

Datu un valodas resursu vākšana: Tehniskie un praktiskie jautājumi Datu pārvaldības darbplūsma

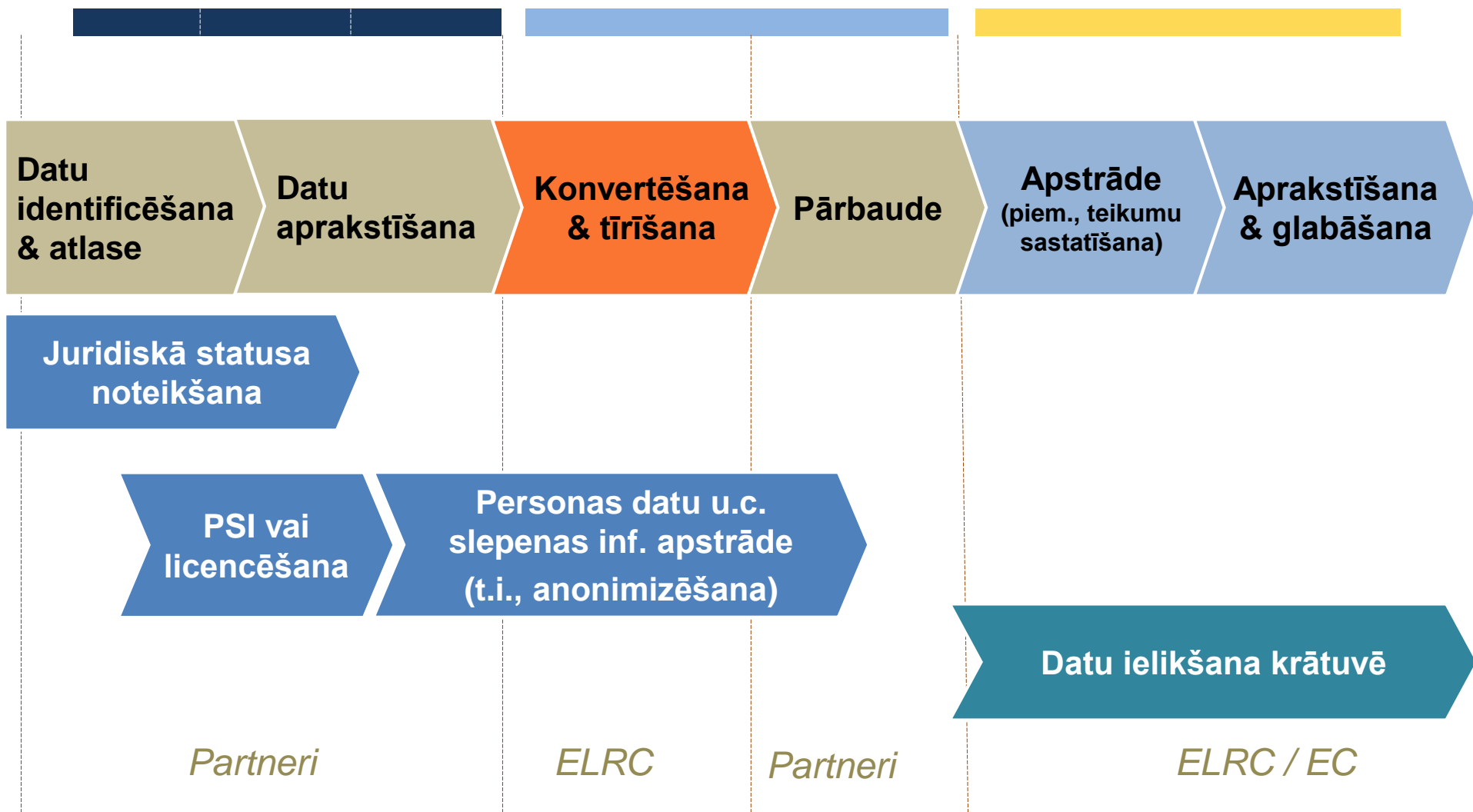
Raivis Skadiņš (ELRC/Tilde)

Khalid Choukri (ELDA)



- Par kādiem datiem mēs runājam?
(iepriekšējās prezentācijas)
- Mašīntulkošana *iemācās* tulkot no datiem
- Ir nepieciešami dati visās valodās
- Kur šos datus atrast?
 - Redzami dati, piemēram, tīmeklī (tīmekļa vietnes, atskaites, ziņojumi, u.tml.)
 - Neredzami dati: arhīvi, iekštīkli, iekšējas dokumentu krātuves u.tml.
 - Tulkošanas pakalpojumu sniedzēji

Kā dati kļūst par izmantojamiem valodas resursiem



- Datu avotu identificēšana, datu kopu identificēšana un atlase (neapstrādāti dati)
- Jau pašā sākumā jā rūpējas par privātuma un ētikas jautājumiem
(ja nepieciešams jāapsver datu anonimizēšana, varbūt dati nav izmantojami vispār)
- Datu pamatinformācijas aprakstīšana
(valodas, joma, gads, ...)
- Kā datus var piegādāt (datu formāts, datu nesējs u.tml.)

Tehniskie jautājumi datu sagatavošanas posmā



- Datu formatējuma tīrīšana (formatējuma atmešana, piem., treknraksts, slīpraksts, rakstzīmju kodējums - UTF8, attēli, reklāmas, tabulas, html tagi, u.tml.)
- Failu formātu konvertēšana (XML, XLIFF, u.tml.)
- Datu sagatavošana MT vajadzībām (piem., sastatīšana)
- Rezultāta kvalitātes kontrole (formāts, saturs, glabāšana)
- Valodas resursu apraksts (metadati)
- Datu sapakošana un piegāde EC (datu krātuve)

Ikviens digitāls datu formāts derēs





28.09.2015

Ministru prezidents

Jāstiprina Latvijas un Itālijas ekonomiskie sakari
Pirmdien, 28.septembrī, Ministru prezidente Laimdota Straujuma atvadu vizītē pieņēma Itālijas vēstnieku Latvijā Džovanni Polici (Giovanni Polizzi), izsakot pateicību par ieguldīto darbu Latvijas un Itālijas attiecību stiprināšanā.

Sarunas laikā abas puses vienojās, ka jāturpina aktīvs valstu divpusējais dialogs, tajā skaitā veicinot ekonomiskos sakarus.

L.Straujuma un Dž.Polici pārrunāja Latvijas un Itālijas sadarbību un aktualitātes trio valstu (Itālija, Latvija, Luksemburga) prezidentūrā ES Padomē.

Ministru prezidente atzinīgi novērtēja sadarbību NATO ietvaros, kā arī pateicās Itālijai par tās ieguldījumu Baltijas gaisa telpas aizsardzībā.

Home > Aktu



Share

Tweet 0

+1 0

28.09.2015

Ministru prezidents

Economic relations between Latvia and Italy to be strengthened
On Monday, September 28, Prime Minister Laimdota Straujuma held a farewell meeting with Giovanni Polizzi, Ambassador of Italy to Latvia, and expressed gratitude for his contribution to strengthening the relations between Latvia and Italy. During the meeting the parties agreed that the active bilateral dialogue between the countries should be continued, incl. by promoting economic relations.

L.Straujuma and G.Polizzi discussed the cooperation of Latvia and Italy and current developments in trio countries' (Italy, Latvia, Luxembourg) presidency of the Council of the European Union. The Prime Minister welcomed the cooperation within NATO framework and expressed gratitude to Italy for its contribution to protection of the Baltic airspace.



RU A A A

Kontakti

akari

izītē pieņēma
u par
as puses
t ekonomiskos

trio valstu

Itālijai par tās





- Juridiskā statusa noteikšana (lēmums par izmantošanu vai atmešanu)
- Vai datu anonimizēšana ir akurāta un pieņemama
- Direktīvas par valsts sektora informācijas atkalizmantošanu (PSI) vai datu licencēšanas piemērošana
- Licencēšana
 - ELRC var palīdzēt ar procedūrām
 - Dažādi licencēšanas veidi
 - Government Open Licenses
 - Standard Re-use Licenses
 - License interoperability



- Tehniskais un juridiskais atbalsts
- Datu krātuve
- Konsultācijas

- Datu avotu identificēšana (neapstrādāti dati)
 - Datus var iegūt no publiskiem avotiem (t.i., ievākt no publiskām tīmekļa vietnēm)
 - Datus var piegādāt valsts sektora institūcijas
 - Valsts sektora institūcijas var palīdzēt atklāt publiskus avotus
- Datu apstrādi ELRC un datu piegādātājs veic kopā



- Jūs zināt savus datus
 - gan publiski pieejamos (redzamos), gan neredzamos
- Piekļuve arhīviem, dziļajiem tīmekļiem, datubāzēm u.tml. resursiem bieži nav pieejama no ārpuses.
- Ne visi dati ir pieejami kā PSI vai ar atļaujošu licenci
- Atvasināto formātu (piemēram, PDF) izmantošana ir neefektīva salīdzinot ar primāro formātu izmantošanu.
- Tulkošanas atmiņu izmantošana ir vēl efektīvāka

- Ko darīt, ja identificētie dati satur informāciju par personām, uzņēmumiem u.tml.?
- Var izmantot nosaukumu atpazīšanas (NER) programmatūru, lai nodzēstu jutīgu informāciju (vārdus, nosaukumus, vietas, datumus, dzimšanas datus, u.tml.) vai to aizstātu ar vispārīgu informāciju.
- Jāpārlicinās, ka rezultāts ir pieņemams !!!
 - Jāatmet dati, ja anonimizācija nav gana pieņemama



- Esošu datu (tekstu tulkojumu) izmantošana jaunam mērķim ir vislabākais veids kā uzlabot automatizētās tulkošanas kvalitāti.
- ELRC var palīdzēt izvērtēt datu piemērotību
- Nenovērtējiet savus valodas resursus par zemu, ieplānojiet šo datu pārvaldības plānu

Palīdzības dienests un atbalsts



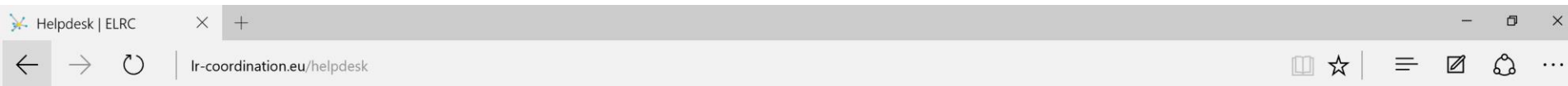
Languages — the heart of Multilingual Europe



The screenshot shows a web browser window displaying the CEF-ELRC Repository website. The browser's address bar shows the URL 194.177.192.69. The website's header is blue and contains the text "The CEF-ELRC META SHARE repository" and navigation links for "Register" and "Login". Below the header, there is a large logo for "European Language Resource Coordination META SHARE" with the tagline "Connecting Europe Facility". Underneath the logo, it states "10 language resources at your disposal". A search bar is present with the placeholder text "Type in your keywords, please..." and a blue "Search" button.

What is the European Language Resource Coordination (ELRC) action?

Public administrations in Europe need to communicate with each other and with the citizens across borders. This communication is often hampered by the language barriers created by the many languages in Europe. The European Commission is supporting public administrations and citizens in their communication tasks by building an **Automated Translation platform** within the Connecting Europe Facility (CEF) Programme.



[Home](#) [About](#) [News](#) [Helpdesk](#) [Events](#) [Resources](#) [Anchor Points](#) [Multilingual Europe](#)

European Language
Resource Coordination



Helpdesk



Helpdesk Dashboard × +

← → ↻ | cef-at-helpdesk.elda.org/overview

Forum Knowledge base Search... Register Log In



European Language
Resource Coordination
Connecting Europe Facility

Automated Translation's HelpDesk

Overview | Legal Technical Metadata General management

Other

All questions Open Closed Unanswered Answered

Overview Section

Welcome on the ELRC Helpdesk!

The ELRC Helpdesk has been set up to answer the questions on Languages Resources and Tools that users (EC data users, data providers (public, commercial, non-governmental organisations), etc.) may want to ask.