EUROPEAN COMMISSION
Directorate-General for Communications Networks, Content and Technology

Data
Accessibility, Multilingualism and Safer Internet

Stakeholder Consultation on Language Technologies in Digital Europe

Brussels 10-11 February 2020

Contents

1. SUMMARY 3

2. METHODOLOGY 6

3. AI CAPACITY BUILDING AND REACHING A CRITICAL MASS 7

3.1. Exploiting language data in EU 7

3.1.1. Challenges and needs 7

3.1.2. Summary of proposed actions 9

3.2. Reaching a critical mass 10

3.2.1. GDPR compliance - Anonymisation 10

3.2.2. Summary of proposed actions on EU level 10

3.2.3. EU as a large scale procurer of language technologies / reference implementation/quality information 10

3.2.4. Summary of proposed actions on EU level 11

3.2.5. Supporting LT SMEs through moonshot projects 11

3.2.6. Summary of proposed action(s) at EU-level 12

3.3. Adaptation of AI technologies to specific languages and domains 12

3.3.1. Challenges and needs 12

3.3.2. Summary of proposed action(s) at EU-level 13

4. DEPLOYMENT OF AI-BASED LANGUAGE TECHNOLOGIES FOR EUROPEAN SMES 14

4.1. Understanding/Addressing the Market 14

4.1.1. Summary of proposed action(s) at EU-level 15

4.2. Financing of LT deployment in European SMEs 15

4.3. European SMEs as a market 17

4.4. Reaching-out to end-users 17

4.5. European Commission's role as market creator/ creating a market 17

5. MOBILISING AND SUPPORT MEASURES FOR LANGUAGE TECHNOLOGIES COMMUNITY 18
5.1. ELG project and platform approach 18
5.1.1. Platform approach from a market perspective, advocated by LT-Innovate 19
5.1.2. Platform from the provider perspective 20
5.1.3. Platform approach from the solution builder/"integrator" perspective 21
5.1.1. Cloud and High-Performance Computing 21
5.2. Standardisation 22
5.3. Role of and collaboration with Member States 22
6. OTHER ISSUES DISCUSSED 23
6.1. Further development of eTranslation 23
6.2. Language technologies and accessibility 24
7. ANNEX 1 – AGENDA 25
8. ANNEX 2: PARTICIPANT LIST 26
9. ANNEX 3: PARTICIPANTS' INDIVIDUAL CONTRIBUTIONS (SEE SEPARATE FILE) 26

Disclaimer:

The information and views set out in this report are those of the workshop participants and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the information included in this report. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

1. SUMMARY

The workshop gathered 50 representatives of the European LT community, mainly developers and providers of LT software and services but also LT users from different sectors for 1,5 days of intense discussions and team work.

The objective of the workshop was to identify current and future needs and challenges related to the deployment of language technologies in Europe with the aim to help prepare and scope the EC intervention concerning language technologies in the early work programmes of the upcoming Digital Europe programme.

State of the art language processing solutions based on AI technologies rely on the quality of an underlying model which in turn relies on language data. More high quality data always gives better results in neural language modelling. In this context the multilingual diversity can be an EU advantage, if all data sets available in Europe could be utilised.

The main challenges related to the exploitation of language data in Europe for the purpose of developing language technology (LT) applications are related to privacy (GDPR), anonymization, the amounts of data needed, copyright, the cost of preparation and annotation, metrics of quality and workflows, discoverability and findability as well as the fragmented data landscape with multiple data platforms and portals.

To address the data fragmentation and readiness a "structure" available to the community at large to efficiently obtain/buy language resources is needed. Interoperability of data and processing platforms and infrastructures must be considered at different levels. Specifying a space that is dedicated for commercial data use (business dimension) would also be needed.

As data creation is costly, data creation and annotation efforts must be focussed and coordinated. Joint efforts on data creation, cleaning and annotation should be facilitated. Crowd-sourcing/community projects to ensure diversity and scale could be initiated or supported by EC. Incentives for encouraging private sector and public sector to share their data are needed. In addition, the LT industry participants highlighted that intellectual property rules (IPR) are an obstacle for the aggregation of language resources that are essential to build interoperable languages tools. To further unlock the potential of language data in Europe, the text and data mining exception should consider that research and innovation is also happening within commercial organizations/industry.

While there was some discussion on protecting the market through protecting data, there seemed to be an agreement among the participants that any data resulting from EU funding should be made widely available.

The discussion on developing AI-based language technologies focussed on reaching a critical mass and adaptation of AI-based technologies to various European languages.

As current processes in the LT industry are often not GDPR compliant, development and industrialisation of secure/anonymised translation pipelines, infrastructure and workflows for text which cannot be anonymised or until the anonymization issue can be solved would be helpful for the LT industry. By building GDPR compliant translation tools, anonymization solutions and pipelines, Europe could assert itself as the global leader in this area. Furthermore, legal validation of translation tools and workflows as "GDPR compliant", for example through certification by an EU body or an ISO type standard

could be envisaged to reassure buyers of LT services that they are paying for a GDPR compliant service.

The EU could establish best practices and standards through procurement, for example it could procure a reference implementation of an industrialised translation and anonymization pipeline that assures GDPR compliance or reference information (e.g. a list of all first names used across the EU) that can be used by others.

Adaptation of AI technologies to specific languages and domains is challenging due to the fragmented market. For the smaller markets there is often no return on investment. It is costly and inefficient to develop and maintain/improve deep learning technologies. Furthermore, scaling to another country requires contacts and a network. The challenge is the cultural difference, visibility and difficulty to form partnerships and find and engage beta-testers and customers. An important challenge is also quality assurance for solutions that are meant for commercial use.

The inefficiency of training deep learning algorithms must be addressed. Adaptation of AI-based technologies to other languages and domains can be facilitated by providing baseline solutions, APIs, data and standards. To address the market gap, i.e. the difficulties faced by SMEs to reach large clients and the public sector, the community must be fostered to help LT providers find each other, share solutions and develop joint offerings. Collaboration/connectedness of European LT SMEs must be facilitated also so that they can share effort in developing and maintaining the technology.

A flagship project with an emotional traction could be envisaged (this is beyond DEP and would need the support and leadership of several or all Member States outside of EU funding programs) in order to make available to the community the largest amounts of language data. For example, the EU could initiate a community effort to collaboratively create a generic data set/a certified reference dataset (e.g., have all versions of Wikipedia articles translated in all European languages and annotated). The EU could also propose and co-fund a flagship multi-billion euro moonshot project to create a European Digital Public Space – many scholars have noted the obvious lack of a joint public sphere in Europe. Filling this gap with a digital platform that supports all EU languages and that enables the European citizens to communicate with one another in their native languages using automated translation and that, furthermore, provides access to audio and video content (including automated translation, subtitling etc.), would surely have positive effects on the political situation in Europe, among others. Such a digital platform would, in addition, bootstrap the European LT industry and make it a market leader in AI and machine learning.

An SME-friendly funding scheme is needed to support the adaptation of AI-based LT to other languages/domains. A support programme for SMEs who want to use and implement language technologies would be welcome to further boost the take up of the technology. It should be lean, efficient and market-driven.

Many participants found that the EU should let the "*1000 flowers blossom*", i.e. support the many small SMEs in developing successful products. By doing so we create value for the citizens because many small companies which work in different markets will create better solutions than one big company will create for a big market. EU should plant "small seeds" and let the markets find their ways. The economic power of public investment was also highlighted.

A platform-approach to addressing challenges faced by LT SMEs was acknowledged and the European Language Grid (ELG) initiative received a lot of interest from participants. All participants agreed that an API-infrastructure/platform with European endpoints/APIs that can be used in any language would benefit the European LT ecosystem. It was agreed that ELG is well positioned to become a European platform providing easy access to language resources, tools and services for European languages. However it must be scaled up, whilst the integration and back-end compatibility with other LT and AI initiatives and platforms (both European and non-European) must be considered and ensured.

One of the most important aspects is connecting to the users. In any case, both from the user and provider perspective, a flexible system is needed that can easily integrate different technologies of different providers. Ease of use, easy plug-and-play of individual components into company workflows must be possible, as well as the trial and comparison of different options in order to facilitate reaching out to and convincing customers.

Communication to different stakeholders about the use they can make of the platform and the benefits it can bring to them is another key success factor. Likewise, governance is critical, since by putting the right framework in place, conflicts can be avoided.

Alignment and interoperability with sectorial AI-platforms and the sectorial data spaces proposed in the DEP programme is essential. The European Language Grid should seek complementarity with the other current or future (commercial) platforms in various domains. It has to be compatible with those platforms, feeding value into them and adding-value also for the bigger players.

Easy access to provider-agnostic cloud facilities would also help increase deployment of LT solutions in Europe. LT SMEs would welcome a solution whereby the SMEs in Europe do not need to choose between different providers, as long as some basic pre-conditions are fulfilled, e.g. data stays in Europe.

Creation and promotion of standards was considered important. For example, standards for procurement for virtual assistants in public administrations would be helpful to increase adoption of this technology. In general, standardized APIs would be helpful for both users and integrators. A platform where you can demo services could drive standardisation.

The EU could also help the LT industry through policies and regulatory measures. It can create legal regulations to ensure language equality across the EU. This would support LT work at a local level by smaller players, as these markets are often ignored by the big players. Legislative measures like the Web Accessibility Directive is a great example. Legislative measures for inclusiveness and accessibility could be extended, for example by setting a requirement to provide public online service in all EU languages, to provide subtitles for all TV broadcasts, to enable text-to-speech for people with disabilities or to promote the development of the European Digital Public Sphere including a novel, interactive, multilingual European Media Platform.

Funding of a thorough market study of all the verticals and with good publicity would help a lot smaller LT SMEs to understand the needs, the gaps and the players in the market.

Finally, Europe needs to develop a strategy. Europe must be clear on what we want to achieve. A strategy requires clear ex ante definition of targets, capabilities and building blocks. We need to put our resources where we are strong and can win. The EU has huge data resources. They should be released.

The EC support for language technologies should consider and at the same time guide Member States' efforts in this area. LT services and data are being funded and made available through national initiatives. Making national services and data available also through a European grid would ease development of LT tools and services covering those languages by SMEs in Europe. Furthermore, a continued collaboration/forum/exchange of best practices at European level is needed. Guidelines and documents from the European Commission and EC-funded initiatives are helpful for national decision-makers. Possibly a dynamic barometer which assesses the LT readiness in Member States could be envisaged.

As regards further development of eTranslation, the service should be made available to the private sector. In particular, an eTranslation plug-in could be developed and opened up for use by language service providers (LSPs). LSPs and other companies could integrate the eTranslation API into their own workflows and return the data to the EC for uploading to ELRC repository or the European Language Grid. Finally it was considered that European LT industry should be more closely involved in further developments of the eTranslation platform.

2. METHODOLOGY

The workshop started by an EC presentation which set the scene by presenting the ongoing LT actions, recent SME consultation results and the Digital Europe Programme Orientations. This was followed by the presentation of the European Language Grid initiative and the LT industry perspective on the challenges at hand by LT-Innovate.

In the second half of the morning and early afternoon, each workshop participant presented a two-minute elevator pitch accompanied by a one-slide presentation related to one or more of the following issues:

Language Technology and AI. This covered questions such as: what are your needs to develop AI-based products/services? What can you provide? As well as examples of how LT was successfully integrated in AI-based products or services.

Exploiting language data in EU. This covered questions such as: what are your needs in terms of language resources? What can you provide? What would help you in your data search activities? As well as examples of useful data sources.

Financing of LT deployment in European SMEs. This covered questions such as: how to support efficient development and deployment of language technologies in Europe? What financing mechanisms would help to convince SMEs to deploy LT? What other incentives would help? As well as examples of effective financing schemes which helped in developing or deploying a new product, finding new clients etc.

Infrastructure requirements. This covered questions such as: what infrastructure you need to develop and deploy your services/products? Which needs are currently not addressed or should be better addressed? What are your requirements or preferences with regard to HPC and Cloud infrastructures? How could the DEP programme best serve you?

Interoperability and standardisation. This covered questions such as: should the EU intervene? Is a "platform" approach suitable to overcome the fragmentation in Europe?

LT for European languages: a shared responsibility of EU and the Member States. This covered questions such as: how to ensure synergies between national and European

activities/programmes on LT? What should be done on national and on EU level? As well as examples of successful national activities related to language technologies.

In the afternoon of the first day and in the morning of the second day participants took part in thematic sessions related to the above subjects.

The workshop was concluded by presentations of the results of the discussions and feedback on the results of all thematic sessions.

3. AI CAPACITY BUILDING AND REACHING A CRITICAL MASS

3.1. Exploiting language data in EU

Since 2017 there has been a global shift in how we approach language processing. Any solution heavily relies on the quality of a language model. The language models in turn rely on data. More high quality data always gives better results in neural language modelling.

Big players like Google or Baidu started by acquiring lots of data and having data pipelines. In Europe, we must secure a pipeline of data acquisition. Currently there is a "rat race" among companies who look for data. There are hardly any large data/speech to text repositories on which SMEs can build.

In this context the multilingual diversity can be an EU advantage, if we could utilise all data sets available in Europe. The EU could enforce/foresee measures to gather whatever public documents the public administrations in Europe create. Public administrations do not realize that they generate useful data. While the Connecting Europe Programme was good in raising the visibility and understanding about the importance of language data, more should be done. Non-governmental organisations should also be considered as potential data contributors, as they often have clean parallel data. Data sharing initiatives among the private players should also be supported. The challenge is to gather more data in various fields and enable their effective and efficient use.

3.1.1. Challenges and needs

The main challenges related to the exploitation of language data in Europe for the purpose of developing language technology applications are related to privacy (GDPR), anonymization, the amounts of data needed, copyright, the cost of preparation and annotation, metrics of quality and workflows. This situation has not changed in the past few years. The recent EU copyright and GDPR regulations have also changed the landscape of data re-use. Moreover, the more platforms and portals that are developed, the more fragmented the market becomes. The discussions in the workshop mainly focussed on text data.

A lot of data is needed and while the data is out there, accessing it is a problem. For example, there are enough resources/data for creating machine translation engines but data is widely dispersed. There are many repositories which were created during EU projects but after the project ends there is mostly no way to access and even upload new data to these repositories. Sustainable data pipelines must be ensured. The data needs to be "consolidated" in one place and made ready for use (data readiness). Here licensing conditions must be considered, too.

To address the data fragmentation and readiness not only is a budget needed but a **"structure" to efficiently obtain language resources**. This language data market place must make it possible for buyers to assess the quality, understand the cost and the differences across languages. It should also address the issue on how to capitalize models across the industry based on confidential data that cannot be shared. While it must be possible to buy static or historical data, accessing "live" data should also be considered as this would be useful for processing news and media, for example.

Furthermore, to make data acceptable and useful for use in LT applications it needs to be processed. This is beyond the financial means of many SMEs.

GDPR and copyright regulations are reshaping data sharing. Data anonymization is a response to GDPR but not to copyright rules. Whereas US has the "fair use" principle, and the "fair dealing" policy exists in the UK, copyright regulations are stricter in the EU as far as the development of language technology applications for commercial purposes is concerned.

While Web crawling and data mining actions are commonly used for scientific purposes, participants felt that the EU copyright rules should allow the re-use the resulting data for commercial purposes, e.g. in language technology products. They considered that language resources that are needed to make languages and tools interoperable should be exempt from copyright rules. For example, if you use the text of books from a library to train models you are not going to publish the books as books, extracts or compilations, so you have not infringed the primary objective of the author.

The participants considered that the industry side should be more proactive to express and present their view towards a revision of the copyright directive.

Protecting language data. There was some discussion on protecting the market through protecting data. It was claimed that Facebook/Google are prohibited in China not for censorship reasons but for data reasons. The Chinese are not allowed to use the services because of the data that goes to the "US giants".

There seemed to be an agreement among the participants that any data resulting from EU funding should be made widely available. Limiting the availability of data to European players only (e.g. issuing licences to reuse this data to EU companies only) to give them a competitive advantage would be counterproductive. The main aim is to make the data available to ease the development of products covering the smaller EU languages. Language data is an infrastructure similar to a highway which requires an ongoing investment. If an infrastructure is created all players should be able to benefit from it.

Interoperability of data infrastructures: A large number of infrastructures/platforms exist, e.g.: ELRC-SHARE, CEF AT services offered by European companies, ELG, CLARIN… All these cater for visibility, discovery, use and reuse. There is a variation in architectures because the infrastructures respond to specific target communities and stakeholders. Cross-links and connections between these must exist and be enabled so that these infrastructures can talk to each other. Interoperability comes at different level and depths: technical, operational, legal… a business dimension can be added. While syntactic interoperability can be tackled, semantic interoperability is more difficult to crack. For all these you need standards. Measures for continuous support of data aggregation and processing infrastructures (not building new infrastructures) are needed.

3.1.2. Summary of proposed actions

The EU should foresee actions regarding language resources. These actions must happen at EU-level because the language coverage in Europe is an EU-wide issue. Ensuring adequate coverage for all European languages and a level playing field for Europe's multilingual society can only happen with EU coordination and support. Language resources should be considered as public resources necessary to build a public infrastructure therefore should not be subject to copyright (as foreseen for public sector information under the 'Open Data Directive' (Directive (EU) 2019/1024).

Joint efforts: As data creation is costly, data creation and annotation efforts must be focussed. Facilitating joint efforts on data creation, cleaning and annotation was repeatedly raised, for example EC funding for platforms for crowdsourced annotation for multilingual data sets; creating a big data set on which people work together. Test suites, annotated data, and benchmarks, are needed.

A best practice example could be the language bank of Finland. A state report produced for the Finnish government where 50 commercial and public organisation in Finland were interviewed revealed a need for a large, balanced, annotated corpus of everyday speech available for commercial purposes. Currently there is an initiative to decide/determine where such a corpus should be located. Model contracts are being established on how this data should be collected. This could be a blueprint for a similar effort in other countries.

Other examples of **crowd-sourcing/community projects to ensure diversity and scale** include the Mozilla project or the exercise that Deutsche Welle did in a Mongolian University to collect transcribed speech data. Bangor University did the same for Welsh.

A bottom-up initiative could foresee collecting everything we produce. The EU has so many resources that we could produce this kind of data store and make it freely available for reuse. This should happen not just at EU institution level but at Member State public administration level, too.

Vouchers could be foreseen to create language resources for the community. However, funding small, short-term projects which then lie stagnant at the end of the project lifecycle should be avoided. EU funded projects should be obliged to make the data produced in their projects available for reuse. Interoperability should be a "must have" in a project.

All the language data related legal issues can perhaps not be addressed in the context of the Digital Europe Programme but they will need to be addressed elsewhere. A strongly position paper is required with the backing of various interested parties, including the politicians. Incentives for **encouraging private sector and public sector to share their data** are essential. They could be monetary or in kind. Many national language institutes and other academic organisations possess vast amounts of manually annotated data that should not be forgotten in this context. Most of these data sets are locked away, access is strictly regulated, commercial use usually forbidden. As regards the public administration scenario, the use of CAT tools in public administrations should imply the requirement to share translation memories. Furthermore, the participants supported the revision of **the copyright directive, allowing language data to be more easily extracted and exploited for commercial purposes**.

3.2. Reaching a critical mass

The discussion focussed on large integration projects and industrialisation in order to better place European LSPs and SMEs in the market.

3.2.1. GDPR compliance - Anonymisation

The global AI market is largely unregulated, however in Europe GDPR is a huge issue holding us back compared to US and China. Current processes in the LT industry are often not GDPR compliant. For example, an insurance company outsources translations to a language service provider (LSP) which then keeps private information in their translation memories (TMs). This constitutes a continuous breach of GDPR.

Anonymisation would solve the GDPR problem in Europe if implemented as a process (e.g. in a data pipeline) not as a one-off. This problem is work in progress and has not been solved.

There is a lack of guidance on how to properly anonymise language data so that the whole translation pipeline is considered GDPR compliant. There needs to be a legal validation system.

At the same time, lack of data sharing due to privacy concerns is a problem to improving language technology quality for European languages.

3.2.2. Summary of proposed actions on EU level

The following actions/support by the European Commission could make a difference:

Development and industrialisation of secure/anonymised translation pipelines, infrastructure and workflows for text which cannot be anonymised or until the anonymization issue can be solved. By building GDPR compliant translation tools, anonymization solutions and pipelines, Europe can assert itself as the global leader in this area and provide solutions to other regions (e.g. California). Building a certified translation/language workflow pipeline should be possible with adequate funding and support within the timeframe of DEP.

Legal validation of translation tools and workflows as "GDPR compliant", for example through certification by an EU body or an ISO type standard to reassure buyers of LT services that they are paying for a GDPR compliant service. An open question of legal nature concerns the anonymization percentage which would be enough to consider a tool GDPR compliant, especially if you can prove you have done all that is possible to anonymise.

3.2.3. EU as a large scale procurer of language technologies / reference implementation/quality information

The EU is a massive buyer of translation services. The translations are done by in-house staff as well as though outsourcing and procuring of LT tools and services. The EU could use this power to help the European LT industry. DGT is the largest "translation agency" in the world, but it does not procure any ground-breaking LT services. EU does not procure for the future (build me something that will be state of the art in 2030, e.g. like DARPA), but only buys existing "old school" technology. For example, since 25 years the EU is asking for the same thing (CAT tool). Instead of driving innovation, it is asking for

the same thing again and again. In Dubai, the government is procuring things that will be feasible in 10 years' time, not what existed 25 years ago.

EU could establish best practices and standards through procurement. If it comes from EU it can be considered trustworthy and compliant. The EU could develop an EU/ISO "standards" certified reference implementation that can be used by others. The LT industry could also re-implement it to provide a higher quality competing service based on a standardised technology.

3.2.4. Summary of proposed actions on EU level

The following actions/support by the European Commission could make a difference:

EU institutions should make a **call for procurement to "provide an industrialised translation and anonymization pipeline that assures GDPR compliance"**. The EU could also take a risk where the EU market is not willing to. The EU could procure "deep language understanding" technology which will be state of the art in 10 years' time.

EU could **provide reference quality information**. For example, a list of all first names in Europe would be interesting. But you need to know it is complete – ALL the first names across Europe. Providing reference quality info is something only a public body can do.

3.2.5. Supporting LT SMEs through moonshot projects

LT SMEs struggle to industrialise due to the fragmented market. SMEs can step in to solve market failure in terms of providing solutions for small languages but they need support. Expanding their solutions to other languages and/or domains is difficult.

Another challenge is uptake from large industries. SMEs struggle to get work from large industries or public administrations who prefer to deal with more established/larger players. Additionally, it doesn't help that SMEs are often prevented by non-disclosure agreements to advertise their existing client base to others potential clients.

Finally, it is costly to develop and maintain deep learning technologies. Development and improvement of deep learning technologies are inefficient. Currently, companies/organisations receive grants, build up technology but cannot grow/scale-up beyond their local markets/languages. First a company needs quality data to develop hand-crafted language model. A lot of money is invested in preparing the same data set. Companies invest a lot of money in exactly same work – preparing data sets. They struggle to keep the technology up-to-date with the latest developments. For example, they have semi-products that can be used in chat box or language models that need adapting based on recent advances in research. For example, the BERT model needs to be updated regularly. They need to find partners and agree on how to maintain these semi-products/models.

The participants considered that the inefficiency of training the deep learning algorithms must be addressed. European companies are ready to cooperate with other European companies to share effort in developing and maintaining the technology, but if they cannot find European partners, they will go global with non-European partners, as well as open-source the technology.

In this context the "high-efficiency" approach to language technology was discussed. The Chinese and US giants took the "brute force" approach to language technology (i.e. if you want to improve the system, you throw more data at it). It does not make sense to fight with the brute force approach. In future, we will not be able to sustain building language

models based on the brute force approach. “High-efficiency AI initiative” actions would be for example to create open-source reference language model, or translate Wikipedia in all EU languages and provide this data to the community to work on. This data reflects our reality now and there is no privacy/GDPR concern.

The EU should invest in datasets (not in the algorithms) in order to make available to the community highest amounts of data. Data should not be a decisive element in how good the technology works. If data is available, the differentiation will be the quality of the algorithm not the amount of data you have at your disposal. Also, the SMEs/organisations would not have to invest each time in data gathering, annotation and preparation.

3.2.6. Summary of proposed action(s) at EU-level

Building critical mass in AI is not just a matter of finance, but also a matter of access to high quality data and resources. A **flagship project** could be envisaged (this is beyond DEP and would need the support and leadership of several or all Member States outside of EU funding programs). EU could coordinate actions across its Member States and pool resources from Member States to have bigger impact, if they are willing to act together.

A **flagship project with an emotional traction** would be needed. People would organise themselves in response to such a project. For example, there could be a community effort to collaboratively create a generic data set or a certified reference dataset (e.g. have all versions of Wikipedia articles translated in all European languages and annotated). The EU could contribute for example by providing a number of translation staff contributing a certain percentage of their working time to this.

To share the effort in data gathering, an idea for Amazon Turk for SME was also mentioned: “We are planning to have this effort on data, who wants to join in and share the effort?”

EU could call for flagship moonshot projects, e.g. EU could propose and co-fund a flagship multi-billion euro moonshot project that bootstraps the European LT industry and makes it a market leader in AI and machine learning (like putting a man on the moon “at any cost”). For example, in Europe the problem is that we do not really have a joint public sphere. There have been some pushes in the past from some Member States. The channel ARTE, a joint German-French television station is an example for this. EU could call for a flagship moonshot project to create a European Public sphere in the digital age. Such a European Digital Public Sphere, where EU citizens can come to discuss their matters, would have to be multilingual. People would discuss in their language with live translation for everyone to understand. This would bring real benefit to the European Project.

3.3. Adaptation of AI technologies to specific languages and domains

3.3.1. Challenges and needs

There are 24 EU official languages and solutions need to be adapted for all the languages. Big companies hold back from small countries because they represent smaller markets. For the smaller markets there is often no RoI. It is a big investment to adapt a technology to a language. Consequently there is a discrepancy between the tools that we can provide for each language. They vary greatly from language to language. For example in case of subtitling: if there is no demand (measured, for example, by the number of TV stations in a given country that could be interested in buying the subtitling solution) then it is not worth it for companies to invest.

Furthermore, it is a challenge to adapt technologies for local languages because data and knowledge resources are needed. Collecting data can be expensive compared to the size of the market. Also knowledge resources are not always existing. Data requires a lot of work (cleaning, annotating, etc.) to get knowledge from it. Matching cross-border systems is not a translation problem, but a knowledge problem. With knowledge resources, you can then develop products on top of it.

In order to deal with multiple languages, companies either find multiple partners or they build the corpora themselves (training material). Building data is costly, involves discussions with linguists etc. Some limitations are inherent by the size of the speaking community. There are datasets scattered across all Member State and there are different platforms for each. Furthermore, there is a lack of any evaluation framework.

Finally there is a challenge in reaching the non-LT SMEs and communication: people/SMEs are afraid of using AI. They need to be convinced. Often travel is necessary. As long as that happens within one country, it is OK but scaling to another country requires more contacts and a network. The challenge is also the cultural difference. One way to succeed in a different country is to find a partner in the other country.

Amazon services are useful for LT SMEs and integrators who otherwise wouldn't be able to cover a small market such as the Hungarian language, for example.

3.3.2. Summary of proposed action(s) at EU-level

An **SME-friendly funding scheme** is needed to support adaptation of AI-based LT to other languages and domains. Currently it is really difficult for an SME to fill in a whole proposal for funding. Something like a voucher system or funding for the buyer side, would help the situation. Grants for funding would help SMEs also to go for things for which they have no incentive to go for, like social rather than market necessities.

Adaptation of AI-based technologies to other languages and domains can be facilitated by **providing baseline solutions, APIs, data and standards**. To address the market gap, i.e. the difficulties faced by SMEs to reach large clients and the public sector, **a community** should be fostered **to help LT providers find each other, share solutions and develop joint offerings**.

As regards data and tools measures should be provided that add value and make it interesting to share. No one wants to share for free. Specifying **a space that is dedicated for commercial data use** would also be needed, so that LT solution developers do not need to determine for every dataset what the legal implications are for using it and in which situations it can be used. A platform that shows which public data is available would be helpful. It would save a lot of time. Creating strong, reliable data pipelines would help localise technologies to different languages and domains.

High-performance computing (HPC) needs to be made easily available to train data intensive models.

EC should also provide an umbrella vision and long term goals (a roadmap).

4. DEPLOYMENT OF AI-BASED LANGUAGE TECHNOLOGIES FOR EUROPEAN SMES

4.1. Understanding/Addressing the Market

It is difficult to talk about the language technology market in Europe. Because of the diversity of languages, there are many markets. Some European LT companies have the capacity to work world-wide but not in Europe. As regards the fragmented Language Technology scene in Europe, there are indeed language silos, cultural fragmentation and niche content markets, however there is also European industry which operates globally, e.g. SDL, Bertin IT or Systran. Some companies have large presence in US but find it difficult to work in Europe.

EU institutions should help create the language technology market in Europe by creating the necessary research and empowering companies to work on this fragmented market. For example, Artificial Solutions offers a conversational AI platform which supports 36 languages of which 34 are European. Artificial Solutions has proprietary NLU, NLP pipelines developed internally. If they could do this effectively for the smaller languages then it would be easier for them to convince their clients to cover also these languages.

Market size. It is difficult to size the LT industry thought there are some recent studies that offer partial answers. There is hardly any proper data. LT-Innovate estimates the Language Technology market at 10 bn EUR. The market of Language Intelligence services (LSP) sitting on top of LT is larger and estimated at 50 bn EUR. The Language Intelligence Application domain is estimated at 200 bn EUR. There are big opportunities, however we need to understand the markets which we are addressing better.

There are no reliable figures available on the different markets that LTs address. Large US companies do market research for their clients. Smaller companies do not have the means to pay for market research. Market research is about finding new opportunities and determining the needs in the market.

Market research is key. We shall continue investing in the technology development, even though China and the US are investing much more, but there has to be a clear idea of the markets we are addressing. There is a need for thorough market research. Integrators who integrate all the different elements offered by the "vendors/providers", for the end-users/public are essential. A clear mapping is needed saying who are the integrators? Who are the users? A systematic analysis is needed.

Funding a market study would help a lot smaller SMEs to understand the needs, the gaps and the players in the market. It would also be of interest to the Commission to know the market, in order to strategically navigate the waters in relation to the US and China and to answer the question on the strategic focus of the activities. Finally, there needs to be more visibility about which value the EC support will create for each segment/market.

***Generic technologies* vs. *specialisation*:** Big players like Google or IBM have generic models, biased towards B2C or search. With such models it is difficult to address a vertical market. EU is a strong player in niche markets. In general, people will not rely on Google for certain combinations of language domains and sectors. However a recent trend shows that these big players are addressing vertical markets, such as the media market (i.e. Google), trying to sell to this industry the ability to transcribe, analyse, etc. This is something new. They are also addressing the financial market. They want to equip the contact centres, the agents. Google and IBM are also addressing the defence and security market. This should not be happening in Europe.

Prototype projects can use Google, but if they want to do something specialised they need to adapt the model and need to have it locally, so Google is no good. They need to modify the data. For example, chat and voice bots start first with Google, and later once they have the customers (and they know their markets) they move to European companies because they want to be GDPR compliant, have a domain-adapted technology etc.

Finally, there is a **difference between** "***technology***" and "***applications***" created based on the technology. *Technology* becomes successful, when *applications* of it are created which are then widely used. The idea is NOT the alpha and omega. It is the SOLUTION responding to a specific need that makes the difference. Google has given some of the most fantastic solutions to us all.

Pushing technologies does not work. The markets are looking for solutions to specific problems – these are the opportunities for the European LT industry. The new programme should focus on solutions to specific problems. We need to know our customers. Everything needs to start from this. "Technologies" need to be productised – they need to be of European and global scale.

At the same time, while everyone is investing in technology in terms of commercial applications, the technology for real social needs is not so easy to be funded. The EU and governments should support technologies which alone cannot survive, also for where there is a social need. Social applications of language technologies may not be not be profitable and so they can only be built if (continuously) subsidised.

4.1.1. Summary of proposed action(s) at EU-level

Funding of a **thorough market study** of all the verticals and with good publicity. The market analysis could be carried out by an innovation programme. A Coordination and Support action funded by the EC could do market research and make the results available to everyone. Another possibility could be to start small, i.e. take a smaller set of companies, asking them what their key markets are. Then others would be interested in joining and the study could be expanded. Establishing a National Committee to report on the LT market to the Commission could also be helpful in getting a lot of information upstream. There could also be a committee of people at the national level (one per country) that examines the LT market and then reports to the Commission.

A **certification of "language technology compliance"**. Certifying that a company is "language access compliant". This is something that goes beyond the DEP programme, and DG CNECT. A certification body would be needed.

EU should **facilitate the collaboration/connectedness of European LT SMEs**. The EC should create an interface (forum) between SMEs and LT, where they can meet and cooperate.

Finally, Europe needs to develop a **strategy**. Europe must be clear on what we want to achieve. A strategy requires clear ex ante definition of targets, capabilities and building blocks. We need to put our resources where we are strong and can win. The EU has huge data resources. They should be released.

4.2. Financing of LT deployment in European SMEs

A **support programme for SMEs** who want to use and implement language technologies is needed. It should be **lean, efficient and market-driven**. SMEs shouldn't be pushed into 5-7 years plans.

The SME portfolio model implemented in Flanders whereby an SME who wants to expand gives 40% topped up by 60% from which international expansion costs may be funded on an ongoing basis could be a model for EC. A Dutch funding scheme where the SMEs can directly apply through a dedicated webpage for microfinancing and with e.g. 60 % of the costs funded was also mentioned, along with Spanish R&D tax incentives and training funding. The example of Enterprise Ireland could also be followed: Invest and then provide capital for SMEs, let them test their technologies. Micro-funding was mentioned with the objective of reaching out to a very high number of LT and non-LT SMEs. A specific LT funding webpage could be foreseen where SMEs in need can apply based on their requirements.

The access to financing should be easy for SMEs. Administrative burden was identified as a challenge to overcome for LT SMEs. In their opinion, European funding only helps them if it is easy to apply for.

EC-funded actions could support vertical ecosystems. **Targets** in those domains/ecosystems must be defined, as well as the **leaders** in each ecosystem. Companies must be put in the driving seat of innovation actions. EC support should only be granted to projects that measure usage.

In the participants' opinion grants must be allocated to the **best European solutions**. Investment must be in successful solutions. Financing of LT SMEs could be linked to the number of non-LT SMEs having adopted LT solutions. E.g.: if an LT SME have deployed solutions to XX non-LT SMEs, they should get XXXX euro funding.

Projects with **clear European added-value** should be funded. The Member States spend money on national languages, so the EU should go beyond that approach. In the financing there must be an EU-level added-value, otherwise it is only a redistribution of money to Member States.

There was some discussion on whether Europe should strive to establish (a) big player(s) in Europe or support an ecosystem of **many small LT SMEs**. A range of aspects and challenges has to be considered when promoting a large number of small players e.g. management of huge data by small players, deployment costs, promotion and visibility challenges, productising and commercialising products, etc.

A general agreement seemed to emerge that in Europe the EU should **let the "*1000 flowers blossom*"**, i.e. support the many small SMEs in developing successful products. By doing so we create value for the citizens because many small companies which work in different markets will create better solutions than the big companies create for the big market. EU needs to support/plant "small seeds" and markets will find their ways. We should become market driven.

To accomplish this, Digital Europe Program could include funding to support European LT start-ups. The funding could be divided among key areas such as:

a) promoting European LT companies and increasing their attractiveness to consumers and investors;

b) direct micro-grants for start-ups that increase EU competitiveness;

c) deploying LT accelerators;

d) funds to incentivize venture capital and private equity groups to engage more with the European start-ups, and to create 50/50% co-investment opportunities.

Organising competitions/prize awards among LT SMEs delivering the best solution and winning a prize could also contribute to the rise of the whole LT industry. This would work as a high-level promotion and incentive. This way the EC could create a meeting platform to exchange ideas which could then also be deployed in other EU regions after adaptation.

A kind of financing resembling **bank loans could be considered**. The advantage of this would be that there is no need for project evaluations and reviews, as the funding must be paid back.

The participants insisted that there is **no need to bring in a middle-man or anchor point** into the funding scheme between LT SMEs and SMEs in need.

4.3. European SMEs as a market

Focussing on non-LT SMEs as deployment targets was questioned. The political goal of deploying LT by European SMEs was discussed. From the perspective of the LT providers, SMEs as a target group may not be the smartest move. According to some participants, concentrating too much on the SMEs aspect may not be the optimal approach to raise the competitiveness of the European LT industry.

4.4. Reaching-out to end-users

In order to facilitate reaching out to and convincing customers, **easier try-out of technologies** must be addressed and information on success stories must be provided.

For example, the adoption of conversational solutions is poor due to lack of awareness. Client SMEs need to be provided with information on what has been deployed where and with what results. Making it possible for the customers to easily try out and compare technologies (for example conversational AI) and services of different providers would be helpful in convincing them to deploy these technologies. Companies don't understand what can be done with conversational AI and what ROI they can expect. They need to try to use multiple proof-of-concept solutions (PoCs) with different vendors in order to make the right business choice. Supporting pilot and short trials would be helpful so that companies understand what can be done.

4.5. European Commission's role as market creator/ creating a market

The European Commission can intervene through policies and regulatory measures.

The EU can create legal regulations to ensure language equality across the EU. This would support LT work at a local level by smaller players, as these markets are often ignored by the big players. Participants clearly expressed that the EU should become a market maker in LT: they want the EU to become a strategy maker.

The Commission can **create markets through legislative measures**. Accessibility is a great example. The Commission created a market where there was none. But at the same time, if you open such a market you risk to benefit Google, because it is the only one that can meet such standards.

Legislative measures for inclusiveness and accessibility could be extended, for example by setting **a requirement to provide public online services in all EU languages**, **to provide subtitles for all TV broadcasts, or enable text-to-speech for people with disabilities**. EU level regulation could contribute to the rise of the whole LT industry: e.g. parking rules should be accessible in all EU language everywhere through a QR code. This would then trigger a huge need in AT/LT. The EU could recommend that any big company providing digital services should do this in all EU official languages. Some participants considered that the EU needs to create legislation for digital anti-trust to allow for competition.

Finally, the EU could create policies that ensure SMEs have a chance to bid for certain work.

5. MOBILISING AND SUPPORT MEASURES FOR LANGUAGE TECHNOLOGIES COMMUNITY

Questions which were discussed concerned infrastructural needs of the LT community to efficiently develop and provide LT services to European SMEs. In this context a platform approach as well as further development of the European Language Grid (ELG) project were discussed.

In the context of developing and deploying language technologies a broader definition of ***infrastructure*** was adopted at the workshop, encompassing not only the physical infrastructure (computing facilities, data centre) but also language data and software and services, which are as important.

An important issue from the LT provider perspective was visibility and the challenges related to reaching the market. This included the challenge to form partnerships and find and engage beta-testers & customers. An important challenge is also quality assurance for solutions that are meant for commercial use.

5.1. ELG project and platform approach

The European Language Grid, being developed in the ongoing H2020 project running until December 2021, is currently conceived as a market place between "providers" and "solution builders and integrators". A stated ambition of the European Language Grid is to become the primary platform and market place to address market fragmentation in Europe, a European Amazon Web Services (AWS) optimized for European languages, providing easy plug-n-play solutions.

A platform-approach to addressing challenges faced by LT SMEs was acknowledged and the European Language Grid initiative received a lot of interest from participants. All participants agreed that an API-infrastructure/platform with European endpoints that can be used in any language would benefit the European LT ecosystem, as API culture helps to buy services at a low cost, on a usage/needs basis and existing platforms do not sufficiently address European languages. For example, Google tools use English/US named entity recognition (NER) systems, so they don't work for smaller regions.

The discussion concerned the issues that need to be considered and challenges to be addressed for ELG to become a one-stop shop for European LT.

It was agreed that ELG is well positioned to become a "European highway for language technologies" providing easy access to language resources, tools and services for European

languages. However it must be scaled up and integrated with other infrastructures. There is a recognised need and ongoing discussions to ensure the collaboration and compatibility with the AI4EU platform and other AI platforms. **The Relationship with other/big LT/AI providers/initiatives/platforms** (European and non-European) must be considered and back-end compatibility must be ensured.

An important aspect is **connecting to the users.** The users of ELG currently targeted are the providers and "LT integrators and solution builders". It is not yet fully clear whether and how other users (end-users) will be reached. It is an open question whether and how ELG can serve other types of stakeholders or "end-users" in a specific market, for example LSPs, media, car industry etc. If other types of users are to be targeted the strategies need to be developed, visibility aspects must be considered and positioning vis-à-vis other sectorial platforms must be clarified (See also the section "platform approach from the market perspective" below.)

In any case, **a flexible system** will need to be considered with plug-ins for some services. For other services integrators will be needed. Some tools, e.g. a terminology extractor could be useful as a stand-alone tool for any end-user. If you need to build a workflow, then there needs to be an "integrator" in-between. For specific vertical segments, e.g. the car industry, the ELG approach may not be specific enough. Something will be needed in-between, which makes specialised models/builds workflows. An LSP could be interested to use for example an MT tool (e.g. eTranslation) integrated through API plus something that could be used to personalise eTranslation.

For outreach and visibility reasons it was considered necessary for ELG to provide some **visible solutions**, e.g. plug-n-play solution for website translation allowing for a flexible choice of MT providers or a spell-checker could be something for everyone to use.

As regards **data**, the current focus of the ELG project is on free data only. Datasets that are not freely available will be made available at a later stage. One participant mentioned the need for the platform to support live data in view of the current trends on the market. The need to process live data versus processing cold/daily data needs to be considered from the architectural point of view.

The discussions on a sustainable operational model for ELG are still ongoing. Currently a **non-profit company/association** is considered.

Communication: communication to different stakeholders about the use they can make of the platform and the benefits it can bring to them is key to success. For example, for big players ELG may be a new channel for promoting their solutions and enabling rapid partner integration.

Governance: by putting the right framework in place conflicts can be avoided. Long-term governance has to be put in place. Community governance should include other "big players". It could have other (big) platforms and clients acting as an advisory board.

Branding: Funding for marketing/visibility will be necessary.

5.1.1. Platform approach from a market perspective, advocated by LT-Innovate

The big picture of the language technology market and stakeholders is complex, where the definition and usage of terms "developer", "provider/supplier", "integrator", "LT users",

"end-user" may be used differently depending on the context and perspective. In future, language technologies will be used more, consequently there will be more providers and more middlemen. It is necessary to define the users of the different platforms and their needs.

Also the community creation and maintenance effort must take into consideration that there is currently not a clear *LT community* in Europe because it is already difficult to explain what a language technology company is per se. For example, Google could be defined as such but it is not.

It is not easy to apply generic language technologies across the board. We need to get to know the specifics of our customers' sectors. American technologies are responding to specific business processes. In Europe, we need technology platforms responding to business processes of a specific client segment.

Consequently, a platform approach for the language technologies could consist of:

1) A publicly maintained infrastructure (meta-platform) with a simple governance model, which feeds/supports the commercial platforms with basic language resources. This piece of infrastructure, "language super highways", should not be in hand of private companies. The utility/success of the meta-platform would be tested by the commercial platforms.

2) Commercial technology platforms, owned by companies. EC support for platforms should be aimed to make these (privately-owned) platforms more successful. EC could fund an innovation project to boost a platform to do this and that.

The EC-funded initiatives must not overlap with private-sector initiatives that the private sector is making with their investment.

Alignment and interoperability with sectorial AI-platforms and the sectorial data spaces proposed in the DEP programme must be considered. The conclusion of the discussion with regard to the European Language Grid was that ELG should not replace the other current or future platforms in various domains. Rather, the ELG has to be compatible with those platforms. ELG has to feed value into (other) commercial platforms and add-value for the bigger players.

5.1.2. Platform from the provider perspective

Small companies providing LT technologies/services see the platform approach as beneficial because it gives them visibility. Bigger players who have already invested in their own platforms need to be appropriately considered and involved. For bigger players, ELG could be useful as the meta-platform for some services.

Local LT provider SMEs often want to form a consortium to develop bigger/new systems and they look for providers of different components, addressing specific languages, e.g. speech technology companies. The SMEs would like to check what they provide, what is the quality, pricing conditions etc. Therefore easy access to other providers, common billing models would be helpful.

Providers include universities who have different tools (e.g. summarisation) but don't know how to put them to use. A system is needed that can easily integrate different technologies of different providers. Providers don't want to develop for one user but for many users. The system must be flexible.

An infrastructure is needed where SMEs can deploy the results of the research. This will help them go to the market.

5.1.3. Platform approach from the solution builder/"integrator" perspective

"Users of LT technologies" and integrators who implement LT technologies in their IT software and services need:

(1) **Algorithms**. It was acknowledged that EC must continue supporting research allowing European companies to be involved in and stay informed about the newest research. This is necessary to stay competitive.

(2**) Quality data**

(3) **Infrastructure**. This helps companies to lower costs and increase efficiency. Cheap and energy-efficient data centres and computing infrastructure optimised for language technology is needed. Currently, no-one in Europe can compete with AWS. Small LT SMEs need HPC infrastructure to train language models with large datasets.

The question to be considered for future EC activities in regard to a platform include: should compute infrastructure/data centre/Cloud optimized for LT be part of ELG?

(4) **Flexibility/workflow development**. A media company representative present at the workshop noted the importance of being able to use any technology, plug-and-play it into the company workflows. Therefore a platform is needed that gives you a choice of any language and technology combination. Clients do not want to be on an island, bound to only one company.

There is a difference between all-in-one or black box solutions and solutions whereby customers buy some solutions from one provider, some from others, some they develop themselves. Therefore the possibility to develop "chains" where various things can be combined is attractive.

(5) **Ease of use/try-out and comparison**: From a user (e.g. LSP) point of view, trying out different services available through the Grid must be possible with minimum investment.

5.1.1. Cloud and High-Performance Computing

For many providers of LT technology the preferred way is to deploy it on the cloud (SaaS model). Here the biggest obstacle is regulation which prevents many clients (e.g. financial institutions to use services in the cloud. On-boarding of such clients (i.e. providing an on-premise installation) takes 12 months which is too long for an SME. It is easier to provide services to American clients who do not have regulatory constraints where the service must be deployed. European financial institutions often use "regulation" as a reason not to use the services of an LT company.

In Europe, there is a need for Cloud facilities which are provider agnostic. SMEs in Europe do not need to choose between the clouds, often they do not need to know where the computation is made, as long as some basic pre-conditions are fulfilled, e.g. data stays in Europe.

Language services providers would need cloud services that they can use and reassure their clients that their data is safe. EC should foresee cloud-related measures that help LSPs reduce costs and make their services available to SMEs.

From the LT SME supplier/provider perspective who participated in acceleration programmes in the US with perks of 100.000 dollars to spend on AWS, facilitated access to cloud is something EU could consider. All cloud suppliers have start-up acceleration programmes. However the qualification for the perks and programme depends on the provider (AWS, Google) who have their own criteria. The EU could play an equivalent role in Europe.

High-performance computing (HPC) is needed to train data intensive models.

5.2. Standardisation

The discussion on standardisation addressed questions such as whether *standardisation* efforts in LT in Europe are necessary, which (open) standards would be needed and how they should be developed.

Creation and promotion of standards was considered important. Users and integrators want to develop against one API and not deal with multiple APIs of various providers. It was agreed that while some sectors, e.g. telecoms have strong incentives to get together and talk, the LT community has less incentives and needs to define standards.

Standards concerning data exchange, language grid interoperability, generic language technology services, speech, chat box and dokerization are needed. Interfaces could be taken out of ELG and put to W3C for discussion. Standardised plug-ins are needed. Standardisation on the data could also make sense, but (semi-)standards need cultivating. Standards for procurement of virtual assistants in the public administration scenario would be helpful to increase adoption of this technology, in particular setting KPIs to measure virtual assistant quality.

Community groups within W3C could be used, resulting in "community group documents" which - if there is traction of big players (Google, Sisco etc.) - could make it into a "standard".

It was agreed that a platform where you can demo services would also drive standardisation, i.e. if there is one API for each type of NLP processing, similar to an existing US platform Inten.to[1].

Inten.to - a US company provides general APIs which allow others to adapt them to their more refined capabilities. It helps integrate components and offers users the possibility to compare services of different providers.

The standardisation and interoperability issues could be addressed by a coordinating and support action (CSA, coordinating the actions between Member States and industry).

5.3. Role of and collaboration with Member States

Questions which were discussed included the necessity and modalities of efficient and effective EC-Member State collaboration. It was discussed that while Member States are mainly interested in protecting their own national language, the EU should take the role of protecting *all* EU languages.

[1] https://inten.to/

In some smaller countries, e.g. in Slovenia, language is a *matter of national sovereignty.* There is discussion in some countries whether access to the different language technologies should be considered as part of the *basic human rights*. Language technology services and data are being funded and made available through national initiatives. Slovenia, Denmark and Estonia are investing in LT. Iceland funded LT with 15 million euros, for Icelandic especially. In many cases language technologies are supported from the AI strategies in the Member States. This is the case in Belgium, Estonia, France, Germany and Malta.

Making national services and data available also through a European grid would ease development of LT tools/services covering those languages by SMEs in Europe.

There seemed to be an agreement among the participants that a collaboration/forum/exchange of best practices at European level is needed, similar to or based on the European Language Resource Coordination (ELRC) network or the Language Technology Council (LTC) created in the current CEF and Horizon 2020 programmes. Guidelines and documents from the European Commission and European initiatives help convince and direct national decision-makers. Possibly a dynamic barometer which assesses the LT readiness in Member States could be envisaged to challenge and show to the authorities how they are doing in comparison with other countries and what needs to be done in order not to fall behind. Such dynamic barometers and "charting the national scenes" would be aligned with the European Parliament's request to provide "basic" information on all languages. Another aspect is that Member States need to know what is done at the European level in terms standards.

6. OTHER ISSUES DISCUSSED

6.1. Further development of eTranslation

An API for eTranslation/transforming it into a plug-in for different environments should be made available. It must be possible to integrate eTranslation into translators' workflows. The Commission could also make the eTranslation translation models available. By doing so, it would help the market in Europe.

eTranslation should be made available to the private sector. In particular, the developed eTranslation plug-in could be opened up for use by LSPs. LSPs and private companies could have access to the eTranslation API which they could integrate into their own workflows and return the data to the EC for uploading to ELRC or ELG.

Finally it was considered that industry should be involved in further development of the eTranslation platform. An example of HUGO.LV, a public platform with services in Latvia which is targeted towards the public sector but also helps SMEs, was given.

Based on the experience of the CEF Automated Translation platform development, future EC-funded activities should leave room for more dialogue between the internal EC development team (DGT) and the external EC-funded projects. The sole customer of many activities under the CEF generic services projects was DGT. There is mutual benefit if the project and the actual output that the project will deliver can be fine-tuned. There is room for doing this.

6.2. Language technologies and accessibility

Some participants mentioned that EU should not forget about non-spoken languages such as sign languages or braille, noting that these also vary across the EU. Language technologies for the disabled should always be included when talking about language technologies in Europe. They should be included within all LT actions.

7. ANNEX 1 – AGENDA

Stakeholder Consultation on Language Technologies in Digital Europe

10 February 2020 from 9:00 to 18:00
11 February 2020 from 9:00 to 13:00
Avenue de Beaulieu, 25, Auderghem, Brussels, Belgium

BU25 S1

10 February

08:45-09:00	Registration
09:00-9:45	**Welcome & setting the scene – BU25 S1** ➢ Introduction, context & flow of the day (Philippe GELIN (EC))
9:45-10:30	**State of Play** ➢ Existing European endeavours (contributions by LT-Innovate and H2020 ELG project, followed by discussion)
10:30-11:00	**Coffee break**
11:00-13:00	**Collection of ideas/contributions** ➢ Elevator pitch relating to one or more of the following issues: o Language Technology and AI, o exploiting language data in EU, o financing of LT deployment in European SMEs, o infrastructure requirements, o interoperability or standardisation, o shared MS/EC responsibility for LTs for European languages; *(maximum 1 slide & 2 minutes per participant)*
13:00-14:00	**Lunch break**
14:00-14:30	**Presentation, discussion and decision on the break-out sessions**
14:30-16:00	**Break-out sessions I** ➢ break-out groups to discuss & shape one thematic session
16:00-16:30	**Coffee break**
16:30-18:00	**Break-out sessions II** ➢ break-out groups to continue discussing & shaping another thematic session

11 February

09:00 -09:15	**Setting the scene**
9:15-10:45	**Break-out sessions III** ➢ break-out group to continue discussing & shaping a third thematic session
10:45-11:00	**Coffee break**
11:00-12:30	**Presentation of outcomes of the different thematic sessions & foreseen actions** ➢ 15 minutes per group
12:30-13:00	**Conclusions**

8. ANNEX 2: PARTICIPANT LIST

First name	Last name	
Antonella	Abbati	interlanguage
Daniel	Aioanei	Artificial Solutions
Carlos	Amaral	Priberam
Laurențiu	Constantin	NEW COMPASS SERVICES SRL
Adam	Dancewicz	Applica AI
Francisco	De Sousa Webber	cortical.io
Arantza	Del Pozo	Vicomtech
Olivier	Deroo	Acapela group
Konstantin	Dranch	self-employed
Manuel	Herranz	Pangeanic
Theo	Hoffenberg	Reverso-Softissimo
Jochen	Hummel	Esteam
Jurgita	Jaroslavienė	Institute of the Lithuanian Language
Jacek	Kawalec	Voicelab
Simon	Krek	Artificial Intelligence Laboratory, Jozef Stefan Institute
Krister	Lindén	University of Helsinki
Martina	Ljubičić	Ciklopea
Luc	Meertens	Crosslang
Fernando	Moreno-Torres	Berca Translator
Ariane	NABETH HALBER	Bertin IT
João	Neto	Voiceinteraction
Stelios	Piperidis	ILSP
Christoph	Prinz	Sail labs
Georg	Rehm	DFKI
Michael	Rosner	University of Malta
Magnus	Sahlgren	RISE Research Institutes of Sweden
Miloslav	Skovajsa	exe
Elar	Tammeraja	wize ai
Marco	Trombetti	Translated
Christos	Tsalidis	Neurolingo
Georg	Tschare	Sign Time GmbH
Indrek	Vainu	AlphaBlues
Tamás	Váradi	Hungarian Academy of Sciences
Andrejs	Vasiljevs	Tilde
Maarten	Verwaest	Limecraft
Philippe	Wacker	LT-Innovate
Bartosz	Ziółko	Techmo
Diego	Bartolome Calvo (*)	TransPerfect
David	Canek (*)	Memsource
Josef	van Genabith (*)	DFKI

(*) only written contribution

9. ANNEX 3: PARTICIPANTS' INDIVIDUAL CONTRIBUTIONS (SEE SEPARATE FILE)