

CLARIN-EL
Common Language Resources and Technology Infrastructure

Η Ευρωπαϊκή Υποδομή CLARIN και η Εθνική Υποδομή CLARIN EL

Στέλιος Πιπερίδης, Μαρία Γαβριηλίδου,
Πένυ Λαμπροπούλου, Γιούλη
Μπακαγιάννη, Κανέλλα Πουλή

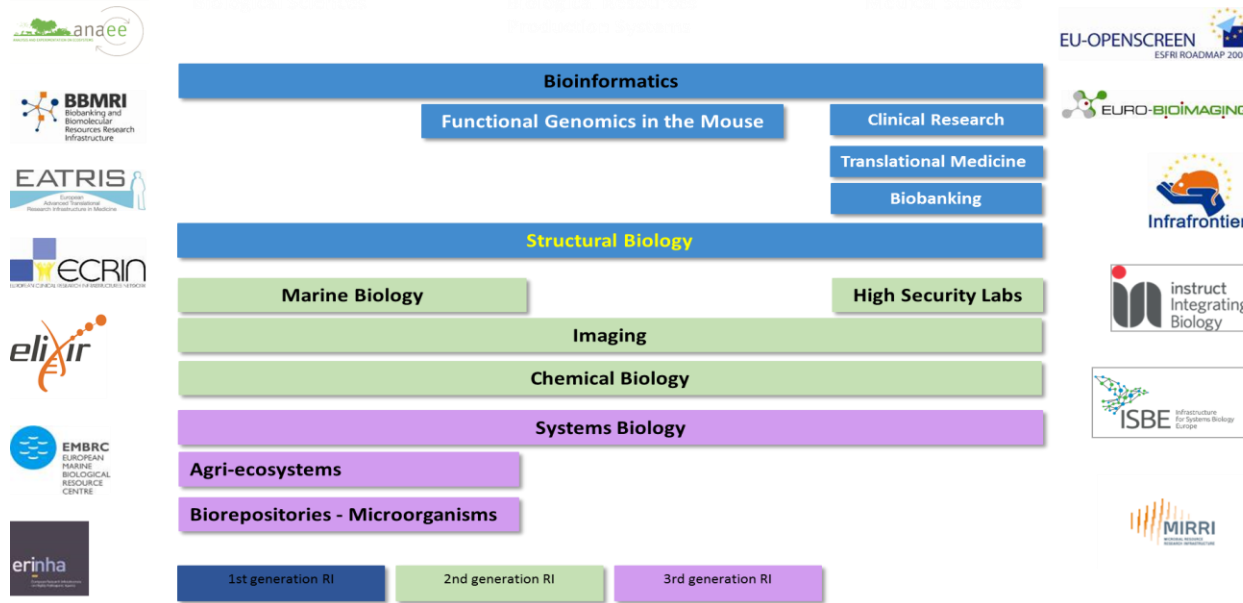
**Ινστιτούτο Επεξεργασίας Λόγου
ΕΚ «Αθηνά»**

Ο ρόλος των (επιστημονικών) δεδομένων

Η δύναμη των δεδομένων

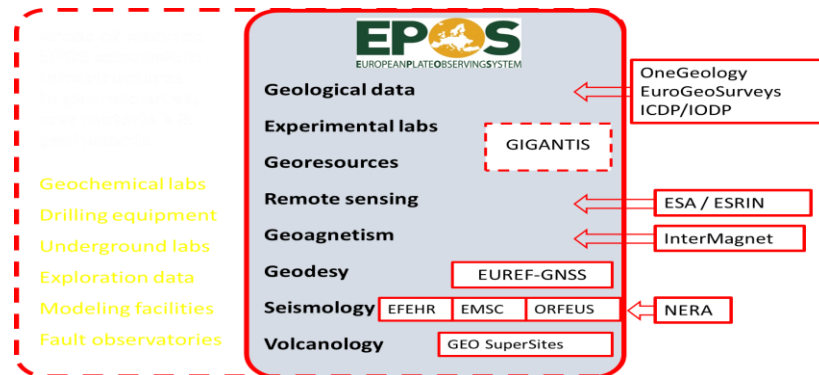
- ◆ Τα (επιστημονικά) δεδομένα μεταμορφώνουν ήδη και βελτιώνουν με ταχύτατους ρυθμούς τη ζωή μας
 - Βιολογικά, γεωγραφικά, μετεωρολογικά, οικονομικά, κ.ά. δεδομένα
- ◆ Η εμπειρία από πολλά επιστημονικά πεδία δείχνει ότι τα δεδομένα και οι υπηρεσίες επεξεργασίας τους αποκτούν ιδιαίτερη αξία όταν διαμοιράζονται και ανοίγονται
 - τόσο για ερευνητικούς και τεχνολογικούς σκοπούς, συμπεριλαμβανόμενης της αξιολόγησης αυτών
 - όσο και για την δημιουργία καινοτομικών εφαρμογών
- ◆ Η πρόσβαση στα αποτελέσματα του Human Genome Project «γέννησε» ~ 3 δις ευρώ σε επενδύσεις E&A, ~ 500 δις ευρώ σε οικονομική δραστηριότητα
- ◆ Η συγκέντρωση γενετικών δεδομένων ασθενών με Alzheimer's οδήγησε στην ανακάλυψη 5 νέων γονιδίων σχετιζόμενων με την θεραπεία της νόσου
- ◆ "Τα επιστημονικά δεδομένα είναι πολύτιμα για να είναι κλειδωμένα στο συρτάρι"

Άνοιγμα και διασύνδεση δεδομένων

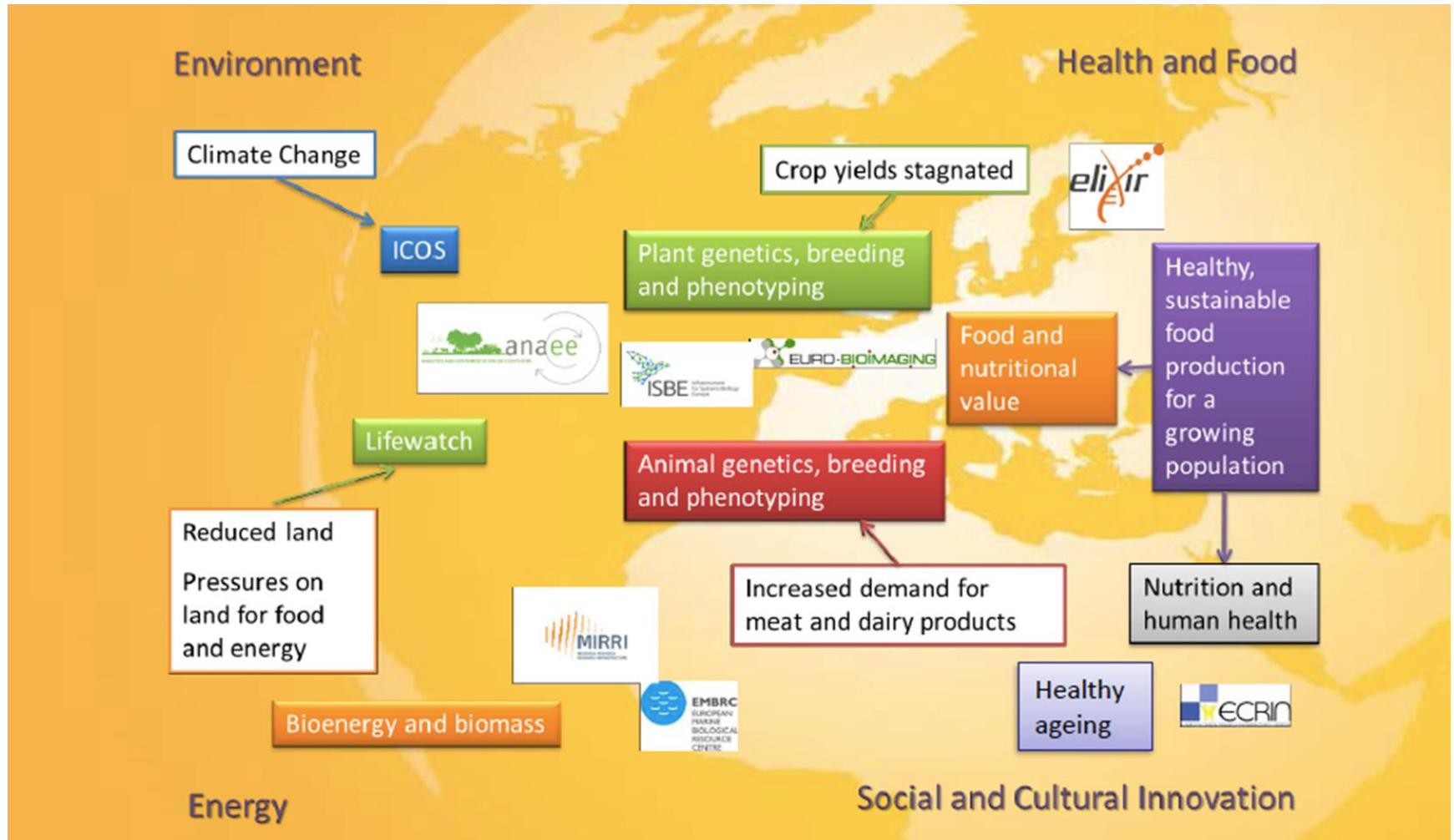


στις επιστήμες
ζωής και υγείας

στις επιστήμες
της γης



Διασύνδεση δεδομένων διαφόρων πεδίων



Γλωσσικά δεδομένα και υπηρεσίες

Γλωσσικά δεδομένα και υπηρεσίες

- ◆ Τι γίνεται όμως με τη γλώσσα και τα γλωσσικά δεδομένα;
- ◆ Με τη γλωσσική τεχνολογία που χωρίς να το συνειδητοποιούμε την χρησιμοποιούμε καθημερινά για να επεξεργαστούμε περιεχόμενο, να αλληλεπιδράσουμε με ανθρώπους και μηχανές;
- ◆ Πολύ συχνά στις ηλεκτρονικές λίστες βλέπουμε ερωτήματα/αιτήματα για δεδομένα/υπηρεσίες
 - Σώμα ελληνικών κειμένων περιόδου X, με επισημείωση σε επίπεδο Ψ
 - Σώμα παράλληλων κειμένων μεταξύ Ελληνικών και γλώσσας X,
 - Στοιχισμένο και ...
 - Εργαλεία μορφοσυντακτικής ανάλυσης ή αναγνώρισης γλώσσας, κ.λπ.
 - Εργαλεία επεξεργασίας video, κ.λπ.

Εξεύρεση Γλωσσικών Πόρων; Διαθεσιμότητα;

- ◆ Σύμφωνα με παλαιότερες αλλά και πρόσφατες μελέτες μόνο ένα μικρό τμήμα των γλωσσικών πόρων (LRs) είναι γνωστό/ ανακοινώνεται/ γίνεται προσβάσιμο/ ανταλλάσσεται/ ...
- ◆ ... παρόλο που η συλλογή δεδομένων και η επιμέλειά τους (καθαρισμός, φροντίδα και συντήρηση) καθώς και η επισημείωσή τους είναι δαπανηρές διαδικασίες
- ◆ Για να επιτευχθεί πρόοδος, να διευκολυνθεί η ανάπτυξη χρήσιμων εφαρμογών, χρειαζόμαστε όλους αυτούς τους επιστημονικούς, τεχνικούς, νομικούς, οργανωτικούς, κοινωνικούς μηχανισμούς που θα επιτρέψουν την πρόσβαση, ανακύκλωση και επαναπροσδιορισμό της χρήσης των απαραίτητων πόρων.



Το πρόβλημα

- ◆ πολλές αρχειακές συλλογές, πόροι, εργαλεία και υπηρεσίες ΓΤ είναι γνωστά μόνο σε ορισμένες ερευνητικές κοινότητες
- ◆ συλλογές, πόροι, εργαλεία και υπηρεσίες είναι ασύνδετα μεταξύ τους και αυτό κάνει την αναζήτηση δύσκολη
- ◆ η εύρεση πόρων, υπηρεσιών, κλπ εξαρτάται από τη γλώσσα τεκμηρίωσης
- ◆ ακολουθούν διαφορετικά πρότυπα τεκμηρίωσης
- ◆ συχνά γλωσσικά δεδομένα και υπηρεσίες/εργαλεία είναι ασύμβατα μεταξύ τους
- ◆ ο ερευνητής (πάροχος ή χρήστης) αντιμετωπίζει ένα νομικό κυκεώνα σχετικά με τα δικαιώματα και τις υποχρεώσεις του
- ◆ δεν υπάρχουν κίνητρα για διαμοιρασμό και διάθεση πόρων

Η Ερευνητική Υποδομή CLARIN

Ο στόχος της ΕΥ CLARIN

- ◆ Η ΕΥ CLARIN (www.clarin.eu) στοχεύει να δημιουργήσει μία ολοκληρωμένη και διαλειτουργική ερευνητική υποδομή Γλωσσικών Πόρων και Τεχνολογιών
- ◆ καταπολεμώντας έτσι την ισχύουσα αποσπασματικότητα
- ◆ και προσφέροντας ένα σταθερό, συνεπές, εύχρηστο και επεκτάσιμο περιβάλλον πρόσβασης σε γλωσσικά δεδομένα
- ◆ στην υπηρεσία όλων των επιστημών και κυρίως των Κοινωνικών και Ανθρωπιστικών Επιστημών (ΚΑΕ)

Η αποστολή του CLARIN

◆ Τι;

- ◆ να δημιουργήσει μια υποδομή που να διαθέσει Γλωσσικούς Πόρους και Τεχνολογίες στους ερευνητές όλων των επιστημών

◆ Πώς;

- ◆ ενοποιώντας ψηφιακές αρχειακές συλλογές σε μία «ομοσπονδία αρχείων» με ενιαία διαδικτυακή πρόσβαση
- ◆ παρέχοντας σχετικές διαδικτυακές υπηρεσίες με τη μορφή γλωσσικών υπολογιστικών εργαλείων που "τρέχουν" πάνω στα γλωσσικά δεδομένα

Δηλαδή...

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

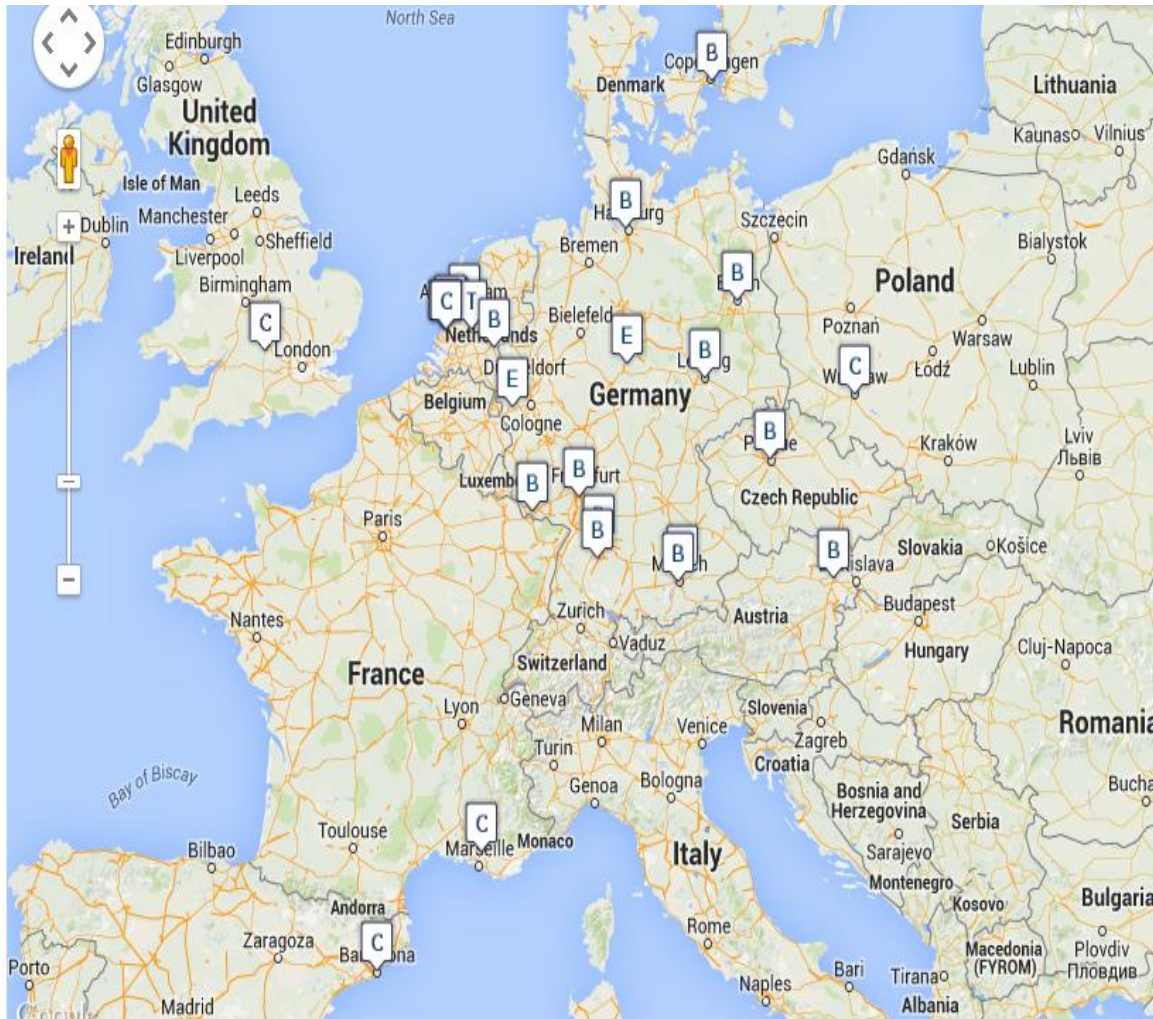
Γιατί μια ευρωπαϊκή υποδομή;

- ◆ μεγάλη αποσπασματικότητα στα έργα του χώρου
- ◆ σχετικές δράσεις παραμένουν άγνωστες
- ◆ απουσία διαλειτουργικότητας
- ◆ μικρή βιωσιμότητα αποτελεσμάτων
- ◆ ύπαρξη τεχνογνωσίας και εμπειρίας αλλά όχι σε όλα τα κράτη
- ◆ εργαλεία ανεξάρτητα γλώσσας μπορούν άμεσα να χρησιμοποιηθούν για άλλους πόρους
- ◆ εργαλεία εξαρτώμενα από γλώσσα συχνά μπορούν να προσαρμοστούν και σε άλλες γλώσσες
- ◆ τα περισσότερα κράτη δεν μπορούν να αναλάβουν μόνα τους το κόστος της προσπάθειας
- ◆ απουσία διακρατικού συντονισμού

Το όραμα του CLARIN

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

Πιστοποιημένα κέντρα CLARIN



14 κράτη
28 πιστο-
ποιημένα
κέντρα

CLARIN ERIC



- About ▾
- Services ▾
- Centres ▾
- Technology ▾
- Events
- Material ▾
- News
- Contact

Search

- Home >
- About >

- About**
- General Information
- Mission
- Governance
- Participants
- National Consortia
- Glossary
- FAQ

- Popular Content**
- Certified centres
- National Consortia
- Centres
- General Information
- Material

General Information

CLARIN is one of the Research Infrastructures
[Roadmap](#) by ESFRI, the European Strategy
 with sites all over Europe. Typical
 have in common that they provide
 and to expertise for researchers
 European level is CLARIN ERIC. An
 Commission in 2009. Its members are gov
 Members and one Observer constitute CLARIN ERIC.

Η Ελλάδα μέλος
 του ERIC από
 19/2/2015

The following thirteen countries are at this moment Members of CLARIN ERIC: [Austria](#), [Bulgaria](#), [Czech Republic](#), [Denmark](#), [Estonia](#), [Germany](#), [Greece](#), [Lithuania](#), [The Netherlands](#), [Poland](#), [Portugal](#) and [Sweden](#). The other member is the [Dutch Language Union](#), which is an intergovernmental body created by the Dutch and the Flemish government, responsible for the maintenance and promotion of the Dutch language. [Norway](#) participates in CLARIN ERIC as an Observer, with a view to joining as a full Member later on. More countries are expected to join in the near future. The goal is to cover all EU and Associated States, but third countries can also join CLARIN ERIC.

It is important to note that CLARIN ERIC's main task is to build, operate, coordinate and maintain the CLARIN infrastructure, and it neither conducts nor funds research activities.

CLARIN ERIC was set up with financial support from the European Commission through the CLARIN Preparatory Phase Project (2008-2011), but is now entirely funded by the participating countries.

Η Ελληνική Υποδομή CLARIN

Ο στόχος μας

- ◆ να ταιριάζουμε τις ανάγκες και τις προσδοκίες των παρόχων γλωσσικών πόρων και των καταναλωτών/χρηστών βελτιώνοντας το κατά πόσο ένας πόρος είναι ορατός (visibility), τεκμηριωμένος (documentation), αναγνωρίσιμος (identification), διαθέσιμος και καλά συντηρημένος (preservation), είτε πρόκειται για γλωσσικά δεδομένα είτε για (βασικά) γλωσσικά εργαλεία.
- ◆ να ταιριάζουμε δεδομένα και εργαλεία/υπηρεσίες και να βελτιώσουμε την δυνατότητα εύκολης χρήσης υπηρεσιών γλωσσικής επεξεργασίας με σκοπό την οργανική ανάπτυξη της υποδομής
- ◆ Είναι ένα μακροπρόθεσμο πολυδιάστατο εγχείρημα βάσει του οποίου οι γλωσσικοί πόροι θα συνεισφέρουν στην προαγωγή της έρευνας, της τεχνολογίας και της καινοτομίας καθώς θα είναι διαθέσιμοι σε ευρεία κλίμακα, συγκεντρωμένοι και ανοικτά προσβάσιμοι για κοινή χρήση

Διάρθρωση του στόχου

- ◆ Δύο στενά διασυνδεδεμένα υποσυστήματα:
 - τεκμηρίωσης, αποθήκευσης, διαμοιρασμού + αναζήτησης, ανάκτησης, καταφόρτωσης γλωσσικών πόρων (resources infrastructure)
 - επεξεργασίας γλωσσικών δεδομένων μέσω διαδικτυακών υπηρεσιών γλωσσικής επεξεργασίας και παραγωγή νέων δεδομένων (processing infrastructure)
- ◆ Σαν πρώτο βήμα, φιλοδοξεί να (προσ)φέρει τους πόρους που χρειάζεται ένας ερευνητής για την έρευνά του
- ◆ σε 5 βήματα, κατά το δυνατόν...

1. Αναζήτηση με λέξεις κλειδιά, π.χ. Ελληνικό σώμα κειμένων σχετικό με τη θεματική περιοχή «νομικά»

The screenshot shows the CLARIN-EL website interface. At the top, there is a navigation bar with a home icon and several menu items: "Browse Resources", "Manage Resources", "Administration", "Community", "Statistics", "Help", "About", and "Your Profile, Juli". A "Logout" button is located in the top right corner. Below the navigation bar, the CLARIN-EL logo is displayed on the left, and a large button labeled "Go to the CLARIN EL Aggregator" is on the right. In the center, it states "66 language resources at your disposal". Below this, there is a search input field containing the text "Greek corpus law" and a "Search" button. The main content area below the search field is titled "What is it? - About the project" and contains two paragraphs of text describing the project's goals and scope.

2. Φυλλομέτρηση αποτελεσμάτων

The screenshot shows the CLARIN-EL web interface. At the top, there is a navigation bar with a home icon, a search bar containing 'Greek corpus law', and a 'Search' button. Below the navigation bar, there are several menu items: 'Browse Resources', 'Manage Resources', 'Administration', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile, Juli'. A 'Logout' button is also present in the top right corner.

The main content area is divided into two columns. The left column is titled 'Filter by:' and contains a list of filter categories, each with a dropdown arrow: Language, Resource Type, Media Type, Availability, Licence, Restrictions of Use, Validated, Foreseen Use, Use Is NLP Specific, Linguality Type, Multilinguality Type, Modality Type, MIME Type, Conformance to Standards/Best Practices, Domain, Geographic Coverage, and Time Coverage.

The right column displays the search results. It starts with the text '66 Language Resources (Page 1 of 4)'. Below this, there are navigation links: '<< Previous | Next >>' and 'Order by: Resource Name A-Z' with a dropdown arrow. The results are listed in a table-like format, with each entry having a title, a list of languages, and icons for download, search, and eye.

Resource Title	Languages	Download	Search	Eye
ACCURAT balanced test corpus for under resourced languages	Croatian, English, Estonian, German, Greek, Modern (1453-), Latvian, Lithuanian, Romanian, Slovenian	0	0	0
ACCURAT corpus of Wikipedia texts	Croatian, English, Estonian, German, Greek, Modern (1453-), Latvian, Lithuanian, Romanian, Slovenian	0	0	0
A parallel corpus collected from the European Constitution	Czech, Danish, Dutch, Flemish, English, Estonian, Finnish, French, German, Greek, Modern (1453-), Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Slovak, Slovenian, Spanish, Castilian, Swedish	0	0	0
A parallel corpus of KDE4 localization files (v.2)	Afrikaans, Arabic, Armenian, Assamese, Asturian, Bable, Leonese, Asturianese, Basque, Belarusian, Bengali, Bokmål, Norwegian, Norwegian Bokmål, Breton, Bulgarian, Catalan, Valencian, Central Khmer, Chinese, Crimean Tatar, Crimean Turkish, Croatian, Czech, Danish, Dutch, Flemish, English, Esperanto, Estonian, Finnish, French, Galician, Georgian, German, Greek, Modern (1453-), Gujarati, Hausa, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Irish, Italian, Japanese, Kannada, Kashubian, Kazakh, Kinyarwanda, Korean, Kurdish, Latvian, Lithuanian, Low German, Low Saxon, German, Low, Saxon, Low, Luxembourgish, Letzeburgesch, Macedonian, Maltese, Malay, Malayalam, Maltese, Marathi, Nepali, Northern Sami, Norwegian Nynorsk, Nynorsk, Norwegian, Occitan (post 1500), Oriya, Panjabi, Punjabi, Pedi, Sepedi, Northern Sotho, Persian, Polish, Portuguese, Pushto, Pashto, Romanian, Russian, Serbian	0	0	0

2α. Αναζήτηση με χρήση φίλτρων

The screenshot shows the CLARIN-EL search interface. At the top, there is a navigation bar with a home icon and buttons for 'Browse Resources', 'Manage Resources', 'Administration', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile, Juli'. A 'Logout' button is in the top right corner. Below the navigation bar is a search bar with a 'Search' button. The main content area is divided into two sections. On the left, there is a 'Filter by:' section with three expandable filter categories: 'Language', 'Resource Type', and 'Domain'. Under 'Language', 'Greek, Modern (1453-)' is selected. Under 'Resource Type', 'Corpus (2)' is selected, with sub-options for 'Annotation Type' and 'Annotation Format'. Under 'Domain', 'Law (2)', 'Politics (1)', and 'Travel (1)' are listed. On the right, the search results are displayed under the heading '2 Language Resources'. An 'Order by:' dropdown menu is set to 'Resource Name A-Z'. Two results are shown: 'Greek Textual Entailment Corpus' and 'OROSSIMO Corpus -Law'. Each result includes a document icon, the resource name, a 'Greek, Modern (1453-)' tag, and download/view icons with counts (0 downloads, 1 view).

3. Επιλογή του πόρου προτίμησης

Home
Logout

CLARIN-EL

Browse Resources
Manage Resources
Administration
Community
Statistics
Help
About
Your Profile, Juli

OROSSIMO Corpus -Law 🇬🇷

▶ View resource name in all available languages

PID: <http://hdl.gnnet.gr/11239/ATHENA-TEST-0000-0000-1702-4-TEST>

A corpus compiled of texts belonging to the Law domain, according to the XCES standard. There is also a terminological resource (see Orossimo terminological resource - Law)

▶ View resource description in all available languages

« Back
Download
Edit Resource

Distribution

Availability

Available - Restricted Use

Licence

CC - BY

Distribution Access/Medium:
Downloadable

Attribution Details: Orossimo Corpus by Athena R.C./ILSP used under CC-BY licence

Contact Person

Penny Labropoulou 🇬🇷

text 📄

Monolingual text corpus

Languages

Greek, Modern (1453-)

Linguality

Linguality type: Monolingual

Size

2,500,000 Words

Character encoding

UTF - 8

Domains

law

Resource Creation

Resource Creator

Institute for Language and Speech Processing / Athena R.C. 🇬🇷

Creation lasted: 01/01/1996 - 12/31/1998

Funding Project

OROSSIMO

URL: <http://www.ilsp.gr/e...>

Funding Type: National Funds

Funding Country: Greece

Project duration: 04/01/1996 - 12/31/1998

Metadata

4. Όροι χρήσης - αδειοδότηση

🏠
Logout

CLARIN-EL

Browse Resources
Manage Resources
Administration
Community
Statistics
Help
About
Your Profile, Juli

OROSSIMO Corpus -Law

Licence Agreement – CC-BY

Creative Commons Corporation (“Creative Commons”) is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an “as-is” basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Creative Commons Attribution 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License (“Public License”). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

I agree to these licence terms and would like to download the resource.

Download

« Resource View
Contact Resource Maintainer

Co-funded by the Greek CLARIN Attiki project (MIS 441451). Support for ESFRI/2006 Research Infrastructures, of the Greek Government.
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Unported License – Terms of Service

5. Καταφόρτωση του πόρου

The screenshot shows the CLARIN-EL website interface. At the top, there is a navigation menu with buttons for 'Browse Resources', 'Manage Resources', 'Administration', 'Community', 'Statistics', 'Help', 'About', and 'Your Profile'. A 'Logout' button is in the top right corner. The main content area displays the title 'OROSSIMO Corpus -Law' and a 'Licence Agreement - CC-BY' section. Below this, there is a text block explaining the Creative Commons Attribution 4.0 International Public License. A 'Download' button is visible at the bottom of the page. A Windows-style dialog box is overlaid on the page, titled 'Εισάγετε το όνομα αρχείου για αποθήκευση...' (Enter the file name for saving...). The dialog shows the 'Save in' location as 'working files' and the 'File name' as 'archive.zip'. The 'Save as type' is set to 'WinZip File'. The dialog also shows a list of folders: 'My Recent Documents', 'Desktop', 'My Documents', 'My Computer', and 'My Network'.

**Πίσω από όλα,
ένα μοντέλο περιγραφής του κόσμου των
γλωσσικών δεδομένων και υπηρεσιών**

Στόχος

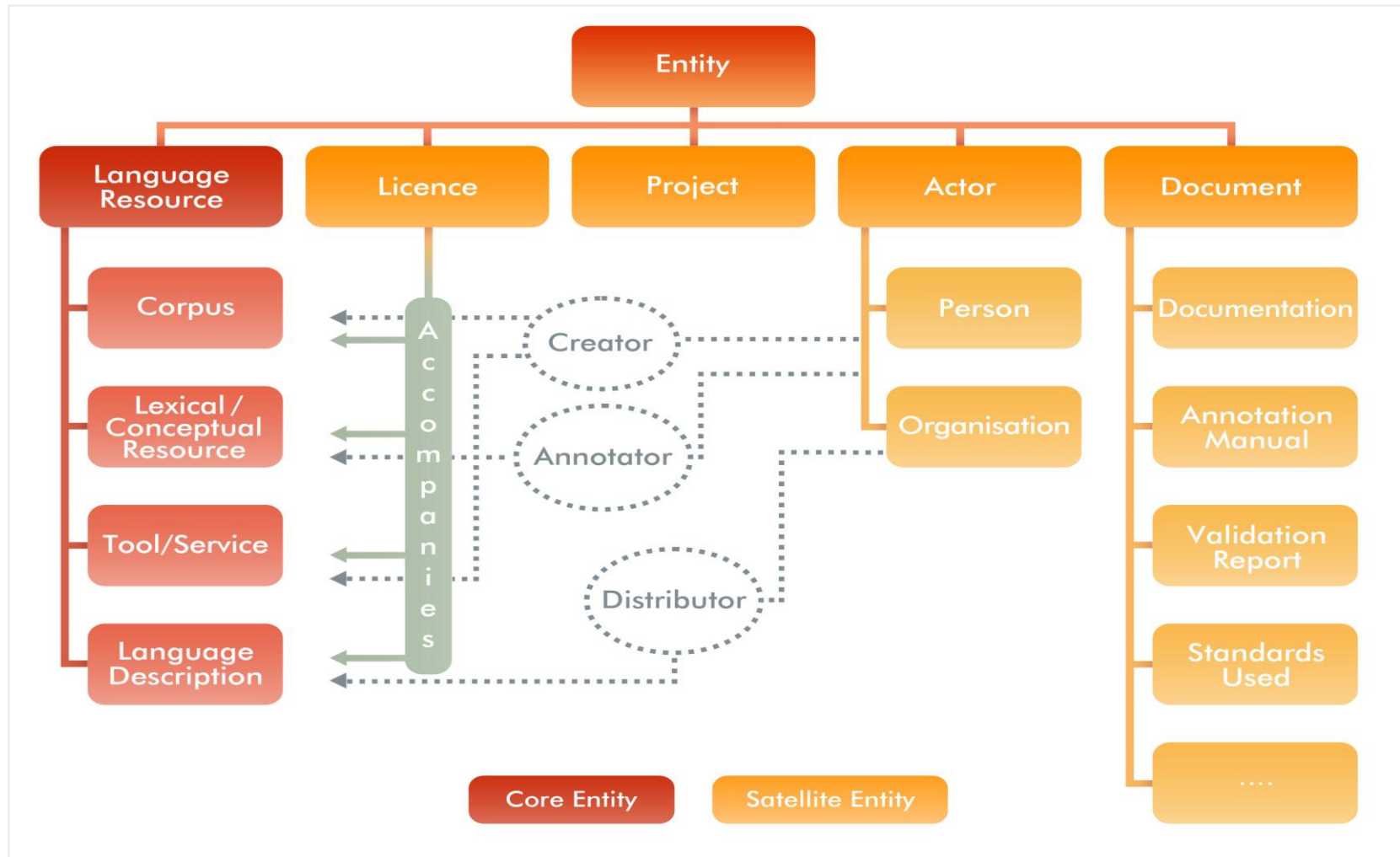
- ◆ Να υποστηρίξει τους χρήστες σε όλες τις ενέργειες τους
 - τεκμηρίωση πόρων + υπηρεσιών (δημιουργία, επικαιροποίηση)
 - αναζήτηση και ανάκτηση,
 - περιήγηση,
 - φόρτωση και καταφόρτωση πόρων
 - συγκομιδή μεταδεδομένων,
 - εποπτεία χρήσης κ.λπ.

Σχήμα μεταδεδομένων - οντολογία

◆ Περιγραφόμενες οντότητες

- Κύρια οντότητα: γλωσσικός πόρος (δεδομένα-εργαλεία/υπηρεσίες)
- Δορυφορικές οντότητες: σχετικά αντικείμενα, π.χ.
 - *δράστης*: άτομα και οργανισμοί που εμπλέκονται ως δημιουργοί πόρων, χρηματοδότες, διανομείς, κλπ.
 - *έγγραφο*: έγγραφα αναφοράς, όπως δημοσιεύσεις που περιγράφουν τον πόρο, αναφορές, εγχειρίδια, κ.λπ.
 - *έργο*: έργα που χρηματοδότησαν τη δημιουργία του πόρου, ή στο πλαίσιο του οποίου χρησιμοποιήθηκε ο πόρος κ.λπ.
 - *άδεια*: η άδεια διάθεσης του πόρου

Οντολογία (απόσπασμα)



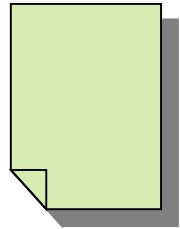
Τυπολογία πόρων (1)

◆ Τυπολογία πόρων

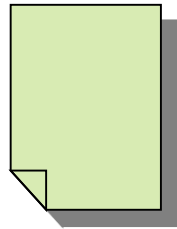
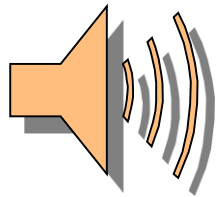
- ταξινόμηση βάσει 2 κύριων κριτηρίων: "τύπος πόρου" (resource type) & "μέσο πόρου" (media type)
- τύπος πόρου
 - **σώμα κειμένων** (γραπτού / προφορικού λόγου, πολυτροπικών/πολυμεσικών κειμένων)
 - **λεξικός / εννοιολογικός πόρος** (π.χ. λεξικό, ορολογικός πόρος, κατάλογος λέξεων, οντολογία κτλ.)
 - **γλωσσική περιγραφή** (π.χ. γλωσσικό μοντέλο, υπολογιστική γραμματική, τυπολογική βάση, κτλ.)
 - **εργαλείο / τεχνολογία** (π.χ. λημματοποιητής, επισημειωτής, εργαλείο αυτόματης μετάφρασης κτλ.)

Τυπολογία πόρων (2)

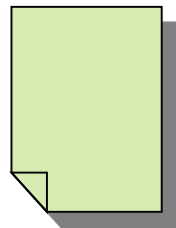
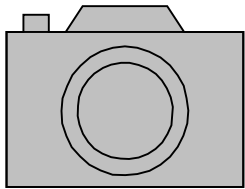
- mediaType: text, audio, image, video



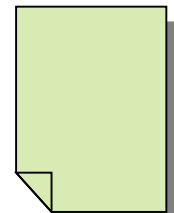
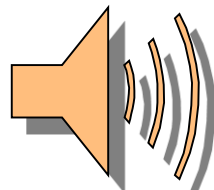
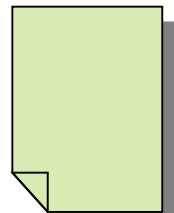
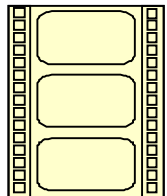
written corpora



spoken corpora



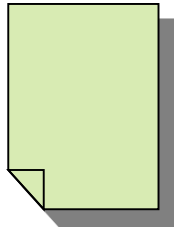
images (multimedia)



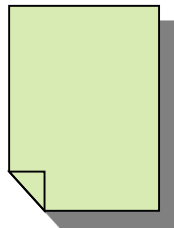
videos
(multimedia)

Τυπολογία πόρων (3)

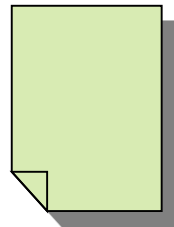
- search for text



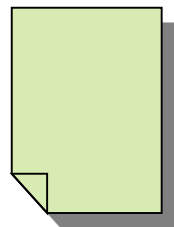
written corpora



spoken corpora

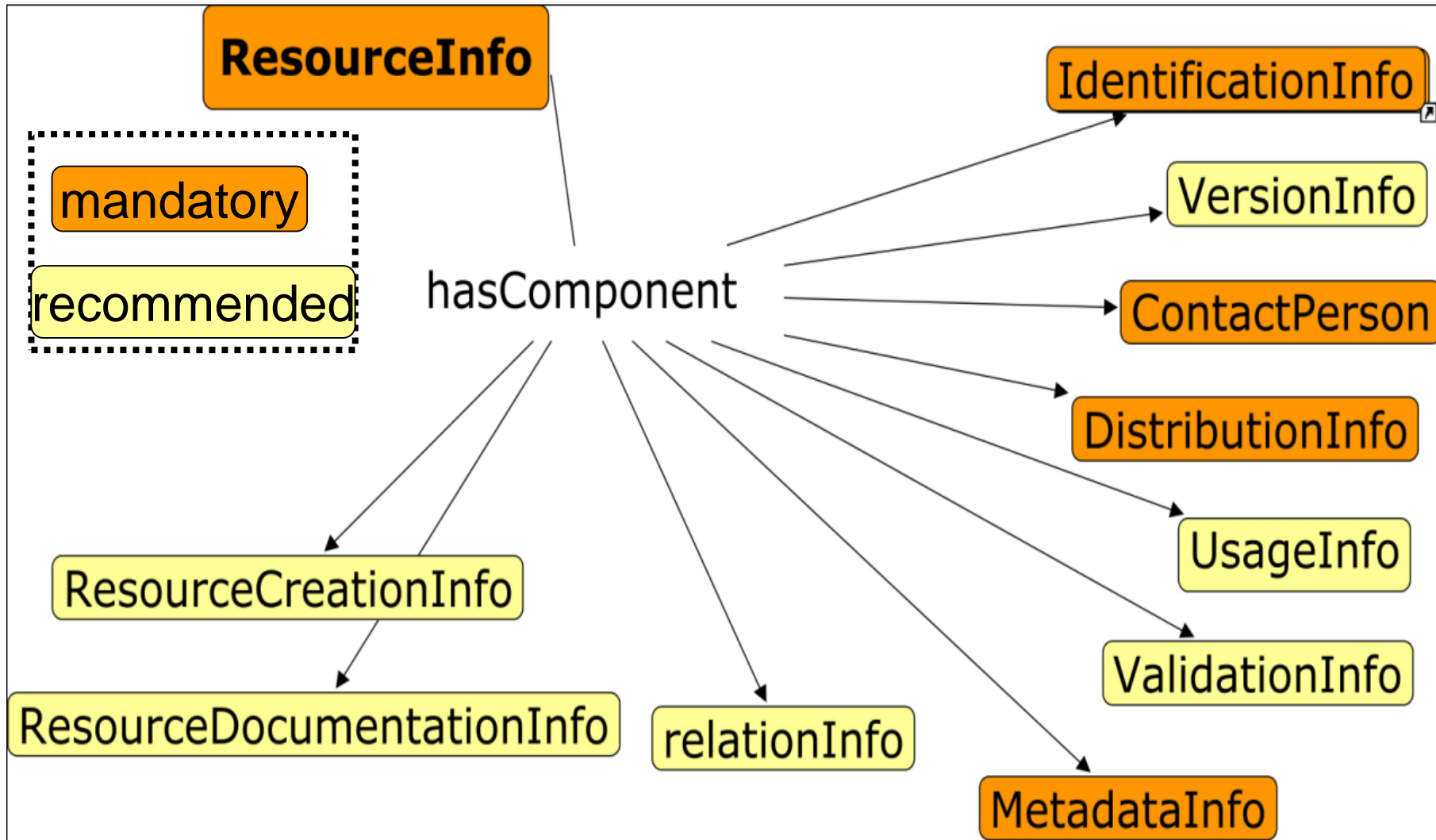


images (multimedia)

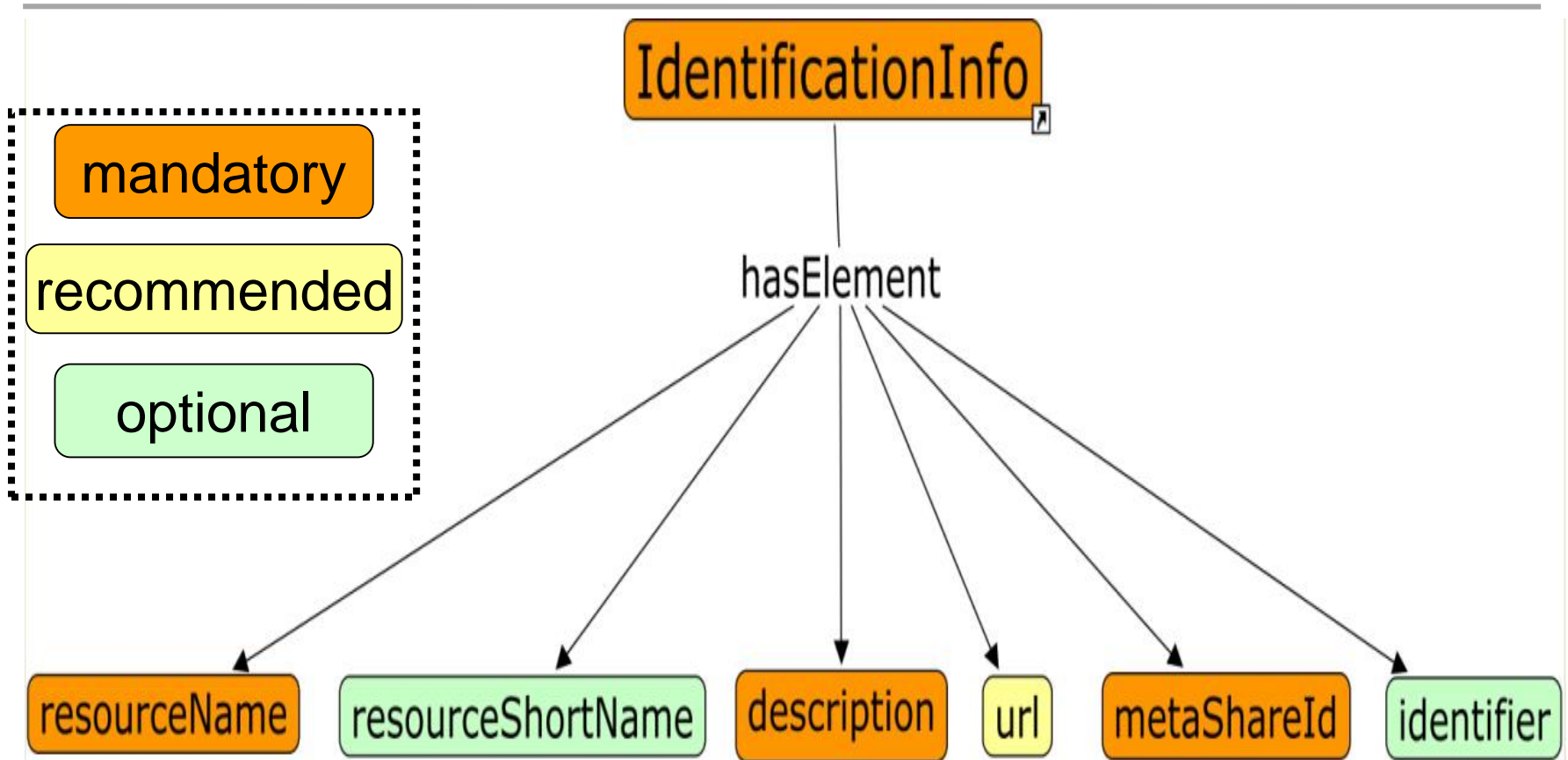


videos
(multimedia)

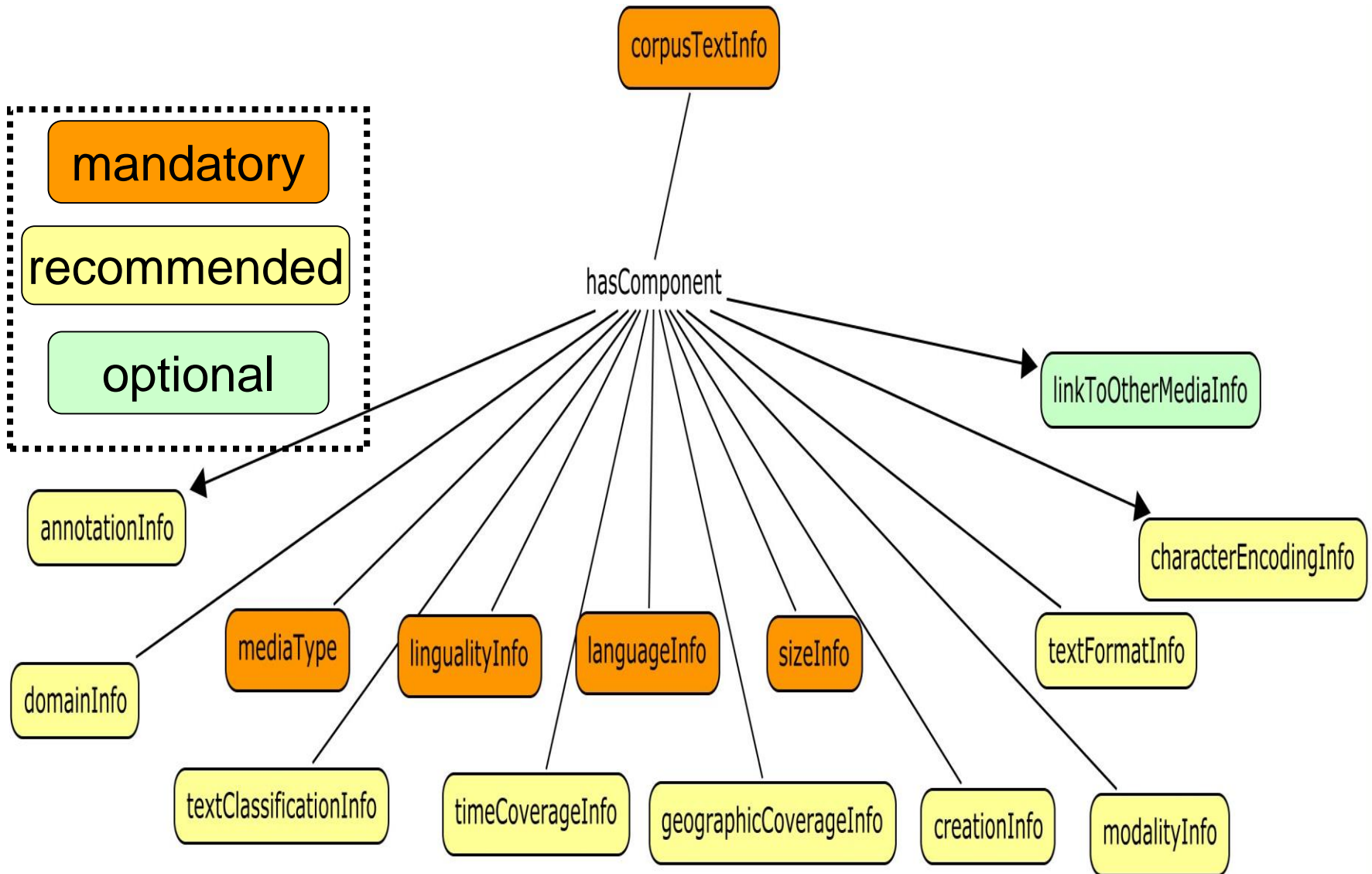
Το βασικό συστατικό περιγραφής



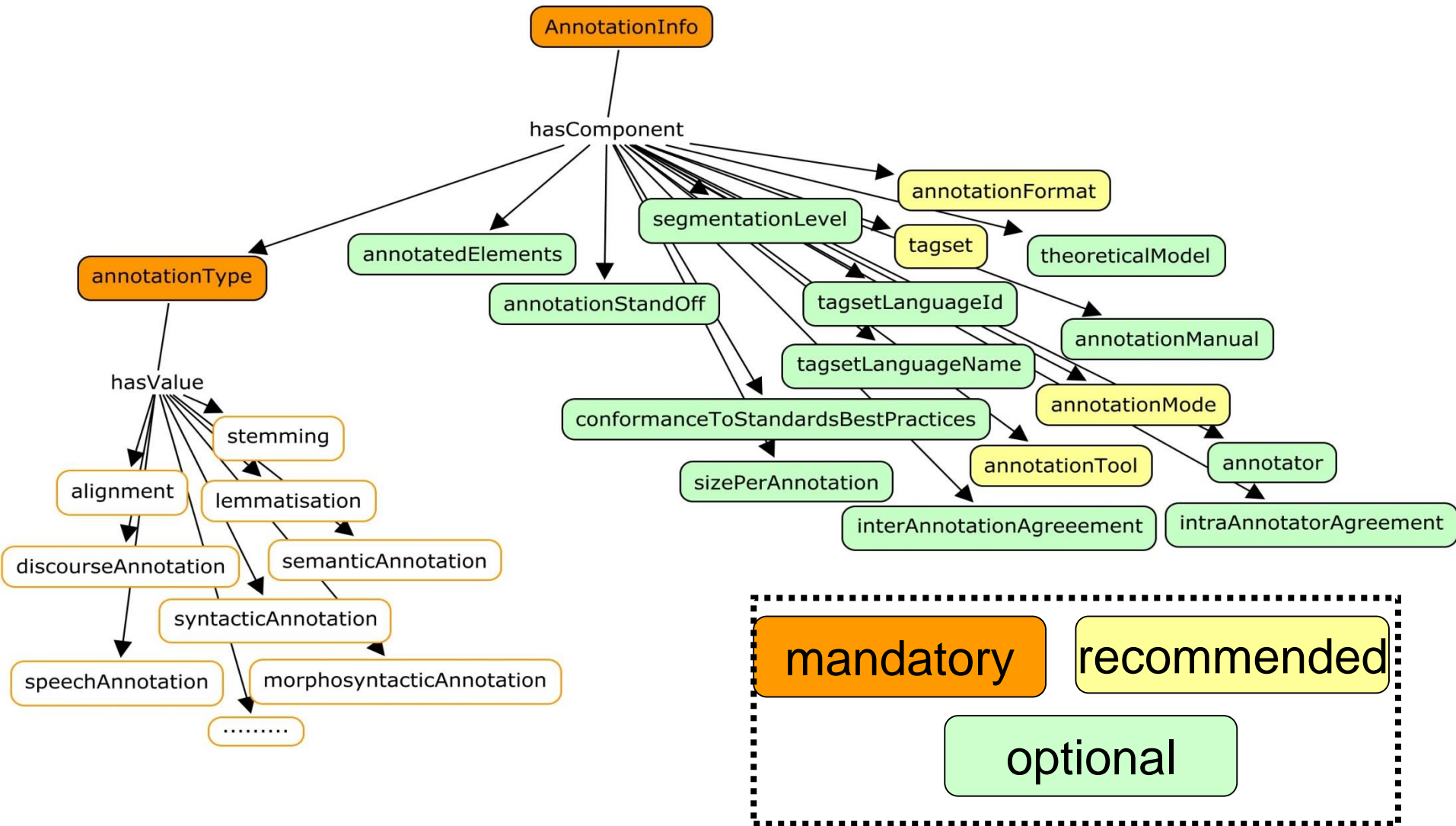
Identification component



corpusTextInfo



Annotation Component



...και σε πιο εύχρηστη μορφή

Τεκμηρίωση μονογλωσσικού ΣΚ

CLARIN-EL

Logout

Browse Resources
Manage Resources
Administration
Community
Statistics
Help
About
Your Profile, Juli

Home > Resources > Resources > Greek Textual Entailment Corpus

Change Resource ⓘ

Fields marked with * are required.

Upload resource
Export Resource Description to XML
History
View on site ➔

Administrative Information

Required

Recommended

Edit Corpus Text Info

Add Corpus Text Info

Required administration information: Identification, Distribution, Contact person, Metadata

Identification ⓘ

<p>* Resource name:</p>	<p>* Ελληνικό Σώμα Κειμενικής Συνεπαγωγής GTEC</p> <hr/> <p>* Greek Textual Entailment Corpus</p> <p style="font-size: x-small; color: green;">+ Add Another Field</p> <p style="font-size: x-small;">The full name by which the resource is known; the element can be repeated for the different language versions using the "lang" attribute to specify the language. ⓘ</p>	<p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <input type="text" value="el"/> <small>(Greek, Modern (1453-))</small>
<p>* Description:</p>	<p>* Το Ελληνικό Σώμα Κειμενικής Συνεπαγωγής (Greek Textual Entailment Corpus, GTEC) αποτελείται από 600 ζεύγη T-H (κείμενα συνεπαγωγής & συνεπαγόμενη υπόθεση) τα οποία έχουν επιστημωθεί ως προς το αν η πρόταση T συνεπάγεται την πρόταση H από ανθρώπινους επιστημωτές σύμφωνα με το σχήμα επιστημωσίας που χρησιμοποιήθηκε στις δοκιμασίες RTE1 & RTE2.</p> <p>* Το δεδομένο είναι οργανωμένο σε τρεις υποομάδες, οι οποίες αντιστοιχούν σε τρεις εφορμογές Γλωσσικής Τεχνολογίας (συστήματα Ερωματοκρίσεων, Συγκρίσιμων Αρχείων και Μηχανικής Μετάφρασης), και ανήκουν σε τρεις γνωστικούς τομείς (νομικό, πολιτική & ταξίδια). Στο σώμα περιλαμβάνονται</p> <p>* GTEC consists of 600 T-H pairs manually annotated for entailment (i.e. whether T entails H or not) by human annotators. The dataset which is tailored to guide training and evaluation of prospect RTE systems, is equally divided in three subsets each one representing the output of a specific HLT application: Question Answering (QA), Comparable Documents (CD) and Machine Translation (MT), and pertaining to specific subject fields (e.g. law, politics, travel). T-H examples that correspond to success and failure cases of the afore-mentioned applications have been included in the corpus. The annotations provided are conformant to the RTE1 and RTE2 challenges.</p> <p style="font-size: x-small; color: green;">+ Add Another Field</p> <p style="font-size: x-small;">Provides the description of the resource in prose; the element can be repeated for the different language versions using the "lang" attribute to specify the language. ⓘ</p>	<p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <input type="text" value="el"/> <small>(Greek, Modern (1453-))</small>
<p>Resource short name:</p>	<p>* GTEC</p> <p style="font-size: x-small; color: green;">+ Add Another Field</p> <p style="font-size: x-small;">The short form (abbreviation, acronym etc.) used to identify the resource; the element can be repeated for the different language versions using the "lang" attribute to specify the language. ⓘ</p>	<p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <input type="text" value="en"/> <small>(English)</small>
<p>Url:</p>	<p>* <input type="text"/></p>	

Τεκμηρίωση μονογλωσσικού ΣΚ (2)

Change Corpus text

Fields marked with * are required.

Information



Required



Recommended


Optional

Required information: Linguality, Language, Size

Linguality



* **Linguality type:** 
Indicates whether the resource includes one, two or more languages 



Multilinguality type: 
Indicates whether the corpus is parallel, comparable or mixed 


Multilinguality type details:
Provides further information on multilinguality of a resource in free text 




Languages




Language: Greek, Modern (1453-) ?

* **Language id:** 
The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) 



* **Language name:** 
A human understandable name of the language that is used in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) 



Language script:
Specifies the writing system used to represent the language in form of a four letter code as it is defined in ISO-15924 


Size per language:  
Provides information on the size per language component 

Language variety:  
Groups information on language varieties occurred in the resource (e.g. dialects) 

Language: #2

* **Language id:** 
The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) 

* **Language name:** 
A human understandable name of the language that is used in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) 

Language script:
Specifies the writing system used to represent the language in form of a four letter code as it is defined in ISO-15924 

Τεκμηρίωση πολυγλωσσικού ΣΚ

[Browse Resources](#) [Manage Resources](#) [Administration](#) [Community](#) [Statistics](#) [Help](#) [About](#) [Your Profile, Penny](#)
Logout

CLARIN-EL

Home > Resources > Resources > A parallel subcorpus collected from the European Constitution (EN-EL) (TMX)

Change Resource
Upload resource
Export Resource Description to XML
History
View on site

Fields marked with * are required.

Administrative Information

Required

Recommended

Edit Corpus Text Info

Add Corpus Text Info

Required administration information: Identification, Distribution, Contact person, Metadata

Identification

<p>* Resource name:</p> <div style="border: 1px solid #ccc; padding: 5px; min-height: 40px;">A parallel subcorpus collected from the European Constitution (EN-EL) (TMX)</div> <p style="text-align: center; font-size: 0.8em; margin-top: 5px;"> Add Another Field</p> <p style="font-size: 0.8em; margin-top: 5px;">The full name by which the resource is known; the element can be repeated for the different language versions using the "lang" attribute to specify the language. </p>	<p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;">en</div> <p>(English)</p>
<p>* Description:</p> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 5px;"> <p>* Το EUconst subcorpus EN-EL (TMX) 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).</p> </div> <div style="border: 1px solid #ccc; padding: 5px;"> <p>* The EUconst subcorpus EN-EL (TMX) is a parallel subcorpus for English and Greek, subset of the EUconst, a parallel corpus collected from the European Constitution.</p> <p>21 languages, 210 bitexts total number of files: 986</p> </div>	<p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;">el</div> <p>(Greek, Modern (1453-))</p> <p>Language of this entry (RFC 3066 code, preferably from ISO 639-1):</p> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;">en</div> <p>(English)</p>

Τεκμηρίωση πολυγλωσσικού ΣΚ (2)

Change Corpus text | Clarin EL backend - Mozilla Firefox

athena-test.clarin.gr/editor/resources/corpusinfotype_model/41/?_popup=1

Change Corpus text

Fields marked with * are required.

Information

Required

Recommended

Optional

Required information: Linguality, Language, Size

Linguality

* **Linguality type:** Indicates whether the resource includes one, two or more languages

Multilinguality type: Indicates whether the corpus is parallel, comparable or mixed

Multilinguality type details: Provides further information on multilinguality of a resource in free text


Languages

Language: English ? Delete

* **Language id:** The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines)

* **Language name:** A human understandable name of the language that is used in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines)

Τεκμηρίωση πολυγλωσσικού ΣΚ (3)


CLARIN-EL
[Logout](#)

[Browse Resources](#)
[Manage Resources](#)
[Administration](#)
[Community](#)
[Statistics](#)
[Help](#)
[About](#)
[Your Profile, Penny](#)

Resource short name: ✘

[+ Add Another Field](#)

The short form (abbreviation, acronym etc.) used to identify the resource; the element can be repeated for the different language versions using the "lang" attribute to specify the language. [?](#)

Language of this entry (RFC 3066 code, preferably from ISO 639-1):

 (English)

Url: ✘

[+ Add Another Field](#)

A URL used as homepage of an entity (e.g. of a person, organization, resource etc.) and/or where an entity (e.g.LR, document etc.) is located [?](#)

Identifier: ✘

[+ Add Another Field](#)

A reference to the resource like a pid or an internal identifier used by the resource provider [?](#)

Τεκμηρίωση πολυγλωσσικού ΣΚ (4)

Change Corpus text | Clarin EL backend - Mozilla Firefox

athena-test.clarin.gr/editor/resources/corpusinfotype_model/41/?_popup=1

Change Corpus text

Fields marked with * are required.

Information

Required

Recommended

Optional

Required information: Linguality, Language, Size

Linguality

* **Linguality type:** Indicates whether the resource includes one, two or more languages

Multilinguality type: Indicates whether the corpus is parallel, comparable or mixed

Multilinguality type details: Provides further information on multilinguality of a resource in free text

Languages

Language: English ? Delete

* **Language id:** The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines)

* **Language name:** A human understandable name of the language that is used in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines)

Τεκμηρίωση πολυγλωσσικού ΣΚ (5)

Change Corpus text | Clarin EL backend - Mozilla Firefox

athena-test.clarin.gr/editor/resources/corpusinfotype_model/41/?_popup=1

Language: Greek, Modern (1453-) ? Delete

* **Language id:**
 The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) ⓘ

* **Language name:**
 A human understandable name of the language that is used in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) ⓘ

Language script:
 Specifies the writing system used to represent the language in form of a four letter code as it is defined in ISO-15924 ⓘ

Size per language:
 Provides information on the size per language component ⓘ

Language variety:
 Groups information on language varieties occurred in the resource (e.g. dialects) ⓘ

Language: #3

* **Language id:**
 The identifier of the language that is included in the resource or supported by the tool/service; an autocompletion mechanism with values from the ISO 639 is provided in the editor, but the values can be subsequently edited for further specification (according to the IETF BCP47 guidelines) ⓘ

* **Language name:**
 A human understandable name of the language that is used in the resource or supported by the

Τεκμηρίωση πολυγλωσσικού ΣΚ (6)

Change Corpus text | Clarin EL backend - Mozilla Firefox

athena-test.clarin.gr/editor/resources/corpusinfotype_model/41/?_popup=1

Language script:
Specifies the writing system used to represent the language in form of a four letter code as it is defined in ISO-15924 ⓘ

Size per language: Provides information on the size per language component ⓘ

Language variety: Groups information on language varieties occurred in the resource (e.g. dialects) ⓘ

[Add another Language](#)

Sizes ⓘ

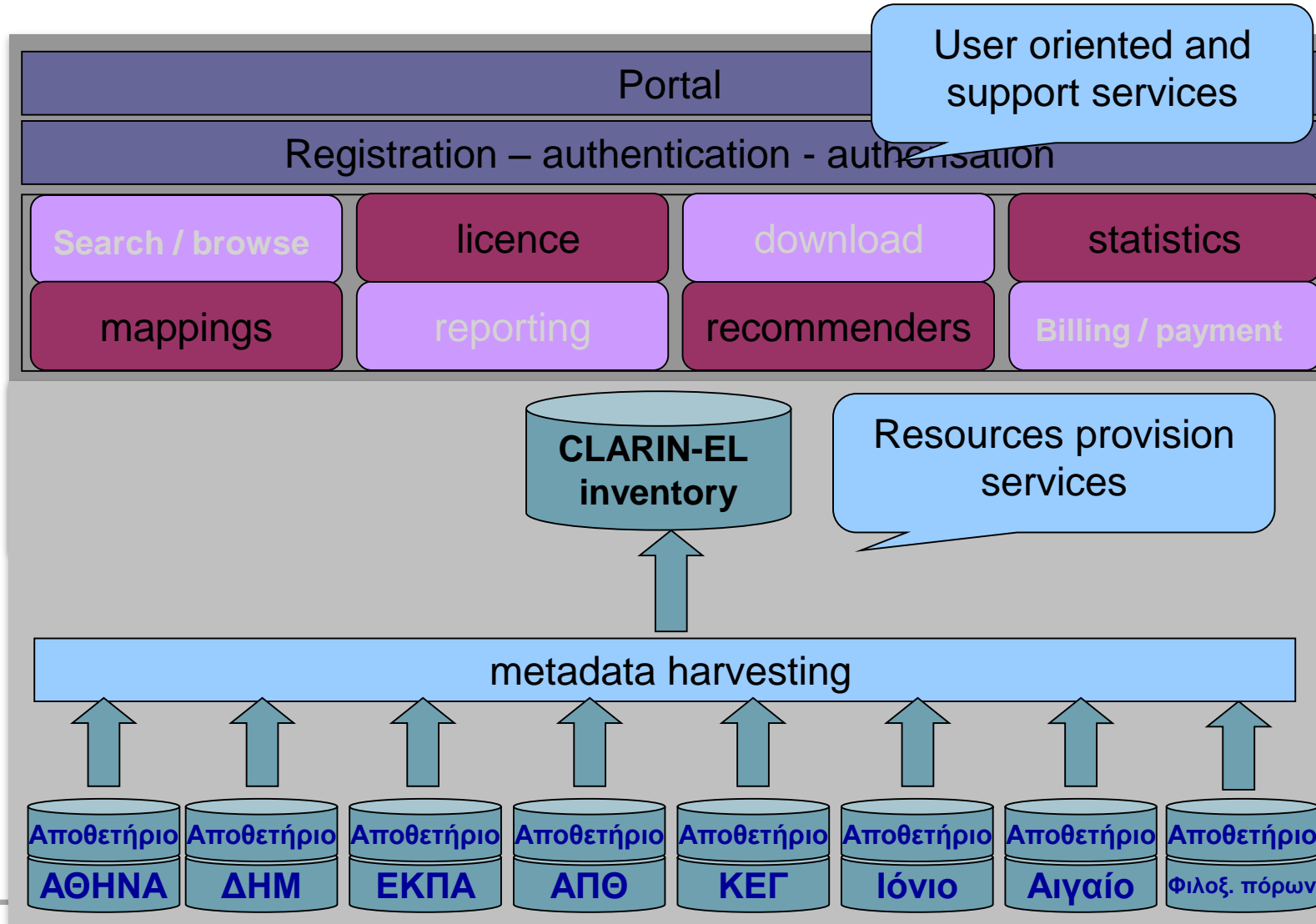
* Size ⓘ	* Size unit ⓘ	Delete?
270.000 Words <input type="text" value="270.000"/>	Words	<input type="checkbox"/>
611 Kb <input type="text" value="611"/>	Kb	<input type="checkbox"/>
<input type="text"/>	-----	<input type="checkbox"/>

[Add another Size](#)

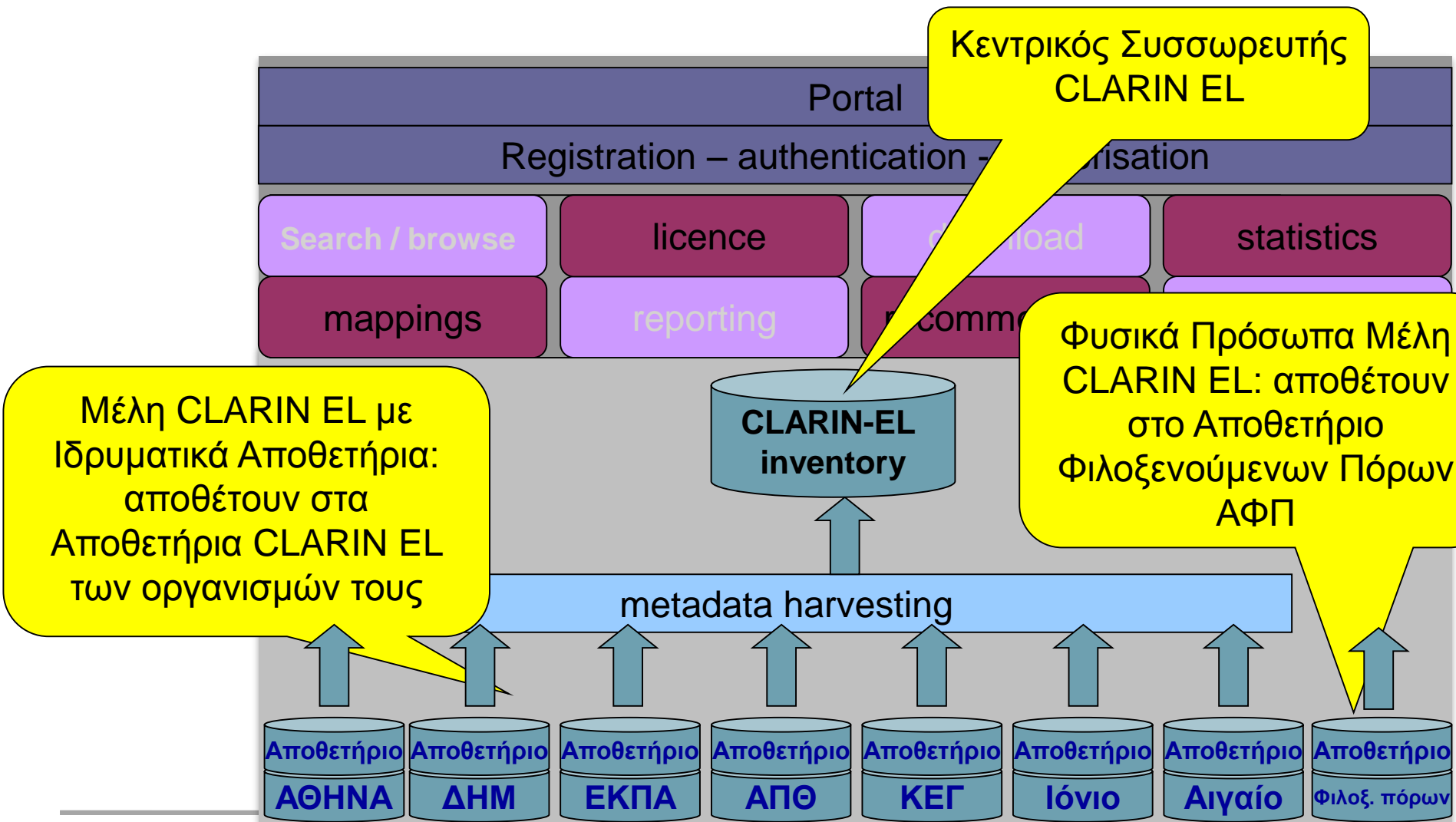
Save and continue editing Cancel Save

Αρχιτεκτονική δικτύου και υποδομής

Αρχιτεκτονική resources infrastructure



Αρχιτεκτονική – Σύσταση Δικτύου



Μέλη Δικτύου

- ◆ Κατασκευαστική Φάση (μέχρι 30/9/2015)
 - Υλοποίηση Υποδομής
 - ΕΚ «Αθηνά»
 - ΕΚΕΦΕ «Δημόκριτος»
 - Εθνικό Δίκτυο Έρευνας Τεχνολογίας (ΕΔΕΤ) ΑΕ
 - Χρήση και εμπλουτισμός
 - ΕΚ Πανεπιστήμιο Αθηνών
 - Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
 - Ιόνιο Πανεπιστήμιο
 - Πανεπιστήμιο Αιγαίου
 - Κέντρο Ελληνικής Γλώσσας
- ◆ Πιλοτική Φάση (από 1/1/2016)
 - Όλοι οι υπόλοιποι ακαδημαϊκοί φορείς (πανεπιστήμια, ερευν. κέντρα)

Τεχνικά συστατικά

- ◆ META-SHARE+ repository building software
- ◆ Πρωτόκολλα συγκομιδής (OAI-PMH, META-SHARE harvesting/syncing protocol)
- ◆ Metadata schemas (META-SHARE, CMDI (CLARIN), DC)
 - Μετασχηματισμοί μεταξύ META-SHARE-CMDI, META-SHARE-DC
- ◆ Αλυσίδες γλωσσικής επεξεργασίας & πρότυπα επισημείωσης και αναπαράστασης
 - http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page
 - <http://www.panacea-lr.eu/>
 - <http://qt21.metashare.ilsp.gr/>

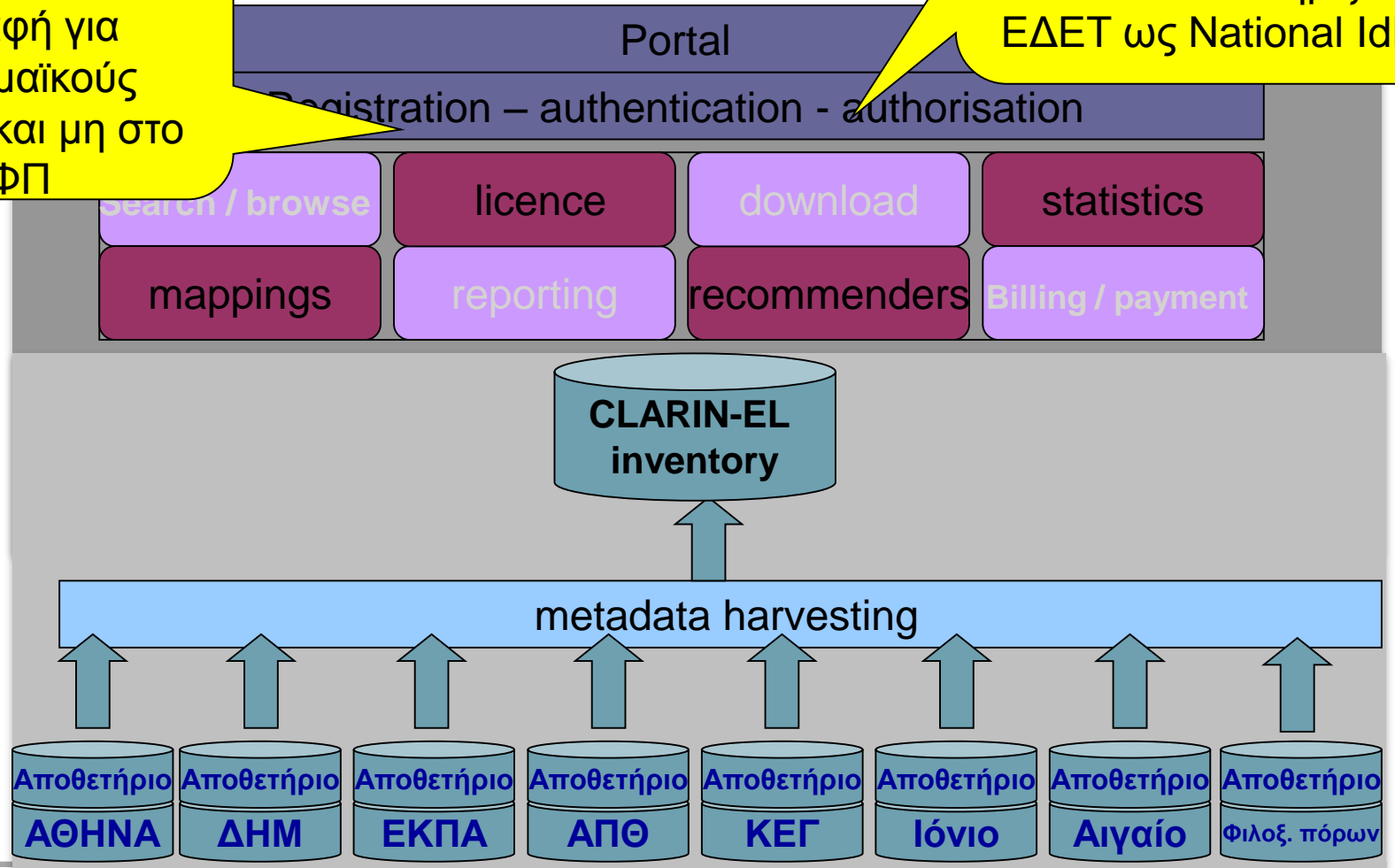
Υπολογιστική Υποδομή

- ◆ Όλα τα συνιστώμενα υποσυστήματα προσαρμόζονται και αναπτύσσονται στην υπολογιστική υποδομή της ΕΔΕΤ
- ◆ Okeanos = cyclades & pithos
- ◆ cyclades : virtual machines and networking
- ◆ Pithos : virtual storage
- ◆ Βέλτιστη διαθεσιμότητα και δυναμική χρήση υπολογιστικών πόρων
- ◆ Ασφάλεια και διατήρηση πόρων, τόσο πρωτογενών (στο περιβάλλον του pithos, όσο και δευτερογενών (metadata databases στο περιβάλλον των cyclades)

Αρχιτεκτονική – Πρόσβαση (1)

«Κλασσική» εγγραφή για ακαδημαϊκούς χρήστες και μη στο ΑΦΠ

Είσοδος με ακαδημαϊκό λογαριασμό (SSO) μέσω της ομοσπονδίας ΑΑΙ που υποστηρίζει η ΕΔΕΤ ως National IdP

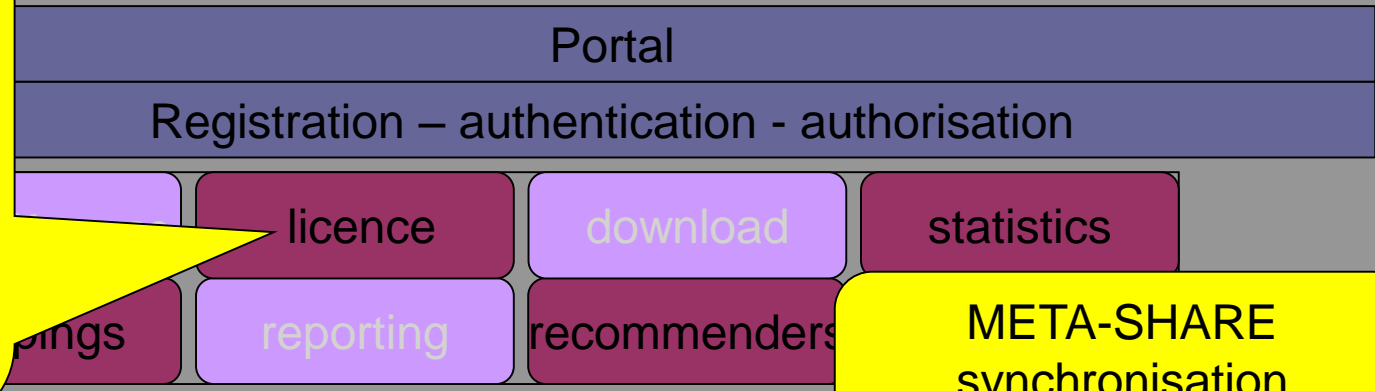


Αρχιτεκτονική – Πρόσβαση (2)

- ◆ Όλοι οι συνεργαζόμενοι πάροχοι (ΕΚΠΑ, ΑΠΘ, ΙΟΝΙΟ, ΑΙΓΑΙΟ, ΚΕΓ) έχουν δικαίωμα ως ακαδημαϊκοί φορείς να είναι μέλη της ΑΑΙ ομοσπονδίας.
- ◆ ΕΚΠΑ, ΑΠΘ, ΙΟΝΙΟ, ΑΙΓΑΙΟ είναι ήδη μέλη της ομοσπονδίας
- ◆ Όλοι οι δυνάμει χρήστες των ΕΚΠΑ, ΑΠΘ, ΙΟΝΙΟ, ΑΙΓΑΙΟ θα έχουν δυνατότητα πρόσβασης με το ακαδημαϊκό τους login.
- ◆ Και μόνον με αυτό!

Αρχιτεκτονική – Πρότυπα

Creative Commons licences 4.0, META-SHARE NoRedistribution licences 1.0, CLARIN licences (forthcoming), FOSS licences



META-SHARE synchronisation protocol, OAI-PMH



META-SHARE metadata schema



Handle based PIDs



Πρόσβαση – Διαχείριση χρηστών

- ◆ Βασικές διχοτομήσεις
 - Πάροχοι
 - Καταναλωτές

- ◆ **Πάροχοι**
 - Ακαδημαϊκοί χρήστες -> Αποθετήρια ακαδημαϊκών οργανισμών (ιδρυματικά αποθετήρια)
 - Μη ακαδημαϊκοί χρήστες -> Αποθετήριο φιλοξενούμενων πόρων

- ◆ **Καταναλωτές**
 - Ακαδημαϊκοί χρήστες και χρήση -> Αναζήτηση, καταφόρτωση και επεξεργασία υλικού χωρίς περιορισμούς ή με τους ελάχιστους δυνατούς
 - Μη ακαδημαϊκοί χρήστες -> Αναζήτηση και καταφόρτωση σύμφωνα με τους περιορισμούς αδειοδότησης και επεξεργασία με περιορισμούς όγκου υλικού

Πρόσβαση – Διαχείριση παρόχων και πόρων

- ◆ Βασικοί ρόλοι παρόχων
 - Διαχειριστής αποθετηρίου (τεχνικός-επιστημονικός)
 - Τεκμηριωτές & ομάδες τεκμηριωτών
 - Συντονιστές ομάδων τεκμηριωτών

- ◆ Βασικές καταστάσεις πόρων
 - Εσωτερικής δημοσίευσης
 - Περιορισμένης δημοσίευσης
 - Πλήρους δημοσίευσης

Υποστήριξη

- ◆ Για τη δημιουργία αποθετηρίου CLARIN EL
 - επικοινωνία στο clarin-el-coordinator@clarin.gr, και clarin-el-techadmin@clarin.gr
 - αυτοματοποιημένη διαδικασία, με σύντομη συνεργασία του Διαχειριστή Αποθετηρίου με τον CLARIN EL Τεχνικό Διαχειριστή της Υποδομής
- ◆ Γραφεία υποστήριξης σχετικά με
 - Νομικά θέματα → legal-helpdesk@clarin.gr
 - Τεχνικά θέματα → technical-helpdesk@clarin.gr
 - Θέματα τεκμηρίωσης → metadata-helpdesk@clarin.gr

Εθνικό δίκτυο CLARIN-EL

CLARIN-EL
Common Language Resources and Technology Infrastructure



Αρχική Σελίδα
CLARIN
CLARIN-EL
Χαρτογράφηση
Συχνές Ερωτήσεις
Δημοσιεύσεις
Παραδοτέα
Επικοινωνία

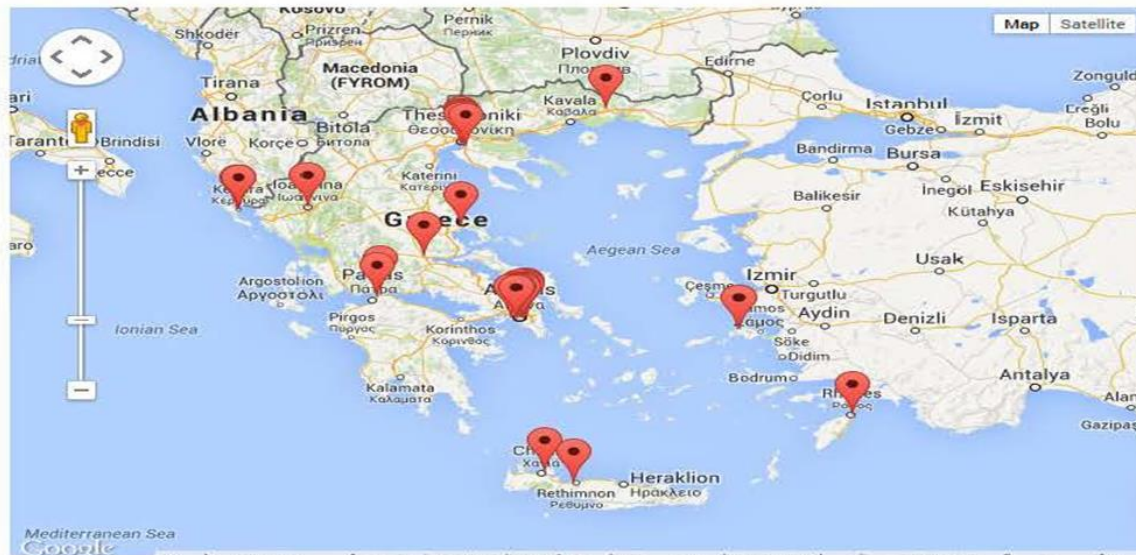
Επιλογές εμφάνισης

Μπορείτε να περιορίσετε τη θέαση των αποτελεσμάτων σε ορισμένες κατηγορίες γλωσσικών πόρων και εφαρμογών, επιλέγοντας τις αντίστοιχες κατηγορίες και επιβεβαιώνοντας την επιλογή σας με το πλήκτρο "Φίλτρο".

- Πόροι πρωτογενούς υλικού
- Πόροι επεξεργασμένου υλικού
- Πόροι αναφοράς
- Εργαλεία/Εφαρμογές Γλωσσικής Τεχνολογίας

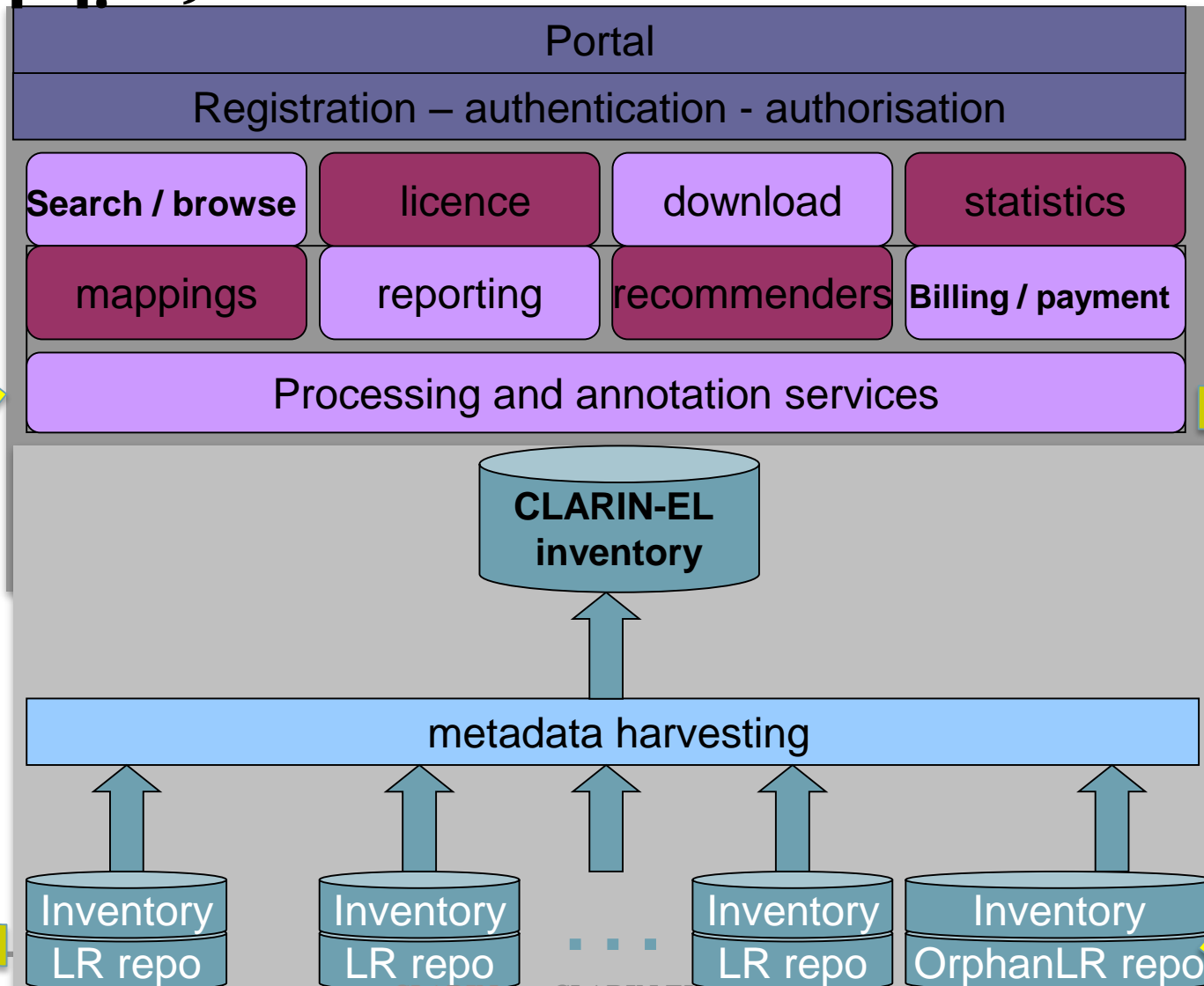
Φίλτρο

Κατάργηση φίλτρου



τι δουλεύουμε τώρα

Αρχιτεκτονική – Υπηρεσίες επεξεργασίας (το 6^ο βήμα)



Υπηρεσίες γλωσσικής επεξεργασίας

- ◆ Υπηρεσίες επεξεργασίας μονογλωσσικών σωμάτων κειμένων (ΣΚ)
 - Αναγνώριση λέξεων και προτάσεων (tokenisation and sentence splitting)
 - Μορφοσυντακτική ανάλυση (part of speech tagging)
 - Λημματοποίηση (lemmatisation)
 - Συντακτική ανάλυση (syntactic parsing)
 - Αναγνώριση και εξαγωγή ορολογίας (term recognition and extraction)
 - Αναγνώριση οντοτήτων (named entity recognition)

- ◆ Υπηρεσίες επεξεργασίας πολυγλωσσικών σωμάτων κειμένων
 - Για Ελληνικά-Αγγλικά
 - Οι παραπάνω για κάθε γλώσσα του παράλληλου ΣΚ
 - Για Ελληνικά-X (=γλώσσα της ΕΕ)
 - Στοίχιση σε επίπεδο πρότασης, ...

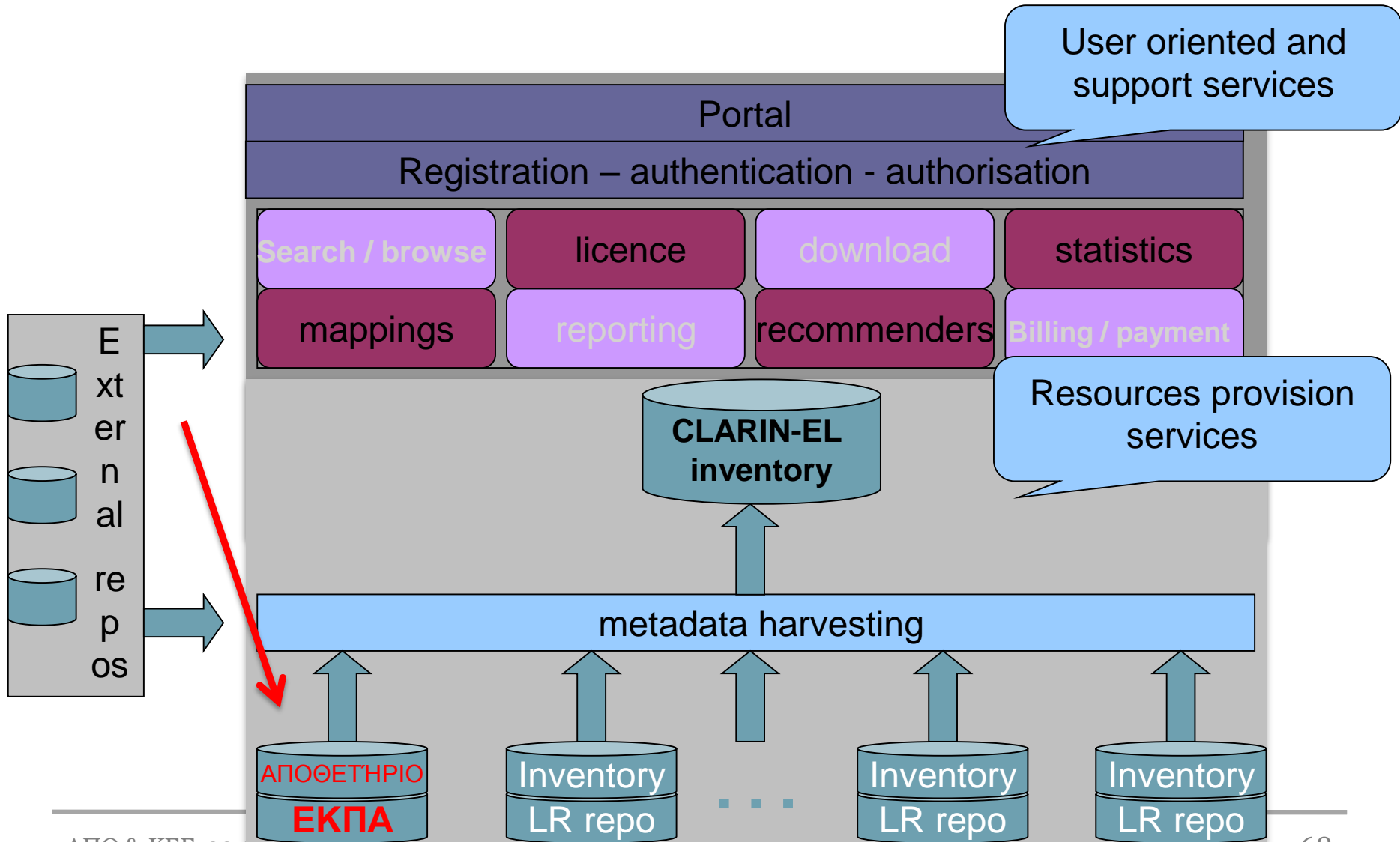
Επεξεργάσιμα δεδομένα

- ◆ Εκείνα που ικανοποιούν τις προδιαγραφές δεδομένων και υπηρεσιών
 - ως προς γλώσσα (Ελληνικά, Αγγλικά, ...)
 - αναμενόμενο μορφότυπο (format) εισόδου
 - αναμενόμενη προεπεξεργασία
 - ένα ήδη μορφοσυντακτικά επισημειωμένο σύνολο δεδομένων μπορεί να
 - αναλυθεί συντακτικά, κλπ.
 - ανοικτά δικαιώματα χρήσης
 - που επιτρέπουν παράγωγα «έργα», δηλ. δημιουργία παράγωγων δεδομένων
 - ανοικτά διατιθέμενες υπηρεσίες επεξεργασίας
- ◆ Στην περίπτωση πιο σύνθετων εργασιών
 - Πλήρης αυτοματοποίηση όλης της ροής επεξεργασίας (workflow)

Επεξεργασία δεδομένα (2)

- ◆ τα προς επεξεργασία δεδομένα μπορούν να
 - βρίσκονται ήδη αποθηκευμένα σε κάποιο Ιδρυματικό Αποθετήριο, ή στο Αποθετήριο Φιλοξενούμενων Πόρων
 - φορτωθούν στην Υποδομή από τον χρήστη
 - προσωρινά
 - να ελεγχθούν ως προς την συμβατότητά τους με τις προδιαγραφές
- ◆ τα επεξεργασμένα δεδομένα συνιστούν νέα δεδομένα
 - που αποθηκεύονται στο ίδιο Ιδρυματικό Αποθετήριο με τα πρωτογενή δεδομένα, και συσχετίζονται οι περιγραφές τους
 - αν τα πρωτογενή τους δεδομένα φορτώθηκαν από τον χρήστη,
 - παραμένουν για 48 ώρες και στη συνέχεια διαγράφονται από την Υποδομή
 - αποθηκεύονται μόνιμα εφόσον ο χρήστης γίνει μέλος

Το ΕΚΠΑ στο CLARIN EL



Εμπλουτισμός του Αποθετηρίου ΕΚΠΑ

Ερευνητική Υποδομή Γλωσσικών Πόρων, Τεχνολογιών και Υπηρεσιών CLARIN EL
Ημερίδα ΕΚΠΑ, 9.3.2015



Στοιχεία ταυτότητας πόρου

όνομα	
ακρωνύμιο (αν υπάρχει)	
περιγραφή (σύντομη περιγραφή σε ελεύθερο κείμενο)	
μέγεθος πόρου (σε όποια μονάδα εξυπηρετεί: bytes, λέξεις, λεπτά...)	
γλώσσες που περιλαμβάνονται	
μορφώματος αρχείων (txt, doc, xml, tnx, mp4, avi, ...)	

Τύπος πόρου

γραπτά κείμενα	
προφορικός λόγος (μεταγραμμένος)	
ήχος	
βίντεο	
λεξικό / γλωσσάρι / θησαυρός	
άλλο (να διευκρινιστεί)	

Στοιχεία διάθεσης πόρου

διαθεσιμότητα (π.χ. ελεύθερα διαθέσιμος, διαθέσιμος με περιορισμούς (ποιους))	
τύπος άδειας (εάν έχει ήδη, π.χ. CC-BY...)	
όνομα κατόχου πνευματικών δικαιωμάτων	

Στοιχεία υπεύθυνου για επικοινωνία

ονοματεπώνυμο	
θέση / ιδιότητα	
σχολή / τμήμα / φορέας	
τηλέφωνο	
email	

Το CLARIN EL στο παρόν και μελλοντικό οικοσύστημα υποδομών

Apollonis

- ◆ Από τον Νοέμβριο 2014, το CLARIN EL έχει ενταχθεί, με τη νέα ονομασία CLARITAS στον Εθνικό Χάρτη Ερευνητικών Υποδομών 2014-2020
- ◆ από κοινού με την ΕΥ DARIAH-ΔΥΑΣ, τη εθνική υποδομή για Έρευνα στις Ανθρωπιστικές Επιστήμες, σχημάτισε την κοινοπραξία Apollonis
- ◆ με στόχο την ενίσχυση της διεπιστημονικής έρευνας, την υποστήριξη της μετάβασης των Κοινωνικών και Ανθρωπιστικών επιστημών στην ψηφιακή εποχή
- ◆ Την υποστήριξη του πολιτισμού και της πολιτιστικής κληρονομιάς

Το CLARIN EL στο οικοσύστημα EY

- ◆ Το CLARIN EL αποτελεί μέρος ενός ευρύτερου οικοσυστήματος ερευνητικών υποδομών και συνεργάζεται στενά με
 - META-NET
 - OPENAIRE
 - OPENMINTED
 - LANGUAGE GRID
 - LAPPS GRID

Προκλήσεις σήμερα και αύριο...

-
- ◆ Πρότυπα κωδικοποίησης και αναπαράστασης γλωσσικών πόρων και γλωσσικής γνώσης με σκοπό τη μεγιστοποίηση της διαλειτουργικότητας μεταξύ τους
 - ◆ Προτυποποίηση και κατά το δυνατόν απλοποίηση του κύκλου εκκαθάρισης πνευματικών δικαιωμάτων
 - ◆ Μηχανισμοί διάδοσης και ενίσχυσης της κουλτούρας διαμοιρασμού πόρων, εργαλείων και ερευνητικών αποτελεσμάτων
 - ◆ Οργανωτικά, επιχειρησιακά και επιχειρηματικά μοντέλα για τη βιωσιμότητα των εθνικών και ευρωπαϊκών υποδομών γλωσσικής τεχνολογίας

E/A

Ευχαριστώ πολύ!