



ELRC Reykjavik Workshop 11.11.2016.

What data is needed for automated translation?

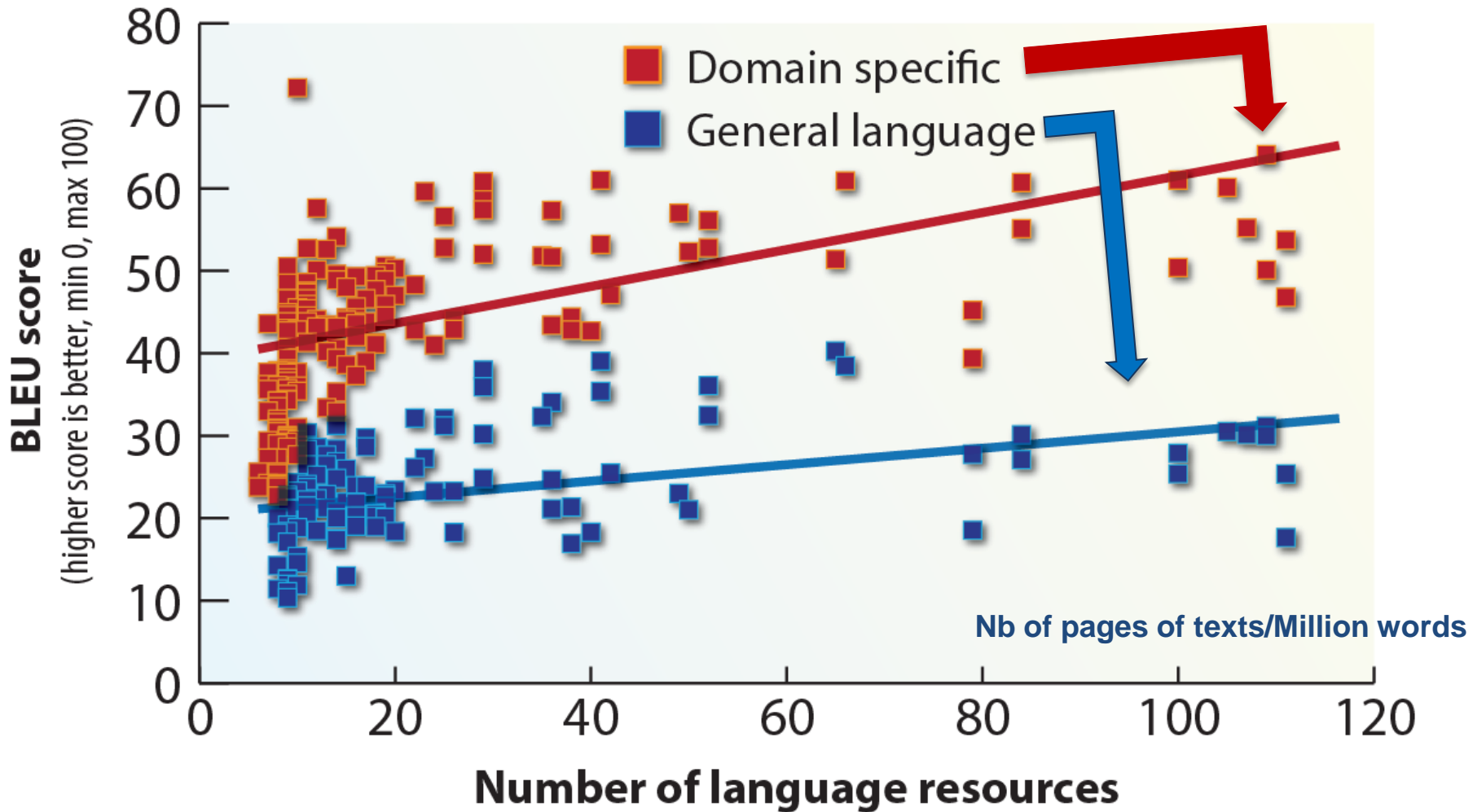
Andrejs Vasiljevs, ELRC/Tilde | Credits: Khalid Choukri, ELRA, Josef van Genabith, DFKI



European Language
Resource Coordination
Connecting Europe Facility



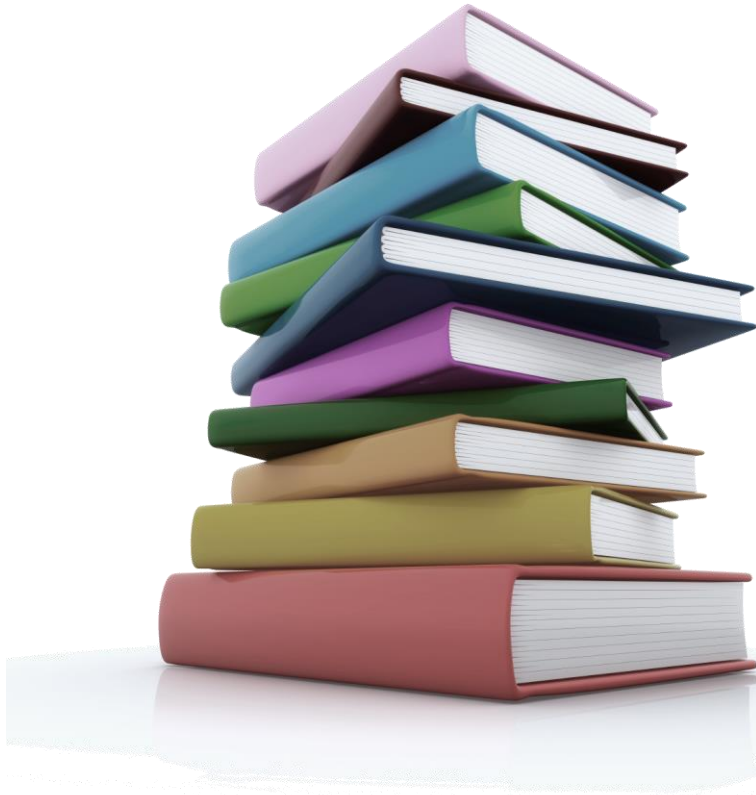
Impact of number of language resources on Statistical MT quality



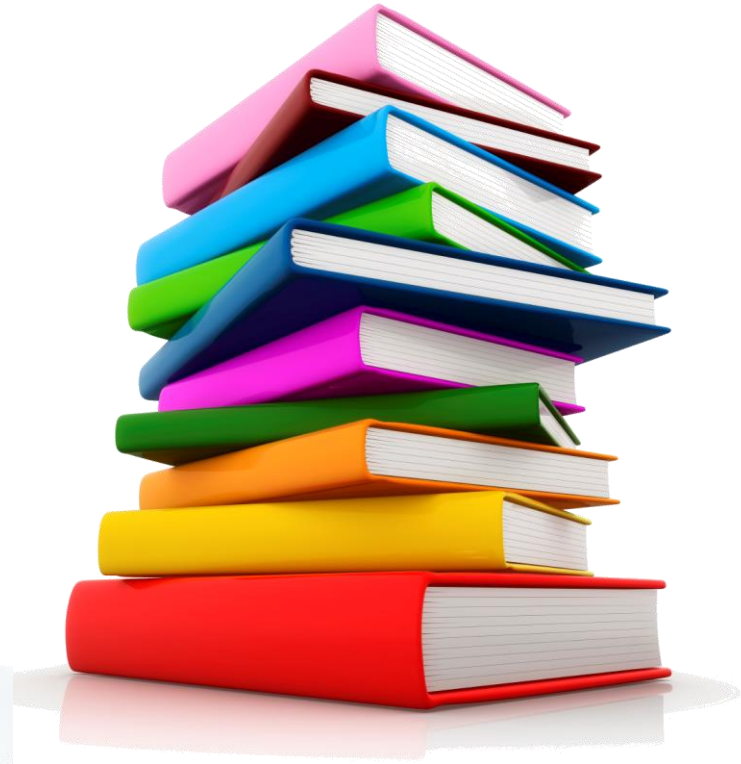


- Anything that contains “words”
- Preferences for “sentences”, even for sentences expressed in multiple languages
- Examples: reports, speeches, documents, web pages, brochures, etc.
- Bags of “words”, “sentences”, multiple bags

What types of data? “Aligned” Translation



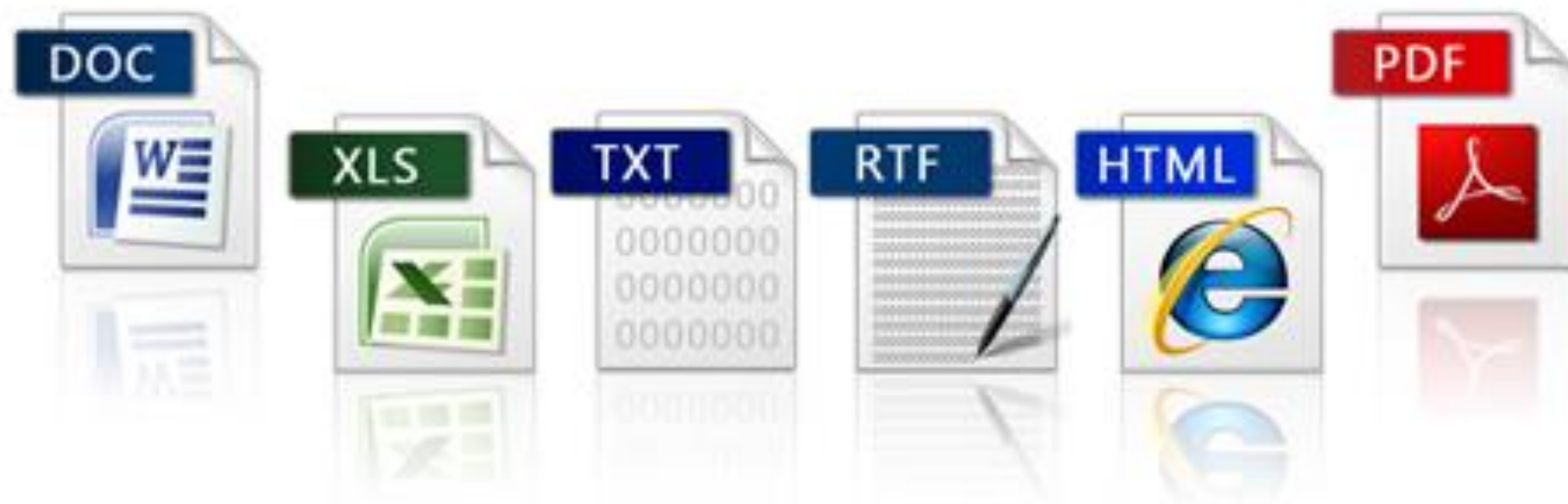
English



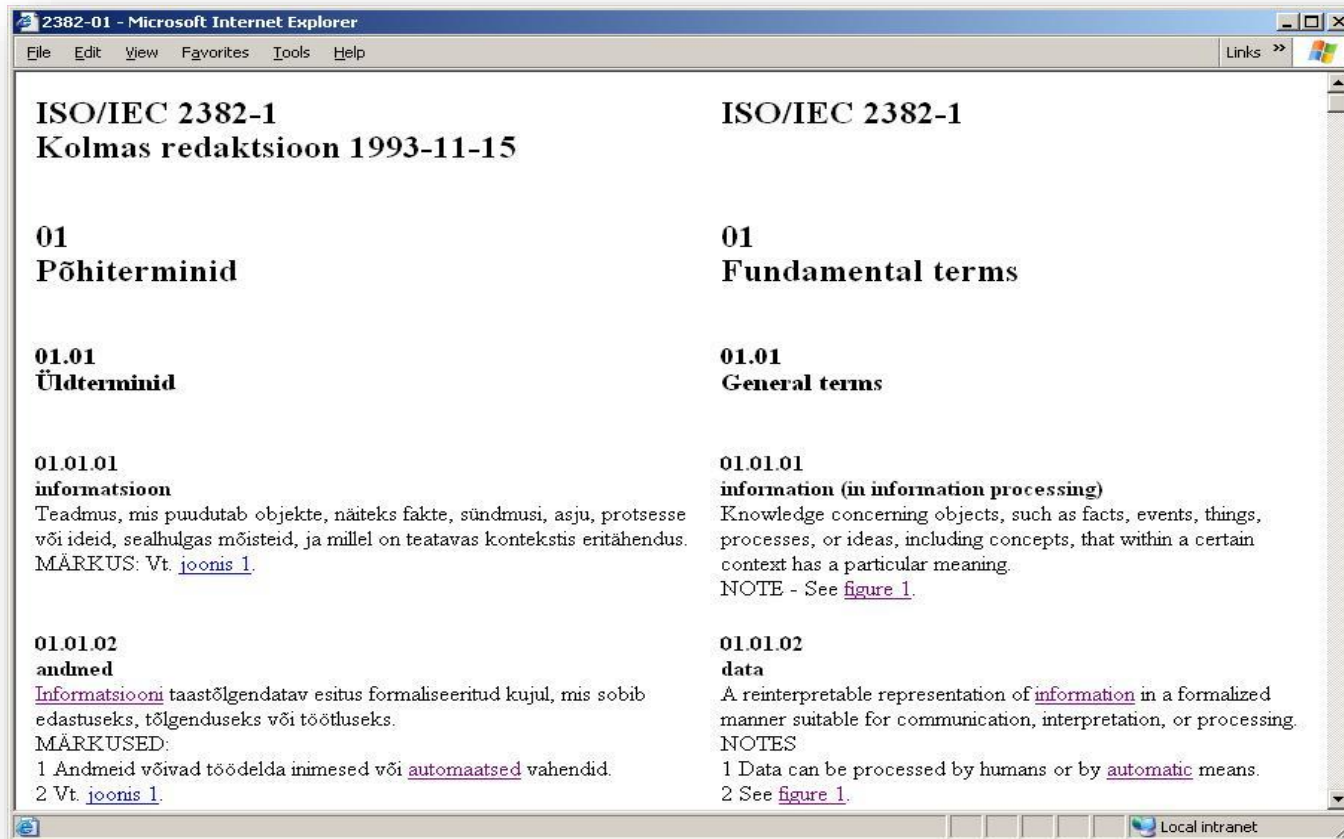
Estonian

Printed sources are hard to process

L 361 lähenema	106	L 388 lülitü	M1 maagiline	107	M 33 majoreerimine
L 361 lähenema piirväärtusele	approach a limit	приближаться к пределу	M		
L 362 lähenemine	approaching	приближение	M 1 maagiline ruut	magic square	магический квадрат
L 363 lähim	closest, nearest	ближайший	M 2 maastikuformaat	landscape size	горизонтальный формат
L 364 lähis-	adjacent	прилежащий	M 3 maatriks	matrix	матрица
L 365 lähis- (= lähend-)	approximate	приблизженный	M 4 maatriks-	matrix	матричный
L 366 lähiskaadet	adjacent side	прилежащий катет	M 5 maatriksalgebra	matrix algebra	матричная алгебра
L 367 lähiskülg	adjacent side	прилежащая сторона	M 6 maatriksesitus	matrix representation	матричное представление
L 368 lähismurd	convergent	подходящая дробь	M 7 maatriksi astak	rank of a matrix	ранг матрицы
L 369 lähte-	initial, original, starting	исходный, начальный	M 8 maatriksi diagonaal	diagonal of a matrix	диагональ матрицы
L 370 lähteandmed (⇒ algandmed)	initial data, source data	исходные данные, начальные данные	M 9 maatriksi elementaartasetused	elementary operations on matrices	элементарные операции над матрицами
L 371 lähtehulk	domain, initial set	множество отправления	M 10 maatriksi jälg	spur of a matrix, trace of a matrix	след матрицы
L 372 lähtepunkt (= alguspunkt)	initial point, starting point	исходная точка, начальная точка, отправная точка	M 11 maatriksi pööramine	inversion of a matrix, matrix inversion	обращение матрицы
L 373 lähtesymbol	initial symbol	начальный символ	M 12 maatriksite liitmine	addition of matrices	сложение матриц
L 374 lähtevõrrand	original equation	исходное уравнение, первоначальное уравнение	M 13 maatriksite ring	ring of matrices	кольцо матриц
L 375 lähtuma	start	исходить	M 14 maatriksmäng	matrix game	матричная игра
L 376 lävi	threshold	порог	M 15 maatriksitähistus	matrix notation	матричная символика
L 377 lühem telg (⇒ väiketelg)	minor axis	малая ось	M 16 maatriksväärtustega funktsioon	matrix-valued function	матрично-значная функция
L 378 lühend	abbreviation	аббревиатура, сокращение	M 17 Mackey topoloogia	Mackey topology	топология Макки
L 379 lühendama	abbreviate, shorten	сокращать, укорачивать	M 18 Maclaurini rida	Maclaurin's series	ряд Маклорена
L 380 lühendamine	abbreviation, shortening	сокращение, укорачивание	M 19 madalaim numbrikoht	least significant digit	младший разряд
L 381 lühendatud	abbreviated, abridged, contracted	сокращенный	M 20 madalamat järku	of lower order	нижнего порядка
L 382 lühendatud korrutamine	abbreviated multiplication	сокращенное умножение	M 21 magasin (= pinu)	stack	магазин, стек
L 383 lühendatud tähis	abridged notation, contracted notation	сокращенное обозначение	M 22 magasinialgoritm	stack algorithm	магазинный алгоритм
L 384 lühike	short	короткий	M 23 magasiniautomaat	push-down automaton	магазинный автомат
L 385 lühim	shortest	кратчайший	M 24 magasinii tipp	top of stack	вершина стека
L 386 lüke (⇒ translatsioon)	shift, slide, translation	перемещение, перенос, сдвиг	M 25 magistraal	turnpike	магистраль
L 387 lüli	link	звено	M 26 mahala tõmbama	cross out	вычеркивать
L 388 lüliti	switch	переключатель	M 27 mahatõmbamine	crossing out	вычеркивание
			M 28 maht (= mahukas)	capacity	ёмкость
			M 29 mahukas	capacious	степенное среднее
			M 30 mahukesmine	power mean	мажоранта
			M 31 majorant	majorant	мажорировать
			M 32 majoreerima	majorize	мажорирование
			M 33 majoreerimine	majorizing	



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



2382-01 - Microsoft Internet Explorer

File Edit View Favorites Tools Help Links »

ISO/IEC 2382-1 Kolmas redaktsioon 1993-11-15	ISO/IEC 2382-1
01 Põhiterminid	01 Fundamental terms
01.01 Üldterminid	01.01 General terms
01.01.01 informatsioon Teadmus, mis puudutab objekte, näiteks fakte, sündmusi, asju, protsesse või ideid, sealhulgas mõisteid, ja millel on teatavas kontekstis eritähendus. MÄRKUS: Vt. joonis 1 .	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 andmed Informatsioon taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötamiseks. MÄRKUSED: 1 Andmeid võivad töödelda inimesed või automaatsed vahendid. 2 Vt. joonis 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 .

Local intranet

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-securite-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 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).

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.

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 2050¹.

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 FAO²) and in the field of nutrition (i.e. additional annual investment of USD 11.8 billion is needed to implement at scale the priority nutrition-specific interventions identified by the Lancet series³ and endorsed by the Scaling Up Nutrition (SUN) movement⁴, according to the World Bank⁵). These investments need to be as well combined with efforts across cross-cutting sectors such as economic empowerment (especially women and young people), education, health, water sanitation and hygiene, etc.

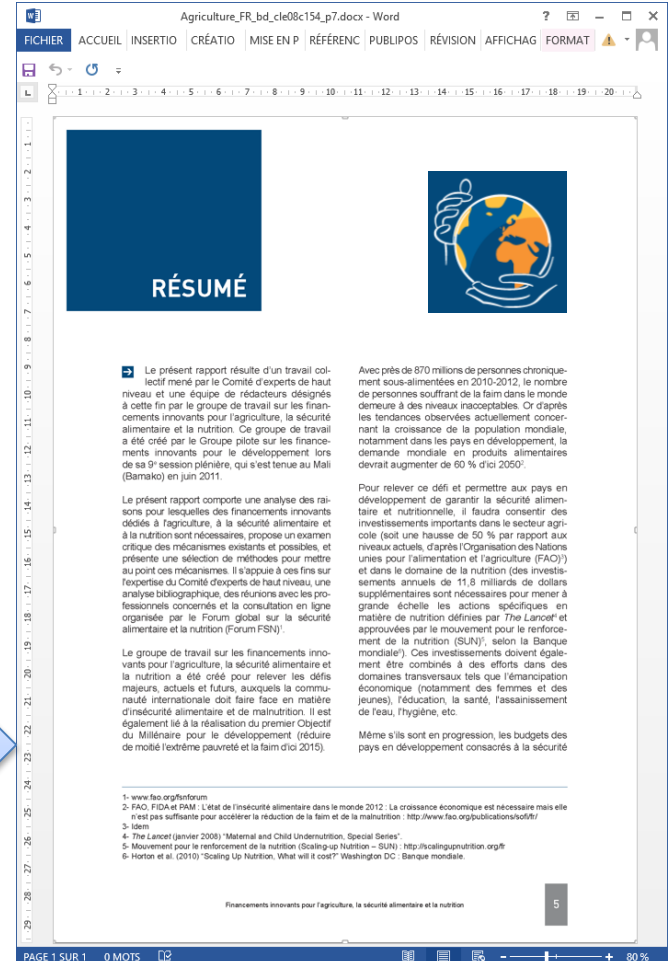
Although they are progressing, budgets for food security, including agriculture and nutrition components, in developing countries are severely constrained. Regarding the agriculture side, the dedicated Official Development Assistance (ODA) has increased in recent years, with higher amounts allocated to emerging middle income countries compared to 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

1- FSN Forum : <http://www.fao.org/fsnforum/>
 2- FAO, IFAD 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- Idem
 4- The Lancet (janvier 2008) 'Maternal and Child Undernutrition, Special Series'
 5- SUN movement: <http://scalingupnutrition.org/>
 6- Horton and al. (2010) 'Scaling Up Nutrition. What will it cost?' Washington DC: World Bank

English version



French version



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 9^e 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).

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 internationale doit faire face en matière d'insécurité alimentaire et de malnutrition. Il est également lié à la réalisation du premier Objectif de Millénaire pour le développement (réduire de moitié l'extrême pauvreté et la faim d'ici 2015).

Avec près de 870 millions de personnes chroniquement sous-alimentées en 2010-2012, le nombre de personnes souffrant de la faim dans le monde demeure à des niveaux inacceptables. Or, d'après 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 2050¹.

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 nutrition définies par The Lancet³ et approuvées par le mouvement pour le renforcement de la nutrition (SUN)⁴, selon la Banque mondiale⁵). Ces investissements doivent également être combinés à des efforts dans des domaines transversaux tels que l'éducation 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

1- www.fao.org/fsnforum/
 2- FAO, IFAD et PAM: L'état de l'insécurité 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- Idem
 4- The Lancet (janvier 2008) '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.

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)¹.

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)¹.

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
[...]



Alignement 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 μακράιωνη συμβολή της στο παγκόσμιο γίνεσθαι, 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 δυνατότητα 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 πρόοδου.
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.	Η Ελλάδα διαθέτει μεγάλο και υψηλής αξίας επιστημονικό δυναμικό, τόσο εντός όσο και εκτός συνόρων.
Greek scientists, with their inventions, innovations and research work, play a leading part in the international scientific community.	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.

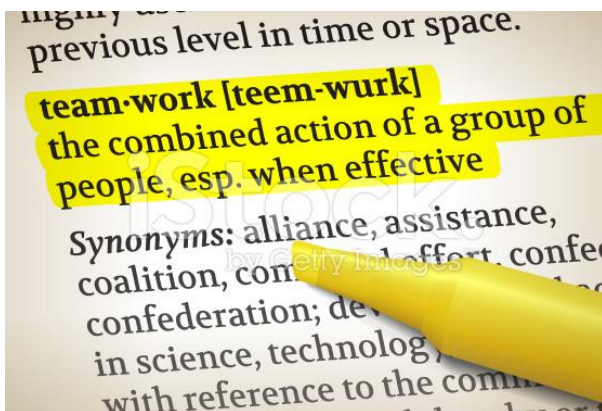
Στους σύγχρονους καιρούς, αυτή η διαδικασία σχεδόν πάντα περιλαμβάνει την αποστολή ηλεκτρομαγνητικών κυμάτων ή ηλεκτρικών σημάτων από κατάλληλες ηλεκτρονικές συσκευές, όπως το τηλέφωνο ή ο ασύρματος, αλλά παλαιότερα περιελάμβανε τη χρήση ακουστικών σημάτων, όπως τυμπάνων, ή οπτικών, όπως ο σηματοφόρος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



5075	Abruzzes	ES	abandono escolar	EL	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



An iceberg floating in the ocean. The small tip above the water represents the public web, while the much larger submerged part represents the deep web. The background is a blue gradient representing the sky and water.

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*



- 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

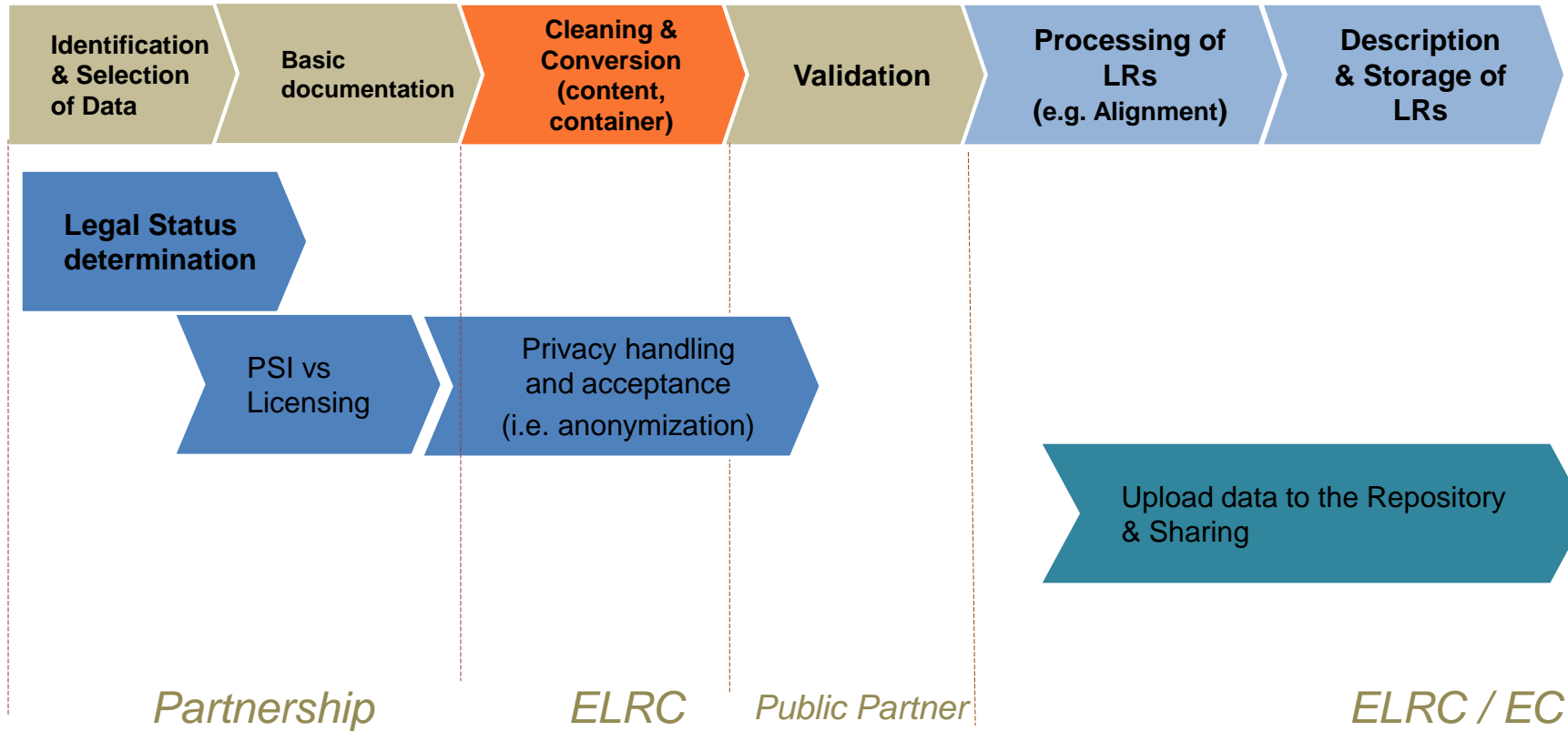
- Identify a large source of data on individuals, organizations etc.
- Use a Named Entity Recognizer (NER) to find and **remove private data** (names, locations, dates, birth information, etc.) and replace with generic placeholders.
- Confirm results meet acceptable requirements
 - Reject data if anonymization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

Illustration of Data Packaging Workflow

Data → LR (Language Resources)

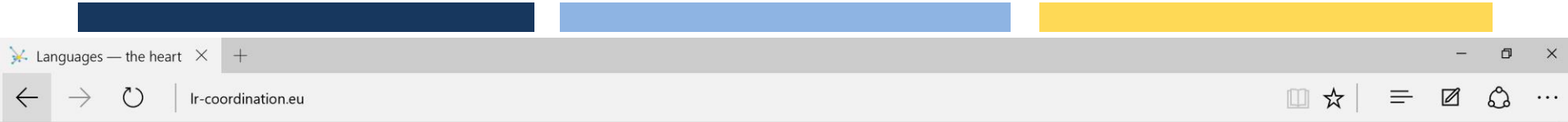


**Value
chain
activity**





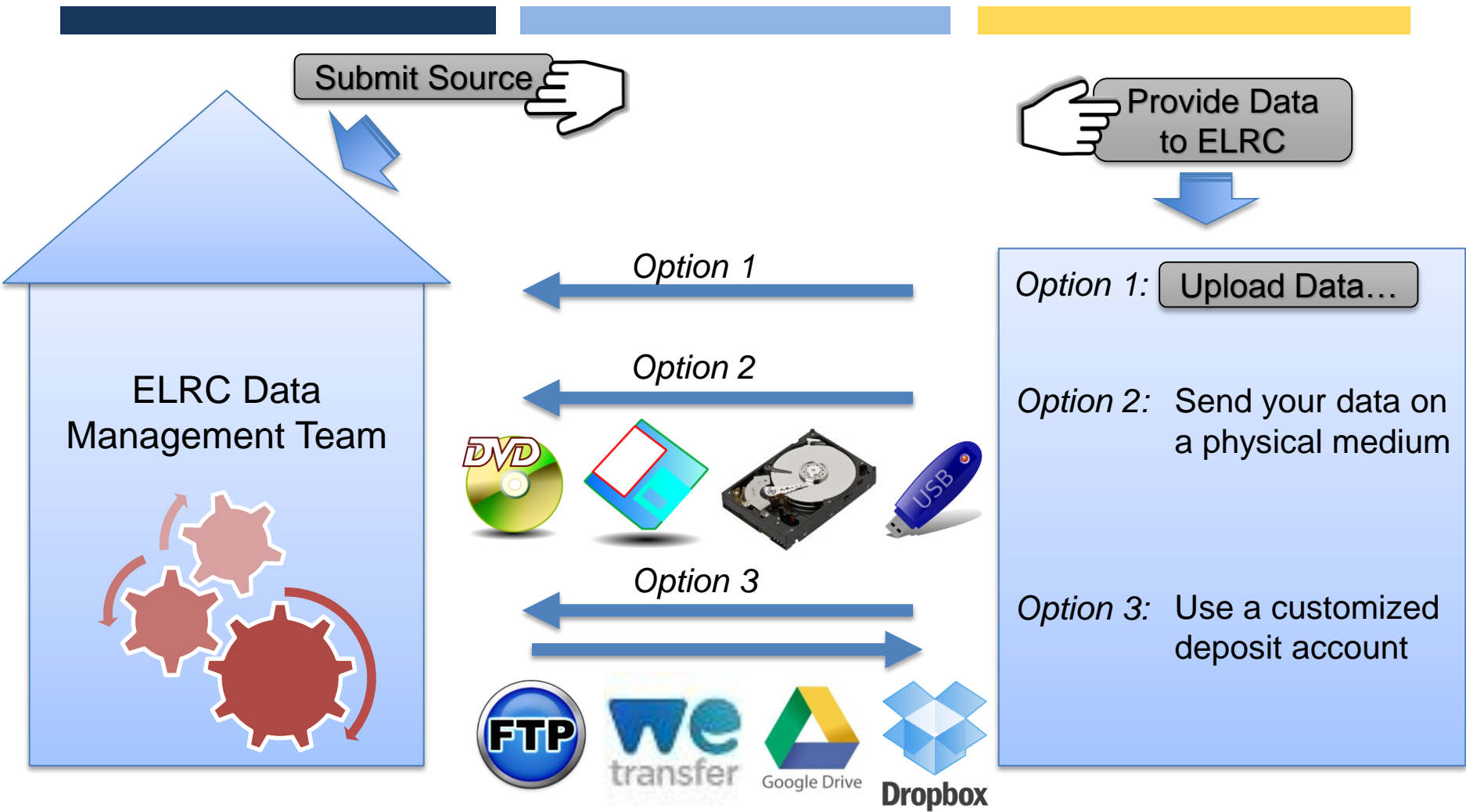
- **Cleaning** of data format
Discarding formatting, encoding character sets e.g. UTF8, formatting features e.g. bold, italic; graphics, ads, tables, html tags, etc.
- File **conversions**
e.g. conversion to XML, XLIFF, etc.
- **Data preparation** for Automated Translation tools
e.g. extraction and alignment
- **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



Languages — the heart of Multilingual Europe



Options for submitting your data



Submit Sources 



URL:

Automatic check if URL already in database. If not, proceed with submission.

Source submission form

Source name:

Languages:

Provider name:

....

Contact name:

Email:



 Submit Data



Data submission form

Data name:

Languages:

Provider name:

Type of data:

....

Contact name:

Email:



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

Follow

Expand

European Data Forum 10 Feb 14
@EUDataForum
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
@publicdataeu
What are you doing for Open Data Day? Inspiration here: blog.okfn.org/2014/02/07/wha... #ODD14 #OpenDataDay #opendata

PublicData.eu 18 Jan 14
@publicdataeu
[PublicData.eu: Providing More Open...](#)

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.



Create your own custom MT with tilde.com/mt

**MACHINE TRANSLATION** **TERMINOLOGY** **LOCALIZATION** 

[Tilde MT](#) [Features](#) [News](#) [Case Studies](#) [Data Library](#)

[Translate](#) [Text Corpora](#) [MT Systems](#) [Administration](#) Super Admins | [Log out](#)

MT Systems

If you have **corpora (translation memories)** that you would like to use you must **UPLOAD HERE** first and only then continue with system creation

1 System properties

Start creating your machine translation system by giving it a name and choosing source and target languages. The corpora (text data for training) which you will choose in the next steps will be filtered by the languages you have selected.

Name / Title *

Subject Domain *

Permissions *

Source language *

Target language *

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



**European Language
Resource Coordination**
Connecting Europe Facility

Takk fyrir!

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



European Language Resource Coordination (ELRC) is a tender operating under the EU's
Connecting Europe Facility SMART 2014/1074 programme.