Strategic Agenda for the Multilingual Digital Single Market

Technologies for Overcoming Language Barriers towards a truly integrated European Online Market

Version 0.5 – April 22, 2015

This document was prepared by the projects CRACKER and LT_Observatory. It represents the current state of discussion within the language technology research, development and innovation community towards developing a full Strategic Research and Innovation Agenda for the Multilingual Digital Single Market.

CRACKER has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645357.

LT_Observatory has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644583.

Executive Summary

The integration of the connected Digital Single Market must address our languages: *the Digital Single Market is a multilingual challenge!* Our treasured multilingualism, one of the main cultural cornerstones of Europe and what it means to be and to feel European, is also one of the main obstacles of a truly connected Digital Single Market. It is the goal of the European Language Technology community – including research, development, innovation and other relevant stakeholders – to provide the technological facilities for a truly connected and integrated *multilingual Digital Single Market*.

We recommend setting up a strategic programme to enable the multilingual Digital Single Market. The approach consists of three different layers: the top layer provides *Technology Solutions for Businesses and Public Services*. These innovative solutions are, in turn, supported, enabled, and driven by a middle layer of *Language Technology Services, Platforms and Infrastructures* that provide services for the translation, analysis, production, generation, enrichment and synthesis of written and spoken language. The bottom layer connects *Priority Research Themes* to the infrastructures. It provides concrete scientific results, approaches, technologies, resources, modules, components, and algorithms etc. that can be used to enable and drive forward the middle and top layers. We plan to intensify work on core resources and technologies for language production and analysis because we need to improve the basic technologies for all languages. In order to equip every language with a set of core resources and technologies, we suggest, among others, intensifying knowledge and technology transfer between larger research centres and groups working on technologies for under-supported languages.

The strategic programme will not only unlock the multilingual Digital Single Market, it will provide the European language technology community and also several different industries with the ability to compete with other markets and subsequently achieve multiple benefits for the European economy and future growth, as well as for society and the citizens.

Only through close cooperation between all stakeholders and tightly coordinated collaboration can we realise the ambitious plan of researching, designing, developing and rolling out platforms, services and solutions that support all businesses, public services and citizens in Europe and beyond, to fully realise the multilingual Digital Single Market.

Awareness, political determination and will are required to take us to a leading position in this technology area through a concerted funding effort. This major dedicated push needs to include the political determination to modify and adopt a shared, EU-wide language policy framework that foresees an important role for language technologies.

As Europeans, we urgently need to ask ourselves some crucial questions: Can Europe afford continued language-blocking, market fragmentation caused by language borders, language discrimination, and, eventually, digital language extinction? Can we afford to have our information, communication and knowledge infrastructure depend so much on monopolistic services provided by foreign, non-European companies, effectively constituting technological lock-in? What is Europe's fallback plan in case the language-related services provided by non-European companies that we rely upon are suddenly switched off or if even more serious access or security issues arise? Is Europe actively making an effort to compete in the global landscape for research and development in language technology? Can we expect third parties from other continents to solve our translation and knowledge management problems in a way that suits our specific communicative, societal and cultural needs?

Language Technology made for Europe in Europe will significantly contribute to future European cross-border and cross-language communication, economic growth and social stability while establishing for Europe a leading global position in technology innovation, securing Europe's future as a world-wide trader and exporter of goods, services and information. Only a large, coordinated push of this magnitude will be able to unlock a truly multilingual Digital Single Market.

Contents

1 The Digital Single Market is a Multilingual Challenge

The Digital Single Market (DSM) holds tremendous potential to transform the European economy and make it more globally competitive. However, *one* digital "European market" as such does not yet exist: it is still a collection of many separate smaller markets, confined by national or regional language boundaries. By contrast, China or the United States represent truly national markets. It is no surprise that most of the pioneering growth in ecommerce has happened in the US, where regulatory barriers are lower and one language can address the vast majority of the market. Europe needs to open up these invisible borders created by our different languages – one of the most treasured pieces of our cultural heritage. All of the languages spoken in Europe are also needed in the Digital Single Market: online shops, information pages, public services, encyclopedias, university pages, company websites, user-generated content, online videos, podcasts, radio stations, and other multimedia content all make use of the official, regional, and unofficial minority languages spoken in Europe.

The European Commission predicts that the transition to the integrated DSM will deliver up to €250 billion in economic growth by 2020. Measures like eliminating mobile roaming charges, improving legislation (especially telecom, copyright, data protection), and making cross-border payments easier are all important and necessary preconditions for the DSM. However, they are not sufficient to accomplish the goal. If customers are hampered by language, online commerce will remain confined to fragmented markets, defined by language silos. Even the unacceptable suggestion for everyone to use English would not deliver a single market, since less than 50% of the EU's population speaks English, and less than 10% of non-native speakers are proficient enough to use English for online commerce. Approximately 60% of individuals in non-Anglophone countries seldom or never make online purchases from English-language sites; the number willing to purchase from sites in non-native languages other than English is much, much lower.[1]

As a result, no single language can address 20% or more of the DSM (German comes closest, as the native language of 19% of the EU's population). Addressing the top four EU languages (German, French, Italian, English) would still address only half of the EU citizens in their native language. Even allowing for second-language speakers, no single language can address more than a fraction of the DSM. Concentrating exclusively on the 24 official EU languages would exclude those European citizens from the DSM who speak regional or minority languages, or languages of important trade partners.

Small and medium-sized European companies are a vital component of the DSM. However, only 15% of European SMEs sell online – and of that 15%, fewer than half do so across borders.[2] SMEs that sell their products and services internationally exhibit 7% job growth and 26% innovate in their offering – compared to a job growth of 1% and 8% innovation for SMEs that do not sell their products and services internationally[3]. Only if Europe accepts the multilingual challenge and decides to design and to implement research and innovation-driven technology solutions as well as a service infrastructure with the goal of overcoming language barriers, can the economic benefits of the Digital Single Market be achieved (**Figure 1**). Enabling and empowering European SMEs to easily use language technologies to grow their business online across many languages is key to boosting their levels of innovation and job creation.

---

1 Common Sense Advisory (2014): "Survey of 3,000 Online Shoppers Across 10 Countries Finds that 60% Rarely or Never Buy from English-only Websites", https://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDet&tabID=64&Aid=21500

2 EC (2015): "How digital is your country? New figures reveal progress needed towards a digital Europe", http://europa.eu/rapid/press-release_IP-15–4475_en.htm.

3 EUbusiness: "Annual Report on European SMEs 2013–14 – A Partial and Fragile Recovery", http://www.eubusiness.com/topics/sme/report-2014.

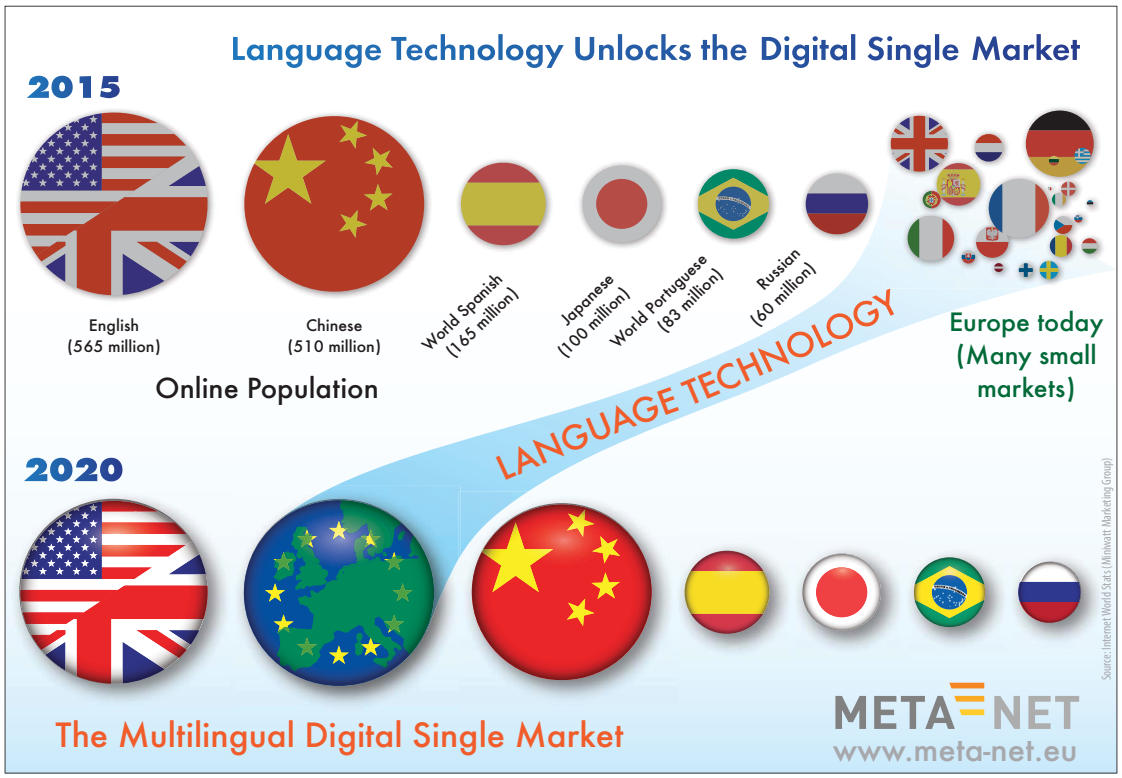


Figure 1: *Language technology unlocks the Digital Single Market*

The European Digital Single Market today would account for approximately 25% of global economic potential. However, if Europe were to overcome the language barriers that hamper intra-European trading, it would also remove barriers to international trade that keep European SMEs from achieving their full economic potential by penetrating markets in other continents beyond our own. Addressing the official and major regional languages of Europe would open access to over 50% of the world's online potential and 73% of the world online market in economic terms, amounting to an online market of approximately €25 trillion in 2013.[4] Most of this increase comes from English, Spanish, French, and Portuguese, but other languages also make significant contributions to world-wide market access. The *global* potential for European businesses exceeds the *continent-internal* opportunities from the DSM by orders of magnitude.

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.

1.1 Overcoming Language Barriers with Technologies

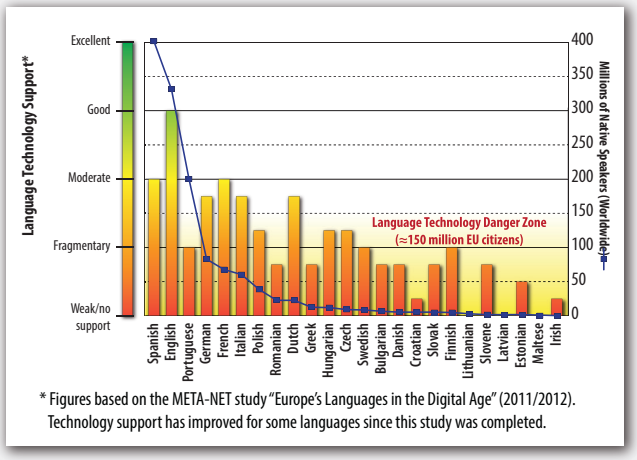
The borders between our beloved languages are invisible barriers at least as strong in their separating power as any remaining regulatory boundaries. They create fragmented and isolated digital markets with no bridges to other languages, thereby hampering the free flow of products, commerce, communication, ideas, help, and thought. Language barriers of this type in the online world can only be overcome

[4] Benjamin B. Sargent, Common Sense Advisory (2013): "The 116 Most Economically Active Languages Online, https://www.commonsenseadvisory.com/AbstractView.aspx?ArticleID=5590.

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.

Online Automatic Translation Quality

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.



* Figures based on the META-NET study "Europe's Languages in the Digital Age" (2011/2012). Technology support has improved for some languages since this study was completed.

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

completely by (1) significantly improving one's own skills in non-native languages, (2) making use of others' language skills, or (3) through using digital technologies. With the 24 official EU languages and dozens of additional languages, relying on the first two options alone is neither realistic nor feasible. For specific types of content and purposes, specialised human language services increasingly assisted by language technology, will continue to play a major role in translating documents, creating subtitles for videos, or localising websites into 20+ other languages. However, relying on human services would exclude most SMEs from the DSM because of the high costs involved. It would create a market that can only be successfully penetrated by large, consolidated enterprises, which is why cost-effective methods must be found to support market access for SMEs and European citizens.

To succeed, any SME must both excel in communicating its expertise in its market niche and be able to engage in two-way conversations with its customers online. The free machine translation services offered by a few US-tech giants are useful for giving users the gist of web content. But they cannot be easily and cheaply tailored to support the niche communication needs between SMEs and their customers. Supplementing this with domain-tailored language services such as content and sentiment analysis, knowledge extraction, and multimodal online engagement is well out of reach for SMEs aiming to engage the half of the EU consumers who do not enjoy English, German, French or Italian as their native language.

The connected and truly integrated Digital Single Market can only exist once all language barriers have been overcome and all languages are connected through technologies. Only advanced communication and information technologies that are able to process and to translate spoken and written language in a fast, robust, reliable, and ubiquitous way, producing high-quality output, can be a viable long-term solution for overcoming language barriers.

Unfortunately, establishing such a technological infrastructure requires an immense collective push that involves designing and implementing technologies, services, and platforms, accelerating innovation, basic and applied research, as well as efficient technology transfer. While a few of our languages are in a moderate to good state with regard to technology support, more than 70% of our languages are seriously under-resourced, actually facing the danger of digital extinction (for example, Maltese, and Lithuanian), even though it must be noted that support for these languages with smaller numbers of speakers is slowly increasing (more details can be found in the Appendix).[5]

[5] See the results of the META-NET White Paper Series, http://www.meta-net.eu/whitepapers.

1.2 Language Technologies Made for Europe – in Europe

Today's IT systems are only just beginning to handle the meaning, purpose, and sentiment behind our trillions of written and spoken words. Language makes up a very large part of our big data treasure. Today's computers cannot understand texts and questions well enough to provide translations, summaries or reliable answers in all languages. Yet in less than ten years such services could be offered for many. Technological mastery of human language can enable a multitude of innovative IT products and services in industry, commerce, government and administration, private and public services, education, health care, entertainment, tourism, and many other sectors.

Language technology is therefore the missing piece of the puzzle that will bring us closer to a fully integrated DSM. But language technology does more than enabling the DSM. It is a key technology for the next generation IT, which will be much smarter and human-centered in its functionality. Almost every digital product uses and is dependent on language – which is why language technology is an absolutely mandatory component! It is the key enabler to boosting growth in Europe and strengthening our competitiveness in a technology sector that has become incredibly critical for Europe's future, considering the significance given to the DSM by the European Union.

Our different European countries and language communities constitute a set of individual, unconnected, fragmented, isolated markets. A truly integrated Digital Single Market that spans our whole continent can never exist if we ignore the "language factor" and the de facto state of play: European citizens are unable to access vast amounts of online content due to *language-blocking*. The European economy is suffering as well because there are no technical means that enable, say, a restaurant owner in Latvia to order ten crates of wine in Portugal if the restaurant owner, who speaks Latvian, is unable to find the website of the vineyard, presented in Portuguese, in the first place. And negotiating and completing a deal would require a translator.

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

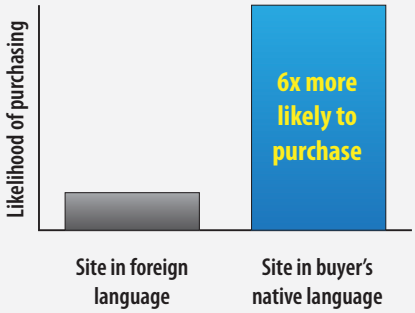
Both geo-blocking and language-blocking are daily problems for tens of millions of EU citizens.

Europe is the most appropriate place for accomplishing the needed breakthroughs in technology by virtue of fundamental and applied research, and even more so in technology development and profitable innovation. Our continent has half a billion citizens who speak one of over 60 European and many non-European languages as their mother tongue. Europe has more than 2,500 small and medium-sized companies in language, knowledge, and interface technologies, and more than 5,000 companies providing language services that can be improved and extended by technology. Europe also has a long-standing research, development, and innovation tradition with over 800 centres performing excellent, highly visible, and internationally recognised scientific and technological research on all European and many non-European languages.

1.3 Online Use of Languages

Current research on the online use of languages demonstrates that there is increasing pressure to overcome language barriers. Online content in hitherto dominant languages is declining and "long-tail" languages are rising.[6] In line with the constant rise of online content, absolute numbers are rising for all languages, and much more significantly so for less common languages. One example in Europe: Basque, Galician, and Catalan all have an increasing share vis-a-vis Spanish; even though the numbers are small, they indicate a long-term shift.

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.

This trend goes hand in hand with increasing public demand for content in regional or local languages due to the increasing availability of broadband as well as high speed mobile connectivity and rising numbers of online users and online services. Europe's citizens are no longer satisfied with using only a few major languages. As a consequence, businesses that cannot provide content in local languages will be global losers. Furthermore, the numbers indicate that market saturation for dominant languages has been reached and that any additional growth is coming from outside the established markets, historically served by a smaller set of languages.[7] If we extrapolate the trends reported by Common Sense Advisory, it only took 37 languages to reach 98% of the world online population in 2009, but already 48 in 2012. The predicted number in 2015 is 62 languages.

More and more citizens are connected and, as a consequence, more and more citizens use – and expect to use – their own native languages in any online activities. However, they are excluded from participating in many online activities due to the fact that language barriers constitute market barriers – especially so with regard to the DSM. True engagement with consumers across language barriers is also deeply entwined with the user's technical, cultural, and individual awareness, preferences, and requirements. The power of personalising any cross-linguistic exchange to an individual user means we should not merely bridge the language barrier but provide the kind of compelling personalised user experience that is key to a vibrant and competitive DSM.

---

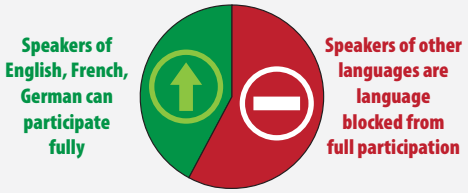
6 Common Sense Advisory (2013): "The Rise of Long-Tail Languages".

7 "Traditional 'power house' languages are seeing some of the biggest drops in overall site support: e.g. German: –11.7%, French: –13.4%, Spanish –14.4%, i.e. a smaller percentage of 'global' sites are supporting these languages, even as the number supporting long-tail languages is increasing." (ibid.)

The impact a truly connected DSM could have is not just felt in terms of sales. Technological integration fails the test if users cannot understand the content available from systems. For example, electronic standards for integrating health records simply add cost without benefit if the recipient is not able to interpret and use those records. If doctors' notes and observations remain in one language and are not accessible, they cannot help doctors in another region, e.g. if a traveller from Poland falls ill while in France. Here the impact of language barriers is measured not just in terms of Euros but in terms of health and, potentially, lives.

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.

1.4 Multilingual Big Data Text Analytics for the European Data Economy

The "language component" is not only a necessary ingredient of the Digital Single Market, it is also a mandatory enabler for the future European Data Economy.

It has been said for a number of years now that data is the oil of the 21st century. Data linking and content analytics are key technologies for refining this oil so that it can drive the engines of understanding – data homogenisation, semantic analysis, enrichment, and repurposing. It is important to note that the large data sets of our Big Data age are never solely numerical data – they always come with natural language components such as, for example, column heads in database tables, free text in table cells, metadata annotations, descriptions, documentation, summaries, links to specific documents etc. In other words, the new Data Economy is not only an integral part of the Digital Single Market. It will require innovative new mechanisms that enable data sets and data value chains to flow freely across language boundaries (**Figure 2**).

In addition to the multilingual challenge, we need to pay attention to the sheer volume of data generated. For example, only one hour of customer transaction data at Walmart, corresponding to 2.5 petabytes of data, is 167 times the amount of data housed for example by the Library of Congress.[8] Data growth keeps rising: 90% of the data available today has been generated in the past two years only.[9] IDC (International Data Corporation) estimates that all digital data created, replicated or consumed will grow by a factor of 30 between 2005 and 2020, doubling every two years. By 2020, it is assumed that there will be over 40 trillion gigabytes of digital data, corresponding to 5,200 gigabytes per person on earth.[10] The Internet of

[8] Beñat Bilbao-Osorio et al. (ed.) (2014): "The Global Information Technology Report 2014 – Rewards and Risks of Big Data", World Economic Forum and INSEAD.

[9] SINTEF (2013): "Big Data, for better or worse: 90% of world's data generated over last two years".

[10] John Gantz and David Reinsel (2012): "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", International Data Corporation (IDC).

Figure 2: *Multilingual data value chains*

Things and Web of Things will add not only more but additional types of data (including large amounts of textual data, of course): Cisco estimates that currently less than 1% of physical objects are connected to computer networks. According to recent estimates by Cisco this number will rise to up to 50 billion devices connected to the Internet by 2020, corresponding to between 6 and 7 devices per person on the planet. So Europe needs a scalable technological infrastructure for handling its big data sets. While the specific Big Data solutions circling around computer science and database technologies will be taken care of by the Big Data Value Contractual Private Partnership (BDV cPPP), the examples given above demonstrate the need for complementary language technologies and how they can create synergies with the BDV cPPP by including robust and precise multilingual text analytics technologies that can perform at web-scale level and, even more crucial, at an Internet of Things level.

Big Data analytics will not just be "slightly better" if we include language technology – it simply won't happen! We cannot download Big Data into a database and then build applications on top of it – we will need to process it sensibly and that sense will need to be based on language. This challenge not only relates to structured Big Data but also to any type of unstructured data (i.e. Linguistic Big Data) including text documents and social media streams, essentially any sequential symbolic process of meaningful information. Language technologies will build bridges from big data to knowledge, from unstructured data to structured data. Language Technology will become the foundation for organising, analysing, and extracting data in a truly useful way, it must be and will become a necessary ingredient in any data value chain.

We suggest engaging in a close, complementary collaboration with the BDV cPPP to ensure that the multilingual big data value chains reflect the subtleties and variety of language in the use of vocabulary, register, idioms, and tone that is distinct to individuals, communities, and domains. The European Big Data Value Strategic Research and Innovation Agenda already mentions the need for complementary research and innovation activities on Linguistic Big Data: "In Europe, text-based data resources occur in many different languages, since customers and citizens create content in their local language. This multilingualism of data sources makes it often impossible to use existing tools and to align available resources, because they are generally provided only in the English language. Thus, the seamless aligning of data sources for data analysis or business intelligence applications is hindered by the lack of language support and availability of appropriate resources." (p. 26).[11]

1.5 EC and Language Technology – Past and Present

Already in the late 1970s the EU realised the profound relevance of language technology as a driver of European unity and began funding its first research projects, such as EUROTRA (1978–1992). After a longer period of sparse funding, the EC set up a department dedicated to language technology and machine translation; this department was later integrated into the new "Data Value Chain" unit in the EC Directorate General for Communications Networks, Content and Technology (DG Connect).

In the past decade or so, the EU has been supporting projects such as EuroMatrix, EuroMatrixPlus (2006–2008, 2009–2012), Let'sMT! (2010–2012), and iTranslate4 (2010–2012), which draw on basic and applied research along with industrial collaboration to generate machine translation resources for many European languages for as the Moses system. More recently, the large-scale META-NET initiative (supported in its first phase by four EU projects), which started in 2010, has assembled the Language Technology community around its core network of excellence which consists of 60 research centres in 34 European countries: META, the Multilingual Europe Technology Alliance, has more than 750 members. META-NET has prepared studies such as its 30-volume White Paper Series, and the META-NET Strategic Research Agenda for Multilingual Europe.[12] The open resource exchange infrastructure META-SHARE provides access to thousands of language resources and technologies. The EU has also facilitated the coalescing of the LT industry through the FP7 support action LT COMPASS. The resulting industry association, LT-Innovate, which currently counts 180 corporate members. LT-Innovate issued a Report on the State of the European Language Technology Industry[13] and an Innovation Agenda[14]. At the beginning of 2015 new projects have been launched, funded through the Horizon 2020-ICT 17 call, "Cracking the Language Barrier". In addition to the large research action QT21, which is working on new paradigms for high-quality machine translation, three innovation actions are adapting and applying new MT methods for industrial and commercial use cases.

In parallel to the research and innovation-oriented activities funded through FP7 and Horizon 2020, the EC is further advancing the Connecting Europe Facility programme (CEF). Part of CEF Digital is the Automated Translation building block that "helps European and national public administrations exchange information across language barriers in the EU" and also to make all of CEF's Digital Ser-

[11] BDVA (Big Data Value Association): "European Big Data Value Strategic Research and Innovation Agenda" (V1.0), http://www.bdva.eu/sites/default/files/europeanbigdatavaluepartnership_sria__v1_0_final.pdf (2015)

[12] See http://www.meta-net.eu/whitepapers and http://www.meta-net.eu/sra.

[13] LT-Innovate Innovation Agenda & Manifesto (2014): "Unleashing the Promise of the Language Technology Industry for a Language-neutral Digital Single Market".

[14] LT2013 (2013): "Status and Potential of the European Language Technology Markets".

vice Infrastructures multilingual.[15] This automated translation service, CEF AT, builds on an existing machine translation system, MT@EC, developed at the EC (DG Translate). It is being implemented on the Moses toolkit, under the Interoperability Solutions for European Public Administrations (ISA) programme. One of the key ideas is to harness the linguistic knowledge embodied in the EC's database of translated documents covering the 24 official languages of the EU. MT@EC is currently only available to staff members of the EC and the EP as well as public administrations of EU member states. Currently a closer collaboration between CEF AT.DSI and the European language technology community is being established, especially with regard to the systematic and coordinated collection and exploitation of language resources in all CEF participating countries.

Looking beyond the EC, research by TAUS[16] has shown that European research funding that fostered the development of the open source machine translation toolkit Moses has opened up new business opportunities in language technology by enabling companies to reduce the cost required to translate content, particularly in fields such as technical support. These cost reductions have helped companies to increase their multilingual reach and engage with customers in language markets inaccessible through traditional translation routes. There is a clear long-term trend to increasing language support and increasing customer engagement via language technologies. According to the report, there are already 22 operative Moses-based MT companies with an estimated market share of about $45 million or about 20% of the entire MT solutions market.

1.6 The Economic Power of Language Technology and Services

In addition to being a key enabling technology for the multilingual DSM, Language Technology comes with a non-trivial economic power itself.

The European market for translation, interpretation, and localisation was estimated to be €5.7 billion in 2008. The subtitling and dubbing sector was at €633 million, language teaching at €1.6 billion. The overall value of the European language industry was estimated at €8.4 billion and expected to grow by 10% per year, i.e. resulting in ca. €16.5 billion in 2015.

The global language technology industry[17] is evaluated at €26.5b in 2015, projected to rise to €65b by 2020. The global speech technology market alone will reach ca. US$20.9 billion by 2015 and ca. US$31.3 billion by 2017.

Yet, this existing capacity is not enough to satisfy current and future needs, e.g. with regard to translation. Today, Google Translate translates the same volume per day as all human translators on the planet translate in one year.

---

15 Connecting Europe Facility (CEF): "Automated Translation", https://joinup.ec.europa.eu/community/cef/og_page/catalogue-building-blocks#AT

16 Achim Ruopp, Jaap van der Meer, TAUS (2015): "Moses MT Market Report", https://www.taus.net/think-tank/reports/ translate-reports/moses-mt-market-report.

17 Figures from "LT2013: Status and Potential of the European Language Technology Markets", April 2013

2 A Strategic Programme for the Multilingual Digital Single Market

The integration of the connected Digital Single Market, by definition, must address our different languages. *The Digital Single Market is a multilingual challenge!* Our treasured multilingualism, one of the cultural cornerstones of Europe and one of the main assets of what it means to be and to feel European, is also one of the main obstacles of a truly connected Digital Single Market (**Figure 3**).

Our goal is to provide the technological facilities for a truly connected and integrated multilingual Digital Single Market. To this end, we – the European Language Technology Community including research, development, and innovation – propose to research, design, and implement a set of technology solutions, services, and infrastructures. As a mission-critical enabling activity, we also propose to increase monolingual, crosslingual, and multilingual technology support required for all languages spoken by a significant part of the population in Europe.

In order to address this challenging goal, we propose a broad-based and strategic programme built with three layers and firmly grounded in Europe's language communities:

- Layer 1: Innovative Technology Solutions for the Multilingual Digital Single Market
- Layer 2: Language Technology Services, Platforms, Infrastructures
- Layer 3: Priority Research Themes

2.1 Layer 1: Innovative Technology Solutions

On the **Solutions Layer** (Layer 1, **Figure 4**) we suggest focusing on technology solutions for businesses and public services to make use of novel technologies in solutions with high economic and societal impact. This should create numerous new business opportunities for European SMEs geared towards the multilingual Digital Single Market. We only briefly list the different solutions here, they are further elaborated upon in Chapter 3.

- **Solutions for Businesses:** Unified Customer Experience; Cross-Cultural Customer Relationship Management; Voice of the Customer; Business Intelligence on Big Data; Content Curation and Production; Multimodal User Experience for Connected Devices; Smart Multilingual Assistants; Translingual Spaces; Digital Translation Centre.
- **Solutions for Public Services:** Voice of the Citizen – Social Intelligence on Big Data; E-Participation; E-Government; Online Dispute Resolution; E-Health; E-Learning.

The multilingual Digital Single Market has a large set of very specific needs in terms of technologies for overcoming language borders. The solutions we propose (see above) are a first suggestion and need to be discussed further with the community and also with the European Commission. The outcome of this discussion, i.e. the final set of technology solutions, will have specific needs and requirements as pull effects with regard to Layers 2 and 3. On Layer 2 we need to foresee the corresponding services for the Solutions layer while Layer 3 needs to provide tangible research results of basic and applied research that, in turn, feed into and enable Layer 2.

2.2 Layer 2: Language Technology Services, Platforms, Infrastructures

The **Language Technology Services Platforms, Infrastructures Layer** (Layer 2) should comprise a set of services (in the sense of cloud-services, web-services, RESTful services, application programming interfaces etc.) that drive the innovative technology solutions (Layer 1). These services can be conceptualised, among others, as Software-as-a-Service (SaaS), but also as components that can be integrated into stand-alone software. This layer needs to start with a small and robust set of clearly defined,

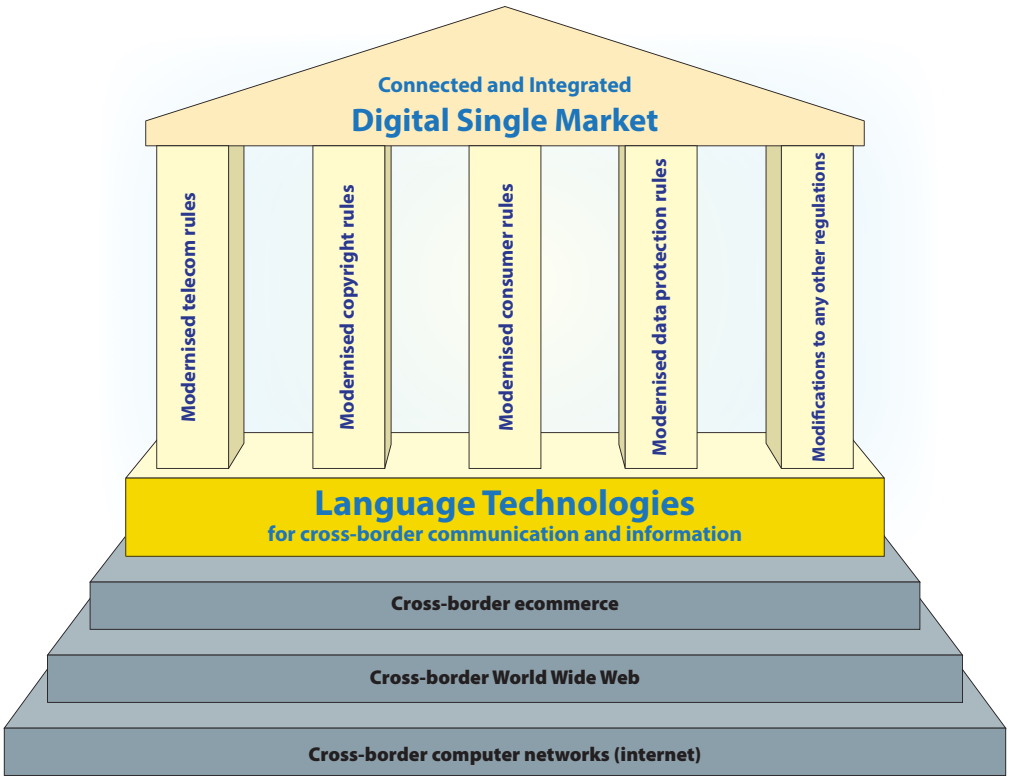


Figure 3: *Language technologies as one of the conceptual foundations of the DSM*

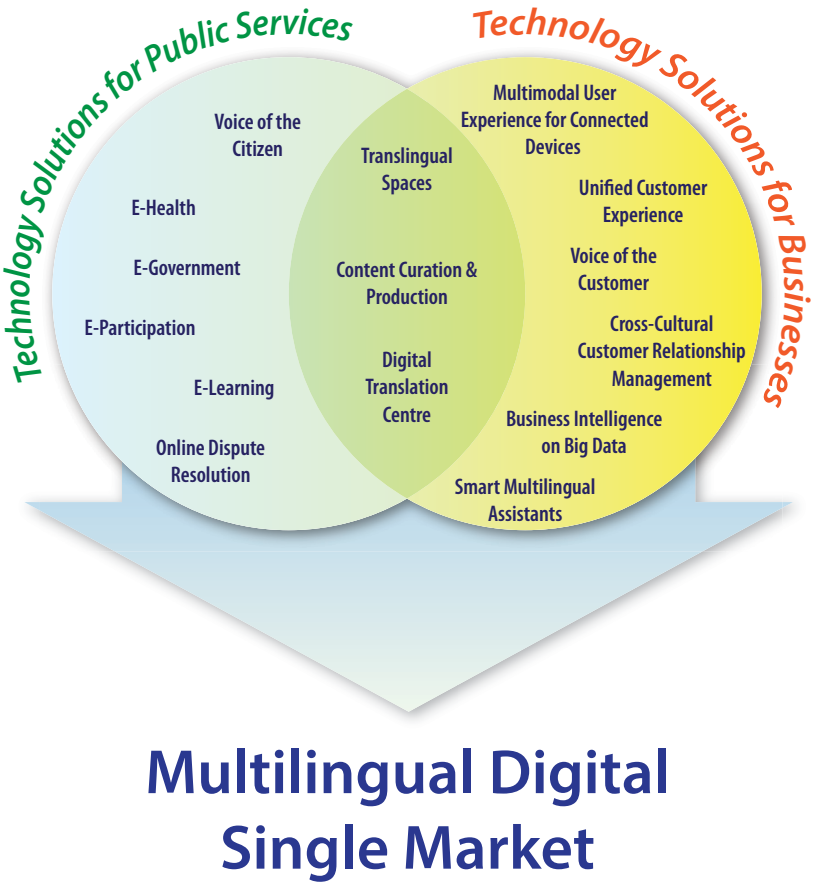


Figure 4: *Technology Solutions for Businesses and Public Services enabling the MDSM*

mission-critical services that are needed by the majority of solutions on the top layer. This initial set of seed services, then, needs to be able to grow organically and scale up into one or more bigger platforms or infrastructures. On this layer it will be important to provide flexibility through a highly innovative ecosystem that enables the emergence of more complex platforms and larger infrastructures (including free services and also for-a-fee commercial services as products).

We can already foresee the need for the emergence of several services, infrastructures, and platforms on this layer. In the following we provide a few examples to illustrate the functionality of this layer: for example, services will be needed that provide flexible multilingual technologies – even including *human* translation. These services would need to be designed from the outset with special emphasis on high-quality output, trust, data security, reliability, privacy, data protection, and confidentiality. We probably also need a bridge to the world of knowledge, data, and meaning through corresponding services. This bridge could provide seamless and ubiquitous access to a multilingual knowledge base that integrates information about products, companies, places, terms, words, and a plethora of other concepts that are important for all monolingual, crosslingual, and multilingual language technology components and data value chains. Designing and implementing such a knowledge service would be a challenge but it can become a reality through the combination of existing repositories such as Wikipedia, Wikidata, BabelNet, DBPedia, Linked Open Data sets and many other language and data resources. Important for industry and e-government will be sector-specific multilingual knowledge systems, which are an important prerequisite for serving a global customer base. Another candidate for one of the sets of generic services is concerned with sophisticated methods for text analytics and production, such as, for example, report generation, text classification, sentiment analysis, and opinion mining. To conclude the list of examples, we could also foresee a platform for natural language interaction methods to assemble APIs for the analysis and synthesis of spoken language such as automatic speech recognition and text-to-speech. All these services would need to be linked up with one another. The speech platform, for example, would need to contain bridges to the translation services. All services would need to have a 24/7 availability and provide web-scale performance and connectivity to carry out their main purpose: support and enable the technology solutions on the top layer and also to support research and innovation by testing and showcasing results as well as providing an environment for hybrid research, i.e. integrating research and operational services. Eventually, the infrastructures will allow providers from industry and research to offer services, resources, and component technologies.

2.3 Layer 3: Priority Research Themes

The **Research Layer** (Layer 3) is meant to support the DSM by enabling the services on Layer 2. The activities subsume both basic and applied research. As already described above, the concrete composition of Layers 2 and 3 largely depend on the set of technology solutions (Layer 1) because the specific solutions selected will formulate needs and demands for the other two layers. In the following we present – as illustrations – a few examples of broad research themes that may be able to drive research and innovation for the multilingual DSM. Independent of the concrete set of themes, it is important to note that all themes need to be tightly intertwined, making use of one another in different application scenarios, especially so when research results, i.e. technologies, are combined on the services and the solutions layer.

Multilingual technologies must be at the core of our technology solutions for the multilingual DSM. This means that one of the research themes may need to tackle high-quality machine translation, including human translation. It would need to provide research results, algorithms, approaches, services, and scientific output to be directly used in Layer 2, i.e. generic and specialised services for reliable spoken and written translation among all European and major non-European languages. A second theme could

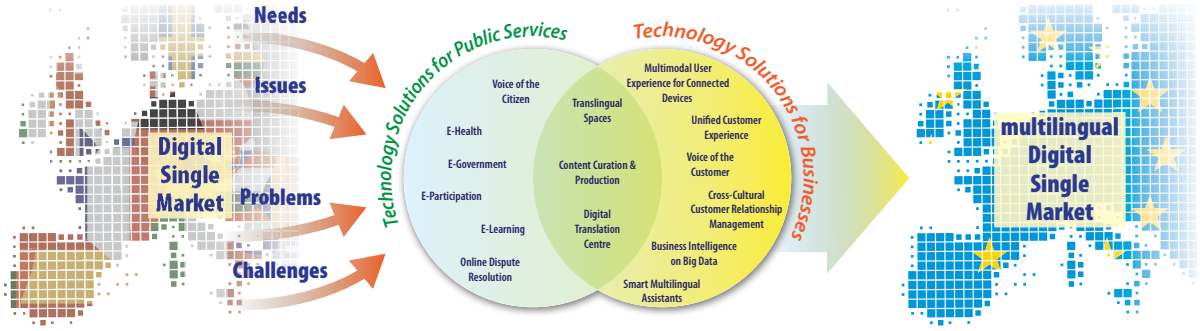


Figure 5: *Making the Digital Single Market multilingual through technology solutions*

handle crosslingual and multilingual big data text and speech analytics research, to provide novel solutions for understanding and dialogue within and across communities of citizens, customers, clients, and consumers. This theme would include, among others, research scenarios for multilingual sentiment analysis, opinion mining, fact mining, rumour and trend detection, and information and relation extraction, as well as components that construct semantics for linguistic analyses – taking into account the multitude of established and emerging online text types and genres. A third theme could concentrate on aspects such as conversational technologies, dialogue systems, and natural language interfaces so as to intensify research on speech interfaces and interactive assistants covering all European languages. Especially with regard to the Internet of Things, and trends such as Wearables and Advanced Manufacturing (Industry 4.0), where a very high demand for spoken natural language interfaces can already now be predicted for the near future. Such spoken language interfaces must be available in all European languages. Furthermore, they could also include socially-aware interactive and pervasive assistants that learn and adapt and that provide proactive and interactive support tailored to the respective user's context. Yet another theme could tackle the increasingly important topic of semantics, knowledge, data, and meaning by providing an umbrella for aligning and harmonising all research activities around monolingual, crosslingual, and multilingual resources, data sets, repositories, and knowledge bases that are needed as background knowledge for all advanced language processing components – from machine translation to text analytics to speech interfaces. This theme could take into account more general repositories such as Linked Open Data sets, Wikidata and Wikipedia, multiple different ontologies, OpenStreetMap, and DBPedia, but also more research-oriented resources such as Yago, WordNet, and BabelNet. All existing and emerging resources would need to be consolidated, rendered interoperable, aligned, and enriched with multilingual information. Additionally, research needs to work on novel approaches for extracting information and knowledge from unstructured text documents and feeding it back into the general knowledge repository. We also need tools for cleaning up data, as well as mechanisms that can aggregate, summarise, and repurpose content. For all applications that interact with data, the regulation of intellectual property rights is an issue that needs to be resolved as soon as possible. The web is a global space, and Europe has to find a legal approach that supports both local research, development, and innovation while fostering global competitiveness. The key recognition that meaning derives from knowledge also supports a recognition that knowledge is contextual, and users must be taken into account in a way that preserves privacy, retains user control, and affords transparent protection of user data. An especially important building block of the three-layer setup is concerned with providing core technologies and resources for Europe's languages. We propose to build a system of shared, collectively maintained, interoperable tools and resources that will ensure that our languages will be sufficiently supported and represented in future generations of IT solutions. This system of shared tools and resources is a crucial prerequisite for the multilingual DSM because it connects the strategic programme to the different languages. Many of these core technologies and resources need to be made available as services.

2.4 Related Areas, Applications, and Societal Challenges

The technology solutions, services, infrastructures, and tangible outcomes of the foreseen research areas will not only have an impact on the multilingual Digital Single Market. Several closely related areas and applications as well as societal challenges that will profit from them as well.

Some of the closely related areas (**Figure 6**) have already been mentioned. Most evident is the complementary connection to the BDV cPPP in terms of technologies for multilingual Linguistic Big Data analytics. There is also a close relationship between interactive and multilingual spoken language interfaces and robots (especially the SPARC Robotics PPP), connected machines (Advanced Manufacturing, Industry 4.0) as well as generic connected devices (Internet of Things, Web of Things). The relationship between multilingual technologies and ecommerce applications is so evident and of such vital importance that we also listed this area, as well as the emerging trend to Smart Cities and Smart Services.

The importance of the languages in our European society has never been in the focus of attention as compared to other highly multilingual societies like South Africa or India where language borders hinder exchange and communication *within* a state. According to the principles of the UN-endorsed World Summit on the Information Society, the "Information Society should be founded on and stimulate respect for cultural identity and cultural and linguistic diversity." Recent scientific work has shown that even our moral decisions are influenced by whether we are speaking our mother tongue or a foreign language.[18]

In fact, the technology solutions detailed in the next chapter address many of the societal challenges specifically to be taken into account by activities under the framework of Horizon 2020.[19] The following list provides several examples:

- *Health, demographic change, and wellbeing* (can be addressed by *Adaptable interfaces for all*, *E-Health*, and *E-Learning* solutions);
- *Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bioeconomy* (can be addressed by the *Digital Translation Centre* solution);
- *Secure, clean, and efficient energy* (can be addressed by the *E-Participation* solution);
- *Smart, green, and integrated transport* (can be addressed by the *Adaptable interfaces for all* solution);
- *Climate action, environment, resource efficiency, and raw materials* (can be addressed by *Digital Translation Centre* solution);
- *Europe in a changing world – inclusive, innovative, and reflective societies* (can be addressed by *Adaptable interfaces for all*, *E-Learning*, *E-Participation* solutions);
- *Secure societies – protecting freedom and security of Europe and its citizens* (can be addressed by *Adaptable interfaces for all* solution).

2.5 Summary

We recommend setting up a large and ambitious strategic programme to enable the multilingual Digital Single Market. The suggested approach consists of three different layers (**Figure 7**): on the top layer we have a set of focused **Technology Solutions for Businesses and Public Services**. These innovative application scenarios and solutions are, in turn, supported, enabled, and driven by the middle layer.

---

[18] A. Costa , A. Foucart, S. Hayakawa, M. Aparici, J. Apesteguia, J. Heafner, B. Keysar (2014): "Your Morals Depend on Language", PLOS One, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094842

[19] European Commission (2014): Horizon 2020, The EU Framework Programme for Research and Innovation, Societal Challenges, http://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges

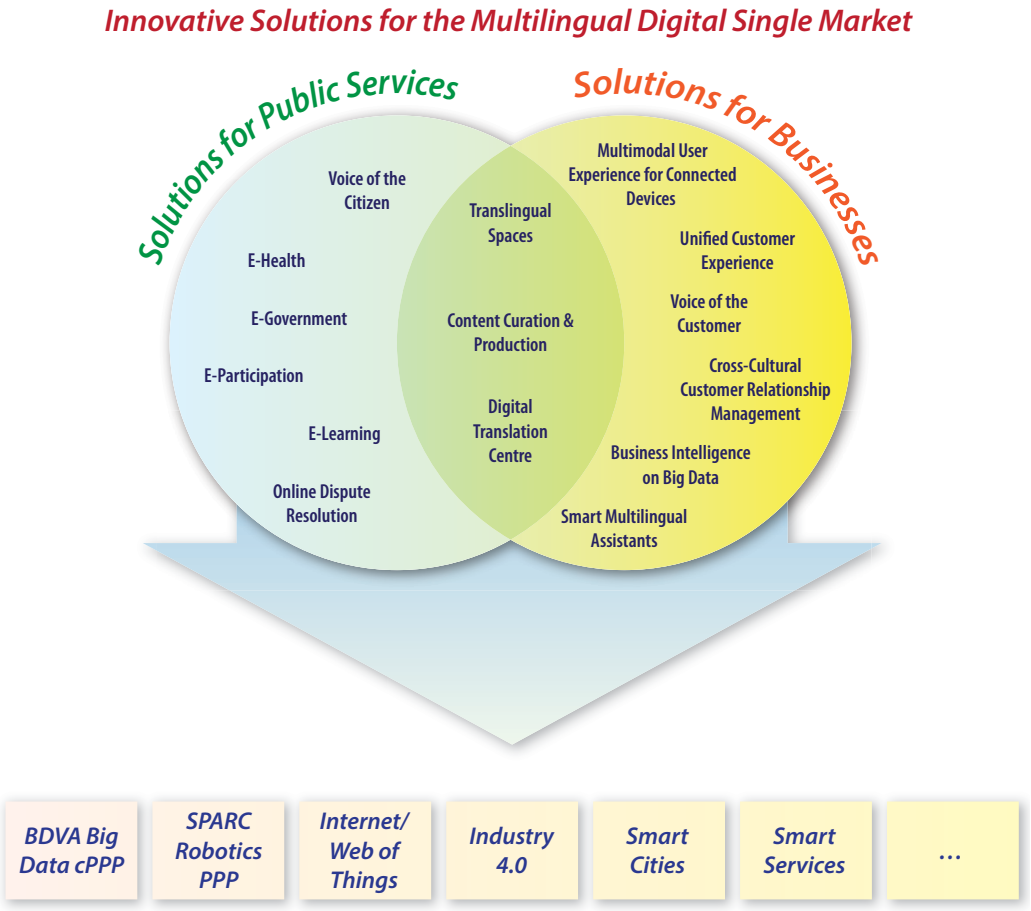


Figure 6: *Closely related areas and applications*

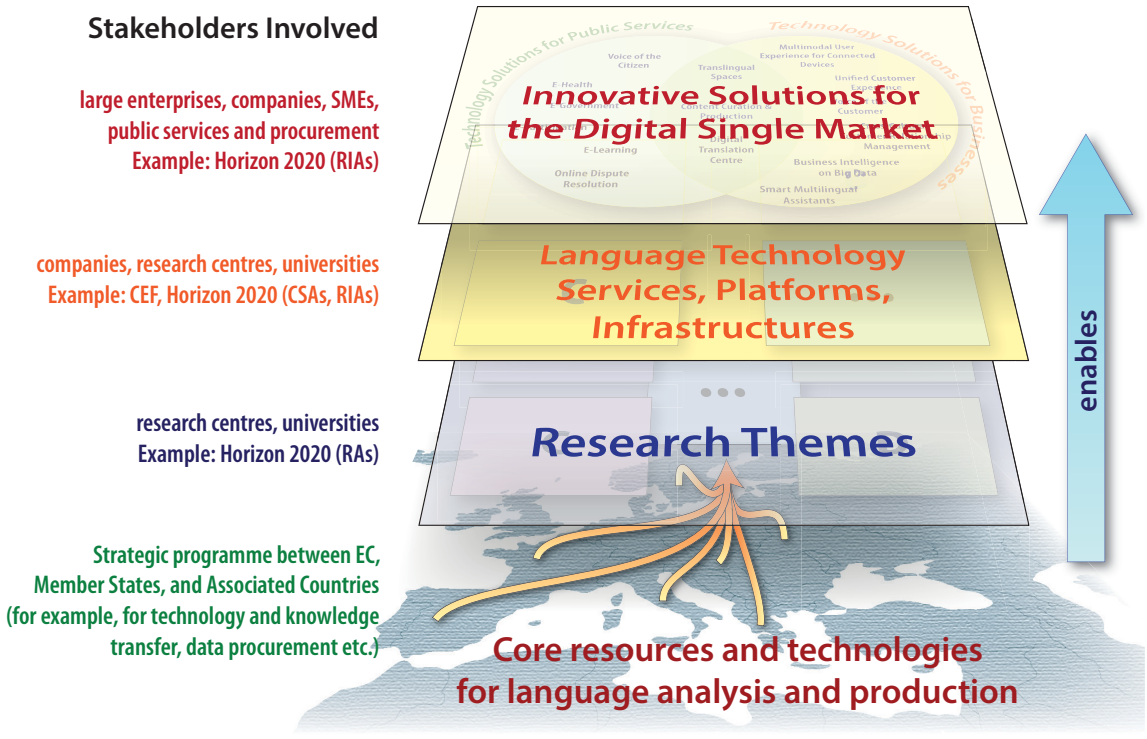


Figure 7: *The strategic programme for the multilingual DSM*

This consists of **Language Technology Services, Platforms, and Infrastructures** that provide, through standardised interfaces, data exchange formats and component technologies, different services for the translation, analysis, production, generation, enrichment, and synthesis of written and spoken language. The bottom layer connects the infrastructures to innovative **Priority Research Themes**. These research themes provide concrete scientific results, approaches, technologies, resources, modules, components, algorithms etc. that can then be used to enable the second and, ultimately, the top layer. Additionally we recommend to significantly intensify work on **core resources and technologies for language production and analysis**. Further to the solutions, services, and research results, we need to improve the basic technologies for the specific languages to be supported through the suggested programme: in order to equip every language with a set of core resources and technologies, we suggest, among others, intensifying knowledge and technology transfer between larger research centres and groups working on technologies for those languages that are in danger of digital extinction.

This three-layer approach will provide European research, development, and innovation in the field of innovative language technologies and also multiple different industries with the ability to compete with other markets and subsequently achieve multiple benefits for European society and citizens as well as an array of opportunities for our economy and future growth.

An integral component of our strategic plans are the member states and associated countries: it is of utmost importance to set up, under the umbrella of the strategic programme proposed in this document and its three-layer approach, **a coordinated initiative both on the national (member states, associated countries, regions) and international level (EC/EU), including research centres as well as small, medium, and large enterprises who work on or with language technologies**. One instrument for such a coordinated initiative could be setting up a contractual Public-Private-Partnership (PPP). A European Flagship Project would also be a good candidate instrument, especially with regard to significantly boosting the development of innovative and novel LT approaches, algorithms, and paradigms as well as fighting digital language extinction.

Only through close cooperation between all stakeholders, tightly coordinated collaboration, and an agreement as well as update of our national and international language policy frameworks will we be able to achieve the ambitious plan of researching, designing, developing, and rolling out platforms, services, and solutions that support all businesses, public services, and citizens of Europe, and beyond, and that also enable and fully realise the multilingual Digital Single Market.

What is missing in Europe is the political awareness, determination, and will to give us a leading position in this technology area through a concerted funding effort. This major dedicated push needs to include the political determination to modify and to adopt a shared, EU-wide language policy that foresees an important role for language technologies.

As Europeans, we urgently need to ask ourselves some crucial questions: Can Europe afford continued language-blocking, market fragmentation caused by language borders, language discrimination, and, eventually, digital language extinction? Can we afford to have our information, communication, and knowledge infrastructure depend so much on monopolistic services provided by foreign, non-European companies, effectively constituting technological lock-in? What is Europe's fallback plan in case the language-related services provided by non-European companies that we rely upon are suddenly switched off or if even more serious access or security issues arise? Is Europe actively making an effort to compete in the global landscape for research and development in language technology? Can we expect third parties from other continents to solve our translation and knowledge management problems in a way that suits our specific communicative, societal, and cultural needs?

Language Technology made for Europe in Europe will significantly contribute to future European cross-border and cross-language communication, economic growth, and social stability while establishing for Europe a worldwide, leading position in technology innovation, securing Europe's future as a worldwide trader and exporter of goods, services, and information. Only a large, coordinated push of this magnitude will be able to unlock a truly multilingual Digital Single Market.

3 Layer 1: Innovative Technology Solutions for the Multilingual Digital Single Market

This chapter describes technology solutions for businesses (Section 3.1) and public services (Section 3.2) that will eventually enable a truly integrated multilingual Digital Single Market. The technology solutions described in this chapter – Layer 1 of the three-layer-setup – are enabled by platforms, infrastructures, and services (Layer 2, Chapter 4) that are, in turn, enabled by priority research themes (Chapter 5) – Layer 3 of the three-layer-setup.

3.1 Technology Solutions for Businesses

All technology solutions for businesses described below contribute to enabling the multilingual Digital Single Market. The "multilinguality component" is an inherent part of all solutions, which is why this specific keyword is not mentioned explicitly in all solution titles. The solutions relate to aspects such as, for example, a unified – i.e. crosslingual, language-crossing – customer experience in ecommerce scenarios, multilingual market research technologies, business intelligence, content curation, and content production. The solutions are meant to be developed, first and foremost, in collaborative European projects by commercial solution providers in tandem with European R&D&I, based on the platforms, infrastructures, and services provided by the second level of the three-level-architecture (see Chapter 2). The primary customers of these solution providers are, again, commercial companies (primarily SMEs) but also public administrations, European institutions, and other organisations that want to participate in the multilingual DSM.

3.1.1 Unified Customer Experience and Cross-Cultural CRM (E-Commerce)

- This technology solution includes applications for crosslingual ecommerce: automatic translation of online shops and other websites (including semi-automatic localization and internationalisation) to enable SMEs to offer their services in more languages, penetrating the whole multilingual DSM
- This technology solution provides contextual, up-to-date and relevant additional information about products or services to users, bringing together content, customer care, customer relationship (CRM), discussion fora, helpdesks etc. in a unified digital (eco)system *across languages*

In today's hyper-connected society, consumers expect to quickly and easily get what they need from a business – anytime, anywhere. This includes access to products and services, but also to information and easy-to-use, powerful self-services. Today, industries interact with their customers on a daily basis. They have to recognise customer needs and intentions in real-time and guarantee the consistency of provided information across channels, audiences, *and languages*.

Automation helps bring together content, product, and customer relationship management in one ecosystem. The goal is a seamless network of content, data, and knowledge that spans multiple modalities and channels (mobile, web, interactive voice response etc.) and incorporates open and closed datasets in a way that is respectful of intellectual property (IP), data privacy, and corresponding licenses. Realising a certain degree of agility at the content level will enable the quick integration of new (external) data resources and will allow marketing experts to dynamically react to changing customer and market needs.

Linked Data technologies can help create a unified information space by bringing together data from different sources, including product data, customer data, and social data. The generation of rich linked knowledge resources enables multimodal and multilingual repurposing of heterogeneous content for different challenges, natural languages, and audiences. Linking resources can enable the visual story generation from multiple sources including text, video, and other modalities, or the creation of semantic user profiles based

on linked information about objects, individuals, groups, intentions, contexts, and cultures. The creation of these resources should be based on standardised ways for representing and linking such information.

In cross-cultural customer relationship management, integrating translation technologies (also supporting niche languages used in micro-domains) into the customer engagement ecosystem will enable companies to efficiently engage with their customers across languages. This will not only allow micro-SMEs in ecommerce to exploit multilingual value chains, making them competitive in market niches, but also help create an extraordinary, contextualised digital experience to all users.

3.1.2 Digital Translation Centre

- This solution relates to providing customisable machine translation services including written and spoken language as well as solutions for specialised micro-domains
- Broad set of target users: businesses, governments, administrations, customers, citizens
- Broad set of use cases: from desktop to mobile to tablet to voice to automatic (via API)

Translation services are moving to cloud-based solutions – generic and specialised federated services for instantaneous reliable spoken and written translation among all European and major non-European languages. Clouds will make it possible to offer different service layers such as a public and an internal service layer for providers with different offerings. This can include a free 24/7 public service of basic automatic services (text translation, term and word translation), professional services available for a fee (including high-quality professional services by human translators, terminology, dictionaries, checking, TMs) and free human translation or post-editing services for special purposes provided by NGO-initiatives such as, e.g. Translators without Borders or the Rosetta Foundation. This set of solutions foresees one or more common, easy-to-use access points for citizens, professionals, businesses, and public organisations providing ubiquitous and instant access to information and communication in any language.

When they travel across borders, products and services are typically tailored to foreign communities and accompanied by documentation covering instructions, insurance, privacy protection, validation forms, after-sales information and more. All this content needs to be adapted to the languages, cultures, measurement systems, safety regulations, and work habits of new customers and end users. Systems need to be engineered to automatically control this process, cut lead times, radically reduce transaction costs, and improve information quality. Such a technological solution will be critical to accelerating the emergence of the multilingual Digital Single Market and other trading platforms.

3.1.3 Content Curation and Content Production

- These technology solutions support intelligent authoring and enrichment of content, making it suitable for linking to other content objects, and making it readable and understandable across language barriers and for machines and humans alike
- These solutions also support the multilingual, automatic or semi-automatic generation of articles and reports based on big data and other sources of structured or unstructured information such as Linked Data

Collecting, organizing, structuring, and displaying information relevant to a particular topic or area of interest is a major task in many areas, including journalism, marketing, and decision-making. Accelerating the process of discovering relevant content is especially crucial for those whose work involves processing large amounts of information in a short time. Technologies for digital content curation reduce

the overall flow of information and make them more targeted to the end user's interests, for example for selecting information that is appropriate for corporate blogs or websites or content for brands to post to social media channels. Language Technology solutions can play a crucial role in this process. Machine translation technology can help handle multilingualism of data sources and facilitate access to multilingual data assets. Semantic technologies are crucial for enabling the semantic interoperability of data sources and help extract and combine content from multiple data sources and across all communication channels (telecommunication, meetings, email, chat etc.).

Language technology can also be of help in various content production tasks. Standardised communication, for example email communication in customer support, can be automated by analysing user feedback at the meaning level and identifying relevant, semantically similar previous communications. Robot journalism can comb structured data for facts and trends and combine them with contextual information to form and string together sentences, enabling the computer-assisted or fully automatic generation of multilingual articles, reports or product websites, also taking into account other data sources such as, for example, website access analytics. Advanced algorithms can adapt perspective, tone, and humour to tailor a story to its audience. In human text generation, authoring support software can flag potential errors, suggest corrections, and use authoring memories proactively to suggest completions of started sentences or even whole paragraphs. Advanced technologies can check for appropriate style according to genre and purpose and help improve comprehensibility.

3.1.4 Virtual and Real Translingual Spaces

- This foreseen solution relates to interactive meeting rooms that provide services like seamless spoken translation, automatic note taking etc.

As businesses and other organisations attempt to cut down on high-carbon travel agendas, they are turning to virtual meetings online as a cost-effective solution for collaborative events. Tremendous value can be added to such meetings by providing automated interpretation solutions for spoken communication or by automatically transcribing and eventually summarising the content of meetings. Incrementally drafted (searchable) summaries can be used for displaying the state of the discussion, including intermediate results and open issues, and to generate meeting minutes. Brainstorming can be facilitated by semantic lookup and structured display of relevant data, proposals, charts, pictures, and maps from databases, enabling participants to make more relevant contributions. Individual realtime translation can simultaneously interpret the contributions of participants, slides, and handwritten text (e.g. on a shared whiteboard) into as many languages as needed. These 'learning' services can streamline the entire meeting process, save time, produce an automatic record, and improve teamwork.

Apart from meeting management, these technologies have the potential to empower and augment human-human communication in general. We envisage applications that help people to be more creative more often (especially in groups), new approaches to social sharing (across languages), design-enabling platforms which enable people to build their own tools, and systems to enable groups (at all scales) to collaborate with shared goals. These applications will facilitate problem solving and provide powerful mechanisms for engagement. Example use cases include social sharing and collaboration platforms; enhanced meetings with text and speech translation; games (entertainment and serious games); shared (task) understanding and communication, for example in multi-discipline professional teams (e.g. medicine); shared learning, and MOOCs.

The underlying technologies will be guided by partial understanding of the contents, i.e. by its semantic association with concepts in semantic models of domains and processes. Reliable dialogue translation

for face-to-face conversation and telecommunication will require high-quality and high-speed machine translation for many languages, across multiple subject fields and text types, both spoken and written.

3.1.5 Voice of the Customer

- This technology solution enables multilingual market research. This means extracting and interpreting the multilingual "voice of the customer" with a high level of accuracy, across languages and modalities, and analysing sentiment as well as opinion at deeper levels beyond mere polarity, including intention recognition.

The recognition of user needs, intentions, and opinions towards products and services is crucial for the success of today's companies. Recognising customer needs and opinions involves the extraction and interpretation of customer interactions with a high level of accuracy and across languages and modalities. The main source of information is user-generated content from social media. Customers and potential customers share their thoughts using blogs (e.g. Twitter), post comments in online forums, or send feedback via email. All these text and voice messages are a valuable source for trending sentiments and opinions about products and services. The "Voice of the Customer" technology solution includes the (targeted) analysis of large volumes of such comments and communications created by citizens, customers, patients, employees, consumers, and other stakeholder communities. Summarising multiple multilingual data streams in real time involves dealing with high-volume, high-velocity data, often of unknown veracity.

Social media analytics builds on improved text analytics methodologies but goes far beyond the analysis. Part of the analysis is directed to the status, opinions, and acceptance associated with the individual information units. As the formation of collective opinions and attitudes is highly dynamic, new developments need to be detected and trends analysed. Emotions play an important part in individual actions such as voting, buying, supporting, and donating, and in collective opinion formation; accordingly the analysis of sentiment at deeper levels is a crucial component of social intelligence. Text analytics will play a crucial role in areas such as analysing the voice and actions of the customer in the context of CRM; brand, product, and reputation management; technology monitoring and competitive intelligence; content management and publishing; search, information access, and question answering.

Automatic summarisation and translation technologies will help monitor, analyse, summarise, structure, document, and visualise social media dynamics and enable multilingual and cross-lingual market research. Technologies such as sentiment analysis, opinion mining, and intention recognition will extract and interpret the voice of the customer with a high level of accuracy and across natural languages and modalities, while providing insights how culture and behaviour affect any conclusion. This technology solution also includes the application of the abovementioned methods to spoken language data, collected, for example, in (automated) call centres.

3.1.6 Business Intelligence using Big Data

- This technology solution allows to analyse large volumes of data, generate clear summaries, detect trends, answer questions, and search concepts (instead of words) etc.

The quality, speed, and acceptance of individual and collective decisions is the single main factor for the success of enterprises (and, likewise, public services, communities, states, and supranational organisations). The growing quantity and complexity of accessible relevant information poses a serious challenge to the efficiency and quality of decision processes. IT provides a wide range of instruments for intelli-

gence applications. Business intelligence, military intelligence or security intelligence applications collect and pre-process decision-relevant information. Analytics applications search data for information and decision support systems evaluate and sort the information and apply problem-specific decision rules.

Language makes (very) big data in the web, intranets, and various databases, user fora, among others. However, although much of the most relevant information is contained in texts, text analytics applications today only account for less than 1% of the more than US$ 10 billion business intelligence and analytics market. Because of their limited capabilities in interpreting texts (mainly business news, reports, and press releases) their findings are still neither comprehensive nor reliable enough. While the main tasks required in text analytics are rather conventional (topic extraction, document classification, entity extraction, relation extraction, event extraction), there is a clear need for analytics solutions that are tuned to the needs of particular domains and are able to generate and incorporate semantic domain-specific knowledge in the form of taxonomies or terminologies to support domain customisation. Summarisation technologies as well as multilingual report generation will help keep up with linguistic big data to be processed. In the other direction, semantic search, question-answering and trend detection services will allow business analysts, decision makers, venture capitalists, and other experts to access complex information in a targeted way. Adaptive techniques are needed to ensure that the responses to the user are relevant: choosing the correct information to provide for that context, and automatically phrasing it in a relevant way that supports their need for further explanation, illustration or clarification.

One of the technical priority themes identified in the European Big Data Value Strategic Research and Innovation Agenda is Deep Analytics. Several of the major expected advanced analytics innovations, such as semantic analysis and multimedia (unstructured) data mining, strongly relate to natural language data. Language technology will be the foundation for organising, analysing, and extracting big data in a truly useful way. To this end, we suggest collaborating closely collaboration with the BDV cPPP to provide this set of technologies, which is complementary to the work foreseen in the BDV cPPP.

3.1.7 Multimodal User Experience for Connected Devices

- This set of technology solutions provides robust spoken-language interfaces and dialogue systems for connected devices, ranging from smart phones to cars, consumer products (televisions, radios) and household appliances
- In addition to spoken language-based interfaces, gesture- and text-based multimodal interfaces are foreseen as well
- These solutions provide an important bridge to the Internet of Things (IoT), Web of Things (WoT) and Industry 4.0 (Advanced Manufacturing) area for which spoken language interfaces will become the norm in the near future

Many everyday objects are already connected to the internet and may even be interconnected with other objects (Internet of Things, Web of Things). Depending on the function, complexity, relevance, and autonomy of these objects, the nature of desired or needed communication can vary widely. Some objects will come with interesting textual information that we would like to query and explore, such as manuals and consumer information. Others will provide information on their state and will have their own individual digital memory. Objects that can perform actions, such as vehicles and appliances, will accept and carry out (multilingual) voice, gesture or eye-tracking commands. Wearable sensors can provide signals about a person's mood or emotional state, offering new affect-focused multimodal interaction with devices. Objects will therefore offer more engaging interaction experience with the person through

the combination and optimisation of the multiple modalities available to present digital content and sensed human engagement.

Robots are currently evolving into collaborative, social machines that will eventually provide useful services to humans in numerous work, medical, educational, and household contexts. Specialised mobile robots will be deployed for personal services, rescue missions, household chores, and tasks of guarding and surveillance. The effort to design social awareness and the capacity to learn into a robot's computer is delivering returns, as humans and their robots can now team up and share tasks more intelligently. Physical environments such as offices, homes, hospitals, and other sites will, like robots, be able to learn about humans and discover unforeseen needs.

Multimodal conversational interfaces have the potential to adapt automatically to the user and to the environment. For instance, interfaces adapted to elderly people will take into account cognitive, auditory, visual, and articulatory ageing; interfaces will adapt to what a user is doing (working in a noisy, hands-free environment, e.g. when rushing for a train); systems for devices where "traditional" interfaces (keyboard, mouse, trackpad, touchscreen, etc.) are not usable (e.g. small wearable devices) or not appropriate ("companion" systems); smart mobile agents which are capable of deeper natural language and multimodal interaction, possibly focused on specific domains, and capable of rich question answering; interfaces will help to rehabilitate users with some kind of disability. Enriched personal assistants would be able to make effective use of environmental, informational, and social context. In particular, this scenario focuses on systems which display social awareness, are able to behave naturally in multiuser situations, and are capable of proactive behaviour. They will also be able to integrate heterogeneous sources of knowledge. Many vertical market sectors with domain-specific assistants exist: shopping, travel, social service planning, learning, and tutoring.

With the evolution of connected devices and robots, enabling communication via natural language commands and dialogues will be the major challenge. Building on top of existing technologies for natural language interaction, including dialogue management and speech technologies, will help create devices that can communicate with us in human language in a user-friendly way. Their wide acceptance will improve productivity, safety, and comfort.

3.1.8 Smart Multilingual Assistants

- This solution relates to smart multilingual assistants embedded in clothing, accessories and other wearables, mobile phones, tablets, connected devices, Internet of Things etc.

With the growing number of assisting software in phones, tablets, and other connected devices, there is a rising need for socially aware, highly personalised assistants that learn and adapt and that provide proactive and interactive support tailored to specific situations, locations, and goals of the user. Voice, gender, language, and mentality of the virtual character need to be adjusted to the user's preferences. The personality and functionality of the interface may also depend on the user type: there may be special interfaces for children, foreigners, and persons with disabilities. Having been trained on the user's behaviour, digital information, and communication space, the assistants can proactively offer valuable unrequested advice.

There will be a competitive landscape of intelligent interfaces to the offered services employing human language and other modes, such as gestures, for communication. Natural language is by far the best medium for interacting with virtual assistants. Integrating multilingual speech technologies will enable assistants to speak in the language and dialect of the user, but also digest information in other languages

and formats, and may even translate or interpret without the user having to request it. Depending on the needed functions and available information, language coverage will range from simple commands to sophisticated dialogues.

The next generation of smart assistants will have to achieve a high out-of-the-box accuracy, but will also have to be able to deal with composite meta-tasks with dynamic collaborative interactions, which will require deep natural language understanding and the ability to reason on knowledge. To achieve such smart assistants and conversational agents it is necessary to better model, synthesise, and understand social speech signals, e.g. laughter, backchannels, pause insertion and duration, intonation, turn taking, etc.

In the future, many providers of information about products, services, or touristic sites will try to present their information with a specific look and feel. New levels of audio and visual resource management and synchronisation will be needed to handle the variety of body and voice features related to the personality and affect (e.g. emotions, laughter) to offer more human-like assistant technology and to open up new horizons in the generation of creative content (e.g. computer games, movies, music, internet). However, challenges remain for speech recognition to deal with noisy environments, multiple speakers, localising the speaker, as well as understanding conversations and non-verbal signals.

Multilingual assistants are closely related to the Internet of Things. Sensors and power-efficient signal processing are critical for real-world usability. This includes intelligent wake-up functions ('always listening'), secure biometrics, convenience, context for accurate interpretation, multi-microphone beamsteering, audio-visual recognition, touch and gestures, location, time, movement, vitals (healthcare applications) etc.

3.2 Technology Solutions for Public Services

3.2.1 Voice of the Citizen – Social Intelligence on Big Data

- This solution includes large-scale, web-scale sentiment analysis, opinion mining, multilingual report generation, trend analysis.
- This solution is, conceptually, a complement to the "voice of the customer" and uses the same technologies.
- Democracy will be enriched by powerful new mechanisms for developing improved collective solutions and decisions (also see E-Participation).

Following the Fukushima incident in 2011 there have been discussions about the dangers of nuclear energy in all European countries. These debates were held in the respective language communities only, there has never been a public European debate about the topic because it is, technically, not yet possible to organize such a debate online. The "Voice of the Citizen" solution is intended to help pave the way to full e-participation by providing technologies for multilingual social media mining. The idea is to analyse social media networks and user generated content in multiple European languages in order to gather concrete numbers and statistics about what Europeans in specific countries or regions think about urgent or important topics such as e-mobility, nuclear energy, climate change etc. Such information can be used to inform European decision support, to increase social reach and also to improve cross-cultural understanding. The goal is to create a "citizen experience" – as a complement to the unified "customer experience" or "user experience" for commercial products, services or offerings.

3.2.2 Online Dispute Resolution

- This solution relates to an online tool where merchants/service providers and consumers can settle their disputes outside courts, across borders, in situations where they do not have a common language.

The solution foresees a multilingual platform to support an interactive and free-of-charge website for Online Dispute Resolution (ODR). ODR aims at resolving contractual disputes from European consumers (B2C) or traders (B2B), which arise from cross-border and domestic online sales or service contracts. Competing for alternative dispute resolution (ADR) models requires not only managing the translation of messages, conversations/mediations flowing among parties: evaluating, sending, and receiving information (especially at cross-border disputes), but also translating documents needed for finding a resolution to the dispute and other needed functionalities of the platform (guidance, easy to fill forms, etc.). We suggest, thus, that ODR uses machine translation to provide a multilingual platform. The EC has a legal obligation (Regulation on consumer ODR and Directive on consumer ADR) to implement this platform in all official languages of the institutions of the EU. The ODR platform presents a unique opportunity. Main multilingual challenges of the platform comprise managing 500 language pairs, a glossary database, spell check, translation of free text fields and different types of languages (formal and everyday languages). The ODR platform not only poses significant challenges for LT, but a high-quality multilingual tool could underpin the uptake and credibility of LT among customers and users. The ODR system aims at boosting online purchases from consumers and traders (especially at cross-border level); visibility of LT could exponentially increase as the ODR platform will be accessible to millions of consumers and thousands of traders using ecommerce in Europe.

3.2.3 E-Participation

Today, collective discussion processes involving large numbers of participants are bound to become non-transparent and incomprehensible, especially as they span the myriad linguistic and cultural boundaries that characterise Europe. Since many discussions will involve participants in several countries, e.g. EU member states, cross-lingual participation needs to be supported. By recording, grouping, aggregating and, counting opinion statements, pros and cons, supporting evidence, sentiments, and new questions and issues, the discussion can be summarised and focused across boundaries to aid engagement across the EU electorate. Decision processes can be structured, monitored, documented, and visualised, so that joining, following, and benefitting from them becomes much easier. The efficiency and impact of such processes can thus be greatly enhanced. Special support will also be provided for participants not mastering certain group-specific or expert jargons and for participants with disabilities affecting their comprehension.

However, meaningful EU-wide citizen engagement in EU issues is not however just restricted by language barriers. Cultural outlook and national viewpoints still dominate public discourse. Promoting deeper cultural and historical understanding across Europe is key to building a common history and identity, especially in confronting the influence of previous strife between national, regional, and ethnic groupings. The EU has successfully invested in promoting public online access to cultural and historical resources through initiatives such as CLARIN and DARIAH, however these resources are still largely contained in linguistic silos. Language technologies such as machine translation and cross-lingual search can assist, but solutions require extremely careful translation of person, place, and event names, e.g. 'Danzig' versus 'Gdansk', or ambiguous word such as the German word 'Gewalt' (meaning legitimate power or violence depending on the context). Existing knowledge resources, such as Wikipedia, can help contextualise the significance of words to different national or cultural audiences, but solutions are

needed to ensure the language resources used by language technology accurately capture and maintain metadata on audience sensitivities.

Social intelligence will support understanding and dialogue within and across communities of citizens and consumers to enable e-participation and more effective processes for preparing, selecting, and evaluating collective decisions. The quality, speed, and acceptance of collective decisions is the single main factor for the success of social systems such as communities, public services, states, and supranational organisations. The growing quantity and complexity of accessible relevant information poses a serious challenge to the efficiency and quality of decision processes.

Social intelligence builds on improved text analytics methodologies but goes far beyond the analysis. One central goal is the analysis of large volumes of social media, comments, communications, blogs, forum postings etc. of citizens, customers, patients, employees, consumers, and other stakeholder communities. Part of the analysis is directed to the status, opinions, and acceptance associated with the individual information units (see Section 3.2.1). As the formation of collective opinions and attitudes is highly dynamic, new developments need to be detected and trends analysed. Emotions play an important part in individual actions such as voting, buying, supporting, and donating, and in collective opinion formation; accordingly the analysis of sentiment at deeper levels is a crucial component of social intelligence.

We envision the emergence of public discussion and opinion formation platforms for Europe-wide deliberations on pressing issues such as energy policies, the financial system, migration, natural disasters, which also supports the proactive engagement of less active parts of the population. Addressing politicians, health providers, manufacturers, the cultural sector, and citizens, it will provide visualisations of social intelligence-related data and processes for decision support. Solutions will be based on the detection and prediction of events and trends from content and social media networks and on mining e-participation content for recommendations and summarisation. Building the underlying technologies will require the usage of high-throughput, web-scale content analysis techniques that can process and extract knowledge from multiple different sources, ranging from unstructured to completely structured data, at different levels of granularity and depth by allowing to trade-off depth for efficiency as required. The success of such a platform will largely depend on the integration of crosslingual technologies to increase the social reach and approach cross-cultural understanding.

3.2.4 E-Government

- This set of solutions is meant further to improve the pan-European cross-border exchange of electronic documents, cross-border communication including legal aspects, specialised free translation services – towards a borderless e-government space in Europe.

The creation of a multilingual Digital Single Market should also include vastly improved cross-border public and government services that interoperate and counter market fragmentation, in particular in the areas of e-government, e-health and e-procurement because these areas can also be an inherent part of multilingual data value chains that flow across borders. This set of solutions foresees, among others, e-procurement platforms in which multilingual language technologies can support the translation of user interfaces, documents, and large narratives that are currently performed manually. Language technologies are needed for concept identification and extraction, matching offer and demand to identify business opportunities and to produce accurate summaries for decision making in tendering processes.

Many of the technologies to be developed for the business-oriented solutions (see Section 3.1) can also be used effectively in e-government scenarios, especially sophisticated high-quality machine transla-

tion methods, or generic text analytics technologies. Specific to e-government are the development of terminologies, linked data sets, and ontologies that harmonise the concepts used in different countries and jurisdictions, as a basis to reach interoperability and to develop a new generation of (public) services that is implemented across countries with multilingual technologies built in. We suggest to design and to deploy an ecosystem of data that is partially open and partially closed but is extended with appropriate provenance and licensing information as well as mechanisms for representing and dealing with trust and confidence, so that the public as well as private companies can exploit the data for their purposes and within their applications. Simplifying access to data by appropriate interfaces, e.g. based on natural language, is a crucial goal to achieve. We also need technologies to generate reports and reviews automatically. It has been estimated that, in five years' time, more data will be generated automatically by machines than by humans. Although much of this content will be low-value advertising or journalism, an increasing proportion of it within the enterprise, hospitals, and government departments will consist of highly actionable summary and review information. Language technology processes can take raw data to transform the numbers and words into succinct reports for later use by specialists. This will save time and money and rapidly inform all stakeholders for further discussion.

3.2.5 E-Health

- Cross-border healthcare scenarios will open new ways for creating a single market for health services where health practitioners, patients, and administrators can communicate seamlessly across language barriers.

When considering cross-border health care, it was shown that challenges go beyond the technical level and include different interactions with health professionals, patients, as part of the interoperable system. Interoperable e-health systems need not only different interfaces to manage data, text or speech, but also to cover different challenges in different levels of the data value chain. Tools and methodologies are needed for high quality translation, codes need to be extended to all EU languages. Automatic translation reliability is needed not only for e-health/medical concepts and terms as defined and modelled by terminologies in a given EU language but also to be understood in the medical domain and/or a given health system context.

Psychological and medical conditions affecting language are among the most severe impairments from which people can suffer. Language technology has been utilised for diagnosing the type and extent of brain damage after strokes. Another application area is the diagnosis and therapy of innate or acquired speech impairments, especially in children. Expected progress in language technology, together with advances in miniaturisation and prosthetics, will open new ways for helping people who cannot naturally enjoy the benefits of communication.

3.2.6 E-Learning

- The combination of life-long learning and massive online training courses with multilingual technologies help self-studying, cross-border migration, training for staff members of pan-European companies etc.

Already today, the software market for computer-assisted language learning (CALL) is growing fast. While current products can help traditional language instruction, they are still limited in functionality because the software cannot reliably analyse and critique the language produced by the learner. This is true for written language and even more so for spoken utterances. Software producers are trying to circumvent the problem by closely restricting the expected responses of the user, something that helps for many

exercises, but still rules out the ideal interactive CALL application: an automatic dialogue partner ready around the clock for error-free conversation on many topics. Such software would analyse and critique the learner's errors and adapt its dialogue to the learner's problems and progress. Current language technologies cannot provide such functionality yet. Its lack of flexibility is the reason why research on CALL applications has not yet come into full bloom. As research on language analysis, understanding, and dialogue systems progresses, we predict a boom in the promising and commercially attractive CALL area.

However, use cases for language technologies in this area are much broader than CALL. Machine translation can help in accessing massive open online courses (MOOCs), virtual assistants can help in tests and learning, language technologies are also essential for multilingual gamification. Speech technologies are crucial for children with dyslexia.

4 Layer 2: Language Technology Services, Platforms, Infrastructures

The Language Technology Services, Platforms, Infrastructures Layer should comprise a set of services that drive innovative technology solutions. These services can be conceptualised as Software-as-a-Service (SaaS), but also as components that can be integrated into stand-alone systems.

This layer needs to start with a small and robust set of clearly defined and mission-critical methods and services that are needed by the majority of solutions on the top layer. This initial set of seed services, then, needs to be able to grow organically and scalably into one or more bigger platforms or infrastructures. On this layer it will be important to provide flexibility through a highly innovative ecosystem that enables the emergence of platforms, services, and infrastructures, as well as products. These are meant to serve as turbo engines for research, innovation, and development for providing ubiquitous resources for the multilingual DSM, public services, and European society. The platforms will be used for fast and economical service delivery to enterprises and end-users as well as for testing, showcasing, proof-of-concept demonstration, avant-garde adoption, and experimental and operational service composition.

Infrastructures and platforms help to reduce complexity on the user side (in this case mostly companies and other organisations that build new products or platforms on top of these enabling services) and support evolution (competition and cross-fertilisation) on the service provider side. The concepts of hybrid research or DevOps, i.e. a tight loop of research, development, and operations that allows for early testing and short development cycles, that has been successful in other areas needs to be adopted by language technology. The small initial set of services, platforms, and infrastructures may or may not share basic services (maintenance, promotion, licensing, payment) and means for quality assurance.

The services and resources provided on this layer need to help implement and support the technology solutions and are driven and supported by the underlying research themes. The range of services needs to include basic low-level technologies such as part-of-speech tagging and high-level (combined) ones such as machine translation including special terminology and human post-editing, automatic generation of spoken usage instructions, or email classification by sentiment and enrichment with background information.

The eventual creation of powerful cloud computing platforms for a wide range of services dealing with human language, knowledge, and emotion will not only benefit the individual and corporate users of these technologies but also the providers. Large-scale ICT infrastructures and innovation clusters such as this suggested platform are also foreseen in the Digital Agenda for Europe.

Users of the technology solutions on the top layer will be able to receive customised integrated services without having to install, combine, support, and maintain the software. They will have access to specialised solutions even if they do not use these regularly. Language technology providers will have ample opportunity to offer stand-alone or integrated services through component technologies or cloud-based application programming interfaces (APIs), respectively. Providers of language services rendered by human language professionals will be able to use platforms for enhancing their services by means of appropriate technology and for providing their services stand-alone or integrated into other application services.

Researchers will have a virtual laboratory for testing, combining, and benchmarking their technologies and for exposing them in realistic trials to real tasks and users. Through the involvement of users, valuable data will be collected within these inherently European platforms (vs. platforms that physically reside on other continents) that can directly feed back into improved services.

Providers of services that can be enabled or enhanced by text and speech processing will utilise the platform for testing the needed LT functionalities and for integrating them into their own solutions.

Corporate users will enjoy the benefits of language technology early and at no (or reasonable) cost through a large variety of generic and specialised services offered at a single source.

In order to allow for the gigantic range of potential solutions, the emerging infrastructures and platforms will have to host (and share) all relevant simple services, including components, tools, and data resources, as well as various layers or components of higher services that incorporate simpler ones. Resource exchange infrastructures such as, for example, META-SHARE can play an important role in the design of the platform.

The initial design and creation of these services and platforms has to be supported by public funding. Because of the demanding requirements about performance, reliability, user support, scalability, and persistence together with data protection and compliance with privacy regulation, the systems need to be established by one or more consortia with strong commercial partners and also be operated by these consortia or commercial contractors.

The infrastructures and platforms are intended for a mix of commercial and noncommercial services. For example, they could be, in one scenario, cost-free for all providers of non-commercial services (cost-free and advertisement-free) including research systems, experimental services, and freely shared resources but would raise revenues by charging a proportional commission on all commercially provided services. In order to reduce dependence on individual companies and software products, base technologies should be supplied by open source toolkits and standards.

The services, infrastructures, and platforms will considerably lower the barrier for market entry for innovative technologies, especially for products and services offered by SMEs. Still, these stakeholders may not have the resources, expertise, and time to create the necessary interfaces to integrate their results into real-life services, let alone the overarching platforms themselves. There is still a gap between research prototypes and products that have been engineered and tested for robust applications. Moreover, many innovative developments require access to special kinds of language resources such as recordings of spoken commands to smartphones, which are difficult to get for several reasons.

The service platforms and infrastructures will be an important instrument for supporting the entire innovation chain, but, in addition, interoperability standards, interfacing tools, middle-ware, and reference service architectures need to be developed and constantly adapted. Many of these may not be generic enough to serve all application areas, so that much of the work in resource and service integration will have to take place.

5 Layer 3: Priority Research Themes

Although we use computers to write and telephones to chat and search the web for knowledge, IT has no direct access to the meaning, purpose, and sentiment behind our trillions of written and spoken words. This is why today's technology is unable to summarise a text, answer a question, respond to a letter or to translate reliably, let alone to implement some of the more complex solutions envisaged above.

Many companies began much too early to invest in language technology research and development and then lost faith after a long period without any tangible progress. During the years of apparent technological standstill, however, research continued to conquer new ground. The results are a deeper theoretical understanding of language, better machine-readable dictionaries, thesauri, grammars and semantic resources, specialised efficient language processing algorithms, hardware with increased computing power and storage capacities for big data, large volumes of digitised text and speech data, and new methods of statistical language processing that could exploit language data for learning hidden semantic regularities governing our language use.

We do not yet possess the complete know-how for unleashing the full potential of language technology as essential research results are still missing (see below), but the speed of research keeps increasing and even small improvements can already be exploited for innovative products and services that are commercially viable. We are witnessing a chain of new products for a variety of applications entering the market in rapid succession. These applications tend to be built on dedicated computational models of language processing that are specialised for a certain task.

But increasingly, we observe a reuse of core components and language models for a wide variety of purposes. It started with dictionaries, spell checkers, and text-to-speech tools. Google Translate, Apple's Siri, Microsoft's Cortana and IBM Watson still do not use the same technologies for analysing and producing language because the generic processing components are simply not powerful enough to meet their respective needs. But many advanced research systems already utilise the same tools for syntactic analysis. This process of consolidation is set to continue. In ten years or less, basic language proficiency is going to be an integral component of any advanced IT.

In the envisaged big push toward delivering the solutions sketched previously through massive research and innovation, our community is faced with three enormous challenges:

1. *Richness and diversity.* A serious challenge is the sheer number of languages, some closely related, others distantly apart. Within a language, technology has to deal with numerous dialects, sociolects, registers, professional jargons, genres, and slangs.
2. *Depth and meaning.* Understanding language is a complex process. Human language is not only the key to knowledge and thought, it also cannot be interpreted without certain shared knowledge and active inference. Computational language proficiency needs semantic technologies.
3. *Multimodality and grounding.* Human language is embedded in our daily activities. It is combined with other modes and media of communication. It is affected by beliefs, desires, intentions, and emotions, and it affects all of these. Successful interactive language technology requires models of embodied and adaptive human interaction with people, technology, and other parts of the world.

It is fortunate for research and the economy that the only way to effectively tackle these three challenges involves submitting the evolving technology continuously to the growing demands and practical stress tests of real world applications. Google's Translate, Apple's Siri, Microsoft's Cortana, Autonomy's text analytics and scores of other products demonstrate that there are plenty of commercially viable appli-

cations for imperfect technologies. Only a continuous stream of technological innovation can provide the economic pull forces and the evolutionary environments for the realisation of the grand vision of a multilingual Digital Single Market.

The research layer is meant to support the DSM by enabling the services on the middle layer. The activities subsume both basic and applied research. The concrete composition of this and the middle layer largely depend on the set of technology solutions (top layer) because the specific solutions selected will formulate needs and requirements for the other two layers. Independent of the concrete set of themes, it is important to note that all themes need to be tightly intertwined, making use of one another in different application scenarios, especially so when research results, i.e. technologies, are combined on the services and the solutions layer.

6 Horizontal Framework Aspects

In this chapter we briefly mention and discuss several horizontal framework aspects that need to be addressed successfully to implement the strategic programme described previously.

6.1 Language Policies and Public Procurement

Technology progress would be even more efficient and effective if the recommended strategic programme could be accompanied by appropriate supportive policy making in several areas. One of these areas is multilingualism. Overcoming language barriers can greatly influence the future of the EU and the whole planet. Solutions for better communication and for access to content in the native languages of the users would not only enable the multilingual Digital Single Market, it would reaffirm the role of the EC to serve the needs of the EU citizens. A substantial connection to the infrastructural program Connecting Europe Facility (CEF) could help to speed up the transfer of research results to badly needed services for the European economy and public. At the same time, use cases should cover areas where the European societal needs massively overlap with business opportunities to achieve funding investment that pays back, for example, Public-Private-Partnerships (PPPs).

Language policies supporting multilingualism can create a tangible boost for technology development. Some of the best results in MT have been achieved in Catalonia, where legislation supporting the use of the Catalan language has created an increased demand for automatic translation.

Numerous US breakthroughs in IT that have subsequently led to commercially successful products of great economic impact were only achieved by a combination of systematic long-term research support and public procurement. Many types of aircraft or the autonomous land vehicle would not have seen the light of day without massive government support – even the internet or the speech technology behind Apple Siri benefited largely from sequences of DARPA programmes often followed by government contracts procuring earlier versions of the technology for military or civilian use by the public sector.

The greed for originality on the side of the public research funding bodies and their constant trial-and-error search for new themes that might finally help the European IT industry to be in time with their innovations has often caused the premature abortion of promising developments, whose preliminary results were more than once taken up by research centres and enterprises in the US. An example in language technology is the progress in statistical machine translation. Much of the groundwork laid in the German government-sponsored project Verbmobil (1993–2000) was later taken up by DARPA research and commercial systems – including Google Translate.

In order to drive technology evolution with public funding to a stage of maturity where first sample solutions can deliver visible benefits to the European citizens and where the private sector can take up technologies to then develop a wide range of more sophisticated profitable applications, we strongly advocate a combination of

1. language policies supporting the status of European languages in the public sector,
2. procurement of solution development by European public administrations,
3. long-term systematic research efforts with the goal to realise badly needed pre-competitive basic services.

European policy making should also speed up technology evolution by helping the research community to gain affordable and less restrictive access to text and speech data repositories, especially to data that have been collected with public support for scientific and cultural purposes.

Today, outdated legislation and restrictive interpretation of existing law hinder the effective use of many valuable data collections such as, for example, several national corpora. The research community urgently needs the help of European and national policy makers for modes of use of these data that would boost technology development without infringing on the economic interests of authors and publishers (also see Section 6.4 below).

6.2 Standards and Interoperability

Especially for the successful design, implementation, deployment, and continuous improvement of the language technology services, platforms, and infrastructure efforts for ensuring the interoperability of methods and services need to be intensified by significantly boosting standardisation activities. Targets of standardisation need to be data exchange formats as well as APIs so that components developed in different organisations can interoperate with one another. The language technology community has already been engaged in standardisation activities for decades (see, for example, the standards produced by ISO TC37, SC4), but these need to be intensified once more, especially in the context of sophisticated computing environments, modern architectures, cloud platforms, and Software-as-a-Service.

6.3 Open Source

While language technology-based industry solutions target an agile high-tech industry, many fields still appear to be dominated by expensive and slow-moving monolithic proprietary software that makes it especially hard for many SMEs to compete with developments. At the same time other areas have shown that massive collaboration in open-source-projects can lead to impressive and future-proof software such as operating systems (e.g. Linux) or CMS systems (e.g. Drupal).

Still, open source projects usually do not run by themselves. They require well conceived forms of organisation fitting the respective community and type of project. Therefore, these developments need to be supported by platforms and funding schemes in their own right.

While we do not want to play off proprietary against open-source software, we do want to support the development of the latter for the language industry. In fact, some tools and standards already exist in the industry and in language technology research, open source development is the normal case. But existing tools are often not mature enough and lack plans for maintenance so that they are only of limited usefulness for the industry and public services.

6.4 Copyright and Data Protection

Research and innovation in language technology depends on language data the way climate research depends on weather data or economic studies depend on financial data. Results derived in language technology research from the analysis of large amounts of texts in areas like machine translation, text and, data mining or text analytics such as statistical models or abstract representations do not interfere with the copyright holders' rights to publish, republish, modify, translate, and otherwise make available the texts in order for someone else to read them as a piece of artwork, document, etc. Still, traditional copyright and half-hearted exceptions for research are experienced as huge obstacles for research and innovation by the European research community. These obstacles come with a threat of severe economic consequences: academic and industrial researchers – already a sparse resource – may leave Europe to pursue their goals in other continents, technology leadership may migrate to the US or Asia, immense opportunities of growth are lost. We are happy that the EC is taking the next steps towards the important and urgent goal of a reform of European copyright law.

7 Conclusions

7.1 Expected Economic Impact

The EC predicts that the transition to the integrated DSM will deliver up to €250 billion in economic growth by 2020. However, this ambitious goal – in fact, even more – can only be reached if the language factor is taken into account! If customers are still hampered by language, online commerce will remain confined to fragmented markets, defined by language silos. Approximately 60% of individuals in non-Anglophone countries seldom or never make online purchases from English-language sites; the number willing to purchase from sites in non-native languages other than English is much, much lower. As a result, no language can address 20% or more of the DSM.

European SMEs are an integral and vital component of the DSM. However, only 15% of them sell online – and of that 15%, fewer than half do so across borders. SMEs that sell their products and services internationally exhibit 7% job growth and 26% innovate in their offering – compared to a job growth of 1% and 8% innovation for SMEs that do not sell their products and services internationally. Only if Europe accepts the multilingual challenge and decides to design and to implement research and innovation driven technological solutions as well as a service infrastructure with the goal of overcoming language barriers, can the economic benefits of the DSM be achieved. Enabling and empowering European SMEs easily to use language technologies to grow their business online across many languages is key to boosting their levels of innovation and jobs creation.

If the strategic programme specified in this Strategic Agenda for the Multilingual DSM is fully realised, we expect economic growth by 2020 to be much higher than the predicted €250 billion since, crucially, we will have successfully enabled many European SMEs to sell online on the *multilingual* Digital Single Market, substantially multiplying their reach. Furthermore, we expect the creation of tens of thousands sustainable new jobs in the medium to long-term. The growth would not stop at the borders of Europe: if the strategic programme is successful, Europe could attempt to offer the solutions to other multilingual societies, for example, to adapt and to export certain parts of the strategic programme to India or South Africa.

The European Digital Single Market today would account for approximately 25% of global economic potential. However, if Europe overcame the language barriers that hamper intra-European trading, it would also remove barriers to international trade that keep European SMEs from achieving their full economic potential by penetrating markets in other continents beyond our own. Addressing the official and major regional languages of Europe would open access to over 50% of the world's online potential and 73% of the world online market in economic terms, amounting to an online market of approximately €25 trillion (sic!) in 2013. *The global potential for European businesses exceeds the continent-internal opportunities from the DSM by orders of magnitude.*

In addition to being a key enabling technology for the multilingual DSM, the field of Language Technology comes with a non-trivial economic power itself. The European market for translation, interpretation, and localisation was estimated to be €5.7 billion in 2008. The subtitling and dubbing sector was at €633 million, language teaching at €1.6 billion. The overall value of the European language industry was estimated at €8.4 billion and expected to grow by 10% per year, i.e. resulting in ca. €16.5 billion in 2015. The global language technology industry[20] is evaluated at €26.5b in 2015, projected to rise to €65b by 2020. The global speech technology market is even bigger, it will reach ca. US$20.9 billion by 2015 and ca. US$31.3 billion by 2017.

[20] Figures from LT2013: Status and Potential of the European Language Technology Markets, April 2013

7.2 Relevance to the EC's Digital Single Market Strategy

On 25 March 2015, the College of Commissioners discussed key aspects of the Digital Single Market and published three main DSM areas with a small number of priorities each.[21] The following table maps key aspects of this Strategic Agenda for the Multilingual DSM (right column) onto the European Commission's current DSM strategy (left column), the final version of which is expected to be available in early May 2015.

"Digital Single Market Strategy: European Commission agrees areas for action" (College of Commissioners, 25 March 2015)	**Relationship of this Strategic Agenda for the Multilingual Digital Single Market to the EC's DSM Strategy**
Area 1: Better access for consumers and businesses to digital goods and services	
Facilitating cross-border e-commerce, especially for SMEs, with harmonised consumer and contract rules and with more efficient and affordable parcel delivery. Today only 15% of consumers shop online from another EU country – which is not surprising, if the delivery charge ends up higher than the actual price of the product (see Factsheet for more figures).	In addition to excessive delivery charges for goods purchased in other countries, language barriers prevent customers from even knowing that potentially relevant shops or companies offer products or services relevant to their need because the customers are unable to find them (moreover, 52% of EU customers do not purchase from English-language sites). Multilingual technology solutions for a unified, language-transparent user experience (i.e. automatic website localisation for ecommerce websites) can help!
Tackling geo-blocking: too many Europeans cannot use online services that are available in other EU countries, often without any justification; or they are re-routed to a local store with different prices. Such discrimination cannot exist in a Single Market.	Geo-blocking is a barrier to access – but so is language-blocking! Language-blocking keeps customers from accessing content in languages they do not speak; customers never even know what they cannot find. Language-blocking is unavoidable: no-one speaks all languages. However, current online translation is insufficient. Language-blocking prevents customers from even *trying* to conduct cross-border commerce . Many of the technology solutions described in this document can help!
Modernising copyright law to ensure the right balance between the interests of creators and those of users or consumers. It will improve people's access to culture – and therefore support cultural diversity – while opening new opportunities for artists and content creators and ensuring a better enforcement of rights.	The European language technology community supports modernising copyright law. A broad and unified copyright reform is needed so that academic, applied, and industrial language technology research are enabled to build modern technologies using publicly available data sets within a safe and secure legal framework.
Simplifying VAT arrangements is important to boost the cross-border activities of businesses, especially SMEs. The cost and complexity of having to deal with foreign tax rules are a major problem for SMEs. The VAT-related costs due to different requirements are estimated at EUR 80 billion.	–

[21] European Commission (2015): "Digital Single Market Strategy: European Commission agrees areas for action", Brussels, Belgium (25 March 2015). http://europa.eu/rapid/press-release_IP-15–4653_en.htm

"Digital Single Market Strategy: European Commission agrees areas for action" (College of Commissioners, 25 March 2015)	Relationship of this Strategic Agenda for the Multilingual Digital Single Market to the EC's DSM Strategy
Area 2: Shaping the environment for digital networks and services to flourish	
All digital services, applications, and content depend on high-speed internet and secure networks: the lifeblood of new, innovative digital services. To encourage investment in infrastructure, the Commission will therefore review the current telecoms and media rules to make them fit for new challenges, in particular relating to consumer uses (for example the increasing number of voice calls made over the internet) and new players in the field.	The European language technology community supports investments into high-speed internet and secure networks. Our applications increasingly depend on fast networks due to the ever increasing use of Software-as-a-Service approaches for language technologies. For example, data centres analyse customer requests made via spoken language input on a mobile phone. Real-time responses of intelligent assistants rely on fast European networks.
Spectrum is the air the internet breathes. Improving coordination among Member States is essential. Europe has witnessed significant delays in the roll-out of the latest 4G technology, as suitable spectrum was not available. Spectrum does not stop at national borders: a European approach to its management is needed to promote a genuine single market with pan-European services.	–
The Commission will look into the growing importance of online platforms (search engines, social media, app stores, etc.) for a thriving internet-enabled economy. This includes looking at how to strengthen trust in online services through more transparency, how to include them in the online value chain, and to facilitate the swift removal of illegal content.	Language technologies can play decisive roles for many different types of online platforms – from search engines to translation platforms (our "Digital Translation Centre" solution) to automatic localisation of arbitrary online platforms (including ecommerce). Europe needs online platforms that support all European languages, driven by European language technologies.
Today, 72% of internet users in Europe are concerned about using online services because they worry that they have to reveal too much personal data online. The swift adoption of the Data Protection Regulation is key to boosting trust.	–
Area 3: Creating a European Digital Economy and Society with long-term growth potential	
Industry is a key pillar of the European economy – the EU manufacturing sector accounts for 2 million companies and 33 million jobs. The Commission wants to help all industrial sectors integrate new technologies and manage the transition to a smart industrial system ("Industry 4.0").	Already in the next few years many machines and manufacturing lines will be connected to the internet through digital technologies. Many of these machines need to be operated through spoken language voice interfaces. The European language technology community can provide smart systems that allow operators to interact with their machines in a robust, efficient, and hands-free way, using all European languages.

"Digital Single Market Strategy: European Commission agrees areas for action" (College of Commissioners, 25 March 2015)	**Relationship of this Strategic Agenda for the Multilingual Digital Single Market to the EC's DSM Strategy**
Standards: ensuring interoperability for new technologies are essential for Europe's competitiveness, they must be developed faster.	The Strategic Agenda for the Multilingual Digital Single Market crucially relies on interoperable software, which requires modern standards. The European language technology community supports the European Commission's push for the faster development of IT standards.
The Commission also wants industry and society to make the most of out of the data economy. Large amounts of data are produced every second, created by persons or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, or GPS signals. Big data is a goldmine, but it also raises important challenges, from ownership to data protection to standards. These need to be addressed to unlock its potential.	Huge quantities of big data sets are actually *linguistic* big data sets, i.e. they consist of shorter or longer natural language texts, often in multiple languages. The European language technology community can help, among others, with the development of sophisticated methods for multilingual big data text analytics. These are needed for applications such as big data-based online marketing, customer relationship management, e-participation and many others. If the goal is to find, online, the opinions of customers on one specific topic, in the future it will be key to aggregate online information in many different languages using cross-lingual and multilingual language technologies for multilingual data value chains in a truly European data economy.
The same goes for cloud computing, the use of which is rapidly growing: the proportion of digital data stored in the cloud is projected to rise from 20% in 2013 to 40% in 2020. While shared networks and resources can boost our economy, they also need the right framework to flourish and be used by more people, companies, organisations, and public services across Europe.	The European language technology community supports the European Commission's push for cloud computing. Many of our solutions already reside in the cloud, future ones increasingly so.
Europeans should also be able to fully benefit from interoperable e-services, from e-government to e-health, and develop their digital skills to seize the opportunities of the internet and boost their chances of getting a job.	In order to enable Europeans fully to benefit from interoperable e-services, from e-government to e-health, the respective services need to be available in all European languages, ideally also across languages. This Strategic Agenda for the Multilingual Digital Single Market foresees a whole set of technology solutions for public services.

7.3 Potential Funding Sources

We suggest setting up, under the umbrella of the strategic programme, a coordinated initiative both on the national (Member States, Associated Countries, regions) and international level (EC/EU), including research centres as well as small, medium, and large enterprises who work on or with language technologies and other relevant stakeholders.

The financial setup of this strategic programme requires a mix of several ingredients. These include the European Union, the Member States, and Associated Countries, as well as industry.

The European Union could support the strategic programme especially through dedicated activities in upcoming Horizon 2020 calls and through Connecting Europe Facility (CEF). Horizon 2020 Research Actions are compatible to our planned activities on Layer 3 while Horizon 2020 Research and Innovation Actions as well as Coordination and Support Actions can be set up for Layer 2. Highly innovative activities with a major commercial impact can be used for Layer 1 – especially on Layers 1 and 2, the European language technology industry will participate (most of these companies are SMEs). Through CEF, technology deployment and innovation actions could be funded, especially with regard to public services. Furthermore, there are horizontal programmes such as, for example, Horizon 2020 Widespread/Teaming that could boost the knowledge and technology transfer between countries that already have excellent research and innovation hubs in language technology and those that do not; the goal would be to enable the less innovative countries to develop technologies for their respective languages (Layer 3 and also the dedicated base layer for the development of core resources and technologies). Similar horizontal programmes to boost SMEs exist as well.

On the national and regional levels, the respective local funding agencies could provide resources, especially to support the development of technologies for their respective national or regional languages (Layer 3). There are also dedicated programmes for supporting national and regional companies becoming more innovative – these programmes are especially adequate for activities on Layer 1.

A European Flagship Project could be one possible candidate instrument for the suggested strategic programme, especially with regard to significantly boosting the development of novel language technology approaches, algorithms, and paradigms. Critically, public procurement can play a decisive role in this strategic programme: if the European Union is willing to invest in the development of multilingual technologies made *in* Europe and apply them *for* Europe, the EU itself would be the perfect reference user of such technologies, setting an example for national or regional governments.

7.4 Next Steps

This version of the Strategic Agenda for the Multilingual Digital Single Market contains an initial suggestion that was prepared by the European language technology community. The document is publicly unveiled at the Riga Summit 2015 on the Multilingual Digital Single Market (April 27–29, 2015, in Riga, Latvia).[22] At the Riga Summit 2015, we will present the document and initiate the first public consultation phase, especially at META-FORUM 2015 (April 27) and at the Riga Summit Plenary Day (April 28). Feedback and additional input gathered during these events will flow back into upcoming versions of the strategic agenda.

As soon as our strategic programme and the proposed technology solutions have been successfully discussed with the European Commission and an agreement has been reached, the strategies and roadmaps need to be further aligned, refined, and specified in the community in one or more public consultation phases, especially with regard to Layer 2 (Infrastructures, Platforms, Services) and Layer 3 (Priority Research Themes). Afterwards, priorities need to be assigned, especially to the technology solutions, from which priorities will follow for Layers 2 and 3, resulting in a more detailed version of this agenda that also includes roadmaps for technology development.

We expect the final version of the Strategic Agenda for the Multilingual Digital Single Market to be available in late 2015.

[22] http://rigasummit2015.eu

Appendix A. Input Documents

The following documents, roadmaps, and presentations have informed the current version of the Strategic Agenda for the Multilingual Digital Single Market.

- Philipp Cimiano (2015): "The LIDER Roadmap in a nutshell", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- Gerald Cultot (2015): "eHealth services – multilingual challenges", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- Andrew Joscelyne (late 2014): "A Strategic Research and Innovation Agenda for a Conversational European Digital Marketplace" (draft position paper).
- Nils Lenke (2015): "Nuance Inc.", DFKI Tech Day, 30 January 2015, DFKI Saarbrücken, Germany.
- Dave Lewis (2015): "Shopping Across the Language Barrier", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- LIDER (10/2014): "Roadmap for the use of Linguistic Linked Data for content analytics"
- META-NET (2013): "Strategic Research Agenda for Multilingual Europe 2020", Georg Rehm and Hans Uszkoreit (eds.), presented by the META Technology Council. Springer.
- MLi (09/2014): "D5.1 – Big and Social Language Data Requirements for the MLi Hub".
- QTLaunchPad (11/2014): "European Quality Translation Research 2015: Ongoing Work and Roadmap".
- Ruben Riestra (2015): "Multilingual data value chains in the Digital Single Market", report presented at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- ROCKIT (10/2014): Roadmap for Conversational Interaction Technologies – Coordination and Support Action, D1.1 Innovation Drivers, future scenarios, and best practice.
- ROCKIT (10/2014): Roadmap for Conversational Interaction Technologies – Coordination and Support Action, D2.1 First Report on Innovation in the ROCKIT Domain.
- ROCKIT (10/2014): Roadmap for Conversational Interaction Technologies – Coordination and Support Action, D3.1 First Report on Research in the ROCKIT Domain.
- ROCKIT (02/2014): Roadmap for Conversational Interaction Technologies – Coordination and Support Action, D4.1 ROCKIT Roadmap Specifications.
- Alan Mas Soro: "Language Technologies for Europe: A way to foster SME internationalization", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- Adomas Svirskas (2015): "Pan-European Electronic Document Platform. Open Interoperable Solution for Europe", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- Hans Uszkoreit (2014): "European Platform(s) for Machine Translation and other Language Technologies", presentation given at the META-NET Platform Strategy Meeting during the Language Resources and Evaluation Conference (LREC), 26–31 May 2014, Reykjavik, Iceland.

- Xenios Xenophontos (2015): "Online Dispute Resolution Platform – Multilingual challenges", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.
- Sonja Zillner (2015): "cPPP Big Data Value-SRIA", presentation given at the Workshop on multilingual data value chains in the Digital Single Market, 16 January 2015, Brussels, Belgium. https://ec.europa.eu/digital-agenda/en/news/workshop-multilingual-data-value-chains-digital-single-market.

Appendix B. Digital Language Extinction in Europe

Most European languages are unlikely to survive in the digital age, a study by Europe's leading Language Technology experts warns. Assessing the level of support through language technology for 30 of the more than 60 European languages, we concluded that digital support for 21 of the 30 languages investigated is "non-existent" or "weak" at best. The study "Europe's Languages in the Digital Age" was carried out by META-NET, a European network of excellence that consists of 60 research centres in 34 countries, working on the technological foundations of multilingual Europe.

Europe must take action to prepare its languages for the digital age. They are a precious component of our cultural heritage and, as such, they deserve future-proofing. The META-NET study shows that, in the digital age, multilingual Europe and its linguistic heritage are facing challenges but also many possibilities and opportunities.

The study, prepared by more than 200 experts and documented in 31 volumes of the META-NET White Paper Series (available both online and in print), assessed language technology support for each language in four different areas: automatic translation, speech interaction, text analysis, and the availability of language resources. A total of 21 of the 30 languages (70%) were placed in the lowest category, "support is weak or non-existent" for at least one area by the experts (**Figure 8**). Several languages, for example, Icelandic, Lithuanian, and Maltese, received this lowest score in all four areas. However, it must be noted that support for some of the languages with smaller numbers of speakers is slowly increasing since the original publication of the META-NET White Paper Series in 2012. At the other end of the spectrum, while no language was considered to have "excellent support", only English was assessed as having "good support", followed by languages such as Dutch, French, German, Italian, and Spanish with "moderate support". Languages such as Basque, Bulgarian, Catalan, Greek, Hungarian, and Polish exhibit "fragmentary support", placing them also in the set of high-risk languages.

The white papers and more details are available at http://www.meta-net.eu/whitepapers.

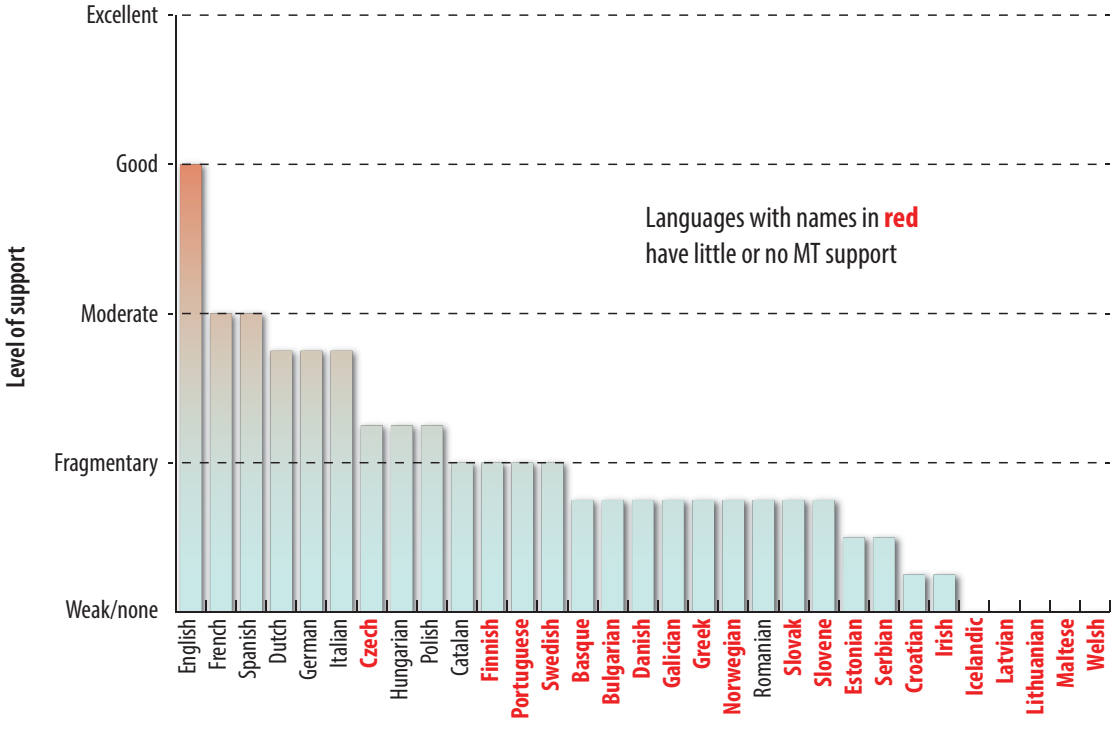


Figure 8: *Language technology support levels from the META-NET White Papers*

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

Unified Customer Experience

- Provides a contextualised experience to users (for eCommerce)
- Brings together content, product, customer care, customer relationship, discussion fora, help-desks, etc.
- Unified digital (eco)system 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 semantification

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 editorial team of this Strategic Research and Innovation Agenda (SRIA) can be reached through Dr. Georg Rehm: georg.rehm@dfki.de.

Preparation of this document has been co-funded through CRACKER and LT_Observatory.