



# The CEF eTranslation platform @ work

Markus Foti eTranslation Programme Manager DGT R.3



#### eTranslation platform at a glance



# **eTranslation**

# MT@EC

- Launched June 2013
- Legalese
- Statistical (Moses)

- Launched July 2017

   (webservice for snippets), Nov. 2017
   (web page for documents)
- Cloud based
- Neural engines

### **CEF.AT**

- More NLPs

   (transliteration,
   named entity
   recognition...)
- Generic services
- Projects (ELRC, market research)



#### eTranslation platform at a glance



#### Available for:

- individuals (submit documents through a web page)
- machine-to-machine use

#### Users:

- Digital Service Infrastructures (EESSI, ODR, Open Data Portal, Europeana, etc.)
- System suppliers (EURLex, N-Lex, Internal Market Information system...)
- Individuals in public administrations

#### • Benefits:

- Increase speed and productivity
- Reduce costs
- Facilitate information exchange



#### MT@EC - Machine Translation



Translate documents	Translate text	My translation requests					My settings	English	<u>▼</u> Lo	gout
		Select at least one source file :	<b>≜</b> Choose files	OF	R	→ Drag files here.				
		Translate from	•							
		Translate into								
		Domain	ALL ▼							
		Output format:	Same as source TMX XLIFF							
			E-mail me my translation.  Delete after download.							
						Translate document				
Selected files										

My settings English Logout



Translate documents

Translate text

My translation requests

Drop files to upload (or click) Supported formats: 🕼 🕼 🕩 🙏 🥯 📋 From\* To \* **Output format** Domain SPD Same as source TMX XLIFF ✓ E-mail me my translation





#### Submit your document...



#### **EUROPEAN COMMISSION**

#### **MEMO**

Brussels, 6 May 2013

# Smarter rules for safer food: Commission proposes landmark package to modernise, simplify and strengthen the agri-food chain in Europe

#### 1) What is "from farm-to-fork" policy?

The EU "from farm-to-fork" policy covers legislation to **protect health** by ensuring that the food we produce and import that ends up on our plates is safe to eat. This body of legislation is a complex, integrated system of rules which takes care of all the links in the food chain from: the protection of plants from pests, the protection of animals and humans from animal diseases (and from such diseases which are transmissible to humans) and the protection of human health from all the hazards that might arise along the agri-food chain. It also includes all those rules which allow the Single Market to function smoothly in this area and provides rules for the protection of consumers.

Official controls are carried out at various stages in the food production chain by EU Member States to ensure that the legislation is being correctly applied.





Get your translation back by e-mail.



#### **COMMISSION EUROPÉENNE**

NOTE

Bruxelles, 6 mai 2013

Règles plus intelligentes pour des de plus sûres»: Paquet décisif Co moderniser, de simplifier et agroalimentaire en Europe

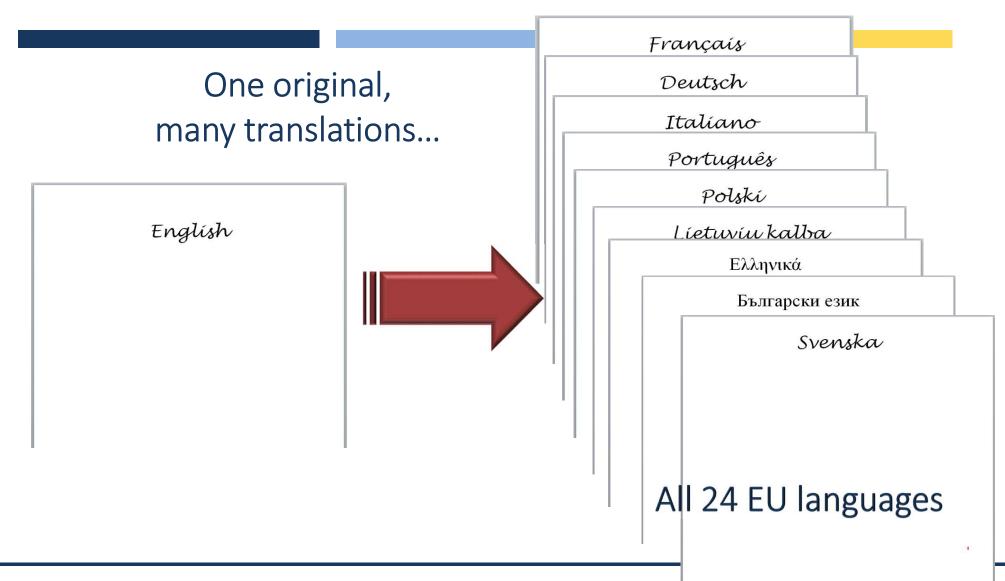
#### 1) Ce qui est «de la fou

L'UE «de la fourche à la four santé en faisant en sorte or que finale sur notre as législatif est un systèr maillons de la chaîne nuisibles, la protect partir de ces mala tous les dangers comprend égaler de manière han consommateurs

Keeps your Keeps your formatting!



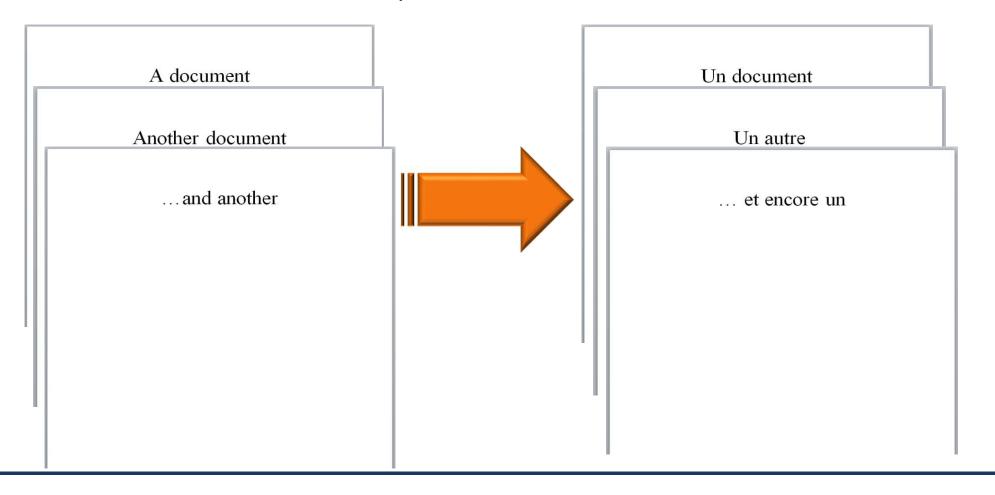








## Many documents at once...







# eTranslation protects your privacy

- All documents deleted after 24 hours or after delivery (on demand)
- IPR rights transferred (Your translation belongs to YOU!)





## No one

# looking

at your translations!



## Who is using eTranslation: DSIs

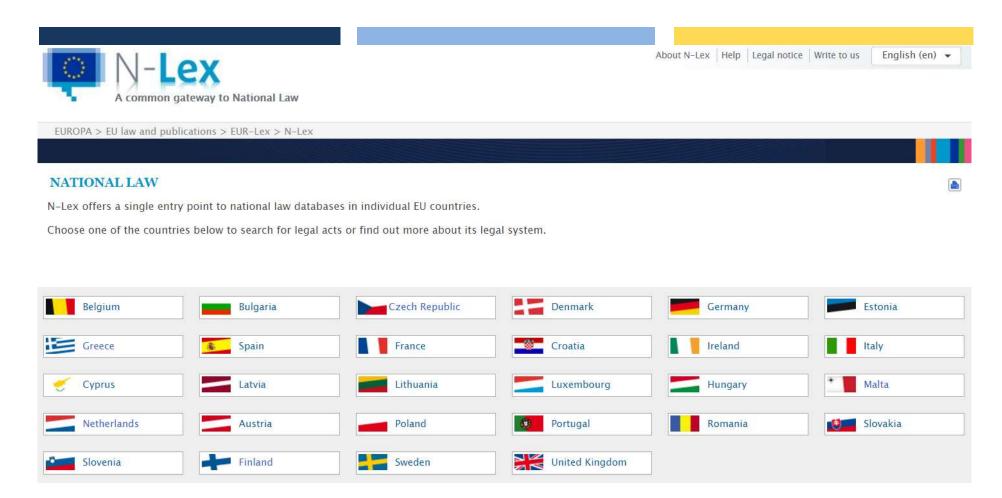


DSI \ BB	eTranslation
BRIS	
Cybersecurity	
EESSI	
Europeana	
ODR	
Open Data	
Safer Internet	
e-Justice Portal	
eDelivery	



# Machine-to-machine services eTranslation on the N-Lex portal

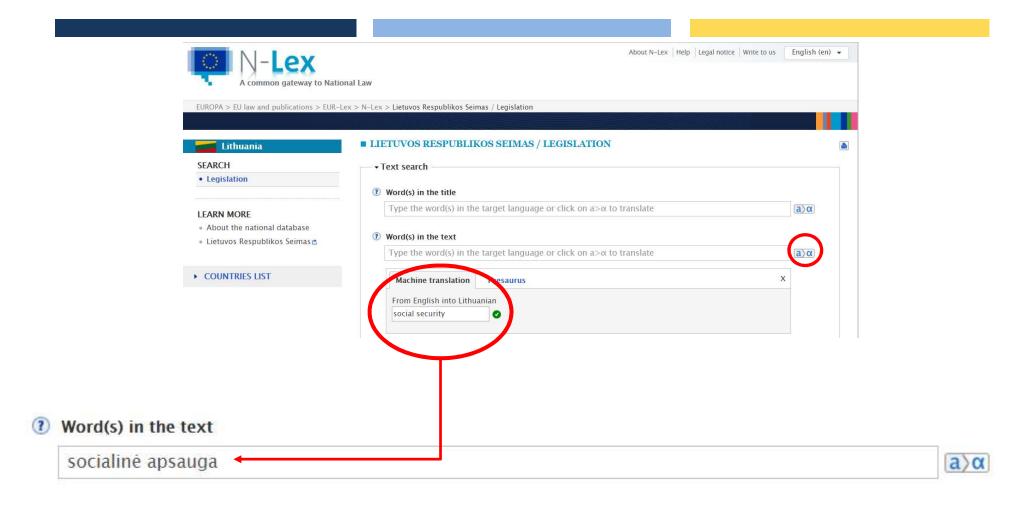






# Machine-to-machine service eTranslation on the N-Lex portal

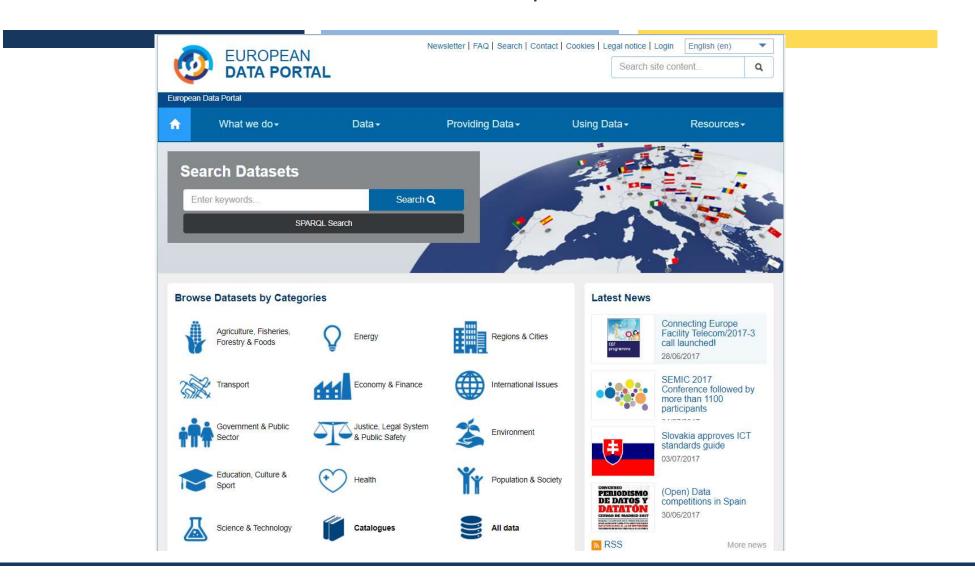






# Machine-to-machine service eTranslation on the European Data Portal

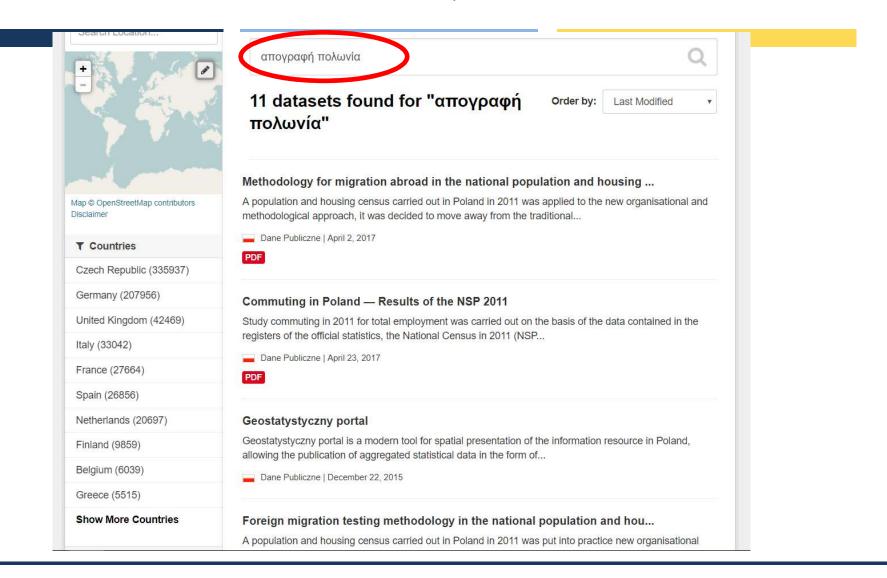






# Machine-to-machine service eTranslation on the European Data Portal

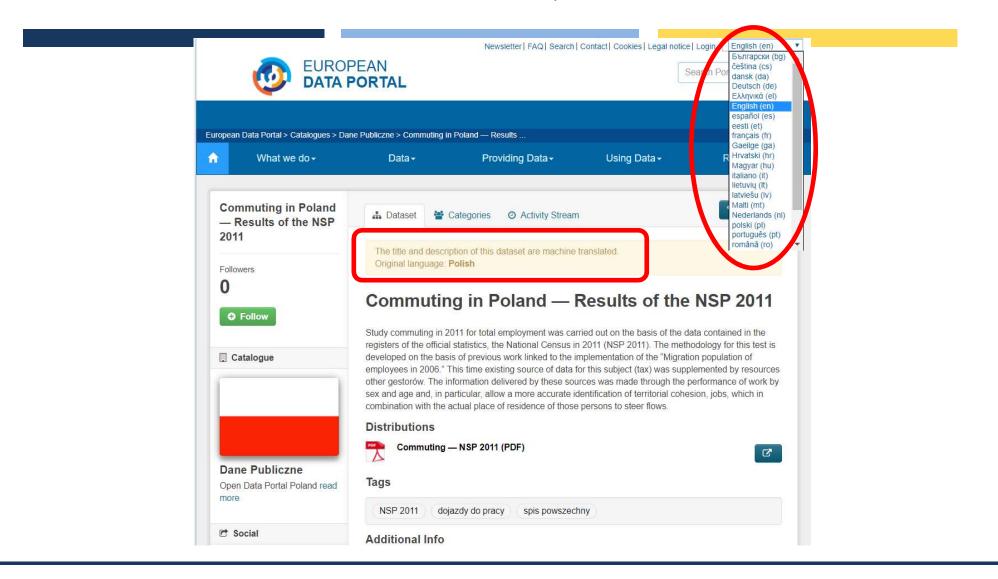






# Machine-to-machine service eTranslation on the European Data Portal







#### How to Connect to eTranslation



- Open to DSIs and public administrations
  - Contact <u>CEF-AT@ec.europa.eu</u> with your request and use case
  - We will provide the technical documentation on how to connect
    - SOAP request or RESTful interface
  - Contact us for credentials
  - Adapt your service to multilingual use!



#### Behind the scenes: how it works



#### Statistical Machine Translation

- MT@EC built only on "EU translations" (Euramis database)
- Covers all 24 languages
- Euramis (and MT@EC) cover EU policies, subjects and language but limited everyday language
- Moses-based engines
- Performs best when trained on large volumes of text pairs (source-translated)
   in specific domains



#### Neural Machine Translation (NMT)



#### What is NMT?

- Machine learning: artificial neural networks trained on existing translations
- The computer devises its own rules on how to translate
- Radical departure from the phrase-based SMT approaches

### Why is it important?

- Translations read better: more fluent and grammatical
- Better able to fill in gaps in data used for training training
- Better for highly inflected languages (e.g. German, Hungarian)

#### This is where the field is headed

— Will SMT become obsolete?





# **Neural Machine Translation Engines**

### • EN into:

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





# **Neural Machine Translation Engines**

#### • EN into:

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





# **Neural Machine Translation Engines**

#### • EN into:

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

24 June 2018

Original Text (http://europa.eu/rapid/press-release IP-18-5709 en.htm):

Starting in 1980, the European Union gradually adopted legislation putting an end to the diverging schedules of the national clock changes. In 2018 however, the purpose of clock changes has become much less relevant, with studies suggesting that energy savings are now marginal and citizens increasingly complaining about negative health impacts.

#### MT@EC (SMT):

От 1980 г. насам Европейският съюз постепенно приетото законодателство, водещо до различни графици на национално часовите промени. През 2018 г. обаче целта на часовите промени има все по-малко уместно, с проучвания, които показват, че икономиите на енергия са незначителни и гражданите все по-често се оплакват от отрицателно въздействие върху здравето им.

#### eTranslation (NMT):

От 1980 г. насам Европейският съюз постепенно прие законодателни актове, с които се слага край на различните списъци с промените в националния часовник. През 2018 г. обаче целта на часовниците получи много по-малко значение, като проучванията сочат, че икономиите на енергия сега са незначителни и гражданите все повече се оплакват от отрицателните последици за здравето.

Original Text (https://www.ft.com/content/0727cf8e-d6cc-11e7-ae3e-563c04c5339a): Visitor surge helps Bulgaria's festivals to flourish Bulgaria, with its diverse natural landscapes and historical sites going back to the Thracians, is starting to sell itself as a place for culture and scenery, shedding a reputation as a venue for Black Sea pub crawls.

#### MT@EC (SMT):

Нарастването на п<mark>осетителите на България</mark> помага фестивалите процъфтяват

България, с неговите разнообразни природни и исторически обекти чрез връщане към траките, започва да продава само по себе си като място за култура и пейзажи, хвърля репутация като място за Черно море експониране на барове.

#### eTranslation (NMT):

Скокът на посетителите помага на българските празници да процъфтяват България, с разнообразния си природен ландшафт и исторически обекти, които се връщат на Thracians, започва да се продава като място за култура и декори, което свързва репутацията като място на виадукти на Черно море.

Original Text (https://www.thebrokebackpacker.com/bulgarian-food-15-dishes-bucket-list/): If you ever venture on a Bulgarian holiday, remember a few things: Bulgarians love to catch up with family, welcome visitors, feed them, and talk to them about their lives and culture. A big part of this bonding happens over meals. Bulgarian food, traditional drinks and Bulgarian cheese all compliment the splendid hospitality and surroundings.

#### MT@EC (SMT):

Ако някога сте предприятие на български празник, не забравяйте няколко неща: Българите обичат да настигнат семейства посрещат посетители ще ги и си поговори с тях относно своя живот и култура. Голяма част от това свързване се осъществява във храна. Българска, традиционните напитки и българските сирена всички допълват чудесния гостоприемство и околностите.

#### eTranslation (NMT):

Ако някога го правите на български празник, си спомняме за няколко неща: Българите обичат да догонват членовете на семейството, да приемат посетители, да се хранят с тях и да говорят с тях за живота и културата си. Голяма част от това свързване се проявява върху храната. Българската храна, традиционните напитки и българското сирене се считат за изостаналост и гостоприемство и среда около тях.

Original Text (https://www.thebrokebackpacker.com/bulgarian-food-15-dishes-bucket-list/): Bulgarian food, traditional drinks and Bulgarian cheese all compliment the splendid hospitality and surroundings.

Българската храна, традиционните напитки и българското сирене се считат за изостаналост и гостоприемство и среда около тях.

#### **BUT:**

Original Text:

Bulgarian food, traditional drinks and Bulgarian cheese all complement the splendid hospitality and surroundings.

Becomes:

Българската храна, традиционните напитки и българското сирене допълват разнасяне на храна, гостоприемство и околна среда

And today:

В България храни, традиционни напитки и българско сирене са допълнение към гостоприемството и околностите.



#### Live (?) Demo



The adapted vs. non-adapted scenario, i.e.

a) General domain vs. EU legislation domain

For as many CEF languages as possible we would need to come up with

- a text in the general domain, e.g. a news article, and
- an in-domain text, e.g. legislation

Show the MT@EC-generated English equivalent for each text

b) General domain vs. DSI-specific domain: for which languages are such engines available?

(if there is no domain adapted engine for the <country's language-EN> pair, you/we may opt to show adapted and non-adapted engines in another language pair, leading to a call for data for the country's language)



### eTranslation EU Legislation



Source: EN

THE EUROPEAN COMMISSION,
Having regard to the Treaty on the
Functioning of the European Union,
[...]

HAS ADOPTED THIS DECISION:

Article 1

The total sum of greenhouse gas emissions covered by Decision No 406/2009/EC for each Member State for the year 2014 arising from the corrected inventory data upon completion of the comprehensive review referred to in Article 19(1) of Regulation (EU) No 525/2013 is set out in the Annex to this Decision.

**Target: BG** 

ЕВРОПЕЙСКАТА КОМИСИЯ, Като взе предвид Договора за функционирането на Европейския съюз, [...]

ПРИЕ НАСТОЯЩОТО РЕШЕНИЕ:

Член 1

Общият сбор на емисиите на парникови газове във всяка държава членка през 2014 г., попадащи в приложното поле на Решение № 406/2009/ЕО, който е получен въз основа на коригираните данни от инвентаризацията след завършването на цялостния преглед по член 19, параграф 1 от Регламент (ЕС) № 525/2013, е посочен в приложението към

настоящото решение.



#### Language resources: the key to success



- More data for all languages
- Better lexical coverage for all languages
- Machine translation adapted to your domain/subject matter



#### Future Improvements



Requested by users such as EDP, e-Justice, Europeana, EESSI, ODR...

- Speed
- Transliteration
- More formats (PDF output, JSON)
- Named-entity recognition
- New languages (AR, RU, TK, ZH)



#### CEF.AT brings...



- Reliable and trustworthy translation for EU and National Public Administrations
- Support for languages with fewer speakers
- Opportunities for private sector, through grants and generic services
- A higher profile for language technologies, thus fostering demand
- Public availability of data collected by ELRC



# Solve your language problems!



