



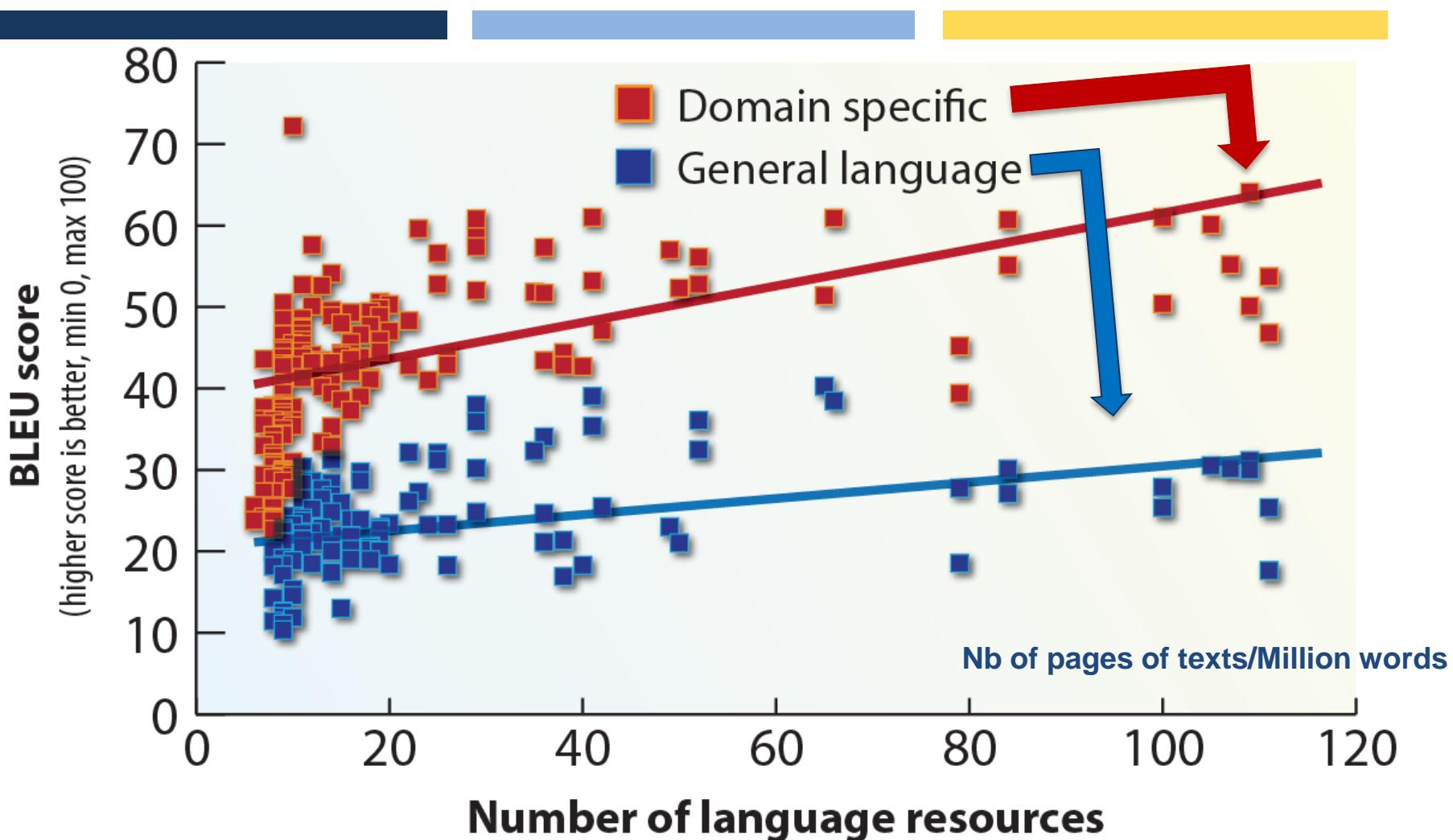
ELRC Oslo Workshop 08.06.2016.

## **How can we engage? – Language Resource Coordination**

Andrejs Vasiljevs, ELRC/Tilde | Credits: Khalid Choukri, ELRA, Josef van Genabith, DFKI



# Impact of number of language resources on Statistical MT quality

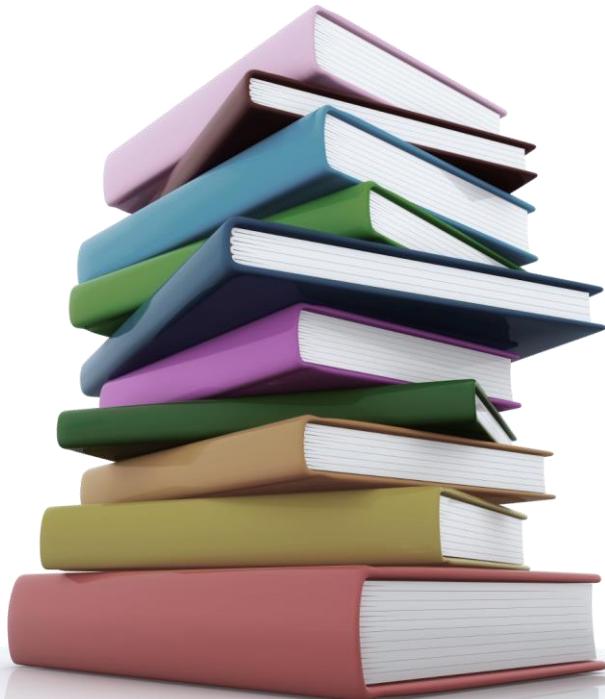


## What Data?

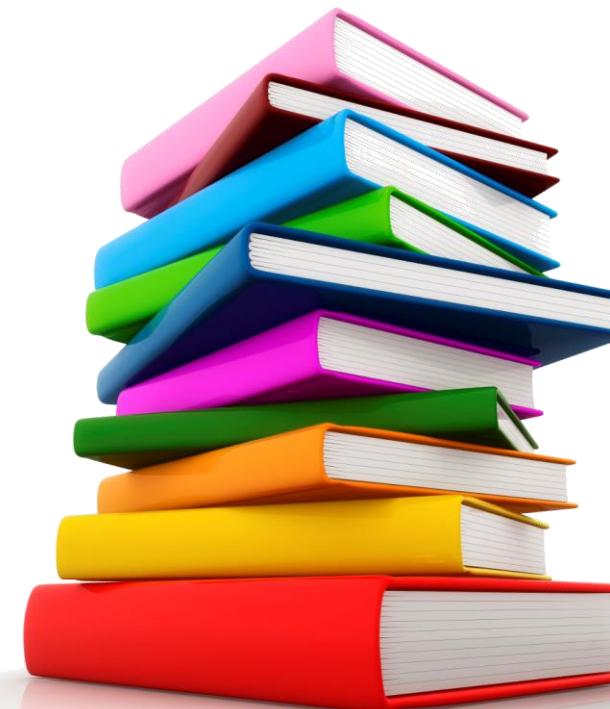


- Anything that contains “words”
- Preferences for “sentences”, even for sentences expressed in multiple languages
- Examples: reports, speeches, documents, web pages, brochures, etc.
- Bags of “words”, “sentences”, multiple bags

# What types of data? “Aligned” Translation



English



Norwegian

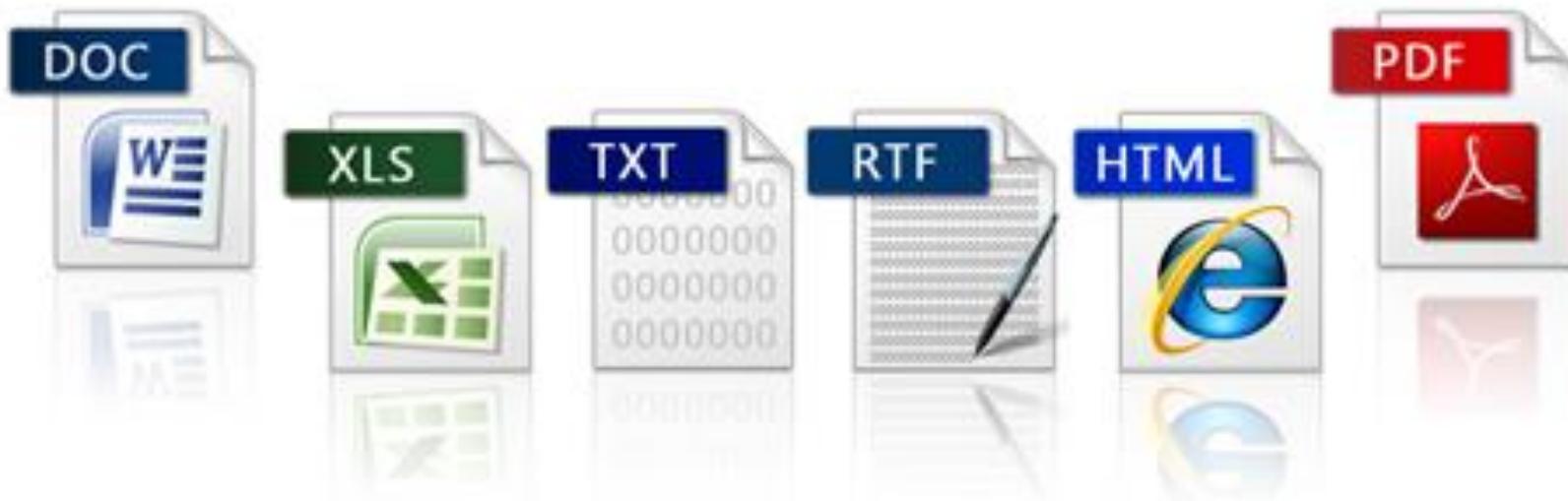


# Printed sources are hard to process

L 361 lähenema	106	L 388 lülitி	
L 361 lähenema piirväärtusele	approach a limit	приближаться к пределу	
L 362 lähenemine	approaching	приближение	
L 363 lähim	closest, nearest	ближайший	
L 364 lähis	adjacent	прилежащий	
L 365 lähis- (= lähind-)	approximate	приближённый	
L 366 lähiskaatet	adjacent side	прилежащий катет	
L 367 lähiskülg	adjacent side	прилежащая сторона	
L 368 lähimurd	convergent	походящая дробь	
L 369 lähte-	initial, original,	исходный, начальный	
L 370 lähteandmed ( $\Rightarrow$ algandmed)	initial data, source data	исходные данные, начальные данные	
L 371 lätehullu	domain, initial set	множество отправления	
L 372 lätepunkt (= alguspunkt)	initial point, starting point	исходная точка, начальная точка, отправная точка	
L 373 lätesimbol	initial symbol	начальный символ	
L 374 lätevõrrand	original equation	исходное уравнение, первоначальное уравнение	
L 375 lähtuma	start	исходить	
L 376 lävi	threshold	порог	
L 377 lühem teig ( $\Rightarrow$ väikestel)	minor axis	малая ось	
L 378 lühend	abbreviation	аббревиатура, сокращение	
L 379 lühendama	abbreviate, shorten	сокращать, укорачивать	
L 380 lühendamine	abbreviation, shortening	сокращение, укорачивание	
L 381 lühendatud	abbreviated, abridged, contracted	сокращённый	
L 382 lühendatud korruatmine	abbreviated multiplication	сокращенное умножение	
L 383 lühendatud tähis	abridged notation, contracted notation	сокращённое обозначение	
L 384 lühike	short	короткий	
L 385 lühim	shortest	кратчайший	
L 386 lüke ( $\leftarrow$ translation)	shift, slide, translation	перемещение, перенос, сдвиг	
L 387 lüli	link	звено	
L 388 lülitி	switch	переключатель	

M 1 maagiline	107	M 33 majoreerimine	
M 1 maagiline ruut	magic square	магический квадрат	
M 2 maastikuformaat	landscape size	горизонтальный формат	
M 3 maatriks	matrix	матрица	
M 4 maatriks	matrix	матричный	
M 5 maatriksalgebra	matrix algebra	матричная алгебра	
M 6 maatriksestitus	matrix representation	матричное представление	
M 7 maatriksi astak	rank of a matrix	ранг матрицы	
M 8 maatriksi diagonaal	diagonal of a matrix	диагональ матрицы	
M 9 maatriksi elementaarisedundused	elementary operations on matrices	элементарные операции над матрицами	
M 10 maatriksi jälg	spur of a matrix, trace of a matrix	след матрицы	
M 11 maatriksi põöramine	inversion of a matrix, matrix inversion	обращение матрицы	
M 12 maatriksite liitmine	addition of matrices	сложение матриц	
M 13 maatriksite ring	ring of matrices	кольцо матриц	
M 14 maatriksmäng	matrix game	матричная игра	
M 15 maatrikstähisustus	matrix notation	матричная символовика	
M 16 maatriksevärtustega funktsioon	matrix-valued function	матрица-функция, матрично-значная функция	
M 17 Mackey topoloogia	Mackey topology	топология Макки	
M 18 Maclaurini rida	Maclaurin's series	ряд Маклорена	
M 19 madalam nimbrikoh	least significant digit	младший разряд	
M 20 madalamat jätku	of lower order	нижнего порядка	
M 21 magasin (= pinu)	stack	магазин, стек	
M 22 magasinalgoritm	stack algorithm	магазинный алгоритм	
M 23 magasinautomaat	push-down automaton	магазинный автомат	
M 24 magasini tipp	top of stack	вершина стека	
M 25 magistraal	turnpike	магистраль	
M 26 maha tömbama	cross out	вычёркивать	
M 27 mahatömbamaine	crossing out	вычёркивание	
M 28 mah (= $m$ ahukus)	capacity	ёмкость	
M 29 mahukas	capacious	ёмкий	
M 30 mahukeskmine	power mean	степенное среднее	
M 31 majorant	majorant	мажоранта	
M 32 majoreerima	majorize	мажорировать	
M 33 majoreerimine	majorizing	мажорирование	



**Digital textual data in various formats –  
valuable language resources**

# Example of bilingual parallel document



ISO/IEC 2382-1  
Kolmas redaktsioon 1993-11-15

01  
Põhiterminid

01.01  
Üldterminid

01.01.01  
informatsioon  
Teadmus, mis puudutab objekte, näiteks fakte, sündmusi, asju, protsesse või ideid, sealhulgas mõisteid, ja millel on teatavas kontekstis eritähendus.  
MÄRKUSED: Vt. [joonis 1](#).

01.01.02  
andmed  
Informatsiooni taastõlgendatavat esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötuseks.

MÄRKUSED:  
1 Andmeid võivad töödelda inimesed või automaatsed vahendid.  
2 Vt. [joonis 1](#).

ISO/IEC 2382-1

01  
Fundamental terms

01.01  
General terms

01.01.01  
information (in information processing)  
Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning.  
NOTE - See [figure 1](#).

01.01.02  
data  
A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.  
NOTES  
1 Data can be processed by humans or by automatic means.  
2 See [figure 1](#).

# Processing of Bilingual Documents for Machine Translation



Word docs from <http://www.diplomatie.gouv.fr/fr/photos-videos-publications/publications/enjeux-planetaires-cooperation/rapports/article/rapports-du-groupe-pilote>,  
Financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition, Ministère des Affaires étrangères et du Développement international



**EXECUTIVE SUMMARY**



This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on innovative financing for agriculture, food security and nutrition, by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011.

The report includes an analysis of the need for investing financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the expertise of a high-level Committee of experts, discussions with relevant professional actors and an on-line consultation on the Global Forum on food security and nutrition (FSN Forum).

Although they are progressing, budgets for food security, including agriculture and nutrition components, in developing countries are severely constrained. Regarding the agriculture side, the dedicated Official Development Assistance (ODA) has increased in recent years, with higher amounts allocated to emerging and income countries, particularly in Sub-Saharan Africa and less developed countries. The challenge is all the more important in Sub-Saharan Africa, the continent where population growth will be the highest, where yields have tended to stagnate

Given the current trends in world population growth, particularly in developing countries, the global demand for food is expected to increase by 60 percent by 2050\*.

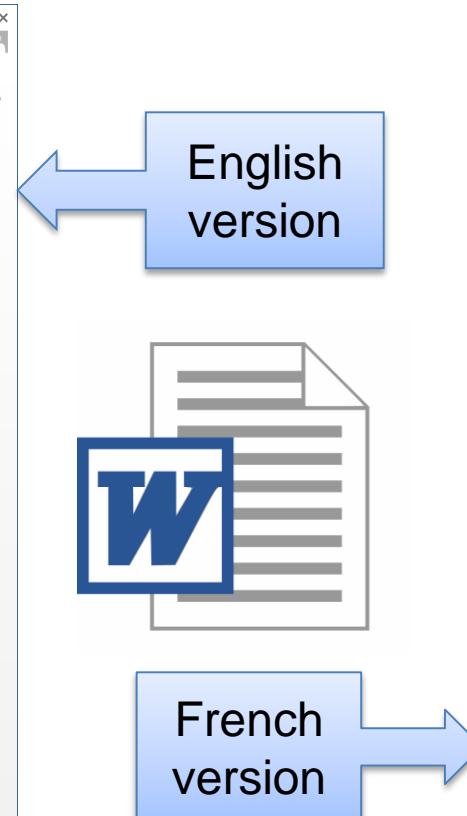
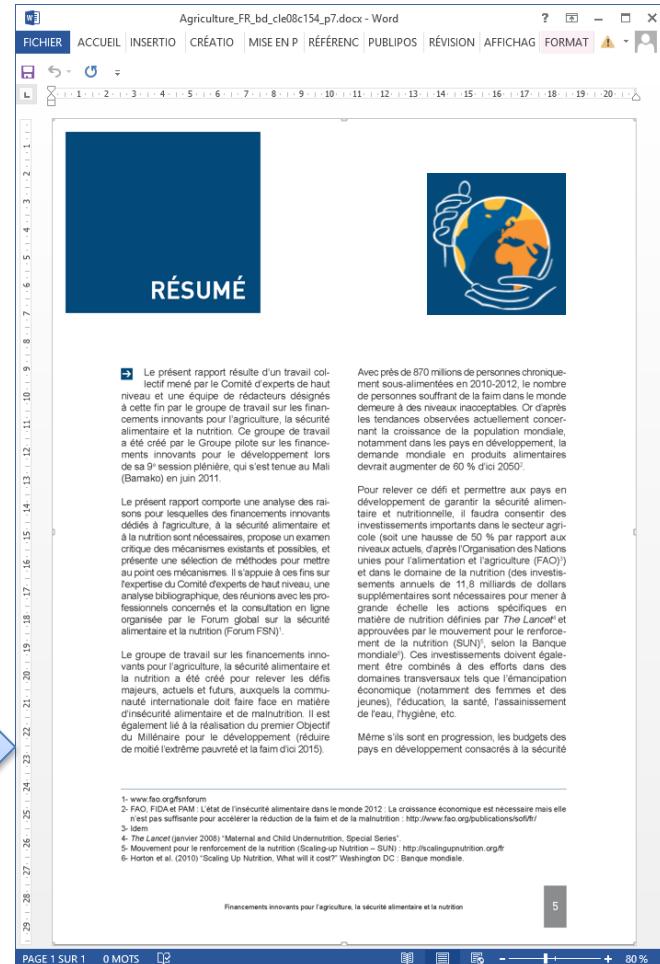
Meeting this challenge and allowing developing countries to ensure food security and nutrition imply an important investment in the agricultural sector (i.e., 50 percent more than the current level according to FAO) and in the field of nutrition\*\*. An additional investment of USD 11.8 billion is needed to implement at scale the priority nutrition-specific interventions identified by the Lancet series\*\*\* and endorsed by the Scaling Up Nutrition (SUN) movement\*\*\*\*, according to the World Bank\*\*\*\*\*. These investments need to be as well combined with other cross-cutting sectors such as economic empowerment (especially women and young people), education, health, water sanitation and hygiene, etc.

The setting up of the Task Force on Innovative Financing for agriculture, food security and nutrition responds to current and future crucial challenges faced by the international community regarding food insecurity and malnutrition and is related to the achievement of the first Millennium Development Goal (MDG 1) (reduction of extreme poverty and hunger by half by 2015).

With almost 870 million chronically undernourished people in 2010-2012, the number of hungry people in the world remains unacceptably high.

1- FSN Forum : <http://www.fao.org/forum/>  
2- FAO, FIDA and WFP: The State of food insecurity in the world (SOFI) 2012: Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition : <http://www.fao.org/publications/sofi/>  
3- FAO, 2011  
4- The Lancet (January 2009) 'Maternal and Child Undernutrition, Special Series'  
5- SUN movement: <http://scalingupnutrition.org/>  
6- Horton et al. (2010) 'Scaling Up Nutrition, What will it cost?' Washington DC: World Bank

Innovative financing for agriculture, food security and nutrition

**RÉSUMÉ**



Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés qui ont effectué un travail sur les financements innovants pour l'agriculture, sécurité alimentaire et la nutrition. Or dans les tendances observées actuellement concernant la croissance de la population mondiale, notamment dans les pays en développement, la demande mondiale en produits alimentaires devrait augmenter de 60 % d'ici 2050.

Pour relever ce défi et permettre aux pays en développement de garantir la sécurité alimentaire et nutritionnelle, il faudra consentir des investissements importants dans le secteur agricole (soit une hausse de 50 % par rapport aux niveaux actuels, d'après l'Organisation des Nations unies pour l'alimentation et l'agriculture (FAO)) et dans le domaine de la nutrition (des investissements annuels de 11,8 milliards de dollars supplémentaires sont nécessaires pour mener à grande échelle les actions spécifiques en matière de sécurité alimentaire et de nutrition). Ces investissements doivent également être combinés à des efforts dans des domaines transversaux tels que l'émancipation économique (notamment des femmes et des jeunes), l'éducation, la santé, l'assainissement de l'eau, l'hygiène, etc.

Même s'ils sont en progression, les budgets des pays en développement consacrés à la sécurité

Avec près de 870 millions de personnes chroniquement sous-alimentées en 2010-2012, le nombre de personnes souffrant de la faim dans le monde devrait augmenter de 60% d'ici 2050. Or dans les tendances observées actuellement concernant la croissance de la population mondiale, notamment dans les pays en développement, la demande mondiale en produits alimentaires devrait augmenter de 60 % d'ici 2050.

Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et recommande des stratégies pour les mettre en œuvre au point ces mécanismes. Il s'appuie à ces fins sur l'expertise du Comité d'experts de haut niveau, une analyse bibliographique, des réunions avec les professionnels concernés et la consultation en ligne organisée par le Forum global sur la sécurité alimentaire et la nutrition (Forum FSN)\*\*\*.

Le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition a été créé pour relever les défis auxquels sont confrontés les pays en développement dans la lutte contre l'in sécurité alimentaire et de malnutrition. Il est également lié à la réalisation du premier Objectif Millénaire pour le développement (réduire de moitié l'extrême pauvreté et la faim d'ici 2015).

1- www.fao.org/forum/  
2- FAO, FIDA et PAM : L'état de l'in sécurité alimentaire dans le monde 2012 : La croissance économique est nécessaire mais elle n'est pas suffisante pour accélérer la réduction de la faim et de la malnutrition : <http://www.fao.org/publications/sofi/>  
3- FAO, 2011  
4- The Lancet (janvier 2009) 'Maternal and Child Undernutrition, Special Series'  
5- Mouvement pour le renforcement de la nutrition (Scaling-up Nutrition - SUN) : <http://scalingupnutrition.org/>  
6- Horton et al. (2010) 'Scaling Up Nutrition, What will it cost?' Washington DC : Banque mondiale

Financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition

# Processing of Bilingual Documents for Machine Translation



## English version – Raw text

### Executive Summary

This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on Innovative Financing for agriculture, food security and nutrition created by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011.

The report includes an analysis of the need for innovating financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the expertise of a high-level Committee of experts, literature review, meetings with relevant professional actors and an on-line consultation on the Global Forum on food security and nutrition (FSN Forum)<sup>1</sup>.

The setting up of the Task Force on Innovative Financing for agriculture, food security and nutrition responds to current and future crucial challenges faced by the international community

[...]

## French version – Raw text

### Résumé

Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés à cette fin par le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition. Ce groupe de travail a été créé par le Groupe pilote sur les financements innovants pour le développement lors de sa 9e session plénière, qui s'est tenue au Mali (Bamako) en juin 2011.

Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et présente une sélection de méthodes pour mettre au point ces mécanismes. Il s'appuie à ces fins sur l'expertise du Comité d'experts de haut niveau, une analyse bibliographique, des réunions avec les professionnels concernés et la consultation en ligne organisée par le Forum global sur la sécurité alimentaire et la nutrition (Forum FSN)<sup>1</sup>.

Le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition a été créé pour relever les défis majeurs, actuels et futurs, auxquels la communauté

[...]

# Processing of Bilingual Documents for Machine Translation



## Alignment of English and French versions

### **S1.** Executive Summary

**S2.** This report is the result of a collective work carried out by the high-level expert Committee and a writing team commissioned by the Task Force on Innovative Financing for agriculture, food security and nutrition created by the Leading Group on Innovative Financing for Development at its 9th plenary session in Mali (Bamako) in June 2011.

**S3.** The report includes an analysis of the need for innovating financing dedicated to the agricultural, food security and nutrition sector, a critical review of existing and possible mechanisms and a proposed selection of avenues for the development of such mechanisms on the basis of the expertise of a high-level Committee of experts, literature review, meetings with relevant professional actors and an on-line consultation on the Global Forum on food security and nutrition (FSN Forum)1.

**S4.** The setting up of the Task Force on Innovative Financing for agriculture, food security and nutrition responds to current and future crucial challenges faced by the international community [...]

### **S1.** Résumé

**S2.** Le présent rapport résulte d'un travail collectif mené par le Comité d'experts de haut niveau et une équipe de rédacteurs désignés à cette fin par le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition.

**S3.** Ce groupe de travail a été créé par le Groupe pilote sur les financements innovants pour le développement lors de sa 9e session plénière, qui s'est tenue au Mali (Bamako) en juin 2011.

**S4.** Le présent rapport comporte une analyse des raisons pour lesquelles des financements innovants dédiés à l'agriculture, à la sécurité alimentaire et à la nutrition sont nécessaires, propose un examen critique des mécanismes existants et possibles, et présente une sélection de méthodes pour mettre au point ces mécanismes.

**S5.** Il s'appuie à ces fins sur l'expertise du Comité d'experts de haut niveau, une analyse bibliographique, des réunions avec les professionnels concernés et la consultation en ligne organisée par le Forum global sur la sécurité alimentaire et la nutrition (Forum FSN)1.

**S6.** Le groupe de travail sur les financements innovants pour l'agriculture, la sécurité alimentaire et la nutrition a été créé pour relever les défis majeurs, actuels et futurs, auxquels la communauté [...]

# Sentence aligned data from parallel documents



en	el
Greece of art and science	Η Ελλάδα των τεχνών και της επιστήμης
Greece is a place of culture, the arts and sciences.	Η Ελλάδα αποτελεί έναν χώρο πολιτισμού, τέχνης και επιστημών.
Its tradition of contribution to global cultural and scientific communities, combined with its outstanding natural beauty and excellent infrastructure, has made it an ideal place in which to hold conferences.	Η μακραίωνη συμβολή της στο παγκόσμιο γίγνεσθαι, σε συνδυασμό με το μοναδικό φυσικό κάλλος και τις άρτιες υποδομές, την καθιστούν ιδανικό τόπο διεξαγωγής συνεδρίων.
Over the last few years, Greece has more and more frequently welcomed people of letters, sciences and the arts, who have participated in symposia, conferences and exhibitions.	Τα τελευταία χρόνια, η ελληνική επικράτεια υποδέχεται όλο και συχνότερα ανθρώπους των γραμμάτων, των επιστημών και των τεχνών, οι οποίοι συμμετέχουν σε συμπόσια, συνέδρια και εκθέσεις.
Athens International Airport 'Eleftherios Venizelos', one of the most modern airports in the world in operation since 2001, greatly boosted the organization of international conferences.	Ο Διεθνής Αερολιμένας Αθηνών «Ελευθέριος Βενιζέλος», ένα από τα πλέον σύγχρονα αεροδρόμια παγκοσμίως, ο οποίος λειτουργεί από το 2001, έδωσε μεγάλη ώθηση στη διοργάνωση διεθνών συνεδρίων.
Conference tourism is extremely interdependent: it requires of course a high level of background support from the host country, and at the same time it can actively contribute to improving the overall standard of services in the region.	Ο συνεδριακός τουρισμός είναι άκρως αλληλεπιδραστικός: απαιτεί, Βέβαια, ένα υψηλού επιπέδου υπόβαθρο από τη χώρα υποδοχής, ταυτόχρονα όμως συμβάλλει ενεργά στην αναβάθμιση της συνολικής ποιότητας μιας περιοχής.
It is logical that a country chosen as a conference location should be involved in the cultural 'product', giving the public, both residents and visitors, the chance to experience human achievement and innovative thought.	Είναι λογικό, ένας χώρος ο οποίος προτιμάται για τη διεξαγωγή συνεδρίων, να μετέχει προνομιακά στο πολιτιστικό «προϊόν», μιας και δίνει τη δύνατότητα σε κοινό, κατοίκους και επισκέπτες, να έρθουν σε επαφή με τα ανθρώπινα επιτεύγματα και τις καινοτομίες.
The Greece of the pre-Socratic philosophers, of the great poets, of Pheidias the sculptor and Asclepius the physician, extends its hospitality and its warmest welcome, honouring people of intellect and creativity, commerce and scientific progress.	Η Ελλάδα των προσωκρατικών φιλοσόφων, των μεγάλων ποιητών, του Φειδία και του Ασκληπιού υποδέχεται φιλόξενα και τιμά τους ανθρώπους του πνεύματος, του εμπορίου και της προόδου.
Scientific conferences in the land that gave birth to science	Συνέδρια στη χώρα που γέννησε τις επιστήμες
Greece has a large number of esteemed scientists, both here in the country and abroad. Greek scientists, with their inventions, innovations and research work, play a leading part in the international scientific community.	Η Ελλάδα διαθέτει μεγάλο και υψηλής αξίας επιστημονικό δυναμικό, τόσο εντός όσο και εκτός συνόρων. Οι Έλληνες επιστήμονες, με τις εφευρέσεις, τις καινοτομίες και το ερευνητικό τους έργο πρωταγωνιστούν στη διεθνή επιστημονική κοινότητα.
Numerous important scientific conferences take place in Greece, reflecting the significance the country places on innovative science.	Τα επιστημονικά συνέδρια που λαμβάνουν χώρα στην Ελλάδα είναι και πολλά και σημαντικά, αντανακλώντας τη σημασία που δίνει η χώρα στις καινοτόμες επιστήμες.
Medical, architectural, natural and humanistic scientific conferences enrich Greece's cultural life, and at the same time give participants the opportunity to experience the	Ιατρικά συνέδρια, αρχιτεκτονικά, φυσικών και ανθρωπιστικών επιστημών, πλουτίζουν την πολιτιστική ζωή της

# Example of Comparable Data



## English

Telecommunication occurs when the exchange of information between two or more entities (communication) includes the use of technology.

Communication technology uses channels to transmit information (as electrical signals), either over a physical medium (such as signal cables), or in the form of electromagnetic waves.

The word is often used in its plural form, telecommunications, because it involves many different technologies.

## Greek

Με τον γενικό όρο τηλεπικοινωνίες, (telecommunications), χαρακτηρίζεται η κάθε μορφής ενσύρματη ή ασύρματη, ηλεκτρομαγνητική, ηλεκτρική, κ.λπ., ακουστική και οπτική επικοινωνία που πραγματοποιείται ανεξαρτήτως απόστασης.

Στους σύγχρονους καιρούς, αυτή η διαδικασία σχεδόν πάντα περιλαμβάνει την αποστολή ηλεκτρομαγνητικών κυμάτων ή ηλεκτρικών σημάτων από κατάλληλες ηλεκτρονικές συσκευές, όπως το τηλέφωνο ή ο ασύρματος, αλλά παλαιότερα περιελάμβανε τη χρήση ακουστικών σημάτων, όπως τυμπάνων, ή οπτικών, όπως ο σηματοφόρος καπνός ή η λάμψη της φωτιάς.

## Spanish

Una telecomunicación es toda transmisión y recepción de señales de cualquier naturaleza, típicamente electromagnéticas, que contengan signos, sonidos, imágenes o, en definitiva, cualquier tipo de información que se desee comunicar a cierta distancia.

Por metonimia, también se denomina telecomunicación (o telecomunicaciones, indistintamente) a la disciplina que estudia, diseña, desarrolla y explota aquellos sistemas que permiten dichas comunicaciones; de forma análoga, la ingeniería de telecomunicaciones resuelve los problemas técnicos asociados a esta disciplina.

Source: First sentences of articles for Telecommunications in the English, Greek and Spanish Wikipedias

# Dictionaries / Terminologies /Ontologies



previous level in time or space.

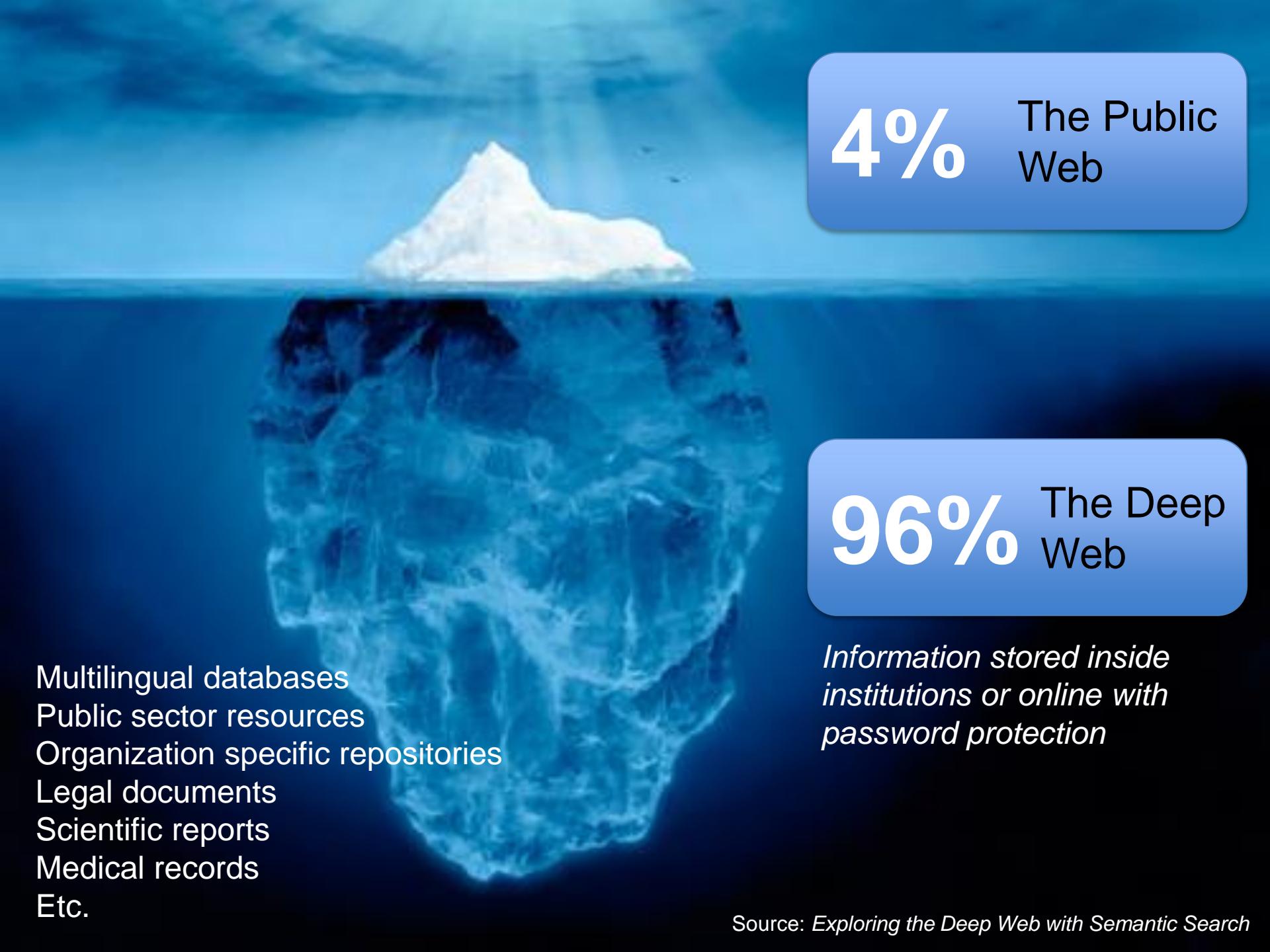
**team-work [teem-wurk]**

the combined action of a group of people, esp. when effective

Synonyms: alliance, assistance, coalition, combination, confederation, confederation; development, especially in science, technology, with reference to the common

		ES	EL
5075	Abruzzes	abandono escolar	διακοπή της σχολικής φοίτησης
5339	absentéisme	despojo	παραπροϊόντα σφαγίων
5984	abstentionnisme	sacrificio de animales	σφαγή ζώων
2	abus de confiance	derogación	κατάργηση
	abus de droit	Abruzos	Αβρουζία
	abus de pouvoir	absentismo	συστηματική απουσία από την εργασία
	accès à l'éducation	abstencionismo	αποχή
	accès à l'emploi	abuso de confianza	απιστία
		abuso de derecho	κατάχρηση δικαιώματος
		abuso de poder	κατάχρηση εξουσίας
		acceso a la educación	πρόσβαση στην εκπαίδευση
		acceso al empleo	πρόσβαση στην αγορά εργασίας





**4%**

The Public  
Web

**96%**

The Deep  
Web

- Multilingual databases
- Public sector resources
- Organization specific repositories
- Legal documents
- Scientific reports
- Medical records
- Etc.

*Information stored inside institutions or online with password protection*

# Public Sector Data Sources



- Visible data e.g. [Web](#)
  - Public websites
  - leaflets, brochures, news etc.
- [Invisible Data](#): archives , hidden web, internal repositories
  - translated documents, reports, speeches, meeting minutes etc.
- Data from outsourced translations:
  - From [Language Service Providers](#) and freelance translators
  - Translation Memory as part of contract delivery
  - Respective provisions in the contracts for outsourced translations



- Legal Issues
  - Legal status determination (accept/reject decision)
  - Accuracy and acceptance of privacy processing (e.g., anonymization)
  - Application of PSI versus need for a License
- Role of the ELRC Consortium
  - Support on practical issues
  - Technical/legal helpdesk, consultancy
  - Model licensing agreements
    - Government Open Licenses
    - Standard Re-use Licenses
    - License interoperability

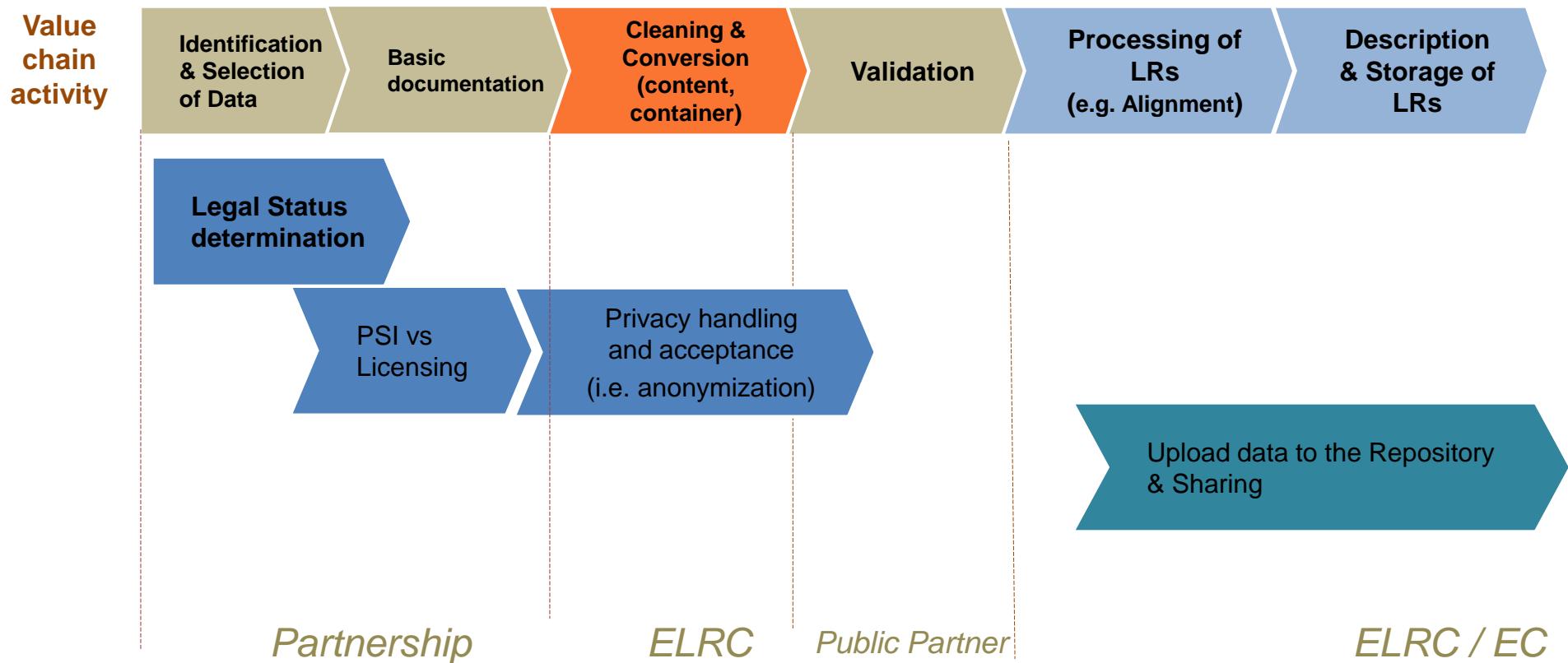
# Data Anonymization



- Identify a large source of data on individuals, organizations etc.
- Use a Named Entity Recognizer (NER) to find and **remove private data** (names, locations, dates, birth information, etc.) and replace with generic placeholders.
- Confirm results meet acceptable requirements
  - Reject data if anonymization not accurate as required



# Illustration of Data Packaging Workflow Data → LRs (Language Resources)



# Technical aspects dealt by ELRC



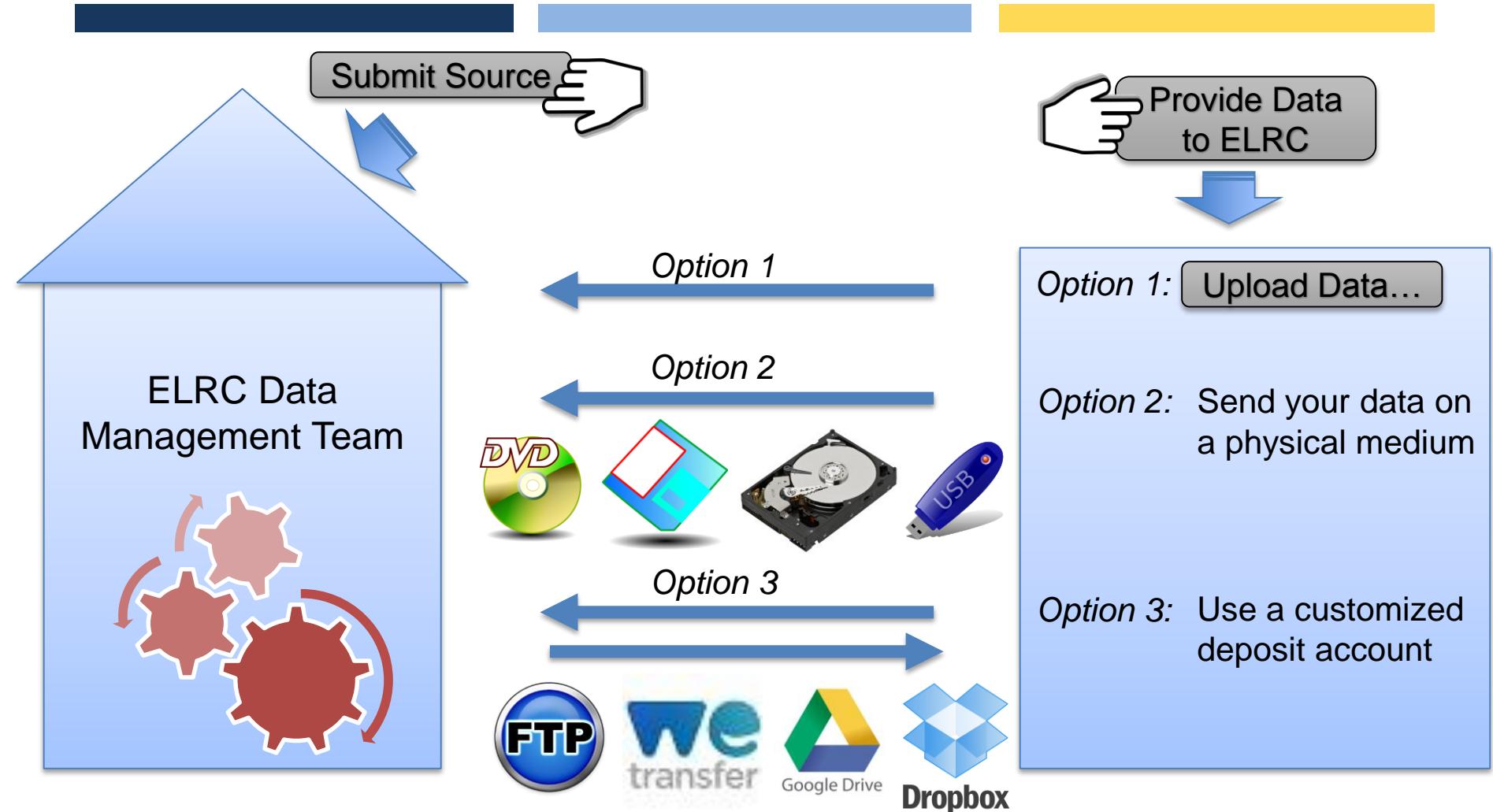
- Cleaning of data format  
*Discarding formatting, encoding character sets e.g. UTF8, formatting features e.g. bold, italic; graphics, ads, tables, html tags, etc.*
- File conversions  
*e.g. conversion to XML, XLIFF, etc.*
- Data preparation for Automated Translation tools  
*e.g. extraction and alignment*
- Validation and Quality Control of the output  
*Language Resource format, content, storage*
- Description of the Language Resource (meta-data)
- Packaging and delivery (data repository with e-sharing) to EC and Owner



A screenshot of a web browser window. The address bar shows 'Languages — the heart X lr-coordination.eu'. The page content includes a navigation menu with 'Home' (underlined), 'About', 'News', 'Helpdesk', 'Events', 'Resources', 'Anchor Points', and 'Multilingual Europe'. To the right is the 'European Language Resource Coordination' logo. The main background image features a modern glass building with many flags flying in front of it.

# Languages — the heart of Multilingual Europe

# Options for submitting your data



# Data submission to ELRC repository



Submit  
Sources



URL:

Submit URL



Submit  
Data



## Data submission form

Data name:

Languages:

Provider name:

Type of data:

....

Contact name:

Email:

Provide Data  
to ELRC

## Source submission form

Source name:

Languages:

Provider name:

....

Contact name:

Email:

Submit Source

Legal notice | Cookies | Contact on Europa | Search on Europa | Login | English (en)

Search



## EUROPEAN DATA PORTAL



**Search Datasets**

Search datasets  Advanced Search (SPARQL)

**What We Do**

Introducing The Portal, The Project & The Team

**Providing Data**

Providing Access To Public Data Across Europe

**Using Data**

Unlocking The Potential Of Public Data

**Training & Library**

Learning More About Open Public Data

**Latest News**

02/09/2015 - 12:30 Digital Single Market

28/08/2015 - 15:45 ODI chairman calls UK govt to urgently open up more health data

[news archive](#)

**Tweets**

European Data Forum 10 Feb 14  
Don't forget to submit your proposals at #EDF2014 by Friday. Call for Exhibits 2014.data-forum.eu/calls/exhibits #opendata #bigdata #linkeddata  
Retweeted by PublicData.eu

PublicData.eu 7 Feb 14  
What are you doing for Open Data Day? Inspiration here: blog.ofn.org/2014/02/07/wha... #ODD14 #OpenDataDay #opendata

PublicData.eu 18 Jan 14  
#PublicData.eu #OpenDataDay #ODD14  
Tweet to @publicdataeu

**Browse Categories**

	Agriculture, Fisheries, Forestry, Foods		Energy		Regions, Cities		Transport		Economy & Finance
	International Issues		Government, Public Sector		Justice, Legal System, Public Safety		Environment		Education, Culture & Sport
	Health		Population & Social Conditions		Science & Technology				

**Articles**



2013 West Nile Virus Data  
The 2013 National West Nile Virus human surveillance data, containing information such as report week, classification, and territory.



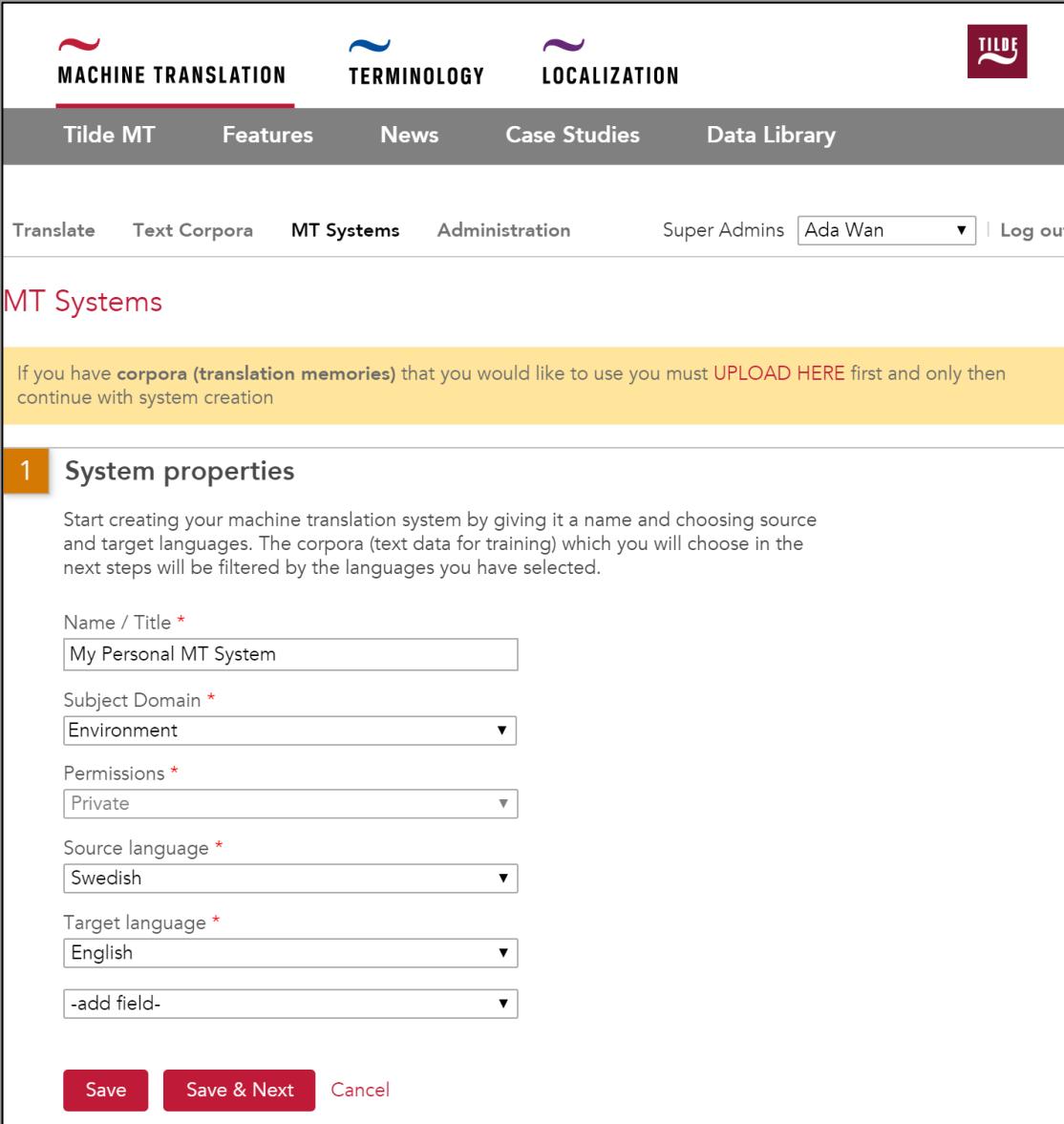
Europe's carbon dioxide emissions  
Our emissions map  
<http://www.sandbag.org.uk/maps/emissions>  
map shows how much carbon dioxide is emitted by factories and power stations in Europe.

**Help us Improve**

Your feedback will help us to improve the overall user experience. Any suggestions?



# Create your own custom MT with [tilde.com/mt](http://tilde.com/mt)



The screenshot shows the Tilde MT web application interface. At the top, there are three main navigation tabs: MACHINE TRANSLATION (highlighted in red), TERMINOLOGY, and LOCALIZATION. Below the tabs is a secondary navigation bar with links: Tilde MT, Features, News, Case Studies, and Data Library. On the far right of the header is the TILDE logo.

In the top right corner, there are user account links: Super Admins (Ada Wan), a dropdown menu, and Log out. Below this, a navigation bar includes Translate, Text Corpora, MT Systems (highlighted in red), and Administration.

The main content area has a yellow banner at the top stating: "If you have **corpora** (translation memories) that you would like to use you must **UPLOAD HERE** first and only then continue with system creation".

The main form is titled "1 System properties". It contains the following fields:

- Name / Title \*: My Personal MT System
- Subject Domain \*: Environment
- Permissions \*: Private
- Source language \*: Swedish
- Target language \*: English
- add field-

At the bottom of the form are three buttons: Save, Save & Next (highlighted in red), and Cancel.

# Create your own custom MT with [tilde.com/term](#)

~  
MACHINE TRANSLATION

~  
TERMINOLOGY

~  
LOCALIZATION

## Quick Look up

Paste a short text below. Hover over highlighted terms to see translations. Max 2000 characters for unregistered users. [Register here.](#)

A computer is a general purpose device that can be programmed to carry out a set of arithmetic or logical operations automatically. Since a sequence of operations can be readily changed, the computer can solve more than one kind of problem. Conventionally, a computer consists of at least one processing element, typically a central processing unit (CPU), and some form of memory. The processing element carries out arithmetic and logic operations, and a sequencing and control unit can change the order of operations in response to stored information. Peripheral devices allow information to be retrieved from an external source, and the result of operations saved and retrieved.

A **computer** is a general purpose **device** that can be programmed to carry out a set of arithmetic or logical operations automatically. Since a sequence of operations can be readily changed, the **computer** can solve more than one kind of problem. Conventionally, a **computer** consists of at least one processing **element**, typically a **central processing unit (CPU)**, and some **form** of memory. The processing **element** carries out arithmetic and logic operations, and a sequencing and **control** unit can change the order of operations in response to stored **information**. **Peripheral** devices **allow** **information** to be retrieved from an external **source**, and the result of operations saved and retrieved.

Look up

# ELRC Portal: Helpdesk



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

European Language  
Resource Coordination

## Helpdesk



### Got a question? We're here to help!

We are happy to answer any questions on the technical or legal aspects related to the use, production, collection, processing, and sharing of language resources.

Please feel free to contact us through one of the following channels:

---

a Web forum

[Web forum](#)

---

Telephone\*

+33 970 440 522

---

reach the Secretariat Support at

+49 681-8575 5285

---

Skype

**CEF-AT-Helpdesk**

---

E-mail

[help@cef-at-helpdesk.org](mailto:help@cef-at-helpdesk.org)

# Thank you!



Looking forward to a fruitful cooperation to  
make Norwegian machine translation better!