



# Automaattinen käännösalusta ja Avoimen datan portaali

## Mitä hyötyä niistä on?

Kimmo Rossi  
Euroopan komissio, CNECT.G3

*Helsinki 19.2.2016*

## Julkiset palvelut ja konekääntäminen

- *Tavoitteena tukea monikielisyyttä*
- *Konekääntämisen rooli on määriteltävä*
- *EU:n nykyinen konekäännösratkaisu: MT@EC*
- *Seuraava vaihe: CEF AT*

# Visio



*Haluan käyttää minkä tahansa EU-maan julkisia verkkopalveluja omasta maastani käsin, omalla kielelläni!*



# Konekääntämisen rooli

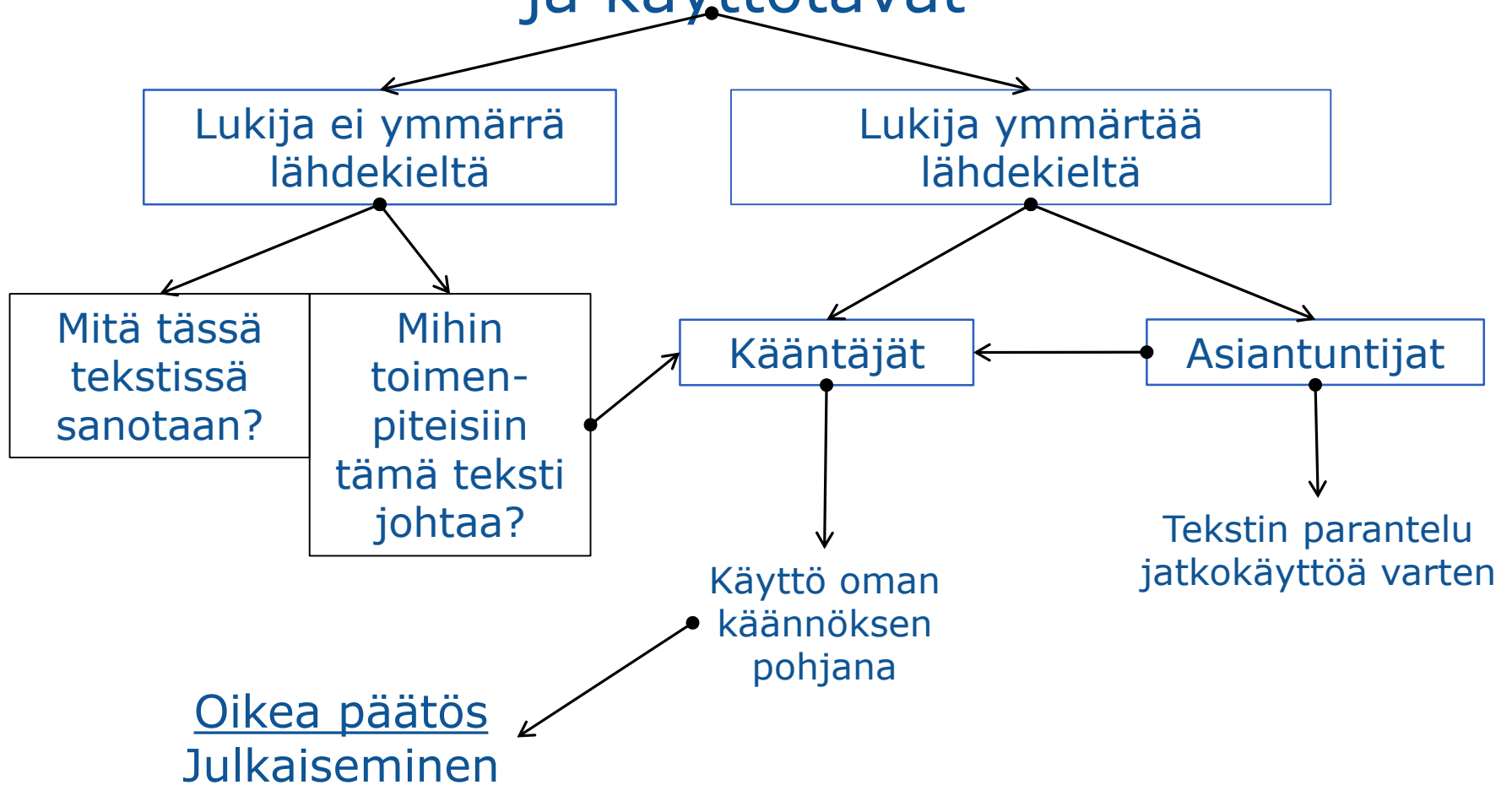
Konekääntäminen voi olla oikea ratkaisu, kun:

- halutaan ottaa selkoa vieraskielisestä asiakirjasta tai verkkosisällöstä nopeasti ja edullisesti
- pitää reagoida vieraskieliseen pyyntöön nopeasti, eikä ole aikaa ihmiskäännökseen
- verkkosivustoja monikielistetään
- *haetaan* verkosta monikielistä aineistoa tai *analysoidaan* suuria aineistoja automaattisesti

Verkkopalvelujen yleistyessä konekääntämisen tarve lisääntyy!



# Konekäännösten käyttäjät ja käyttötavat



# MT@EC on julkishallinnon käytettävissä!

**Tuotantokäyttö alkoi :** 26.6.2013

**Kielet :** Kaikki EU:n 24 virallista kieltä

**Toimintaperiaate:** tilastollinen käännin, perustuu avoimeen MOSES-alustaan, joka on tulosta useasta EU-tekniologiahankkeesta (esim. EuroMatrix, EuroMatrix+, MosesCore)

**Kehittäjä: DGT** vuosina 2010-2013

**Rahoituslähde:** ISA-ohjelma (toimenpide 2.8)

[http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action\\_en.htm](http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm)

# MT@EC kuvaus

**Käyttötavat:** - käyttöliittymä (käyttäjänä ihminen)  
- sovellusliittymä (käyttäjänä tietojärjestelmä)

## Ominaisuudet:

- **Käyttöliittymä** 24 kielellä
- asiakirjan muotoilu säilyy *[ei kuitenkaan pdf-tiedoston muotoilu]*
- Koneluettavat tiedostomuodot: tmx ja xliff
- Käännös voidaan toimittaa myös sähköpostitse
- Mahdollisuus kääntää useita asiakirjoja useille kielille
- Viitteellinen **laatuluokitus** kullekin kieliparille (BLEU)
- Mahdollisuus antaa **palautetta** (EU Survey)

# MT@EC ja tietoturva

- Palvelimet ovat komission datakeskuksessa
- Käyttäjän tunnistustapa: **ECAS** (EC Authentication Service)
- Tiedostojen siirtotavat:
  - **sTESTA\*** - suojattu erillisverkko tietojen siirtämiseen EU:n ja kansallisten julkishallintojen välillä
  - **internet-yhteys** (suojattu **https-yhteys**)

\* Lisätietoja: <https://portal.testa.eu/jetspeed/portal/homepage/about.psml>.





# MT@EC – Konekäännöspalvelu

Kimmo ROSSI  
Euroopan komissio



Tietoa järjestelmästä | Palaute | Ohje

Valitse vähintään yksi lähdetiedosto:

Valitse tiedostot...

TAI

Vedä tiedostot tänne.

Lähdekieli:

EN

Kohdekieli:

BG	CS	DA	DE	EL	ES	ET	FI	FR	GA
HR	HU	IT	LT	LV	MT	NB	NL	PL	PT
RO	SK	SL	SV						

Käännösasiakirjan tiedostomuoto:

Sama kuin lähdetiedoston    TMX    XLIFF

- Lähetä käännös sähköpostiini
- Poista lataamisen jälkeen

Käännä asiakirja

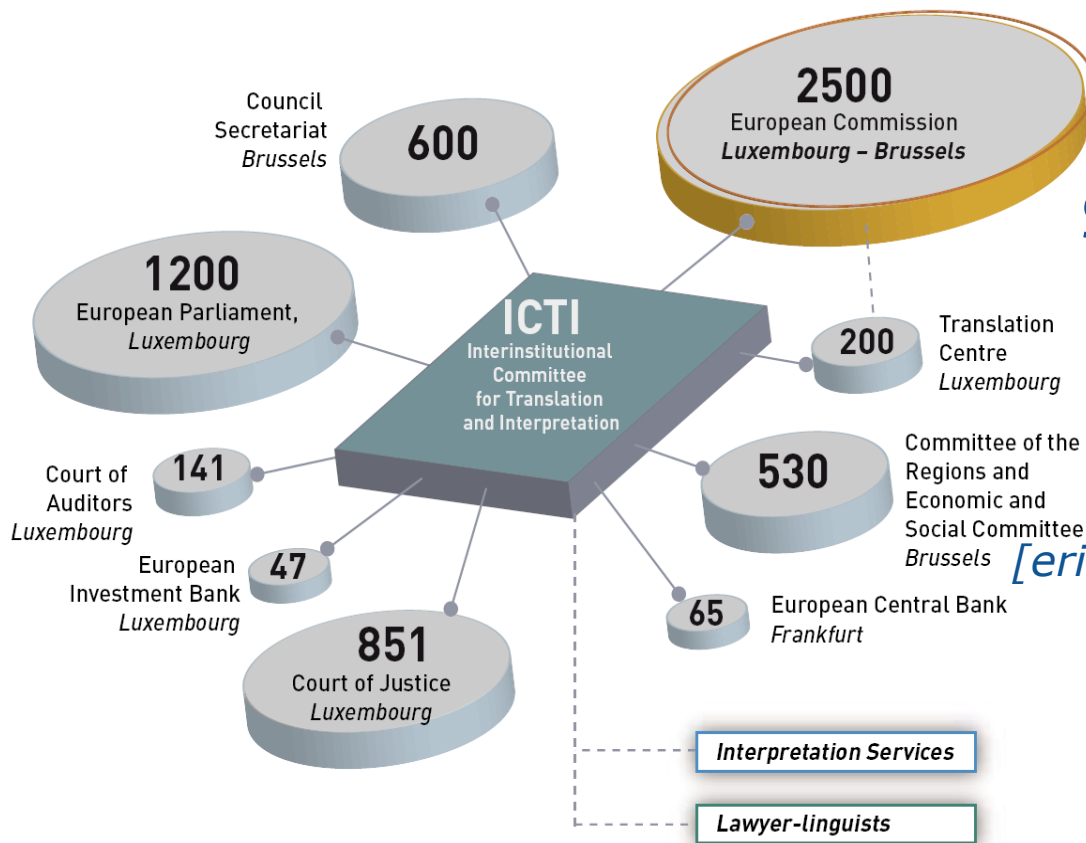
# Kuka voi käyttää MT@EC:tä?

## → *Euroopan unionin toimielimet:*

- |   |  |
|---|--|
| <ul style="list-style-type: none"><li>▪ Euroopan komissio</li><li>▪ Euroopan parlamentti</li><li>▪ EU:n neuvosto</li><li>▪ EU:n tuomioistuin</li><li>▪ tilintarkastustuomioistuin</li></ul> | <ul style="list-style-type: none"><li>▪ talous- ja sosiaalikomitea</li><li>▪ alueiden komitea</li><li>▪ Euroopan keskuspankki</li><li>▪ Euroopan investointipankki</li><li>▪ ...</li></ul> |
|---|--|

→ *Verkkopalvelut, joita EU ylläpitää tai rahoittaa*

→ *EU ja ETA-maiden julkishallinto*



**Euramis**

940 miljoonaa virkettä  
[vuoden 2015 lopussa]

**Kasvunopeus:**  
2,6 miljoonaa/kk

24 EU-kieltä

[eri määrä aineistoa, kielen mukaan]

**sisältö: "EU-tekstejä"**

[aihepiirit ja tekstityyli]

# MT@EC on maksuton julkishallinnon käyttäjille!

- Hanki ensin **ECAS**-tunnukset. Anna yhteystiedoksi työsähköpostiosoitteesi.  
<https://webgate.ec.europa.eu/cas/eim/external/register.cgi>
- Lähetä sähköposti osoitteeseen [DGT-MT@ec.europa.eu](mailto:DGT-MT@ec.europa.eu) ja pyydä MT@EC-käyttäjätilisi aktivointia.
- DGT luo sinulle käyttäjätunnuksen ja ilmoittaa sähköpostitse.

# MT@EC:hen liitetyt verkkopalvelut (esimerkkejä)

Palvelu	Kuvaus
IMI	Sisämarkkinoiden tietojärjestelmä – hallinnon sisäinen verkkopalvelu, johon kuuluu yli 7000 julkishallintoa EU:n alueella <a href="http://ec.europa.eu/internal_market/imi-net/index_en.html">http://ec.europa.eu/internal_market/imi-net/index_en.html</a> → Käyttäjien välisten viestien kääntäminen
SOLVIT	Verkkopalvelu sisämarkkinasääntöjen vastaisesta menettelystä johtuvien ongelmien ratkaisuun <a href="http://ec.europa.eu/solvit/">http://ec.europa.eu/solvit/</a> → Käyttäjien välisten viestien kääntäminen
nLex	Kansallisen lainsäädännön oikeusportaali <a href="http://eur-lex.europa.eu/n-lex/">http://eur-lex.europa.eu/n-lex/</a> → Hakutulosten kääntäminen
CircaBC	Virtuaalinen työtila, viestien ja asiakirjojen vaihto <a href="https://circabc.europa.eu/">https://circabc.europa.eu/</a> → Asiakirjojen kääntäminen

## Mitä tämän jälkeen?

**MT@EC:stä...**

**kohti CEF:n automaattista käännösalausta  
...eli lyhyemmin: CEF AT**

Mikä on **CEF AT**:

- se perustuu MT@EC-palveluun ja laajentaa/kehittää sitä
- tavoitteena **suojattu, laadukas ja muunneltava** käännösalausta EU:n laajuisille verkkopalveluille (ja vähän muillekin...)
- monikielisyyden tukialusta – enemmän kuin konekäännin

# Sovittaminen eri verkkopalveluihin

*mitä verkkopalveluja CEF AT tukee?*

Palvelu	Kuvaus
Europeana	Euroopan digitaalikirjasto
ODP	Avoimen Datan portaali, joka verkottaa kansalliset ja alueelliset avoimen data lähteet.
EESSI	32 valtion välinen järjestelmä sosiaaliturvaan liittyvän tiedon vaihtoon
ODR	Verkkopalvelu esim. verkkokauppaan liittyvien kuluttajariitojen ratkaisemiseen
e-justice	EU:n oikeusportaali
SaferInternet	Verkkopalvelu, jonka tavoitteena on tehdä internet turvalliseksi lapsille

## Mitä hyötyjä CEF AT tuo?

- *Nopeampi ja paremmin suojattu käännöspalvelu*
- *Monikielisyyttä tukevat EU:n laajuiset julkiset verkkopalvelut*

### **sekä:**

- *Parempi konekäännösten laatu*
- *Ottaa huomioon käyttöalan erikoispiirteet*  
(etusijalla CEF-ohjelman kattamat verkkopalvelut)

Tämä edellyttää, että erikielisiä **kieliaineistoja** on käytettävissä palvelun kehittämiseen





## Nykytila: MT@EC kääntää sujuvasti EU-tekstejä (mutta vaihtelevasti eri kielipareja)

- *MT@EC rakentuu EU-käännöksille (Euramis)*
- *Euramiksessa on eri määrä erikielisiä tekstejä (esim 97 miljoonaa englanninkielistä, 58 miljoonaa ranskankielistä, mutta vain 2,8 miljoonaa kroatiankielistä virkettä (tyypillisesti 22-38 miljoonaa))*
- *MT@EC soveltuu EU-aiheisiin liittyvien asiatekstien, mutta ei puhekielen, kääntämiseen*
- *MT@EC kattaa useita aihealueita, mutta sitä ei ole optimoitu mihinkään niistä erityisesti. Aineistoja tarvitaan eri aihealueilta.*



# Avoimen datan portaalit

**Oikeusperusta: direktiivi 2003/98/EY julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (+ muutosdirektiivi 2013/37/EU)**

**Kaksi eri portaalia :**

**1. Euroopan avoimen datan portaali:**

- <http://www.europeandataportal.eu/>
- verkottaa EU:n jäsenvaltioiden, alueiden ja kaupunkien avoimen datan

**2. EU:n *toimielinten* avoimen datan portaali:**

- <https://open-data.europa.eu/>



# EUROPEAN DATA PORTAL (BETA)

## Search Datasets

EUROPEAN DATA PORTAL

**BETA VERSION**

You can access the data tools from here

- Home
- What we do
- Providing Data
- Using Data
- Training & Library

### Latest News

- 08/02/2016  
Digital4EU Stakeholder Forum 2016  
Brussels, 25 February 2016
- 04/02/2016  
Watch the Open Data Forum in Sweden
- 03/02/2016  
Conference: Access to and Reuse of EU Legal Information
- 02/02/2016  
Survey on benefits and use cases relevant to indoor positioning

### Browse Categories

- Agriculture, Fisheries, Forestry & Foods
- Energy
- Regions & Cities
- Transport
- Economy & Finance
- International Issues
- Government & Public Sector
- Justice, Legal System & Public Safety
- Environment
- Education
- Health
- Social Services
- Science



## Translation memories of the official service of translators of IVAP

Followers

0

📖 Catalog



**datos.gob.es**

datos.gob.es - Spain, Spanien, Espagne [read more](#)

📱 Social



Dataset



Categories



Activity Stream

🗉 Feedback

The title and description of this dataset are machine translated.

## Translation memories of the official service of translators of IVAP

On this page several translation memories in TMX format and compressed zip format. This memory easily puedes descargarlas para usarlas back into your computer translation assignments. To this end, it is necessary to provide for and be fitted with a programme that can handle translation memories, as wordfast, Trados, transit or déjà-vu ejemplo.por, therefore you can have easy access to translation memories of the official service of translators, who is the brainchild of translators in the translation service and the rest of the Basque Government, supervised by the markers. It is not, however, a normative nature and 2988888 working tool which can prove useful and aid for many personas.si send us your translation memories, the "will also provide for which could be used by persons who may have an interest here.

### Distributions



Unnamed resource (ZIP)



Unnamed resource (XML)





# Euroopan unionin avoimen datan portaali

EUROPA > Avoimen datan portaali > Data

- Data
- Sovellukset
- Linkitetty data
- Sovelluskehittäjien resurssit
- About



Hae datajoukkoja...



Näytä tulokset, joissa on:  kaikki nämä sanat |  mikä tahansa näistä sanoista |  sanatarkka vastine ?

Datajoukkojen kokonaismäärä: 7943

## Ehdota datajoukkoa

Onko olemassa jokin EU:n datajoukko, jota et löydä tästä portaalista?

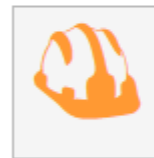
[Pyydä datajoukkoa>>](#)

## Eniten tarkastellut datajoukot

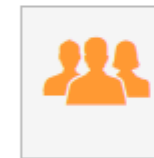
[» näyttää kaikki »](#)

- DGT-Translation Memory**  
(13089 views)
- Elevation map of Europe**  
(9916 views)
- CORDIS – EU research projects under Horizon 2020 (2014–2020)**  
(9306 views)
- EuroVoc**

## Selaa datajoukkoja aiheittain tai ryhmittäin



Työllisyys ja työ



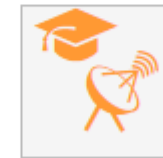
Yhteiskunnalliset kysymykset



Talouselämä



Finanssiasiat



# Avoimen datan portaalit

**Soveltuvat myös kieliaineistojen julkaisuun ja levittämiseen**

**Tavoitteena on kuvausdatan (=metadatan) harmonisointi ja (monikielisen) tiedonhaun helpottaminen**

**Tietojen tallennus, säilytys ja ylläpito pysyvät tiedon "omistajien" tehtävänä ja vastuulla**

**CEF AT –kieliaineistojen taltiointiin tarvitaan erillisratkaisu (pilvipalvelun koekäyttö alkaa 2016)**



European  
Commission



Kiitos!