



Cracking the Language Barrier for a Multilingual Europe

Georg Rehm

META-NET Network Manager – CRACKER Coordinator

DFKI, Germany
georg.rehm@dfki.de

EFNIL Conference 2015

Helsinki, Finland, October 08, 2015



META-NET has received funding from the EU's Horizon 2020 research and innovation programme through the contract CRACKER (grant agreement no.: 645357). Formerly co-funded by FP7 and ICT PSP through the contracts T4ME (grant agreement no.: 249119), CESAR (grant agreement no.: 271022), METANET4U (grant agreement no.: 270893) and META-NORD (grant agreement no.: 270899).

Outline

1. META-NET – recap and overview
2. CRACKER
3. META-FORUM 2015 and Riga Summit 2015
4. Cracking the Language Barrier
5. Connecting Europe Facility – Automated Translation (CEF AT)

META-NET and META

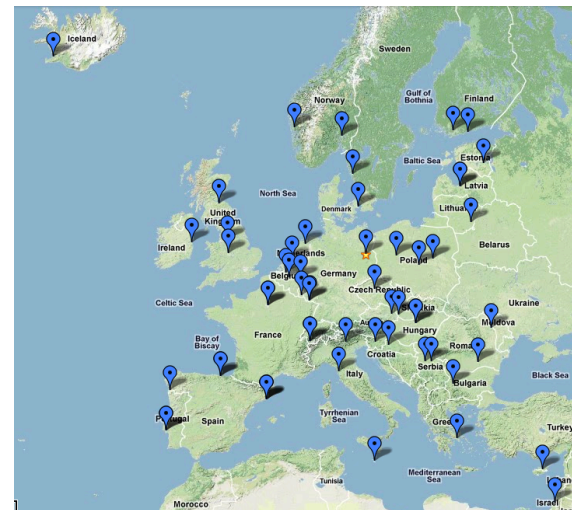


META-NET

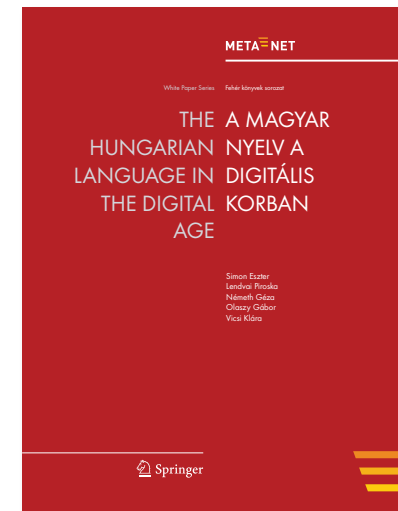
60 research centres in 34 countries
(via four EU-funded projects: T4ME,
CESAR, METANET4U, META-NORD).
New chair of Exec. Board: Jan Hajic.

META

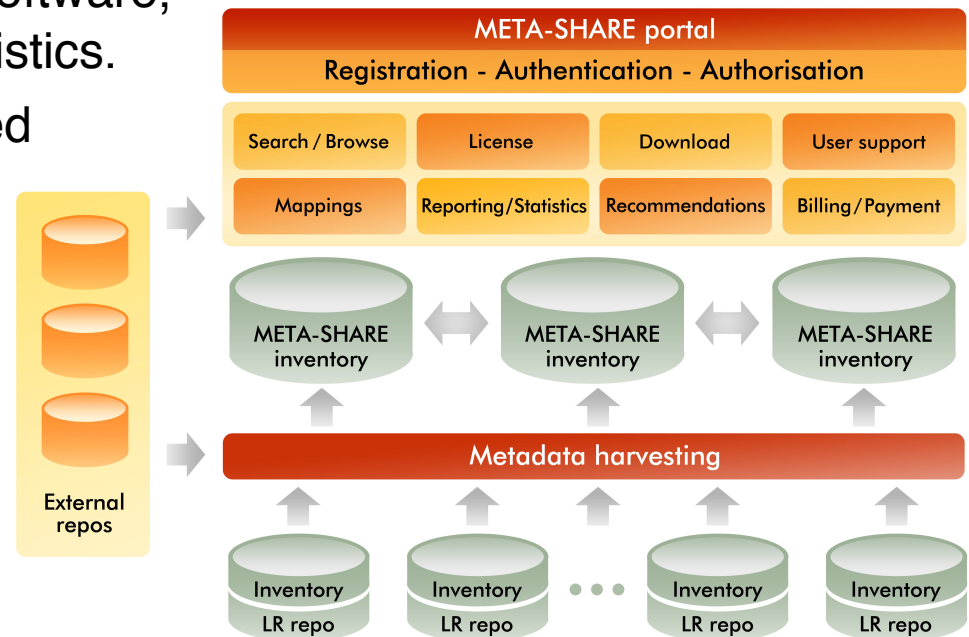
Multilingual Europe Technology Alliance,
800+ members in 70 countries



<http://www.meta-net.eu/members>



- ❑ Pan-European infrastructure, bringing together providers and consumers of language data, tools and services.
- ❑ LRs are documented, uploaded, stored, catalogued, downloaded, shared – to improve visibility, documentation, identification, availability, interoperability.
- ❑ META-SHARE includes repository software, a metadata model, licensing kit, statistics.
- ❑ 29 distributed repositories maintained by 37 organisations in 25 countries.
- ❑ 2.500+ resources (corpora: 49%, lexical: 38%, tools/services: 12%), covering ca. 100 languages.
- ❑ 7.000+ downloads in total; ca. 70% of all LRs have been downloaded.



About the project

META-NET is designing and implementing META-SHARE, a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be both open and with restricted access rights, free and for-a-fee. META-SHARE targets existing but also new and emerging language data, tools and systems required for building and evaluating new technologies, products and services.



About the partners

META-SHARE will start by integrating nodes and centres represented by the partners of the META-NET consortium. It will gradually be extended to encompass additional nodes/centres and provide more functionality with the goal of turning into an as largely distributed infrastructure as possible.

Select network node

Please select one of the following META-SHARE network nodes to proceed. For an explanation of the differences between META-SHARE Managing Nodes and other META-SHARE nodes, you can visit this page.

META-SHARE Managing Nodes

 CNR – National Research Council of Italy	 DFKI – Deutsches Forschungszentrum für künstliche Intelligenz	 ELDA – Evaluations and Language resources Distribution Agency	 FBK – Fondazione Bruno Kessler	 ILSP – Institute for Language and Speech Processing	 IPIPAN – Institute of Computer Science, Polish Academy of Sciences
 Tilde					

Other META-SHARE Nodes

 Budapesti Műszaki és Gazdaságtudományi Egyetem	 Filozofski fakultet Sveučilišta u Zagrebu	 Institute for Bulgarian Language, Bulgarian Academy of Sciences	 Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences	 Lithuanian Language Institute	 National Library of Norway
 Research Institute for Linguistics, Hungarian Academy of Sciences	 Romanian Academy Center for Artificial Intelligence	 Synthema	 University of Belgrade	 University of Copenhagen	 University of Gothenburg
 University of Helsinki	 University of Lisbon	 University of Lodz	 University of Tartu	 University of the Basque Country	 University of Vigo
 Universitat Politècnica de Catalunya					
 Universitat Pompeu Fabra					

Register Login
Home Browse Resources Community Statistics Help About

2,583 language resources at your disposal

What is it? - About the project

META-SHARE, the open language resource exchange facility, is devoted to the sustainable sharing and dissemination of language resources (LRs) and aims at increasing access to such resources in a global scale.

META-SHARE is an open, integrated, secure and interoperable sharing and exchange facility for LRs (datasets and tools) for the Human Language Technologies domain and other applicable domains where language plays a critical role.

META-SHARE is implemented in the framework of the META-NET Network of Excellence. It is designed as a network of distributed repositories of LRs, including language data and basic language processing tools (e.g., morphological analysers, PoS taggers, speech recognisers, etc.).

Go to top ▲

The META-SHARE Infrastructure

META-SHARE aims to provide a multi-layer infrastructure that:

- makes available quality LRs and related metadata over the META-SHARE network;
- ensures that such LRs and metadata are properly preserved and maintained;
- makes a set of services available to all META-SHARE members and users;
- promotes the use of widely acceptable standards for language resource building which ensures the maximum possible interoperability of LRs;
- allows associated third parties to easily index their LRs over the META-SHARE network;
- allows any potential user of the LRs to easily and legally acquire the requested high-quality LRs.

Go to top ▲

Connect and Share! - Our vision

META-SHARE Members set up and maintain their own Language Resource Repositories, the META-SHARE Network Nodes, acting as Repository Service Providers. They can do so by using the open source META-SHARE software which is freely available at GitHub.

META-SHARE Members who do not wish to set up their own language resource repository can deposit their language resources to META-SHARE repositories maintained by other META-SHARE Network Members who undertake to host and/or be responsible for those LRs (hosting repositories).

META-SHARE Managing Nodes are META-SHARE Network Members that provide core services to all META-SHARE Network Members, acting as Core Service and User Support Service providers.

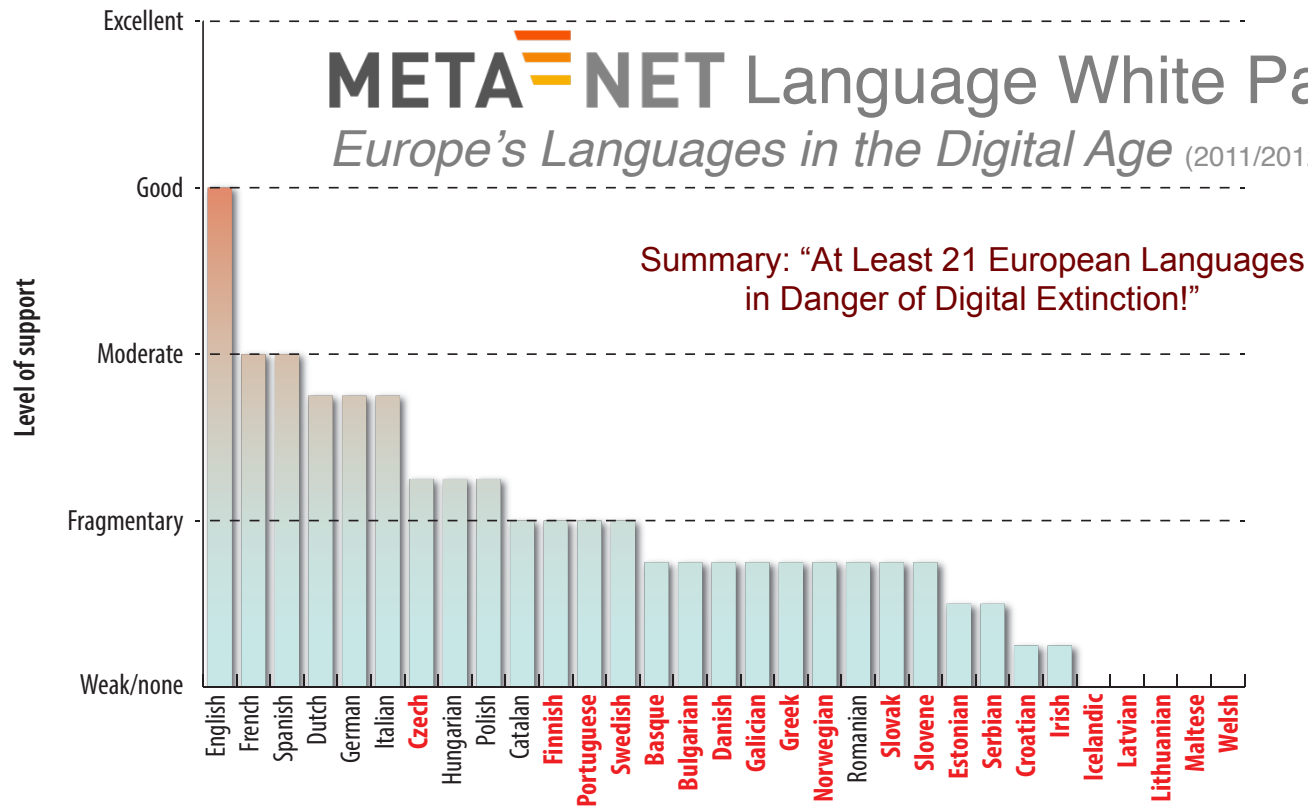
Third Parties are able to use the search services and easily acquire LRs from the META-SHARE network in accordance with the relevant licensing terms and conditions.

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts TAME (grant agreement no. 249119), CESAR (grant agreement no. 274023), META-NET/AL (grant agreement no. 270893) and META-NORD (grant agreement no. 270899).
Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License - Terms of Service

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts TAME (grant agreement no. 249119), CESAR (grant agreement no. 274023), META-NET/AL (grant agreement no. 270893) and META-NORD (grant agreement no. 270899).
Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License - Terms of Service

META³NET Language White Paper Series

Europe's Languages in the Digital Age (2011/2012)



MT

Resources

	excellent	good	moderate	fragmentary	weak or no support
MT		English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish, Welsh
Resources		English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Irish, Latvian, Lithuanian, Maltese, Welsh

CRACKER (CSA, 2015–2017)



CRACKER – Cracking the Language Barrier: Coordination, Evaluation and Resources for European MT Research

1	DFKI	Germany	Georg Rehm (Coordinator)
2	CUNI	Czech Republic	Jan Hajic
3	ELDA	France	Khalid Choukri
4	FBK	Italy	Marcello Federico
5	ATHENA RC	Greece	Stelios Piperidis
6	UEDIN	UK	Philipp Koehn
7	USFD	UK	Lucia Specia

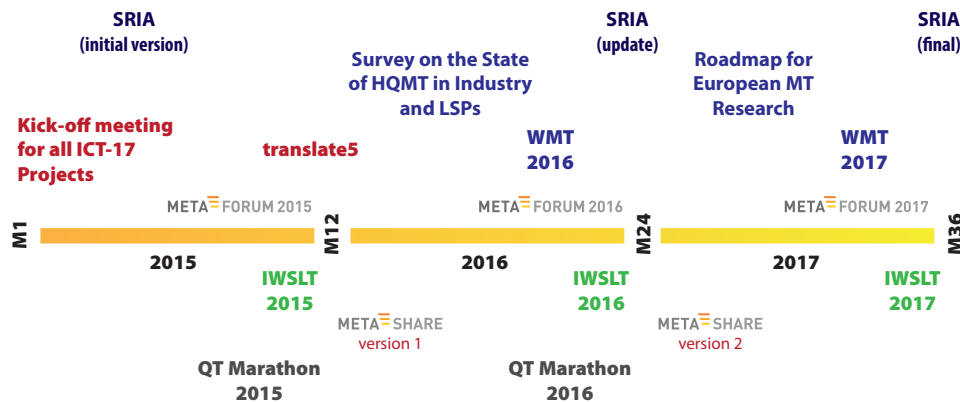


Communities



- META-NET incl. META-SHARE and META
- MT evaluation initiatives – WMT, IWSLT, MT Marathons
- MT and other LT industry
- Language resources – META-SHARE, ELRA
- HT/MT evaluation tool – translate5
- Translation industry and translation profession
- MT user communities

Selected Activities



- **Production of resources** (e.g., for WMT 2016 and 2017, IWSLT 2015-2017)
- **Tools** (quality control, evaluations; towards the idea of a smart workbench for translators)
- **Strategies and roadmaps** (SRIA, Roadmap for European MT Research)
- **Exchange and sharing facility for resources** (META-SHARE)

META= SHARE

Maintenance of Operations and Outreach

- Provide services, adapt them to evolving user requirements and licensing landscape
 - adapt, streamline and extend the metadata schema;
 - adapt licensing toolkit to new international licensing setups;
 - streamline and simplify operations for repository providers and data depositors.
- Technical support and bug fixing

Horizon 2020-ICT-17



META-NET	
META-SHARE META-RESEARCH META-VISION	
2010	
2011	
2012	
2013	

Funded actions in the call H2020-ICT-17 “Cracking the Language Barrier”:

- One Research and Innovation Action
 - **QT21** (Quality Translation 21) – DFKI
- Three Innovation Actions
 - **HimL** (Health in my Language) – UEDIN
 - **MMT** (Modern Machine Translation) – Translated
 - **TraMOOC** (Transl. of MOOCs) – HUB
- Two Coordination and Support Actions
 - **CRACKER** – DFKI
 - **LT_Observatory** – Zabala

META-FORUM 2010-2015

- ❑ **META-FORUM 2010** – Nov. 17/18, Brussels, Belgium
Challenges for Multilingual Europe
- ❑ **META-FORUM 2011** – June 27/28, Budapest, Hungary
Solutions for Multilingual Europe
- ❑ **META-FORUM 2012** – June 20/21, Brussels, Belgium
A Strategy for Multilingual Europe
- ❑ **META-FORUM 2013** – Sept. 19/20, Berlin, Germany
Connecting Europe for New Horizons
- ❑ **META-FORUM 2015** – April 27, Riga, Latvia
Technologies for the Multilingual Digital Single Market

META-FORUM 2015



- ❑ META-FORUM 2015
- ❑ April 27 in Riga, Latvia
- ❑ Riga Summit 2015 on the Multilingual Digital Single Market
- ❑ Highly compressed programme (one day instead of two)
- ❑ Two important components:
 - Emerging SRIA
 - Further community fusing
- ❑ <http://www.meta-forum.eu>

The screenshot shows the website for META-FORUM 2015. The header features the META logo and the tagline 'A Network of Excellence forging the Multilingual Europe Technology Alliance'. The main content area is titled 'META-FORUM 2015 – Technologies for the Multilingual Digital Single Market' and includes a navigation menu, a search bar, and a list of links for the forum. The main text describes the forum as an international conference on language technologies, highlighting the special theme of 'Technologies for the Multilingual Digital Single Market'. It also mentions the registration process and provides a summary of the programme, including topics like Horizon 2020 and Connecting Europe Facility (CEF). The website footer includes logos for META-NET, the European Union, and other partners, along with a small text block about funding from the EU's Horizon 2020 research and innovation programme.

META= FORUM 2015

Technologies for the Multilingual Digital Single Market



... and many more research centres, companies, EU projects etc.

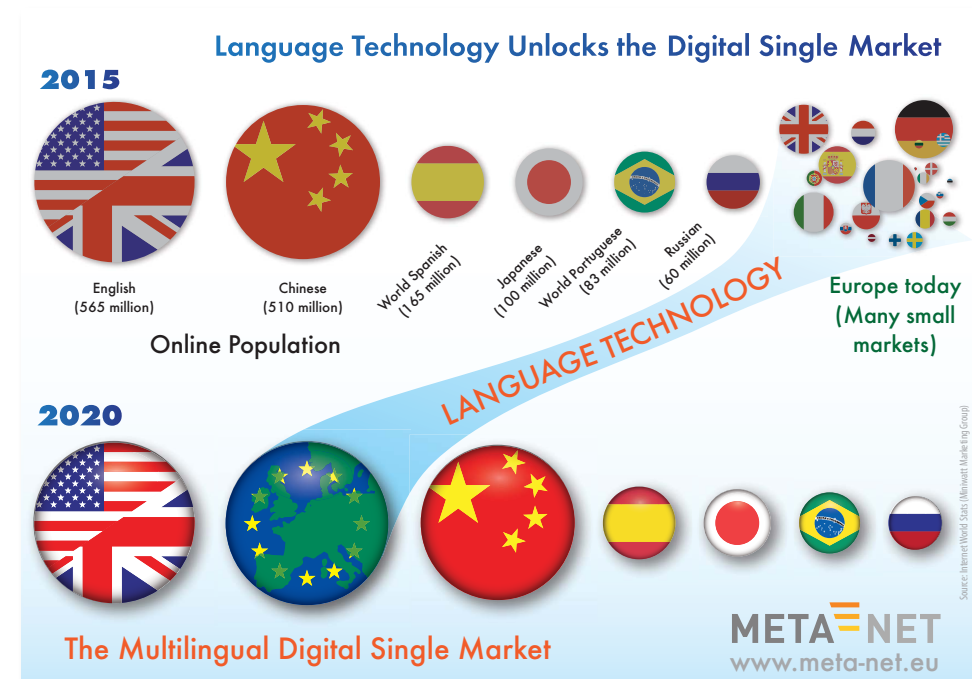
Digital Single Market

- ❑ The DSM is one of the most important priorities of the EU.
- ❑ Expected to add 340b€ to European GDP, 100.000s of new jobs.
 1. Better access for consumers and businesses to digital goods and services (harmonised consumer rules, geo-blocking, copyright etc.)
 2. Shaping the environment for digital networks and services to flourish
 3. Creating a European digital economy and society with long-term growth potential (smart industrial system, standards, cloud computing etc.)
- ❑ Language topic not included in early drafts.
- ❑ EC presented the DSM strategy on May 05, 2015.
- ❑ Unfortunately, language was a complete and utter no show.



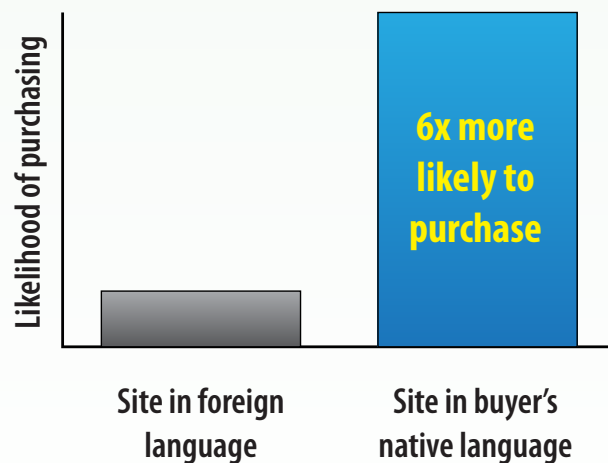
DSM must be multilingual!

- ❑ Europe's DSM must be multilingual to be integrated and unified.
- ❑ Otherwise it will consist of 20+ isolated markets that don't have any bridges between them – language barriers are market barriers!
- ❑ Multilingualism is at the very heart of the European idea.
- ❑ 24 official and many more regional, minority, immigrant languages as well as those of important trade partners.
- ❑ DSM cannot be hampered by language barriers!



Facts and Figures

Customers are **six times more likely to buy** from sites in their native language.



English is not the answer

52% of EU customers **do not purchase** from English-language sites.

Adding even a few languages to an SME's website beyond English can have a **major impact on revenue**. Large organizations today often localize products and websites into fifty or more languages to increase market share.

Most EU languages address less than 3% of the market, fundamentally **limiting SMEs** operating in countries where those languages are spoken.

Geo-blocking and language-blocking are barriers to access

Geo-blocking:

- keeps customers from accessing content due to nationality, location, or residence
- can be worked around by tech-savvy customers
- prevents some cross-border commerce



Language-blocking:

- keeps customers from accessing content in languages they do not speak
- customers never even know what they cannot find
- is unavoidable: no-one speaks all languages; however, current online translation is insufficient
- prevents customers from even *trying* to conduct cross-border commerce
- disproportionately impacts speakers of less common languages

Both geo-blocking and language-blocking are daily problems for tens of millions of EU citizens.

The MDSM Fact Sheet



Why Europe needs a Multilingual Digital Single Market

Europe's 24 official languages present a tremendous opportunity for European business

Removing language barriers within Europe would open access to 73% (with >€25 trillion in annual revenue) of the world's digitally accessible market to European enterprise.

Europe today is not a single market: it is a separated into 20+ small language markets.

Current eCommerce growth within Europe is about half that of the US, due partially to a lack of language coverage from European SMEs.

Less than 5% of European SMEs currently sell cross-language.

No single language accounts for more than 20% of the potential Multilingual Digital Single Market.

Most account for less than 3% of the DSM.

Without a solution, the European Digital Single Market will remain fragmented.

THREE PRIORITY AREAS FOR ACHIEVING THE MULTILINGUAL DIGITAL SINGLE MARKET

1 Multilingual access to all digital goods and services across Europe

Customers are six times more likely to buy from sites in their native language.

English is not the answer

52% of EU customers do not purchase from English-language sites.

Adding even a few languages to an SME's website beyond English can have a major impact on revenue. Large organizations today often localize products and websites into fifty or more languages to increase market share.

Language can be expensive for SMEs

Online businesses face around €5,000 in up-front costs for each new language they translate their websites into, plus similar ongoing costs for maintenance. These figures do not include support and marketing costs.

Even when sites are translated, the vast majority of SMEs cannot respond to support requests or customer feedback in other languages. Such responsiveness is needed to achieve customer satisfaction and build brand loyalty.

Lack of language technology support (automatic translation, tools to assist human translators, and multilingual support in content-creation tools) restricts customer access to content and excludes millions of individuals from the online audience of European businesses.

Geo-blocking and language-blocking are barriers to access

Geo-blocking:

- keeps customers from accessing content due to nationality, location, or residence
- can be worked around by tech-savvy customers
- prevents some cross-border commerce

Language-blocking:

- keeps customers from accessing content in languages they do not speak
- customers never even know what they cannot find
- is unavoidable: no one speaks all languages; however, current online translation is insufficient
- prevents customers from even trying to conduct cross-border commerce
- disproportionately impacts speakers of less common languages

2 Boosting commerce through multilingual technologies

The native languages of approximately 140 million EU citizens are in the Language Technology Danger Zone, where language technology is inadequate to support the DSM.

Current online automatic translation provided by US tech giants does not solve the "language problem": less than 30% of automatically translated content is truly useful for online commerce.

Only three European languages (Spanish, English, and French) meet at least the "moderate" level of language technology support.

Translation opens 20 times its cost in revenue opportunity. However, translation remains too expensive for many European SMEs, blocking this opportunity and limiting economic growth in Europe. Lowering these costs is a strategic opportunity for high-quality automatic translation.

3 Connecting citizens to European digital public services

290 million EU citizens excluded

Over half of EU citizens are language blocked from interacting with the European Commission's web resources for citizen participation.

Without Language Technology, the European Commission has no way to respond effectively to citizen participation.

Current language technology is inadequate for over half of the EU official languages to help the European Commission solve its citizen engagement problem.

Most local governmental services are monolingual only. This poses a problem for tourists, expatriates, and linguistic minorities. Language technology can provide the next generation of technology solutions for public services.

Multilingual eParticipation can help build the European Identity

In the medium-to-long term, multilingual language technologies are needed to enable all European citizens to communicate with one another in their respective native languages with sophisticated machine translation working behind the scenes. Only when EU citizens can interact in their own languages will they truly develop a sense of European identity and community.

Technology Solutions

Investment in the following solutions will help achieve the Multilingual Digital Single Market

Unified Customer Experience (for eCommerce)

- Brings together content, product, customer care, customer relationship, discussion fora, help-desks, etc.
- Unified digital (ecosystem across languages)

Multimodal User Experience for Connected Devices

- Multilingual speech, text, and gesture interfaces
- For connected devices such as robots, cars, household appliances, and consumer products (Internet of Things)

Voice of the Customer

- Comprehensive methods for multilingual market research
- Connects business to customer opinion and experience across borders and languages

Content Curation and Production

- Smart multilingual authoring support
- Multilingual and multimodal report generation, cross-lingual linking, enrichment, and semanticification

Digital Translation Centre

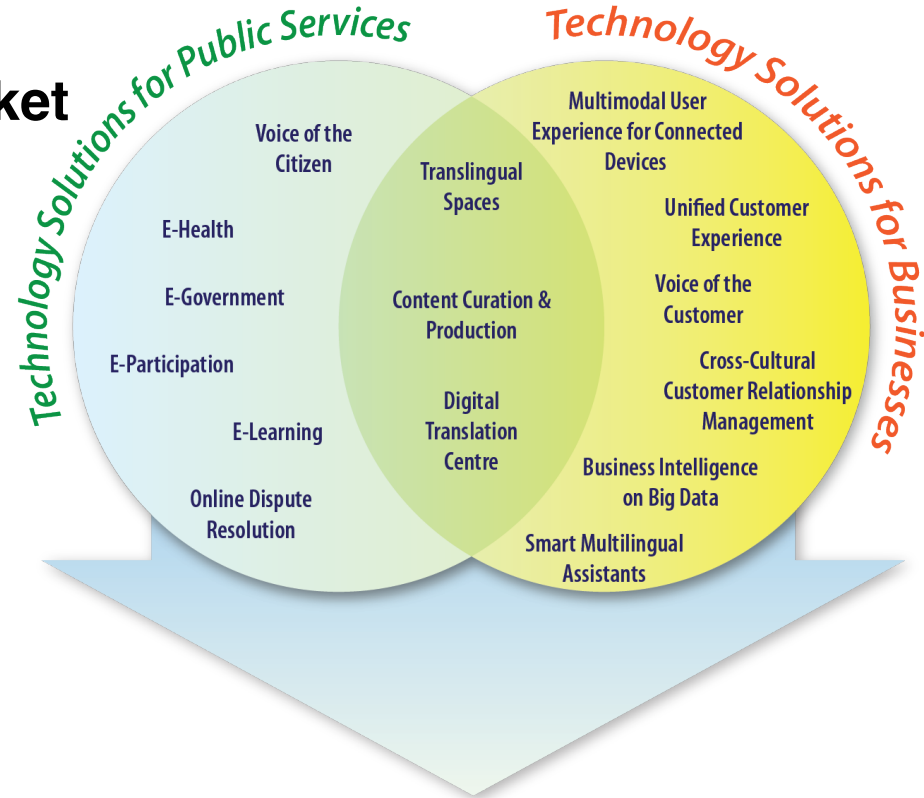
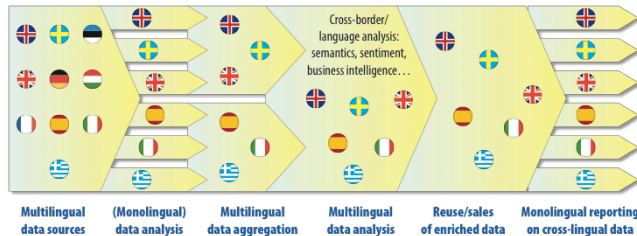
- Automatic translation services
- Free (for the citizen) or for a fee (specialised HQ services)
- To and from businesses, governments, customers, citizens

The forthcoming Strategic Agenda for the Multilingual Digital Single Market will provide additional details on these and other solutions for the needs of the Multilingual Digital Single Market.

Download this fact sheet from <http://cracker-project.eu>. For more information contact Dr. Georg Rehm (DFKI) at georg.rehm@dfki.de.

A Strategy for the MDSM!

- ❑ Strategic R&I Agenda for the Multilingual Digital Single Market
- ❑ Core: Technology Solutions
- ❑ Data economy is an inherent component – LT for effective multilingual data value chains.

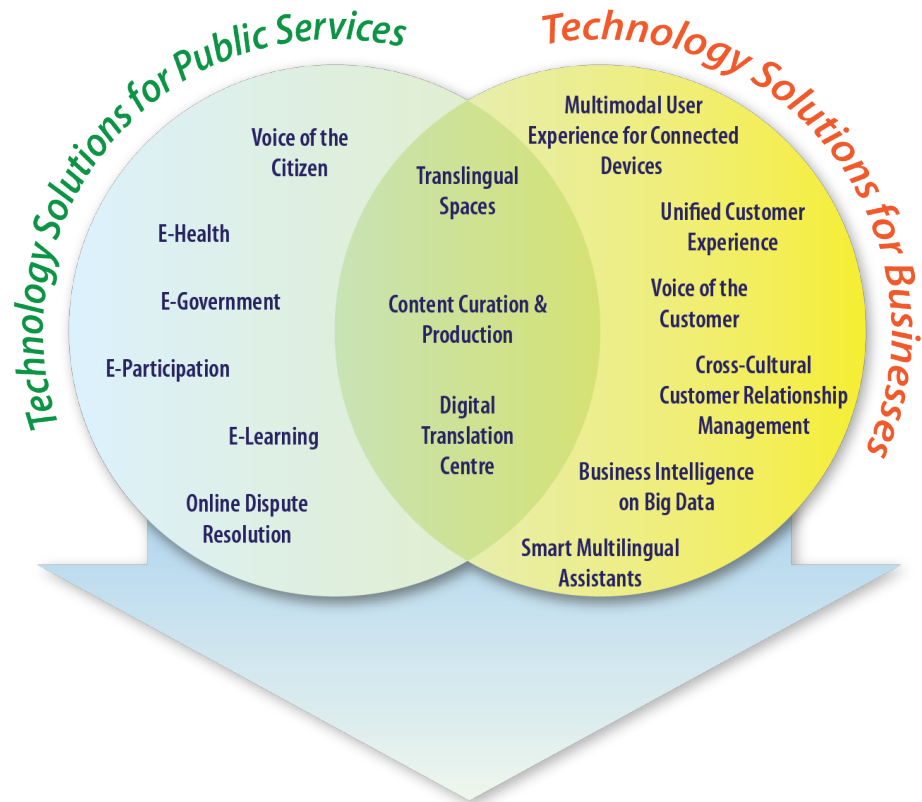


Multilingual Digital Single Market

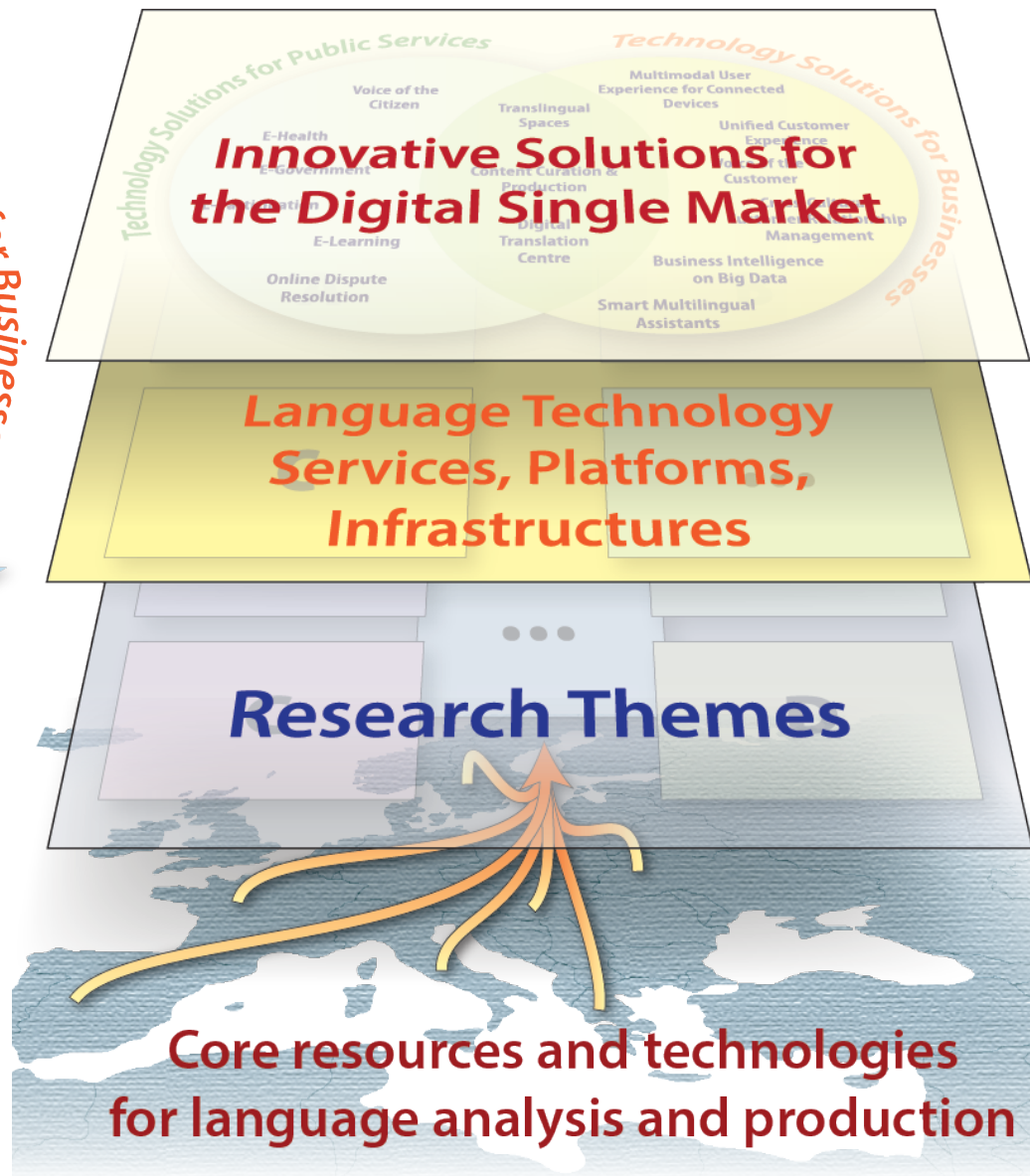
Strategic Agenda for MDSM

- ❑ Presented at META-FORUM 2015 and Riga Summit for the first time.
- ❑ Version 0.5 – *work in progress*
- ❑ Builds upon many strategy papers and roadmaps prepared by several European projects, incl. the META-NET SRA (2013).
- ❑ Input and feedback collected at the Riga Summit 2015 to be used for upcoming versions.





Multilingual Digital Single Market



Joint EFNIL and NPLD Panel



- ❑ Joint EFNIL and NPLD panel at META-FORUM 2015.
- ❑ A joint position paper was presented.



Joint NPLD/EFNIL Position Paper on the Multilingual Digital Single Market

"Languages are not only a means of communication. They also have embedded in them people's values, aspirations and hopes." (European Roadmap for Linguistic Diversity 2015, NPLD)

"Many European languages run the risk of becoming victims of the digital age as they are under-represented and under-resourced online. Huge regional market opportunities remain untapped because of language barriers." (Multilingual Europe: A challenge for language tech. MultiLingual. April/May 2011, page 51/52)

Initially presented at META-FORUM 2015 and the Riga Summit 2015
on the Multilingual Digital Single Market, April 27, 2015

www.rigasummit2015.eu

Open Letter to the EC

- On Friday, March 20, 2015, we published an open letter to the EC on <http://multilingualeurope.eu>.

Published online on Friday, 20 March 2015 17:00 CET

Open Letter to the European Commission

Europe's Digital Single Market must be multilingual!

The Digital Single Market strategy must address the challenge of multilingualism to provide equal digital opportunities in all EU official languages

Read the letter and leave a comment on the European Commission's [Digital4EU platform](#).

- 5 Members of the European Parliament
- 150+ high-level representatives from industry (CxO level)
- 1200+ professors
- 400+ project or research managers
- 20+ entrepreneurs and founders
- hundreds of language and language technology professionals, officials, researchers, administrators and representatives from related stakeholder groups

- On Monday, March 23, 2015, we informed President Juncker and all Commissioners about the campaign and the 1300+ signatures.
- By now more than **3500 signatures!**

Who signed?





Cracking the Language Barrier

- A federation of European projects and organisations working on technologies for a multilingual Europe.
- Multi-lateral Memorandum of Understanding; 8 organisations and 19 projects on board already (including FP7 and H2020-ICT15).
- Selected areas of collaboration: data management and repositories (incl. Data Management Plan), tools, shared tasks and evaluations.
- Goal: provide *one* platform and umbrella organisation for the whole community.

<http://www.cracker-project.eu> • <http://www.meta-net.eu>

The screenshot shows the homepage of the Cracking the Language Barrier initiative. At the top, there is a navigation menu with links for About, Projects, Organisations, Documents, Events, and Contact. The main header features the title "Cracking the Language Barrier" and a subtitle: "A federation of European projects and organisations working on technologies for a multilingual Europe". Below this, a section titled "The Initiative" provides a brief description of the project's goals and collaboration areas. A grid of logos follows, categorized into "Organisations" and "Projects". The "Organisations" row includes logos for CITIA, CLARIN, EFNIL, GALA, LT-INNOVATE.EU, META-NET, npld.eu, and IITAUS. The "Projects" section contains logos for CRACKER, EUMSSI, FALCON, FREME, HimL, KCONNECT, mulider, LT-OBSERVATORY, MixedEmotions, MLI, MMT, MULTIFED, MultiSensor, PHEME, and QT=21. At the bottom, there is a sign-up form for receiving updates, with a text box for the email address and a "Sign up" button. A small disclaimer at the very bottom states that the initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 642537 (CRACKER).



Cracking the Language Barrier



- Website: information about all projects and organisations.
- Downloadable documents.
- List of events.
- Upcoming: news, blog etc.
- Several new members will join the initiative soon.

<http://www.cracking-the-language-barrier.eu>

Organisations



Projects



<http://www.cracker-project.eu> • <http://www.meta-net.eu>



- Connecting Europe Facility (CEF Digital) is a set of pan-European Digital Service Infrastructures (e.g., cross-border e-justice, e-procurement, e-health, Europeana, Open Data Portal)
- CEF AT to provide multilinguality for these Digital Service Infrastructures via the MT@EC system
- CEF AT goals:
 - To make public services usable by all EU users, irrespective of their working language and language skills
 - To facilitate cross-border information exchange in public administration
- Free access to machine translation tool (MT@EC):
 - Apply for ECAS account:
<https://webgate.ec.europa.eu/cas/eim/external/register.cgi>
 - Send an email to DGT-MT@ec.europa.eu asking for activation of MT@EC



- Identify language resources relevant to CEF AT in national public service administrations across Europe
- Provide language resources relevant to CEF AT to the EC
- Provide best services back by adapting CEF AT according to the needs of national public service administrations
- Optimize Europe's Digital Service Infrastructures according to the needs of national public service administrations
- Contribute to the success of the Multilingual Digital Single Market (MDSM)

- ELRC is organising workshops in all countries participating in CEF.
- Target audience: representatives of public administrations and public services (data holders).
- All delegates of the EFNIL 2015 conference who want to know more about ELRC and CEF AT are invited to contact my colleague Andrea Lösch <andrea.loesch@dfki.de>.
- More information: <http://www.lr-coordination.eu>

Summary and Conclusions

- ❑ Europe is only starting to recognise the potential/importance of LT.
- ❑ Language Technology will be a key ingredient of our future IT.
- ❑ EU has a unique opportunity for strategic investment into our future.
- ❑ EU needs to decide if it wants to contribute – or if it can afford not to.
- ❑ Multilingual public services play an increasingly important role (CEF).
- ❑ Strong support of open letter a big surprise – maybe an opportunity?
- ❑ We must increase our collaboration to be able to make a difference.
- ❑ Next steps:
 - (a) updated (and simplified) version of the MDSM SRIA;
 - (b) extension of the Cracking the Language Barrier initiative;
 - (c) maybe try to get the support of the European citizen?

Thank you!

META=FORUM 2015



available at the registration desk



<http://www.meta-net.eu>

<http://www.meta-forum.eu>

<http://multilingualeurope.eu>

<http://rigasummit2015.eu>

<http://www.cracker-project.eu>

<http://www.cracking-the-language-barrier.eu>

<http://www.lr-coordination.eu>