC400</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Geography & Travel</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC910</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Music</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC780</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">History</ms:categoryLabel>
    <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
    >DDC900</ms:DomainIdentifier>
</ms:domain>
<ms:domain>
    <ms:categoryLabel xml:lang="en">Literature, Rhetoric &

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        Criticism</ms:categoryLabel>
        <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
        >DDC800</ms:DomainIdentifier>
    </ms:domain>
    <ms:domain>
        <ms:categoryLabel xml:lang="en">Computer Science, Information &u
↪General
        Works</ms:categoryLabel>
        <ms:DomainIdentifier
ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
↪DDC_classification"
        >DDC000</ms:DomainIdentifier>
    </ms:domain>
    <ms:resourceCreator>
        <ms:Organization>
            <ms:actorType>Organization</ms:actorType>
            <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
Λόγου</ms:organizationName>
            <ms:organizationName xml:lang="en">Institute for Language and Speech
Processing</ms:organizationName>
            <ms:website>http://www.ilsp.gr/</ms:website>
        </ms:Organization>
    </ms:resourceCreator>
    <ms:fundingProject>
        <ms:projectName xml:lang="el">Τρίγλωσσο Ορολογικό Λεξικό</ms:projectName>
        <ms:projectName xml:lang="en">Trilingual Terminological Dictionary</
↪ms:projectName>
        <ms:website>https://bit.ly/2V4hWLe</ms:website>
        <ms:website>https://www.ilsp.gr/projects/tol/</ms:website>
        <ms:fundingType>http://w3id.org/meta-share/meta-share/euFunds</
↪ms:fundingType>
        <ms:fundingType>http://w3id.org/meta-share/meta-share/nationalFunds</
↪ms:fundingType>
        <ms:funder>
            <ms:Organization>
                <ms:actorType>Organization</ms:actorType>
                <ms:organizationName xml:lang="en">Ministry of Education and
↪Religious
                Affairs</ms:organizationName>
            </ms:Organization>
        </ms:funder>
        <ms:funder>
            <ms:Organization>
                <ms:actorType>Organization</ms:actorType>
                <ms:organizationName xml:lang="el">Ευρωπαϊκή Επιτροπή</
↪ms:organizationName>
                <ms:organizationName xml:lang="en">European Commission</
↪ms:organizationName>
                <ms:website>https://ec.europa.eu/info/index_en</ms:website>
            </ms:Organization>
        </ms:funder>
    </ms:fundingProject>
    <ms:LRS subclass>
        <ms:LexicalConceptualResource>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

    <ms:lrType>LexicalConceptualResource</ms:lrType>
    <ms:lcrSubclass>http://w3id.org/meta-share/meta-share/
↪terminologicalResource</ms:lcrSubclass>
    <ms:encodingLevel>http://w3id.org/meta-share/meta-share/other</
↪ms:encodingLevel>
    <ms:LexicalConceptualResourceMediaPart>
      <ms:LexicalConceptualResourceTextPart>
        <ms:lcrMediaType>LexicalConceptualResourceTextPart</
↪ms:lcrMediaType>
        <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
        <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪multilingual</ms:lingualityType>
        <ms:language>
          <ms:languageTag>el</ms:languageTag>
          <ms:languageId>el</ms:languageId>
        </ms:language>
        <ms:language>
          <ms:languageTag>en</ms:languageTag>
          <ms:languageId>en</ms:languageId>
        </ms:language>
        <ms:language>
          <ms:languageTag>tr</ms:languageTag>
          <ms:languageId>tr</ms:languageId>
        </ms:language>
      </ms:LexicalConceptualResourceTextPart>
    </ms:LexicalConceptualResourceMediaPart>
    <ms:DatasetDistribution>
      <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
      <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
      <ms:accessLocation>http://fixme.com</ms:accessLocation>
      <ms:distributionTextFeature>
        <ms:size>
          <ms:amount>5224.0</ms:amount>
          <ms:sizeUnit>http://w3id.org/meta-share/meta-share/entry</
↪ms:sizeUnit>
        </ms:size>
        <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Xml</
↪ms:dataFormat>
      </ms:distributionTextFeature>
      <ms:licenceTerms>
        <ms:licenceTermsName xml:lang="en">Creative Commons
↪Attribution Non
        Commercial No Derivatives 4.0 International</
↪ms:licenceTermsName>
        <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc-nd/4.0/legalcode</ms:licenceTermsURL>
        <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
↪nc-nd/4.0/</ms:licenceTermsURL>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪nonCommercialUse</ms:conditionOfUse>
        <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪noDerivatives</ms:conditionOfUse>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
<ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
<ms:LicenceIdentifier
  ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
  >CC-BY-NC-ND-4.0</ms:LicenceIdentifier>
</ms:licenceTerms>
<ms:attributionText xml:lang="el">Τρίγλωσσο Ορολογικό Λεξικό
↪(ΤΟΛ) .
  Δημιουργός: Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↪Κέντρο Αθηνά.
  Άδεια: Creative Commons Attribution Non Commercial No
↪Derivatives 4.0
  International
  (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode,
  https://creativecommons.org/licenses/by-nc-nd/4.0/). Πηγή:
  http://hdl.handle.net/11500/ATHENA-0000-0000-5837-0
  (CLARIN:EL) </ms:attributionText>
<ms:attributionText xml:lang="en">Trilingual Term Dictionary
↪(TTD) by
  Institute for Language and Speech Processing - Athena
↪Research Center
  used under Creative Commons Attribution Non Commercial No
↪Derivatives
  4.0 International
  (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode,
  https://creativecommons.org/licenses/by-nc-nd/4.0/). Source:
  http://hdl.handle.net/11500/ATHENA-0000-0000-5837-0
  (CLARIN:EL) </ms:attributionText>
</ms:DatasetDistribution>
<ms:personalDataIncluded>>false</ms:personalDataIncluded>
<ms:sensitiveDataIncluded>>false</ms:sensitiveDataIncluded>
</ms:LexicalConceptualResource>
</ms:LRS subclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```

26.3 3. Εργαλεία/Υπηρεσίες

26.3.1 Μεμονωμένο εργαλείο

```

<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
↪gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2015-09-14</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
  <ms:actorType>Person</ms:actorType>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:surname xml:lang="en">Person_Surname</ms:surname>
<ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">Αναγνωριστής γλώσσας ψηφιακού κειμένου
      IEA</ms:resourceName>
    <ms:resourceName xml:lang="en">ILSP Language Identification System</
↪ms:resourceName>
    <ms:resourceShortName xml:lang="en">ILSP LangId</ms:resourceShortName>
    <ms:description xml:lang="el">Ο Αναγνωριστής γλώσσας του IEA είναι ένα
↪εργαλείο που
      χρησιμοποιείται για την αυτόματη αναγνώριση της γλώσσας ενός ψηφιακού
↪κειμένου. Το
      εργαλείο αναγνωρίζει την ελληνική, αγγλική, γερμανική, γαλλική, ολλανδική
↪γλώσσα
      καθώς και για τα greeklish, ενώ δίνεται και η δυνατότητα προσθήκης
↪επιπλέον γλωσσών
      με την προσθήκη συμπληρωματικών αρχείων για κάθε γλώσσα.</ms:description>
    <ms:description xml:lang="en">The ILSP Language Identification System is a
↪tool used for
      language identification in digital texts. The tool performs language
↪identification
      for Greek, Greeklish, English, German, Dutch and French; it can also be
↪used for
      other languages upon provision of specific supplementary external files
↪for each new
      language.</ms:description>
    <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
      >http://hdl.handle.net/11500/ATHENA-0000-0000-23E7-4</ms:LRIdentifier>
    <ms:version>1.0.0 (automatically assigned)</ms:version>
    <ms:additionalInfo>
      <ms:email>person@phs.uoa.gr</ms:email>
    </ms:additionalInfo>
    <ms:contact>
      <ms:Person>
        <ms:actorType>Person</ms:actorType>
        <ms:surname xml:lang="en">Person_Surname</ms:surname>
        <ms:givenName xml:lang="en">Person_Name</ms:givenName>
      </ms:Person>
    </ms:contact>
    <ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
↪Κέντρο
      Αθηνά (2015). Αναγνωριστής γλώσσας ψηφιακού κειμένου IEA. Version 1.0.0

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

(automatically assigned). [Software (Tool/Service)]. CLARIN:EL.
http://hdl.handle.net/11500/ATHENA-0000-0000-23E7-4</ms:citationText>
<ms:citationText xml:lang="en">Institute for Language and Speech Processing -
↪Athena
Research Center (2015). ILSP Language Identification System. Version 1.0.0
(automatically assigned). [Software (Tool/Service)]. CLARIN:EL.
http://hdl.handle.net/11500/ATHENA-0000-0000-23E7-4</ms:citationText>
<ms:keyword xml:lang="en">text</ms:keyword>
<ms:resourceCreator>
  <ms:Organization>
    <ms:actorType>Organization</ms:actorType>
    <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
      Λόγου</ms:organizationName>
    <ms:organizationName xml:lang="en">Institute for Language and Speech
      Processing</ms:organizationName>
    <ms:website>http://www.ilsp.gr/</ms:website>
  </ms:Organization>
</ms:resourceCreator>
<ms:intendedApplication>
  <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageTechnology</ms:LTCClassRecommended>
</ms:intendedApplication>
<ms:intendedApplication>
  <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageIdentification</ms:LTCClassRecommended>
</ms:intendedApplication>
<ms:isDocumentedBy>
  <ms:title xml:lang="el">Αναγνώριση γλώσσας ηλεκτρονικού κειμένου</
↪ms:title>
  <ms:title xml:lang="en">Language identification in digital texts</
↪ms:title>
  <ms:DocumentIdentifier
    ms:DocumentIdentifierScheme="http://purl.org/spar/datacite/url"
    >http://www.ilsp.gr/homepages/protopapas/pdf/Protopapas_2004_
↪LangIDsubm.pdf</ms:DocumentIdentifier>
  </ms:isDocumentedBy>
  <ms:LRS subclass>
    <ms:ToolService>
      <ms:lrType>ToolService</ms:lrType>
      <ms:function>
        <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LanguageIdentification</ms:LTCClassRecommended>
      </ms:function>
      <ms:SoftwareDistribution>
        <ms:SoftwareDistributionForm>http://w3id.org/meta-share/meta-
↪share/sourceCode</ms:SoftwareDistributionForm>
        <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
        <ms:licenceTerms>
          <ms:licenceTermsName xml:lang="en">BSD 2-Clause "Simplified"
            License</ms:licenceTermsName>
          <ms:licenceTermsURL>https://opensource.org/licenses/BSD-2-
↪Clause</ms:licenceTermsURL>
          <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:conditionOfUse>
          <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
        <ms:LicenceIdentifier
            ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
            >BSD-2-Clause</ms:LicenceIdentifier>
        </ms:licenceTerms>
        <ms:attributionText xml:lang="el">Αναγνωριστής γλώσσας ψηφιακού
↪κειμένου
        IEΛ. Δημιουργός: Ινστιτούτο Επεξεργασίας του Λόγου -
↪Ερευνητικό Κέντρο
        Αθηνά. Αδεια: BSD 2-Clause "Simplified" License
        (https://opensource.org/licenses/BSD-2-Clause). Πηγή:
        http://hdl.handle.net/11500/ATHENA-0000-0000-23E7-4
        (CLARIN:EL) </ms:attributionText>
        <ms:attributionText xml:lang="en">ILSP Language Identification
↪System by
        Institute for Language and Speech Processing - Athena
↪Research Center
        used under BSD 2-Clause "Simplified" License
        (https://opensource.org/licenses/BSD-2-Clause). Source:
        http://hdl.handle.net/11500/ATHENA-0000-0000-23E7-4
        (CLARIN:EL) </ms:attributionText>
    </ms:SoftwareDistribution>
    <ms:languageDependent>true</ms:languageDependent>
    <ms:inputContentResource>
        <ms:processingResourceType>http://w3id.org/meta-share/meta-share/
↪corpus</ms:processingResourceType>
        <ms:language>
            <ms:languageTag>el-Latn</ms:languageTag>
            <ms:languageId>el</ms:languageId>
            <ms:scriptId>Latn</ms:scriptId>
            <ms:languageVarietyName xml:lang="en">Greeklish</
↪ms:languageVarietyName>
        </ms:language>
        <ms:language>
            <ms:languageTag>el-Grek</ms:languageTag>
            <ms:languageId>el</ms:languageId>
            <ms:scriptId>Grek</ms:scriptId>
        </ms:language>
        <ms:language>
            <ms:languageTag>fr</ms:languageTag>
            <ms:languageId>fr</ms:languageId>
        </ms:language>
        <ms:language>
            <ms:languageTag>en</ms:languageTag>
            <ms:languageId>en</ms:languageId>
        </ms:language>
        <ms:language>
            <ms:languageTag>de</ms:languageTag>
            <ms:languageId>de</ms:languageId>
        </ms:language>
        <ms:language>
            <ms:languageTag>nl</ms:languageTag>
            <ms:languageId>nl</ms:languageId>
        </ms:language>
    </ms:language>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
  </ms:inputContentResource>
  <ms:outputResource>
    <ms:processingResourceType>http://w3id.org/meta-share/meta-share/
↪corpus</ms:processingResourceType>
    <ms:language>
      <ms:languageTag>el-Latn</ms:languageTag>
      <ms:languageId>el</ms:languageId>
      <ms:scriptId>Latn</ms:scriptId>
      <ms:languageVarietyName xml:lang="en">Greeklish</
↪ms:languageVarietyName>
    </ms:language>
    <ms:language>
      <ms:languageTag>el-Grek</ms:languageTag>
      <ms:languageId>el</ms:languageId>
      <ms:scriptId>Grek</ms:scriptId>
    </ms:language>
    <ms:language>
      <ms:languageTag>fr</ms:languageTag>
      <ms:languageId>fr</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>en</ms:languageTag>
      <ms:languageId>en</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>de</ms:languageTag>
      <ms:languageId>de</ms:languageId>
    </ms:language>
    <ms:language>
      <ms:languageTag>nl</ms:languageTag>
      <ms:languageId>nl</ms:languageId>
    </ms:language>
    <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
  </ms:outputResource>
  <ms:evaluated>>false</ms:evaluated>
</ms:ToolService>
</ms:LRSubclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>

```

26.3.2 Συνδυασμοί εργαλείων

```

<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.
↪clarin.gr/metadata-schema/CLARIN-SHARE.xsd">
  <ms:metadataCreationDate>2019-04-02</ms:metadataCreationDate>
  <ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
  <ms:metadataCurator>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:actorType>Person</ms:actorType>
<ms:surname xml:lang="en">Person_Surname</ms:surname>
<ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="en">Voyant Tools</ms:resourceName>
    <ms:description xml:lang="el">Τα Voyant Tools είναι ένα διαδικτυακό
    ↪ περιβάλλον ανοιχτού
      κώδικα που χρησιμοποιείται για την ανάγνωση και ανάλυση κειμένου ή/και
    ↪ σωμάτων
      κειμένων. Το περιβάλλον των Voyant Tools σχεδιάστηκε και υλοποιήθηκε από
    ↪ τους Stéfan
      Sinclair και Geoffrey Rockwell το 2016 με στόχο να διευκολύνει τους
    ↪ φοιτητές και
      ακαδημαϊκούς των ψηφιακών ανθρωπιστικών επιστημών αναφορικά με τις
    ↪ πρακτικές
      ανάγνωσης και ερμηνείας κειμενικού υλικού. Συγκεκριμένα, μπορείτε να
    ↪ χρησιμοποιήσετε
      τα Voyant Tools για: - να δείτε πώς επιτυγχάνεται η κειμενική ανάλυση με
    ↪ τη βοήθεια
      υπολογιστικών εργαλείων, - να μελετήσετε κείμενα που βρίσκετε στο
    ↪ διαδίκτυο ή
      κείμενα που έχετε επεξεργαστεί και τα οποία βρίσκονται αποθηκευμένα
    ↪ τοπικά στον Η/Υ
      σας, - να προσθέσετε επιπλέον λειτουργικότητες στις διαδικτυακές σας
    ↪ συλλογές,
      περιοδικά, ιστολόγια ή ιστότοπους, ώστε να μπορεί κανείς να δει το υλικό
    ↪ αυτό με τη
      βοήθεια εργαλείων ανάλυσης, - να προσθέσετε διαδραστικά στοιχεία στα
    ↪ κείμενα ή στα
      άρθρα που δημοσιεύετε στο διαδίκτυο ή ακόμα και να προσθέσετε διαδραστικά
    ↪ πάνελ στις
      ερευνητικές σας εκθέσεις (εφόσον αυτές μπορούν να δημοσιευθούν στο
    ↪ διαδίκτυο), ώστε
      οι αναγνώστες σας να μπορούν πολύ γρήγορα να βρουν τα βασικά σημεία και τα
      αποτελέσματα της έρευνάς σας, - να αναπτύξετε τα δικά σας εργαλεία
    ↪ χρησιμοποιώντας
      τις λειτουργικότητες και τον κώδικα των Voyant Tools.</ms:description>
    <ms:description xml:lang="en">Voyant Tools is a web-based text reading and
    ↪ analysis
      environment. It is a scholarly project that is designed to facilitate
    ↪ reading and
      interpretive practices for digital humanities students and scholars as
    ↪ well as for

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

the general public. What you can do with Voyant: --Use it to learn how
computers-assisted analysis works. Check out our examples that show you
↳how to do real academic tasks with Voyant. --Use it to study texts that you find on
↳the web or texts that you have carefully edited and have on your computer. --Use it
↳to add functionality to your online collections, journals, blogs or web sites so
↳others can see through your texts with analytical tools. --Use it to add interactive
↳evidence to your essays that you publish online. Add interactive panels right into
↳your research essays (if they can be published online) so your readers can
↳recapitulate your results. --Use it to develop your own tools using our functionality
↳and
code.</ms:description>
<ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
>http://hdl.handle.net/11500/ATHENA-0000-0000-5827-2</ms:LRIdentifier>
<ms:version>1.0.0 (automatically assigned)</ms:version>
<ms:additionalInfo>
  <ms:landingPage>https://voyant-tools.org/</ms:landingPage>
</ms:additionalInfo>
<ms:contact>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:contact>
<ms:contact>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:contact>
<ms:citationText xml:lang="en">Rockwell, Geoffrey; Sinclair, Stéfan (2019).
↳Voyant Tools. Version 1.0.0 (automatically assigned). [Software (Tool/Service)].
↳CLARIN:EL.
http://hdl.handle.net/11500/ATHENA-0000-0000-5827-2</ms:citationText>
<ms:keyword xml:lang="en">text</ms:keyword>
<ms:resourceCreator>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:resourceCreator>
<ms:resourceCreator>
  <ms:Person>
    <ms:actorType>Person</ms:actorType>
    <ms:surname xml:lang="en">Person_Surname</ms:surname>
    <ms:givenName xml:lang="en">Person_Name</ms:givenName>
  </ms:Person>
</ms:resourceCreator>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

    </ms:Person>
  </ms:resourceCreator>
  <ms:LRS subclass>
    <ms:ToolService>
      <ms:lType>ToolService</ms:lType>
      <ms:function>
        <ms:LTCClassRecommended>http://w3id.org/meta-share/omtd-share/
↪LinguisticAnalysis</ms:LTCClassRecommended>
      </ms:function>
      <ms:SoftwareDistribution>
        <ms:SoftwareDistributionForm>http://w3id.org/meta-share/meta-
↪share/webService</ms:SoftwareDistributionForm>
        <ms:executionLocation>https://voyant-tools.org/</
↪ms:executionLocation>
        <ms:webServiceType>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:webServiceType>
        <ms:licenceTerms>
          <ms:licenceTermsName xml:lang="en">GNU General Public License
↪v3.0 or
            later</ms:licenceTermsName>
          <ms:licenceTermsURL>https://www.gnu.org/licenses/gpl-3.0-
↪standalone.html</ms:licenceTermsURL>
          <ms:licenceTermsURL>https://opensource.org/licenses/GPL-3.0</
↪ms:licenceTermsURL>
          <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪unspecified</ms:conditionOfUse>
          <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
          <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
          <ms:LicenceIdentifier
            ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
            >GPL-3.0-or-later</ms:LicenceIdentifier>
          </ms:licenceTerms>
          <ms:licenceTerms>
            <ms:licenceTermsName xml:lang="en">Creative Commons
↪Attribution 4.0
              International</ms:licenceTermsName>
            <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/legalcode</ms:licenceTermsURL>
            <ms:licenceTermsURL>https://creativecommons.org/licenses/by/4.
↪0/</ms:licenceTermsURL>
            <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
↪attribution</ms:conditionOfUse>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsDirectAccess</ms:licenceCategory>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪allowsProcessing</ms:licenceCategory>
            <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
↪public</ms:licenceCategory>
            <ms:LicenceIdentifier
              ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
↪meta-share/SPDX"
              >CC-BY-4.0</ms:LicenceIdentifier>
            </ms:licenceTerms>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:attributionText xml:lang="el">Voyant Tools. Δημιουργός:↵
↵Geoffrey
Rockwell and Stéfan Sinclair. Άδεια: GNU General Public↵
↵License v3.0 or
later (https://www.gnu.org/licenses/gpl-3.0-standalone.html,
https://opensource.org/licenses/GPL-3.0) and Creative Commons
Attribution 4.0 International
(https://creativecommons.org/licenses/by/4.0/legalcode,
https://creativecommons.org/licenses/by/4.0/). Πηγή:
http://hdl.handle.net/11500/ATHENA-0000-0000-5827-2
(CLARIN:EL) </ms:attributionText>
<ms:attributionText xml:lang="en">Voyant Tools by Geoffrey↵
↵Rockwell and
Stéfan Sinclair used under GNU General Public License v3.0 or↵
↵later
(https://www.gnu.org/licenses/gpl-3.0-standalone.html,
https://opensource.org/licenses/GPL-3.0) and Creative Commons
Attribution 4.0 International
(https://creativecommons.org/licenses/by/4.0/legalcode,
https://creativecommons.org/licenses/by/4.0/). Source:
http://hdl.handle.net/11500/ATHENA-0000-0000-5827-2
(CLARIN:EL) </ms:attributionText>
</ms:SoftwareDistribution>
<ms:languageDependent>false</ms:languageDependent>
<ms:inputContentResource>
  <ms:processingResourceType>http://w3id.org/meta-share/meta-share/
↵corpus</ms:processingResourceType>
  <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↵ms:mediaType>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Pdf</
↵ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Rtf</
↵ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Xml</
↵ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/ConllU</
↵ms:dataFormat>
  <ms:dataFormat>http://w3id.org/meta-share/omtd-share/Html</
↵ms:dataFormat>
</ms:inputContentResource>
<ms:outputResource>
  <ms:processingResourceType>http://w3id.org/meta-share/meta-share/
↵corpus</ms:processingResourceType>
  <ms:mediaType>http://w3id.org/meta-share/meta-share/image</
↵ms:mediaType>
</ms:outputResource>
<ms:outputResource>
  <ms:processingResourceType>http://w3id.org/meta-share/meta-share/
↵corpus</ms:processingResourceType>
  <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↵ms:mediaType>
</ms:outputResource>
<ms:evaluated>false</ms:evaluated>
</ms:ToolService>
</ms:LRSubclass>
</ms:LanguageResource>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```
</ms:DescribedEntity>
</ms:MetadataRecord>
```

26.4 4. Γλωσσικές περιγραφές

```
<?xml version="1.0" encoding="utf-8"?>
<ms:MetadataRecord xmlns:ms="http://w3id.org/meta-share/meta-share/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://w3id.org/meta-share/meta-share/ https://inventory.clarin.
gr/metadata-schema/CLARIN-SHARE.xsd">
<ms:metadataCreationDate>2015-09-10</ms:metadataCreationDate>
<ms:metadataLastDateUpdated>2021-05-28</ms:metadataLastDateUpdated>
<ms:metadataCurator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCurator>
<ms:compliesWith>http://w3id.org/meta-share/meta-share/CLARIN-SHARE</ms:compliesWith>
<ms:metadataCreator>
  <ms:actorType>Person</ms:actorType>
  <ms:surname xml:lang="en">Person_Surname</ms:surname>
  <ms:givenName xml:lang="en">Person_Name</ms:givenName>
</ms:metadataCreator>
<ms:sourceOfMetadataRecord>
  <ms:repositoryName xml:lang="el">Αποθετήριο ΕΚ ΑΘΗΝΑ</ms:repositoryName>
  <ms:repositoryName xml:lang="en">ATHENA RC Repository</ms:repositoryName>
  <ms:repositoryURL>http://inventory.clarin.gr</ms:repositoryURL>
</ms:sourceOfMetadataRecord>
<ms:DescribedEntity>
  <ms:LanguageResource>
    <ms:entityType>LanguageResource</ms:entityType>
    <ms:resourceName xml:lang="el">PANACEA σώμα ελληνικών n-γραμμάτων (n-grams)
    περιβαλλοντικού τομέα</ms:resourceName>
    <ms:resourceName xml:lang="en">PANACEA Environment Corpus n-grams EL
    (Greek)</ms:resourceName>
    <ms:description xml:lang="el">Ο συγκεκριμένος πόρος περιλαμβάνει ελληνικά n-
    γράμματα
    (για n = 1 - 5) λέξεων και n-γράμματα λέξης/μορφοσυντακτικού
    <ms:description>
      <ms:description xml:lang="en">PANACEA Environment Corpus n-grams EL (Greek) 1.
      0 contains
```

<ms:resourceName xml:lang="en">PANACEA Environment Corpus n-grams EL
 (Greek)</ms:resourceName>
 <ms:description xml:lang="el">Ο συγκεκριμένος πόρος περιλαμβάνει ελληνικά n-
 γράμματα
 (για n = 1 - 5) λέξεων και n-γράμματα λέξης/μορφοσυντακτικού
 χαρακτηρισμού/λήμματος
 στον τομέα του περιβάλλοντος, τα οποία συνοδεύονται από τις αντίστοιχες
 συχνότητες εμφάνισης στο σώμα κειμένων. Ο πόρος αναπτύχθηκε στο πλαίσιο του έργου
 PANACEA
 (http://www.panacea-lr.eu), που χρηματοδοτήθηκε από το 7ο ΠΠ. Βασίστηκε σε
 ιστοσελίδες που ανακτήθηκαν αυτόματα από το διαδίκτυο με χρήση του
 εργαλείου FPC,
 αφού αναγνωρίστηκε αυτόματα ότι τα κείμενα είναι γραμμένα στην ελληνική
 και είναι
 σχετικά με το περιβάλλον. Το κειμενικό σώμα αποτελείται από περίπου 31,71
 εκατ.
 μονάδες. Η συλλογή των δεδομένων έγινε το καλοκαίρι του 2011.</ms:
 description>
 <ms:description xml:lang="en">PANACEA Environment Corpus n-grams EL (Greek) 1.
 0 contains

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

Greek word n-grams and Greek word/tag/lemma n-grams in the "Environment"
 ↳ (ENV) domain. N-grams are accompanied by their observed frequency counts. The
 ↳ length of the n-grams ranges from unigrams (single words) to five-grams. The data
 ↳ were collected in the context of PANACEA (<http://www.panacea-lr.eu>), an EU-FP7
 ↳ Funded Project under Grant Agreement 248064. The n-gram counts were generated
 ↳ from crawled Web pages that were automatically detected to be in the Greek language
 ↳ and were automatically classified as relevant to the ENV domain. The collection
 ↳ consisted of approximately 31.71 million tokens. Data collection took place in the
 ↳ summer of 2011. </ms:description>
 <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/handle"
 >http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3</ms:LRIdentifier>
 <ms:version>1.0</ms:version>
 <ms:additionalInfo>
 <ms:landingPage>http://nlp.ilsp.gr/panacea/D4.3/data/201209/gms/env_el/
 ↳ README.txt</ms:landingPage>
 </ms:additionalInfo>
 <ms:contact>
 <ms:Person>
 <ms:actorType>Person</ms:actorType>
 <ms:surname xml:lang="en">Person_Surname</ms:surname>
 <ms:givenName xml:lang="en">Person_Name</ms:givenName>
 </ms:Person>
 </ms:contact>
 <ms:contact>
 <ms:Person>
 <ms:actorType>Person</ms:actorType>
 <ms:surname xml:lang="en">Person_Surname</ms:surname>
 <ms:givenName xml:lang="en">Person_Name</ms:givenName>
 </ms:Person>
 </ms:contact>
 <ms:citationText xml:lang="el">Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό
 ↳ Κέντρο
 Αθηνά (2015). PANACEA σώμα ελληνικών n-γραμμάτων (n-grams)
 ↳ περιβαλλοντικού τομέα.
 Version 1.0. [Model (n-gram model)]. CLARIN:EL.
 <ms:citationText><http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3></ms:citationText>
 <ms:citationText xml:lang="en">Institute for Language and Speech Processing -
 ↳ Athena
 Research Center (2015). PANACEA Environment Corpus n-grams EL (Greek).
 ↳ Version 1.0.
 [Model (n-gram model)]. CLARIN:EL.
 <ms:citationText><http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3></ms:citationText>
 <ms:keyword xml:lang="en">monolingual</ms:keyword>
 <ms:domain>
 <ms:categoryLabel xml:lang="en">environment</ms:categoryLabel>
 <ms:DomainIdentifier
 ms:DomainClassificationScheme="http://w3id.org/meta-share/meta-share/
 ↳ ClarinEL_domainClassification"

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

        >Clarín_Domain002</ms:DomainIdentifier>
    </ms:domain>
    <ms:resourceCreator>
        <ms:Organization>
            <ms:actorType>Organization</ms:actorType>
            <ms:organizationName xml:lang="el">Ινστιτούτο Επεξεργασίας του
                Λόγου</ms:organizationName>
            <ms:organizationName xml:lang="en">Institute for Language and Speech
                Processing</ms:organizationName>
            <ms:website>http://www.ilsp.gr/</ms:website>
        </ms:Organization>
    </ms:resourceCreator>
    <ms:creationStartDate>2011-06-01</ms:creationStartDate>
    <ms:creationEndDate>2011-08-31</ms:creationEndDate>
    <ms:fundingProject>
        <ms:projectName xml:lang="en">Platform for Automatic, Normalized_
↪Annotation and
            Cost-Effective Acquisition of Language Resources for Human
            Language</ms:projectName>
        <ms:website>http://www.panacea-lr.eu</ms:website>
        <ms:website>http://panacea-lr.eu/</ms:website>
        <ms:grantNumber>ICT-248064</ms:grantNumber>
        <ms:fundingType>http://w3id.org/meta-share/meta-share/euFunds</
↪ms:fundingType>
        <ms:funder>
            <ms:Organization>
                <ms:actorType>Organization</ms:actorType>
                <ms:organizationName xml:lang="el">Ευρωπαϊκή Επιτροπή</
↪ms:organizationName>
                <ms:organizationName xml:lang="en">European Commission</
↪ms:organizationName>
                <ms:website>https://ec.europa.eu/info/index_en</ms:website>
            </ms:Organization>
        </ms:funder>
    </ms:fundingProject>
    <ms:creationDetails xml:lang="en">automatic web crawling, automatic language_
↪detection,
        data preprocessing (boilerpipe filtering, lemmatization &
        tagging)</ms:creationDetails>
    <ms:isCreatedBy>
        <ms:resourceName xml:lang="en">boilerpipe library</ms:resourceName>
        <ms:LRIdentifier ms:LRIdentifierScheme="http://purl.org/spar/datacite/url"
            >https://code.google.com/archive/p/boilerpipe/</ms:LRIdentifier>
        <ms:version>unspecified</ms:version>
    </ms:isCreatedBy>
    <ms:isCreatedBy>
        <ms:resourceName xml:lang="en">ILSP Lemmatizer</ms:resourceName>
        <ms:version>unspecified</ms:version>
    </ms:isCreatedBy>
    <ms:isCreatedBy>
        <ms:resourceName xml:lang="en">ILSP Feature-based multi-tiered POS
            Tagger</ms:resourceName>
        <ms:version>unspecified</ms:version>
    </ms:isCreatedBy>
    <ms:isDocumentedBy>
        <ms:title xml:lang="en">PANACEA Environment Corpus n-grams EL 1.0 README</
↪ms:title>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:DocumentIdentifier
  ms:DocumentIdentifierScheme="http://purl.org/spar/datacite/url"
  >http://nlp.ilsp.gr/panacea/D4.3/data/201209/gms/env_el/README.txt</
↪ms:DocumentIdentifier>
  </ms:isDocumentedBy>
  <ms:LRSubclass>
    <ms:LanguageDescription>
      <ms:lrType>LanguageDescription</ms:lrType>
      <ms:LanguageDescriptionSubclass>
        <ms:NGramModel>
          <ms:ldSubclassType>NGramModel</ms:ldSubclassType>
          <ms:baseItem>http://w3id.org/meta-share/meta-share/word</
↪ms:baseItem>
            <ms:order>5</ms:order>
          </ms:NGramModel>
        </ms:LanguageDescriptionSubclass>
        <ms:LanguageDescriptionMediaPart>
          <ms:LanguageDescriptionTextPart>
            <ms:ldMediaType>LanguageDescriptionTextPart</ms:ldMediaType>
            <ms:mediaType>http://w3id.org/meta-share/meta-share/text</
↪ms:mediaType>
              <ms:lingualityType>http://w3id.org/meta-share/meta-share/
↪monolingual</ms:lingualityType>
              <ms:language>
                <ms:languageTag>el</ms:languageTag>
                <ms:languageId>el</ms:languageId>
              </ms:language>
            </ms:LanguageDescriptionTextPart>
          </ms:LanguageDescriptionMediaPart>
          <ms:DatasetDistribution>
            <ms:DatasetDistributionForm>http://w3id.org/meta-share/meta-share/
↪downloadable</ms:DatasetDistributionForm>
            <ms:downloadLocation>http://www.hiddenLocation.org</
↪ms:downloadLocation>
              <ms:samplesLocation>http://nlp.ilsp.gr/panacea/D4.3/data/201209/
↪gms/env_el/ENV_EL_1000.3gms.sample</ms:samplesLocation>
              <ms:samplesLocation>http://nlp.ilsp.gr/panacea/D4.3/data/201209/
↪gms/env_el/ENV_EL_wpl_1000.3gms.sample</ms:samplesLocation>
              <ms:distributionTextFeature>
                <ms:size>
                  <ms:amount>14954020.0</ms:amount>
                  <ms:sizeUnit>http://w3id.org/meta-share/meta-share/five-
↪gram</ms:sizeUnit>
                </ms:size>
                <ms:size>
                  <ms:amount>13683940.0</ms:amount>
                  <ms:sizeUnit>http://w3id.org/meta-share/meta-share/four-
↪gram</ms:sizeUnit>
                </ms:size>
                <ms:size>
                  <ms:amount>3860716.0</ms:amount>
                  <ms:sizeUnit>http://w3id.org/meta-share/meta-share/bigram
↪</ms:sizeUnit>
                </ms:size>
                <ms:size>
                  <ms:amount>9767383.0</ms:amount>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```

<ms:sizeUnit>http://w3id.org/meta-share/meta-share/trigram
</ms:sizeUnit>
</ms:size>
<ms:size>
  <ms:amount>435189.0</ms:amount>
  <ms:sizeUnit>http://w3id.org/meta-share/meta-share/unigram
</ms:sizeUnit>
</ms:size>
<ms:dataFormat>http://w3id.org/meta-share/omtd-share/Text</
</ms:distributionTextFeature>
<ms:licenceTerms>
  <ms:licenceTermsName xml:lang="en">Creative Commons
Attribution Share
  Alike 4.0 International</ms:licenceTermsName>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
sa/4.0/legalcode</ms:licenceTermsURL>
  <ms:licenceTermsURL>https://creativecommons.org/licenses/by-
sa/4.0/</ms:licenceTermsURL>
  <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
attribution</ms:conditionOfUse>
  <ms:conditionOfUse>http://w3id.org/meta-share/meta-share/
shareAlike</ms:conditionOfUse>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
allowsDirectAccess</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
allowsProcessing</ms:licenceCategory>
  <ms:licenceCategory>http://w3id.org/meta-share/meta-share/
public</ms:licenceCategory>
  <ms:LicenceIdentifier
    ms:LicenceIdentifierScheme="http://w3id.org/meta-share/
meta-share/SPDX"
    >CC-BY-SA-4.0</ms:LicenceIdentifier>
  </ms:licenceTerms>
  <ms:attributionText xml:lang="el">PANACEA σώμα ελληνικών n-
γραμμάτων
  (n-grams) περιβαλλοντικού τομέα. Δημιουργός: Ινστιτούτο
Επεξεργασίας του
Λόγου - Ερευνητικό Κέντρο Αθηνά. Άδεια: Creative Commons
Attribution
  Share Alike 4.0 International
  (https://creativecommons.org/licenses/by-sa/4.0/legalcode,
  https://creativecommons.org/licenses/by-sa/4.0/). Πηγή:
  http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3
  (CLARIN:EL) </ms:attributionText>
  <ms:attributionText xml:lang="en">PANACEA Environment Corpus n-
grams EL
  (Greek) by Institute for Language and Speech Processing -
Athena
  Research Center used under Creative Commons Attribution Share
Alike 4.0
  International (https://creativecommons.org/licenses/by-sa/4.0/
legalcode,
  https://creativecommons.org/licenses/by-sa/4.0/). Source:
  http://hdl.handle.net/11500/ATHENA-0000-0000-23DA-3
  (CLARIN:EL) </ms:attributionText>

```

(συνέχεια στην επόμενη σελίδα)

(συνεχίζεται από την προηγούμενη σελίδα)

```
        </ms:DatasetDistribution>
        <ms:personalDataIncluded>false</ms:personalDataIncluded>
        <ms:sensitiveDataIncluded>false</ms:sensitiveDataIncluded>
    </ms:LanguageDescription>
</ms:LRS subclass>
</ms:LanguageResource>
</ms:DescribedEntity>
</ms:MetadataRecord>
```

Το σχήμα μεταδεδομένων

Στο σχήμα μεταδεδομένων ορίζονται τα δομικά στοιχεία των μεταδεδομένων, τα **χαρακτηριστικά** (attributes) και **στοιχεία** (elements), ο βαθμός αναγκαιότητάς τους για την περιγραφή (π.χ. *υποχρεωτικά*, υπό συνθήκη υποχρεωτικά, προαιρετικά κτλ.), η επαναληψιμότητά τους, οι επιτρεπτές τιμές για τη συμπλήρωσή τους, καθώς και άλλοι κανόνες και περιορισμοί. Το σχήμα παρουσιάζεται αναλυτικά εδώ:

- Εγγραφή Μεταδεδομένων (Metadata record, Base item)
- Γλωσσικό Πόρος (Language Resource)
 - Εργαλείο/Υπηρεσία (Tool/Service)
 - Σώμα κειμένων (Corpus)
 - Γλωσσική περιγραφή (Language description)
 - Λεξικός/Εννοιολογικός πόρος (Lexical/Conceptual resource)

Όλα τα **χαρακτηριστικά** (attributes) και **στοιχεία** (elements) παρουσιάζονται σε αλφαβητική σειρά.

CLARIN:el XML Schema Documentation

Table of Contents

- [Schema Document Properties](#)
- [Global Declarations](#)
 - [Attribute: AccessRightsStatementScheme](#)
 - [Attribute: AudioGenreClassificationScheme](#)
 - [Attribute: ClassificationScheme](#)
 - [Attribute: DocumentIdentifierScheme](#)
 - [Attribute: DomainClassificationScheme](#)
 - [Attribute: FunderIdentifierScheme](#)
 - [Attribute: IdentifierScheme](#)
 - [Attribute: ImageGenreClassificationScheme](#)
 - [Attribute: LicenceIdentifierScheme](#)
 - [Attribute: LRIdentifierScheme](#)
 - [Attribute: MetadataRecordIdentifierScheme](#)
 - [Attribute: OrganizationIdentifierScheme](#)
 - [Attribute: PersonalIdentifierScheme](#)
 - [Attribute: ProjectIdentifierScheme](#)
 - [Attribute: RepositoryIdentifierScheme](#)
 - [Attribute: ResourceIdentifierScheme](#)
 - [Attribute: socialMediaOccupationalAccountType](#)
 - [Attribute: SpeechGenreClassificationScheme](#)
 - [Attribute: SubjectClassificationScheme](#)
 - [Attribute: TextGenreClassificationScheme](#)
 - [Attribute: TextTypeClassificationScheme](#)
 - [Attribute: VideoGenreClassificationScheme](#)
 - [Element: abstract](#)
 - [Element: accesses](#)
 - [Element: accessLocation](#)
 - [Element: accessRights](#)
 - [Element: AccessRightsStatementIdentifier](#)
 - [Element: Actor](#)
 - [Element: actualUse](#)
 - [Element: actualUseDetails](#)
 - [Element: additionalHWRequirements](#)
 - [Element: additionalInfo](#)
 - [Element: address](#)
 - [Element: addressSet](#)
 - [Element: affiliatesOrganization](#)
 - [Element: affiliation](#)
 - [Element: age](#)
 - [Element: ageGroup](#)
 - [Element: ageGroupOfParticipants](#)
 - [Element: ageRangeEndOfParticipants](#)
 - [Element: ageRangeStartOfParticipants](#)
 - [Element: algorithm](#)
 - [Element: algorithmDetails](#)
 - [Element: alias](#)
 - [Element: alternativeTitle](#)
 - [Element: amount](#)
 - [Element: annotatedElement](#)
 - [Element: annotation](#)
 - [Element: annotationEndDate](#)
 - [Element: annotationMode](#)
 - [Element: annotationModeDetails](#)
 - [Element: annotationReport](#)

Αν θέλετε να μάθετε περισσότερα για κάποιο στοιχείο (element), κάντε κλικ πάνω του και θα δείτε την πλήρη περιγραφή του. Στο ακόλουθο παράδειγμα, παρουσιάζονται δύο στοιχεία, τα **download location** (τοποθεσία λήψης/καταφόρτωσης) and **duration of audio** (διάρκεια ηχητικού δεδομένου).

Element: **downloadLocation**

Name	downloadLocation
Type	ms: httpURL
Nullable	no
Abstract	no
Documentation	A URL where the language resource (mainly data but also downloadable software programmes or forms) can be downloaded from
Application Data	<identifier> http://w3id.org/meta-share/meta-share/downloadLocation </identifier> <label xml:lang="en"> download location </label> <note xml:lang="en"> rule: must be filled in when SoftwareDistributionForm = downloadable; must not be filled in for web services; optional in all other cases </note>

XML Instance Representation

```
<ms:downloadLocation ms:httpURL "</ms:downloadLocation">
```

Schema Component Representation

```
<xsl:element name="downloadLocation" type="ms:httpURL"/>
```

Element: **durationOfAudio**

Name	durationOfAudio
Type	ms: Duration
Nullable	no
Abstract	no
Documentation	Specifies the duration of the audio recording including silences, music, pauses, etc.
Application Data	<identifier> http://w3id.org/meta-share/meta-share/durationOfAudio </identifier> <label xml:lang="en"> duration of audio </label>

XML Instance Representation

```
<ms:durationOfAudio>  
  <ms:amount> ... </ms:amount> [1]  
  <ms:durationUnit> ... </ms:durationUnit> [1]  
</ms:durationOfAudio>
```

Schema Component Representation

```
<xsl:element name="durationOfAudio" type="ms:Duration"/>
```

Τι ενέργειες μπορώ να κάνω πάνω στους πόρους;

28.1 I. Ταξινόμηση ενεργειών βάσει του σταδίου στο οποίο βρίσκεται ο πόρος & του τύπου χρήστη

Για να δείτε τις **ενέργειες** (actions) που μπορείτε να κάνετε πάνω σε κάθε έναν από τους πόρους στους οποίους έχετε κάποια μορφή δικαιοδοσίας, πρέπει να μεταβείτε στο *dashboard* (πίνακας ελέγχου). Εκεί επιλέξτε:

- *View my resources* (Προβολή των πόρων μου) από το *My resources* (Οι πόροι μου), αν είστε *επιμελητής* ή έχετε ταυτόχρονα και κάποιον άλλο ρόλο (*υπεύθυνος αποθετηρίου* ή *επικυρωτής*) αλλά θέλετε να δείτε **τους πόρους που εσείς δημιουργήσατε** (C, curator),
- *View my supervision tasks* (Προβολή των εργασιών που επιβλέπω) από το *My repository* (Το αποθετήριό μου), αν είστε *υπεύθυνος αποθετηρίου* (S) - σε αυτή τη λίστα βλέπετε τους πόρους που έχουν δημιουργήσει όλοι *επιμελητές* του αποθετηρίου σας,
- *View my validation tasks* (Προβολή των εργασιών επικύρωσής) από το *Validation tasks*, αν είστε *επικυρωτής* (V).

Η λίστα των πόρων χωρίζεται σε νοητές γραμμές κάθε μία εκ των οποίων είναι αφιερωμένη σε έναν μόνο πόρο και έχει ένα κουμπί **Actions** (Ενέργειες), όπως φαίνεται στην εικόνα. Όταν κάνετε κλικ σε αυτό, θα δείτε μια αναπτυσσόμενη λίστα που περιέχει όλες τις επιτρεπόμενες ενέργειες ανά πόρο.

Lexical/Conceptual Resource (Mandatory elements)

<input type="checkbox"/>	1.0.0 (automatically assigned)	curator	curator_hosted@gmail.com	Actions ▼ syntactically valid Edit Metadata Copy record Submit for publication Delete Metadata
	Lexical/Conceptual resource	supervisor	supervisor_hosted@gmail.com	
	Hosted Resources Repository			
	Created: 30 May 2021			
	Updated: 09 June 2021			

Testrecognizer

<input type="checkbox"/>	Tool/Service	curator	curator_hosted@gmail.com	draft syntactically valid
	Hosted Resources Repository			
	Created: 08 June 2021			
	Updated: 08 June 2021			

democorpus

<input type="checkbox"/>	1.0.0 (automatically assigned)	curator	curator_hosted@gmail.com	syntactically valid
	Corpus	supervisor	supervisor_hosted@gmail.com	
	Hosted Resources Repository			
	Created: 08 June 2021			
	Updated: 08 June 2021			

has data

Εναλλακτικά, μπορείτε να επιλέξετε έναν πόρο κάνοντας κλικ στο όνομά του και να δείτε τις διαθέσιμες ενέργειες από τη *σελίδα θέασης* του¹.

draft syntactically valid submitted approved published

Lexical/Conceptual Resource (Mandatory elements)

LexicalConceptualResource

Version: 1.0.0 (automatically assigned)

This is a resource with only the mandatory (and mandatory upon condition) elements filled in.

Language English Modern Greek (1453-)	Keyword mandatory, testing
----------------------------------------------------	--------------------------------------

Overview Access

clarin:el

Actions ▼

Edit Metadata
Copy record
Submit for publication
Delete Metadata

Οι ενέργειες που βλέπετε κάθε φορά είναι οι ενέργειες **που εσείς έχετε το δικαίωμα να εκτελέσετε** ανάλογα με τον *ρόλο* σας και το στάδιο στο οποίο βρίσκεται ο πόρος. Στον ακόλουθο πίνακα φαίνονται συγκεντρωτικά όλες οι ενέργειες που έχει δικαίωμα να κάνει κάθε τύπος χρήστη σε κάθε στάδιο του κύκλου ζωής του πόρου. Κάντε κλικ στην ενέργεια που σας ενδιαφέρει και θα μεταφερθείτε στην αντίστοιχη ενότητα για να μάθετε περισσότερα για αυτήν.

¹ Δεν θα μπορείτε να δείτε τη σελίδα θέασης του πόρου μέχρι να συμπληρώσετε όλα τα *υποχρεωτικά μεταδεδομένα*. Μόνο τότε η εγγραφή μεταδεδομένων, η οποία είναι συντακτικά έγκυρη, μπορεί να **αποθηκευτεί** (save) και να παρουσιαστεί ως σελίδα θέασης.

Action/Status	draft	synt.valid	submitted	approved	published	requested ^{Σελίδα 191, 2}	unpublished
<i>edit metadata</i>	C	C, S	S	S	n/a	n/a	S
<i>copy record</i>	n/a	C, S	C, S	C, S	C, S	C, S	C, S
<i>submit for publication</i>	n/a	C, S	n/a	n/a	n/a	n/a	n/a
<i>assign supervisor</i>	n/a	n/a	S	n/a	n/a	n/a	n/a
<i>assign validator</i>	n/a	n/a	S	n/a	n/a	n/a	n/a
<i>validate</i>	n/a	n/a	V	n/a	n/a	n/a	n/a
<i>publish</i>	n/a	n/a	n/a	S	n/a	n/a	S
<i>unpublish</i>	n/a	n/a	n/a	n/a	S	S	n/a
<i>request to unpublish</i>	n/a	n/a	n/a	n/a	C	n/a	n/a
<i>create new version</i>	n/a	n/a	n/a	n/a	C, S	C, S	n/a
<i>delete metadata</i>	C	C, S	S	S	n/a	n/a	S

- n/a = not applicable

Μπορείτε να κάνετε κάποια ενέργεια πάνω σε **έναν** μόνο πόρο, από την **σελίδα θέασής** του, ή σε **περισσότερους**, από την λίστα των πόρων στους οποίους έχετε **δικαιώματα**. Για να επιλέξετε περισσότερους από έναν πόρους κάντε κλικ στο πλαίσιο στα αριστερά του ονόματός τους και επιλέξτε μια ενέργεια από το πλαίσιο **action** στο επάνω μέρος της σελίδας. Εάν επιλέξετε πόρους **των οποίων η κατάσταση είναι διαφορετική**, θα δείτε **μόνο τις διαθέσιμες ενέργειες** για όλους τους πόρους. Στην περίπτωση που δείχνει η ακόλουθη εικόνα, μόνο η **εξαγωγή μεταδεδομένων** (export metadata) επιτρέπεται εφόσον έχει επιλεγθεί ένας πόρος του οποίου η δημοσίευση έχει γίνει αίτημα να αρθεί (requested to be unpublished) και ένας πόρος που έχει υποβληθεί για δημοσίευση (submitted).

The screenshot shows the CLARIN interface with a list of resources. On the left, there are filters for Curator, Resources (Corpus, Lexical/Conceptual, Tool/Service, Language description), Status (draft, submitted, syntactically valid, published, unpublished), and Has data (no). The main table has columns for Resource name, Actions, and Status. Two resources are listed: 'Language Description (Mandatory elements)' and 'testing'. Both have a green checkmark in the Status column. The 'Language Description' resource has an 'Actions' dropdown menu open, showing 'Export metadata'. The 'Status' column for this resource shows 'requested for unpub...'. The 'testing' resource has a 'submitted' status.

Στη σελίδα που παρουσιάζεται η λίστα των πόρων σας, υπάρχουν επίσης φίλτρα για να σας διευκολύνουν. Ο παρακάτω πίνακας παρουσιάζει τα διαθέσιμα φίλτρα ανά τύπο χρήστη. Κάντε κλικ στο φίλτρο που σας ενδιαφέρει και θα μεταφερθείτε στην αντίστοιχη ενότητα για να μάθετε περισσότερα για αυτό.

² Πόροι για τους οποίους έχει γίνει αίτημα να αρθεί η δημοσίευσή τους.

Filter	Curator	Supervisor	Validator
<i>resource type</i>	X	X	X
<i>status</i>	X	X	X
<i>has data</i>	X	X	X
<i>processable</i>	X	X	X
<i>action required</i>		X	
<i>metadata valid</i>		X	X
<i>legally valid</i>		X	X
CLARIN:EL compatible service		X	X
<i>requested for unpublish</i>		X	
<i>validator assignment required</i>		X	
<i>for information</i>		X	
<i>metaresources</i>		X	

28.2 II. Παρουσίαση και ανάλυση των ενεργειών

28.2.1 Edit (Επεξεργασία)

Κάνοντας κλικ στο **edit metadata** θα μεταφερθείτε στην εγγραφή μεταδεδομένων του πόρου στο *περιβάλλον τεκμηρίωσης και ανάρτησης*³. Μπορείτε να επεξεργαστείτε τα μεταδεδομένα του πόρου όσες φορές θέλετε. Μετά την εκάστοτε ολοκλήρωση της επεξεργασίας μπορείτε να αποθηκεύσετε τον πόρο **ως προσχέδιο** (save draft) ή αν έχετε συμπληρώσει όλα τα *υποχρεωτικά μεταδεδομένα* και είστε ικανοποιημένοι με την περιγραφή του πόρου μπορείτε να κάνετε **τελική αποθήκευση** (save). Κατά την τελική αποθήκευσή όλα τα μεταδεδομένα (υποχρεωτικά και μη) που έχουν συμπληρωθεί ελέγχονται και εφόσον δεν υπάρχουν λάθη/παραλείψεις, το σύνολο των μεταδεδομένων θεωρείται ότι είναι **συντακτικά έγκυρο** (syntactically valid) γιαυτό και ο πόρος αλλάζει στάδιο. Η εγγραφή παραμένει επεξεργάσιμη και πλέον μπορεί η επεξεργασία των μεταδεδομένων να γίνει **και από τον υπεύθυνο αποθετηρίου** (κάτι που δεν ήταν εφικτό όσο ο πόρος βρισκόταν στο στάδιο του προσχεδίου, κατά το οποίο την δυνατότητα αυτή την είχε μόνο ο επιμελητής).

- Εάν είστε **επιμελητής** (curator) μπορείτε να επεξεργαστείτε τον πόρο σας από το στάδιο: *draft* (προσχέδιο) μέχρι το στάδιο: *syntactically valid* (συντακτικά έγκυρος).
- Αν είστε **υπεύθυνος αποθετηρίου** (supervisor) μπορείτε να επεξεργαστείτε έναν πόρο από το στάδιο: *syntactically valid* (συντακτικά έγκυρος) μέχρι το στάδιο: *approved* (εγκεκριμένος) και στη συνέχεια όταν ο πόρος είναι στο στάδιο: *unpublished* (απο-δημοσιευμένος).

Μια εγγραφή μεταδεδομένων θα πρέπει **εκ νέου** να υποστεί επεξεργασία αν ο επικυρωτής μεταδεδομένων την απορρίψει (reject). Η απόρριψη έχει ως αποτέλεσμα ο πόρος να επιστρέφει αυτόματα στο στάδιο: **συντακτικά έγκυρος**. Ως επιμελητής θα λάβετε ένα email με τα σχόλια του επικυρωτή τα οποία θα εξηγούν και τι αλλαγές/διορθώσεις πρέπει να γίνουν. Μόλις τις κάνετε θα χρειαστεί να υποβάλετε τον πόρο για δημοσίευση ξανά.

³ Οι αγγλικοί όροι που αναφέρονται στο περιβάλλον τεκμηρίωσης και ανάρτησης πόρων (**metadata editor** και **editor**) χρησιμοποιούνται εναλλακτικά χωρίς διαφορά στη σημασία.

28.2.2 Copy record (Αντιγραφή εγγραφής)

Η αντιγραφή μεταδεδομένων επιτρέπεται εφόσον ένας πόρος είναι **συντακτικά έγκυρος** (syntactically valid).

Lexical/Conceptual Resource (Mandatory elements)

1.0.0 (automatically assigned) curator

Lexical/Conceptual resource curator_hosted@gmail.com

Hosted Resources Repository supervisor

Created: 30 May 2021 supervisor_hosted@gmail

Updated: 09 June 2021

Testrecognizer

Tool/Service curator

Hosted Resources Repository curator_hosted@gmail.com

Created: 08 June 2021

Updated: 08 June 2021

democorpus

1.0.0 (automatically assigned) curator

Corpus curator_hosted@gmail.com

Hosted Resources Repository supervisor

Created: 08 June 2021 supervisor_hosted@gmail.com

Actions

syntactically valid

Edit Metadata

Copy record

Submit for publication

Delete Metadata

syntactically valid

Κάντε κλικ στην ενέργεια και θα ανοίξει ένα νέο παράθυρο, όπου πρέπει να δώσετε ένα όνομα στον νέο πόρο, που θέλετε να δημιουργήσετε με την αντιγραφή, καθώς και να ορίσετε την έκδοσή του⁴.

Please fill in the following fields in order to create a copy of Lexical/Conceptual Resource (Mandatory elements)

resource name *

LCR (mandatory elements) - COPY

The official name or title of the language resource/technology

Version

2.0.0

The new version of the record that will be created. Recommended format: major_version.minor_version.patch (see semantic versioning guidelines at <http://semver.org>)

cancel Create copy

Όταν τελειώσετε, κάντε κλικ στο **Create copy** (Δημιουργία αντιγράφου). Θα ενημερωθείτε ότι το αντίγραφο της εγγραφής μεταδεδομένων δημιουργήθηκε με επιτυχία, ενώ παράλληλα θα μεταφερθείτε στη σελίδα θέασης του νέου πόρου.

STATUS

You can continue editing your metadata record while the status is draft (syntactically valid); when you are satisfied with it, you can submit it for publication; the resource will be validated and published by the supervisor of the repository, or, if required, you will be contacted for further information

draft syntactically valid submitted approved published

LCR (mandatory elements) - COPY

LexicalConceptualResource

Version: 1.1.0

This is a resource with only the mandatory (and mandatory upon condition) elements filled in.

Language English Modern Greek (1453)

Keyword mandatory, testing

clarin:el

Actions

Success

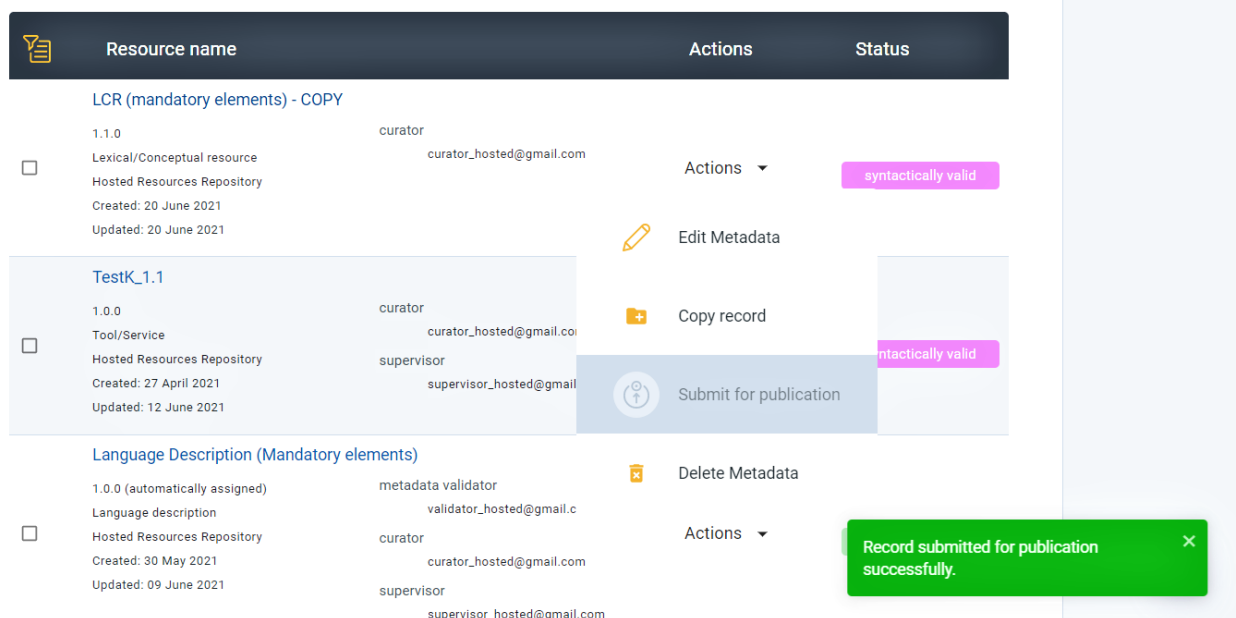
Ο νέος πόρος θεωρείται **συντακτικά έγκυρος** και πρέπει με τη σειρά του να περάσει από τα στάδια που περι-

⁴ Η τιμή 1.0.0 δίνεται αυτόματα όταν ο επιμελητής δεν έχει συμπληρώσει το υποχρεωτικό μεταδεδομένο της έκδοσης (version).

γράφονται στον *κύκλο ζωής του πόρου*.

28.2.3 Submit for publication (Υποβολή για δημοσίευση)

Αυτή η ενέργεια πραγματοποιείται μόνο σε πόρο που είναι **συντακτικά έγκυρος**. Εφόσον έχετε ολοκληρώσει την περιγραφή των μεταδεδομένων ενός πόρου (είτε ως επιμελητής είτε ως υπεύθυνος αποθετηρίου), μπορείτε να τον **υποβάλετε για δημοσίευση**.



Resource name	Actions	Status
LCR (mandatory elements) - COPY 1.1.0 Lexical/Conceptual resource Hosted Resources Repository Created: 20 June 2021 Updated: 20 June 2021	Actions Edit Metadata	syntactically valid
TestK_1.1 1.0.0 Tool/Service Hosted Resources Repository Created: 27 April 2021 Updated: 12 June 2021	Copy record Submit for publication	syntactically valid
Language Description (Mandatory elements) 1.0.0 (automatically assigned) Language description Hosted Resources Repository Created: 30 May 2021 Updated: 09 June 2021	Delete Metadata Actions	Record submitted for publication successfully.

- Αν είστε **υπεύθυνος αποθετηρίου** θα λάβετε ένα email που θα σας ζητά να «*αναθέσετε επικυρωτές στην ακόλουθη εγγραφή μεταδεδομένων, καθώς είναι έτοιμη να ελεγχθεί από τους επικυρωτές*» («*assign validators to the following record, as it is ready for legal and metadata validation*»).

28.2.4 Assign supervisor (Ανάθεση υπεύθυνου αποθετηρίου)

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο για τους **υπεύθυνους αποθετηρίου** και απαιτείται μόνο όταν υπάρχουν **περισσότεροι του ενός** σε ένα αποθετήριο. Εάν υπάρχει μόνο ένας υπεύθυνος στο αποθετήριο, η υποδομή του/της αναθέτει αυτόματα όλους τους πόρους. **Δεν** πρόκειται για τον ορισμό κάποιου χρήστη ως **υπεύθυνου αποθετηρίου** αλλά για την επιλογή του υπεύθυνου αποθετηρίου ως **υπεύθυνου για τον πόρο**.

Μόλις ένας πόρος υποβληθεί για δημοσίευση, θα λάβετε το ακόλουθο email: *Παρακαλούμε ορίστε τον εαυτό σας ως υπεύθυνο για την ακόλουθη εγγραφή και στη συνέχεια αναθέστε της επικυρωτές, καθώς είναι έτοιμη να ελεγχθεί* (Please assign yourself as supervisor to the following record and then assign legal and metadata validators, as it is ready for validation).

Κάνοντας κλικ στην ενέργεια, θα ανοίξει ένα νέο παράθυρο όπου θα πρέπει να επιλέξετε τον χρήστη ο οποίος θα είναι ο υπεύθυνος για τον πόρο.

Αφού γίνει η επιλογή του χρήστη, κάντε κλικ στο submit και θα δείτε να εμφανίζεται ένα μήνυμα ότι η διαδικασία ολοκληρώθηκε επιτυχώς, στο κάτω δεξιά μέρος της σελίδας.



28.2.5 Assign validator (Ανάθεση επικυρωτή)

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο για τους **υπεύθυνους αποθετηρίου**. Για να μπορέσετε να αναθέσετε έναν πόρο στους επικυρωτές μεταδεδομένων και νομικών θεμάτων, πρέπει πρώτα να έχετε αναθέσει αυτούς τους ρόλους σε χρήστες στο αποθετήριο σας. Δείτε [εδώ](#) πώς γίνεται αυτό.

Μόλις ένας πόρος υποβληθεί για δημοσίευση, θα λάβετε ένα email που θα σας ενημερώνει ότι πρέπει να τον αναθέσετε σε επικυρωτές. Εάν ο πόρος **δεν** συνοδεύεται από αρχεία περιεχομένου, θεωρείται αυτόματα **νομικά έγκυρος**⁵ και το μόνο που χρειάζεται είναι να του αναθέσετε έναν επικυρωτή μεταδεδομένων.

⁵ Όταν δεν υπάρχουν αρχεία περιεχομένου, δεν υπάρχει ανάγκη να δηλωθούν *όροι χρήσης* και κατά συνέπεια, ο πόρος θεωρείται αυτομάτως **νομικά έγκυρος** (legally valid).

Εάν ο πόρος **έχει** αρχεία περιεχομένου, πρέπει να αναθέσετε και τους δύο επικυρωτές.

Η διαδικασία πρέπει να γίνει ξεχωριστά για κάθε επικυρωτή. Πρέπει να κάνετε κλικ στην ενέργεια, **ανάθεση επικυρωτή μεταδεδομένων** (assign metadata validator) ή **ανάθεση επικυρωτή νομικών θεμάτων** (assign legal validator) και στη συνέχεια θα εμφανιστεί ένα νέο παράθυρο.

Please select a validator in order to validate TestK_1

Assign validator *

legalValidator_1@email.com
legalValidator_2@email.com

Θα δείτε μια αναπτυσσόμενη λίστα με τους επικυρωτές που υπάρχουν στο αποθετήριο σας (αν είναι περισσότεροι του ενός). Επιλέξτε τον επικυρωτή που θέλετε και στη συνέχεια κάντε κλικ στο **submit** (υποβολή). Θα δείτε να εμφανίζεται ένα μήνυμα ότι η διαδικασία ολοκληρώθηκε επιτυχώς, στο κάτω δεξιά μέρος της σελίδας.

Επιστρέφοντας στη λίστα των πόρων του αποθετηρίου σας, η οποία είναι διαθέσιμη από το [View my supervision tasks](#), θα μπορείτε να δείτε τους επικυρωτές που έχετε αναθέσει σε κάθε πόρο ή την ανάγκη ανάθεσης επικυρωτών, καθώς και τυχόν σχόλια που μπορεί να έχουν γίνει κατά την επικύρωση.

The screenshot displays two resource entries in the CLARIN interface. The first entry, 'testing', is highlighted with a yellow box. It shows a 'Review comments' section with a message: '[21/06/2021]Metadata validation review: The resource is bilingual but the same language has been used twice. Please check the metadata on linguality and languages.' The second entry, 'TestK_1', is highlighted with an orange box. It shows a 'validator assignment required' status with a 'yes' button.

28.2.6 Publish (Δημοσίευση)

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο για τους υπεύθυνους αποθετηρίου.

Μόλις ένας πόρος εγκριθεί (από έναν ή και τους δύο επικυρωτές ανάλογα με την περίπτωση), αλλάζει στάδιο: από **submitted** (υποβληθείς) σε **approved** (εγκεκριμένος). Θα λάβετε email ότι «έχει εγκριθεί από τους επικυρωτές, επομένως είναι έτοιμος για δημοσίευση» («it has been approved by the validators, therefore it is ready for publication»)

Language Description (Mandatory elements)

1.0.0 (automatically assigned)	metadata validator	approved
Language description	validator_hosted@gmail.com	legally valid
Hosted Resources Repository	curator	yes
Created: 30 May 2021	curator_hosted@gmail.com	metadata valid
Updated: 09 June 2021	supervisor	no
metadata validation date: 09 June 2021	supervisor_hosted@gmail.com	
	validator assignment required	

test_upload

1.0.0 (automatically assigned)	legal validator	approved
Corpus	user1@athenarc.gr	gally valid
Hosted Resources Repository	metadata validator	no
Created: 20 April 2021	user1@athenarc.gr	metadata valid
Updated: 26 April 2021	curator	
	admin@email.com	

Actions

- Edit Metadata
- Publish**
- Copy record
- Delete Metadata

Μόλις τον δημοσιεύσετε, θα δείτε να εμφανίζεται ένα μήνυμα ότι η διαδικασία ολοκληρώθηκε επιτυχώς, στο κάτω δεξιά μέρος της σελίδας, και ταυτόχρονα ο επιμελητής θα λάβει με email την ίδια ειδοποίηση.

Record published successfully.

28.2.7 Request to unpublish (Αίτημα για άρση δημοσίευσης)

Εάν είστε **επιμελητής** και πιστεύετε ότι ένας δημοσιευμένος πόρος δεν θα έπρεπε να βρίσκεται στον κεντρικό κατάλογο, μπορείτε να ζητήσετε την άρση της δημοσίευσής του.

Grammar (Mandatory elements)

1.0.0 (automatically assigned)	metadata validator	published
Language description	user1@gmail.com	
Hosted Resources Repository	curator	
Created: 22 June 2021	user1@gmail.com	
Updated: 22 June 2021	supervisor	
	user1@gmail.com	

Actions

- Create new version
- Copy record
- Request to unpublish**
- Unpublish

Όταν κάνετε κλικ στην ενέργεια, ανοίγει ένα νέο παράθυρο που σας ζητά να δηλώσετε τους λόγους για τους οποίους ζητάτε αυτήν την αλλαγή.

Request to unpublish

Unpublication request reason *

The resource contains personal data and must be anonymized.

Please specify the reason for your unpublication request.

cancel

Request to unpublish

Αφού δηλώσετε τους λόγους, πατήστε το **Request to unpublish** (Αίτημα για άρση δημοσίευσης). Θα δείτε ένα μήνυμα, στην κάτω δεξιά πλευρά της σελίδας, ότι το αίτημά σας υποβλήθηκε με επιτυχία και στη λίστα των πόρων θα εμφανιστεί το νέο στάδιο στο οποίο βρίσκεται ο πόρος (requested for unpublishing).

Grammar (Mandatory elements)

1.0.0 (automatically assigned)

Language description

Hosted Resources Repository

Created: 22 June 2021

Updated: 22 June 2021

metadata validator

user1@gmail.com

curator

user1@gmail.com

supervisor

user1@gmail.com

Actions

requested for unpub...

28.2.8 Unpublish (Άρση δημοσίευσης)

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο για τους υπεύθυνους αποθετηρίου.

Μπορεί να γίνει **άρση δημοσίευσης** (unpublish) μόνο για τους πόρους που είναι **δημοσιευμένοι** (published) ή για τους οποίους έχει γίνει **αίτημα άρσης δημοσίευσης** (requested for unpublishing).

Grammar (Mandatory elements)

1.0.0 (automatically assigned)

Language description

Hosted Resources Repository

Created: 22 June 2021

Updated: 22 June 2021

metadata validation date: 22 June 2021

metadata validator

user1@gmail.com

curator

user1@gmail.com

supervisor

user1@gmail.com

validator assignment required

no

Actions

requested for unpub...

legally valid

yes

adata valid

no

Create new version

Copy record

Unpublish

draft

test

Corpus

Hosted Resources Repository

Created: 16 June 2021

Updated: 16 June 2021

curator

user2@gmail.com

validator assignment required

no

Όταν η ενέργεια ολοκληρωθεί επιτυχώς, ο πόρος χαρακτηρίζεται ως **απο-δημοσιευμένος** (unpublished).

Grammar (Mandatory elements)

1.0.0 (automatically assigned)
Language description
Hosted Resources Repository
Created: 22 June 2021
Updated: 22 June 2021
metadata validation date: 22 June 2021

metadata validator

user1@gmail.com

curator

user1@gmail.com

supervisor

user1@gmail.com

validator assignment required

no

Actions

unpublished

legally valid

yes

metadata valid

yes

test maria

Corpus
Hosted Resources Repository
Created: 16 June 2021
Updated: 16 June 2021

curator

user2@gmail.com

validator assignment required

no

draft

Record Unpublished successfully.

28.2.9 Create New Version (Δημιουργία νέας έκδοσης)

Αφού ένας πόρος έχει δημοσιευθεί, μπορείτε να δημιουργήσετε μια νέα έκδοσή του. Η ενέργεια αυτή είναι χρήσιμη σε περίπτωση που πρέπει να γίνει επικαιροποίηση ενός πόρου.

tool1

1.0
Tool/Service
Hosted Resources Repository
Created: 28 April 2021
Updated: 15 June 2021

metadata validator

validator_hosted@gmail.com

curator

curator_hosted@gmail.com

supervisor

supervisor_hosted@gmail.com

validator assignment required

no

Actions

published

legally valid

yes

adata valid

yes

Create new version

Copy record

Unpublish

Κάντε κλικ στην ενέργεια και θα ανοίξει ένα νέο παράθυρο. Συμπληρώστε τον αριθμό της νέας έκδοσης και την ημερομηνία και, στη συνέχεια, κάντε κλικ στο **create new version** (δημιουργία νέας έκδοσης).

Please fill in the following fields in order to create a new version for the record tool1

Version *

2.1.1

The new version of the record that will be created. Recommended format: major_version.minor_version.patch (see semantic versioning guidelines at <http://semver.org>)

version date

2021-06-01

The date of the LRT version (latest update of the particular version if possible)

cancel

Create new version

Θα δείτε ένα μήνυμα ότι η δημιουργία έγινε επιτυχώς στην κάτω δεξιά πλευρά της σελίδας.

Success

28.2.10 Delete metadata (Διαγραφή μεταδεδομένων)

Εάν δεν είστε ικανοποιημένοι με την εγγραφή μεταδεδομένων που έχετε δημιουργήσει για έναν πόρο, μπορείτε να την διαγράψετε.

The screenshot shows a table of records in the CLARIN interface. The first record is 'Corpus for Demo' (version 1.1.1) with curator 'curator_hosted@gmail.com'. The second record is 'Tool Mandatory' (version 1.0.0) with curator 'supervisor_hosted@gmail.com'. Both records have a 'Delete Metadata' button. A 'Mandatory Corpus' button is at the bottom.

Όταν η ενέργεια ολοκληρωθεί, θα δείτε ένα μήνυμα ότι έγινε επιτυχώς στο κάτω δεξιά μέρος της σελίδας. Η εγγραφή μεταδεδομένων που διαγράψατε δεν υπάρχει πλέον.

Record Deleted successfully.

28.2.11 Validate (Επικύρωση)

Προσοχή: Αυτή η ενέργεια είναι διαθέσιμη μόνο για τους **επικυρωτές**.

Όταν σας έχει ανατεθεί ένας πόρος για επικύρωση, θα λάβετε ένα ενημερωτικό email. Από το dashboard (πίνακας ελέγχου) μεταβείτε στο [Validation tasks](#) (Εργασίες επικύρωσης). Εντοπίστε τον πόρο και κάντε κλικ πάνω στο όνομά του. **Δεν μπορείτε** να εκτελέσετε καμία ενέργεια από τη λίστα πόρων, και για τον λόγο αυτό πρέπει να επισκεφθείτε τη σελίδα θέασης του πόρου. Κάντε κλικ στην αναπτυσσόμενη λίστα ενεργειών, όπως φαίνεται στην παρακάτω εικόνα. Θα εμφανιστεί η ενέργεια⁶ επικύρωσης που αντιστοιχεί στον ρόλο σας. Η διαδικασία που περιγράφεται στη συνέχεια είναι ίδια τόσο για την επικύρωση μεταδεδομένων όσο και για τη επικύρωση νομικών θεμάτων.

⁶ Οι πόροι που δεν έχουν αρχεία περιεχομένου θεωρούνται αυτομάτως **νομικά έγκυροι** (legally valid) και ελέγχονται μόνο ως προς τα μεταδεδομένα τους.

STATUS

Your metadata record has been submitted for technical validation by the CLARIN-EL team and can no longer be edited; you will be notified when it is published or, if needed, for further information

draft syntactically valid submitted approved published

testing

Corpus

Version: 1.0.0 (automatically assigned)

Testing resource

clarin:el

Actions



Metadata validate

Language

Aragonese

Keyword

testing, and testing again

Corpus subclass

raw corpus

Όταν κάνετε κλικ στην ενέργεια, θα ανοίξει ένα νέο παράθυρο που θα σας ζητά να **εγκρίνετε** (approve) ή να **απορρίψετε** (reject) την εγγραφή μεταδεδομένων.

Please fill in the following fields in order to validate testing

☐ Approve

☐ Reject

Submit Cancel

- Αν επιλέξετε **έγκριση** και υποβάλετε τη φόρμα, το παράθυρο κλείνει και ο πόρος χαρακτηρίζεται ως **έγκυρος** (valid) για τα μεταδεδομένα.
- Αν επιλέξετε **απόρριψη**, ανοίγει ένα νέο παράθυρο όπου πρέπει να γράψετε τους λόγους απόρριψης του πόρου.

Please fill in the following fields in order to validate testing

☐ Approve

☒ Reject

Reasons *

The resource is bilingual but the same language has been used twice. Please check the metadata on linguality and languages.

Please state the reasons for rejecting this record

Submit Cancel

Όταν επιστρέψετε στη λίστα των πόρων που υπάρχουν στις εργασίες επικύρωσης, θα βρείτε τα σχόλια που έχετε

κάνει.

testing

1.0.0 (automatically assigned)
Corpus
Hosted Resources Repository
submitted: 08 June 2021

metadata validator
validator_hosted@gmail.com
curator
curator_hosted@gmail.com
supervisor
supervisor_hosted@gmail.com

syntactically valid

Review comments

[21/06/2021]Metadata validation review: The resource is bilingual but the same language has been used twice. Please check the metadata on linguality and languages.

Τα σχόλιά σας θα κοινοποιηθούν στον επιμελητή του πόρου μέσω email. Ο πόρος επιστρέφει αυτόματα στο στάδιο κατά το οποίο ήταν **συντακτικά έγκυρος** ώστε να μπορεί ο επιμελητής να κάνει τις διορθώσεις/προσθήκες που επισημάνετε.

Πληροφορίες νομικού χαρακτήρα

Οι Όροι Χρήσης των Υπηρεσιών CLARIN:EL είναι δεσμευτικοί και για το λόγο αυτό παρακαλείστε να τους διαβάσετε προσεκτικά. Συμβουλευθείτε επίσης και την Πολιτική απορρήτου (υπερσύνδεσμοι υπάρχουν στο κάτω μέρος κάθε σελίδας της υποδομής.)



© CLARIN:EL 2021 [Terms of service | Privacy Policy](#)

Need Help ?
CLARIN:EL Helpdesks
FAQs
Bibliography

Πρακτική συμβουλή: Δείτε, επίσης, το Προτεινόμενο σχήμα αδειοδότησης Γλωσσικών Πόρων που έχει δημιουργηθεί με σκοπό να περιοριστεί η περιπλοκότητα της αδειοδότησης.

ΚΕΦΑΛΑΙΟ 30

Δημοσιεύσεις

Μπορείτε να βρείτε στην διαδικτυακή πύλη όλες τις δημοσιεύσεις και παρουσιάσεις για το CLARIN:EL.

Indices and tables

- `genindex`
- `modindex`
- *Αναζήτηση*