LT-INNOVATE.EU

THE FORUM FOR EUROPE'S LANGUAGE TECHNOLOGY INDUSTRY

Language Standards: can Europe take the lead?

LT-Innovate Standards Survey Report

March 2014

Contributors: Bradley Harrad (LTC), Kara Warburton (Termologic, Chair of ISO TC 37), Serge Gladkoff (Logrus International), Andrzej Zydron (XTM International).

Executive Summary

During the course of January 2014, a research survey was conducted to investigate the industry understanding and usefulness of standards. This report delves into unifying efforts for open standards and technology in the language industry in light of the H2020 Program. Observations and some conclusions are made from 162 survey responses to describe the current status, priorities and the potential role of LT-Innovate.

The report analyses responses from both users and developers of language technology. The principal results suggest that many companies work on similar categories of software. Unified action is considered a requirement and useful to bring together the variety and multitude of language-related standards for the future of a multilingual Europe without language barriers. LT-Innovate would like to further develop methods to harmonise the standards arena in Europe.

Supportive efforts in the form of education, clarification and consulting work are seen as beneficial for the bodies and stakeholders that influence standards. More actions that will provide recommendation towards a preferred set of standards are still required to promote the advantageous use of open standards. The potential is increasing for a third party organisation, such as LT-Innovate, to lead the collective action towards a simplified language standards landscape in Europe. The broadcast and efficient coordination of standards will support the use and development of technology and mobilise the industry.

Main Conclusions

The survey responses (see below) have led us to make the following conclusions:

Facilitating development work; role of open standards in R&D curve

There is significant duplication of effort within the language industry, which has limited development resources. Standardisation of data, tools, content and data formats, as well as process, workflow and business practices will help to:

1. reduce duplication of effort, making resources available on new aims rather than on reinventing the wheel;
2. simplifying interoperability of various tools;
3. enabling developers to focus on innovative solutions;
4. will improve the situation of tools adoption;
5. reduce interoperability and cost barriers for both users and developers of language technology.

Open standards are already somewhat popular, but we need to push for more

The research revealed that if TMX and XLIFF are well established standards already, some important standards such as SRX (segmentation rules) are not well broadcast and as a result may be considered misunderstood or underdeveloped. This is something to work on; and nothing will increase adoption of these standards better than making them standards practices and a foundation basis for European Horizon 2020 projects and Language Cloud.

Clearly, merely ***listing*** them as recommended standards for H2020 projects and the future European Language Cloud will significantly improve their penetration and the entire interoperability landscape. It is critically important to standardize the European language landscape so that the array of tools and standards and lack of interoperability will be less confusing and create less duplication of effort.

LT-Innovate is viewed in position not just participate, but lead in this work

The majority of respondents have responded “YES” to the question whether they agree that LT-Innovate is positioned best to carry out this work. This is especially encouraging since the majority of respondents were not LT-Innovate members. As compared to ISO, which is purely a standards body, LT-Innovate can serve as connector between the industry, standards world and the European Commission for H2020 program.

Possible action items

The need is here, and the practical sense is overwhelming.
By taking the role of “unifying body” and creating a list of recommended Open standards for Horizon 2020, LT-Innovate can help to significantly improve the interoperability and, by implication, the productivity of the European language industry.

H2020 needs a reference guide regarding recommended standards and their means of adoption. LT-Innovate via such initiatives as the European Language Cloud can provide this so that the H2020 funded projects, as well as the European language industry in general, can leave a lasting legacy that includes interoperability and allows for the maximum reuse of code and linguistic assets.

Contents

Executive Summary 1

Main Conclusions 2

About LT-Innovate 5

The context and the shape of this research 6

Scope 6

Survey audience and demographics 7

The observations from the survey responses 8

Fixing the problem – self-certification or external certification? 15

Fixing the problem – who will take the lead? 17

Annex: Survey questions 21

Introduction

Europe's competitiveness is evidently linked to its ability to overcome language barriers that impede cooperation and trade. It is important to remove language barriers; however, we need to overcome data barriers first. Different formats and confusion about standards, specifications, technical reports, processes and recommendations prevent interoperability. Various standards organizations such as ISO, ETSI LIS, OASIS and W3C, as well as independent project groups such as Linport for example, work on language standards under their own angle. The cooperation between various standards bodies and groups is insufficient. This is a problem in a world with an increasing and implicit dependence on standards, when it's getting more and more important to ensure progress in the reduction of technical barriers to leverage big data and cloud technologies.

The **Horizon 2020** Research and Innovation program provides funding to projects that further Europe's global competitiveness. Under Horizon 2020 the overall aim of EU research and innovation in ICT is to bring the benefits of progress in these technologies to European citizens and businesses. The goal is to ensure Europe produces world-class science, removes barriers to innovation and makes it easier for the public and private sectors to work together in delivering innovation. These measures will aim at breaking down barriers to create a genuine single market for knowledge, research and innovation.
In the same vein, the goal of the **European Language Cloud**, an LT-Innovate initiative (find out more about the European Language Cloud at www.lt-innovate.eu/elc), is to facilitate the deployment of multi-language and cross-language resources, services, and products supporting all European languages. The European Language Cloud should be based on open standards. Standards are required at various levels in the global hardware and software environment to ensure smooth connectivity and interoperability in a complex world of competing IT products.
Creating a unified resource for all European languages and making this data available to tool providers and researchers is beyond the capabilities of individual national organizations or companies. The analogy is with building and running road networks. The building and maintenance of road networks is undertaken by national governments and local authorities because it is typically beyond the scope of individual companies. Having an effective and efficient road network boosts the economy and brings benefits and prosperity to the state and its citizens. Similarly, in order to create one digital market it is imperative that the creation of the EU Language Cloud be undertaken by the EU and that access to the resultant resources be free for researchers and tool providers. It will result in a substantial improvement in both products and services.
Various projects that will be conceived and implemented on the path to Horizon 2020 goals will have to rely on common standards for data, processes and data formats in all their variety including terminology and adjacent technical specifications. A unified approach and recommended practices and formats will likely be required.

Exactly how ready is the industry for such an effort? In order to assess this LT-Innovate has carried out this research on language standards for Europe, and analysed the results.

About LT-Innovate

LT-Innovate is the Forum for Europe's Language Technology Industry. It gathers companies developing intelligent content, speech and translation technologies. Language Technologies are pervasive and are by nature key enabling technologies for numerous application areas in a multilingual world.

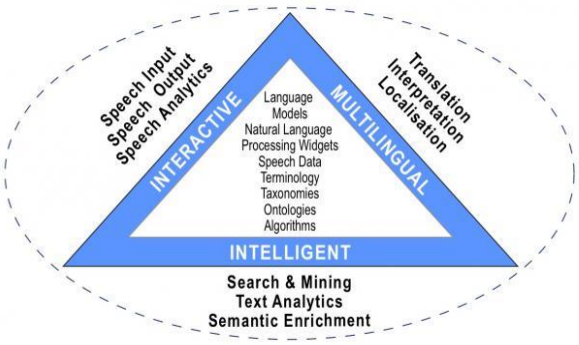


Fig.1 About LT-Innovate.

LT-Innovate's objectives are:

- to unify Europe's Language Technology industry for increased competitiveness in the global market place;
- to promote Language Technology as a key enabling technology for economic success and societal wellbeing;
- to articulate collective interests of LT stakeholders (buyers, researchers, investors and policy makers).

Visit the LT-Innovate website at www.lt-innovate.eu.

Find out more about the European Language Cloud at www.lt-innovate.eu/elc.

The context and the shape of this research

The research on standards for the future multilingual Europe, the survey and analytics on collected data aims to gain a better understanding of the current interest in and adherence of standards for language technologies, and the expected priorities that could shape the standards-related activities within the LT-Innovate forum.

LT-Innovate has conducted this initial research, beginning with a survey, to identify the potential issues, discover the variety of approaches, brainstorm potential solutions and develop methods to harmonize the standards arena for Europe. The questions in the survey are intended to collect information that can benefit many EU Horizon 2020 projects.

LT-Innovate maintains that a clear European strategy with respect to the use of language-related standards is necessary for the development of innovation to support the European economy. LT-Innovate is therefore investigating a potential leadership role in studying language standards needed to enable the European Language Cloud.

With this research, the aim is to gain a better understanding of the current likely adherence to standards, the interest into standards and the expected priorities that could shape the standards interests of the LT-Innovate forum. The particular focus is to develop a pragmatic and collective opinion around standards highly relevant to suppliers of language technology tools (intelligent content, speech interaction and translation technologies).

Scope

The initial data collection will be handled via the survey that is described in this paper.

The survey to collect data for this research has been distributed across LT-Innovate membership and Localization Professional LinkedIn group, with more than 22,000 members worldwide.

The survey aims to identify the standards that LT-Innovate members are implementing today, the possible future uses and priorities.

The survey was created in SurveyMonkey and contained 20 questions, which can be seen in the annex of this document.

Survey audience and demographics

The survey has been distributed among LT-Innovate members and 22,000 members of Localization Professionals group.

The respondent demographics are:

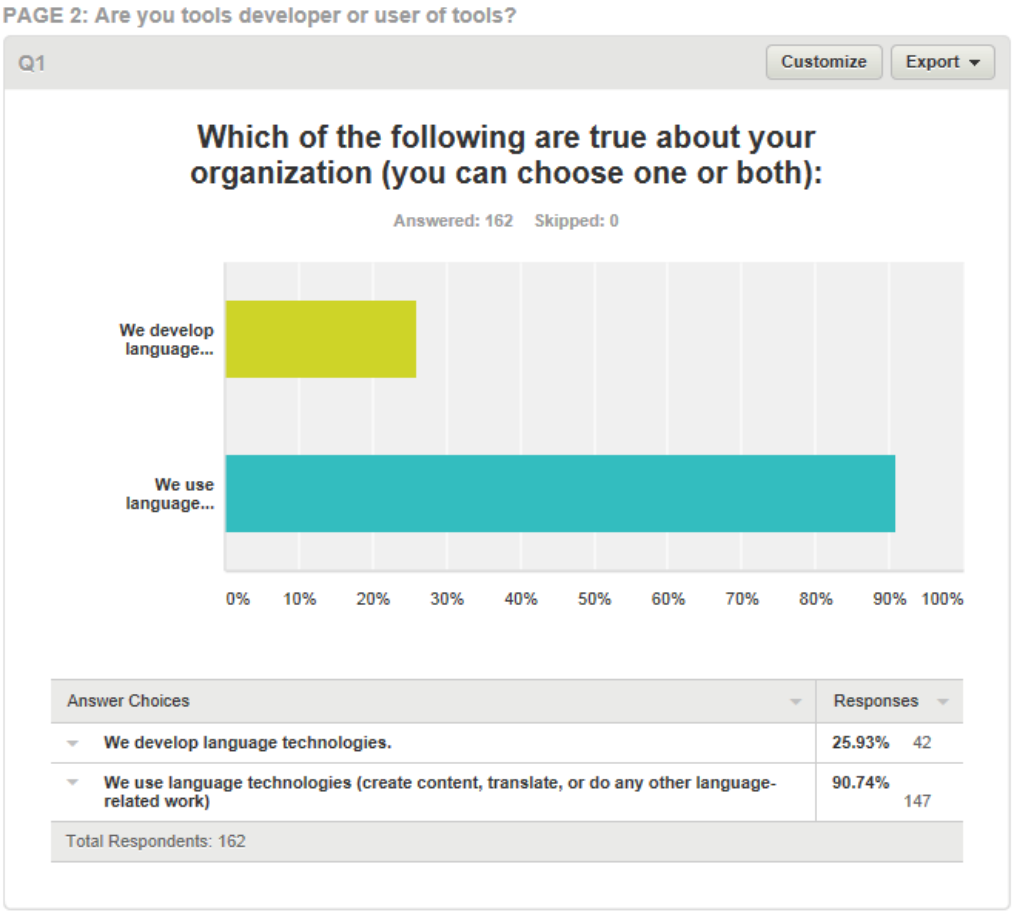


Fig.2 Survey demographics.

162 organizations responded to the survey, including:

Studio Gambit Sp. z o.o.
XTM International
TMServe
SALT Group
Rosetta Stone
Welocalize
Logrus International
Text Case
InfoDutch
ITINEUROPA
Virginia Bordas Language Services SRL
Depictum Osk
Interago Text & Layout
Sperapro
Trans4CN
Datamundi
Aliquantum, Inc.
sQuid
Maxprograms
Synthema
AutomaticTrans
Transline Deutschland Dr.-Ing Sturz GmbH
Dictanova

42 respondents expressed the wish to be included in further discussions; these people may form a wider community to do collective brainstorming along the lines of the survey.

The respondents are both developers and users of language technology; 97% are users and 26% are developers AND users of language technology.

The observations from the survey responses

This section lists some immediate observations about the responses that have been given to the survey.

The types of technology that is now being developed by language technology developers.

It is quite interesting to see what types of technology language technology developers are creating.

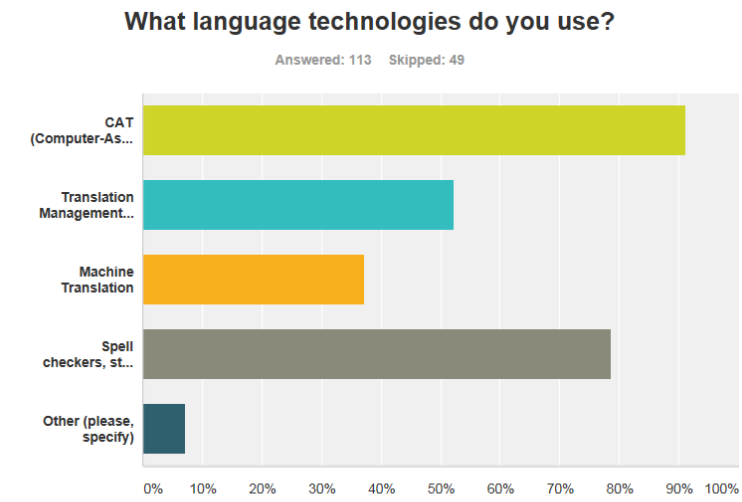
Answer Choices	Responses	
CAT (Computer-Assisted Translation) tools, TM, etc.	42.50%	17
Translation Management Systems (workflow)	40.00%	16
Machine Translation	40.00%	16
Spell checkers, style checkers, term extraction, parsers, etc.	47.50%	19
Other (please, specify)	37.50%	15
Total Respondents: 40		

Fig.3 Types of technology the tools developers are creating.

As seen from this breakdown, the majority of developers create common categories of tools: CAT Tools, Translation Memory Systems (workflow), Machine Translation, Spell checkers, style checkers, term extraction, speech recognition, and parsers – also language quality control and language learning software (around 40% each group); and just about 40% fall into "other" category, which include less conventional categories of software, among them:

i. text analysis for media monitoring
ii. authoring assistant system
iii. online terminology tool
iv. data analysis tool for language industry
v. localization kit (LocKit) validation tools
vi. opinion miners

Among these tools we can see only perhaps opinion minder and text analysis for media monitoring (as well as other data analysis) as something unconventional; the rest of the categories are well known categories of language software.
It is therefore more or less evident that **there's huge duplication of effort currently going on in language technology industry, many companies work on similar categories of software**.
In this context, it is worth asking ourselves, isn't it more or less evident that common, standardized open source tools will reduce duplication of the effort in the areas which are relatively commonplace? And wouldn't this development work be greatly simplified by establishing common open standards for data, workflow and various formats?
With users of language technology the situation with tools is a bit different:



Answer Choices	Responses	
CAT (Computer-Assisted Translation) tools, TM, etc.	91.15%	103
Translation Management Systems (workflow)	52.21%	59
Machine Translation	37.17%	42
Spell checkers, style checkers, term extraction, parsers, etc.	78.76%	89
Other (please, specify)	7.08%	8
Total Respondents: 113		

Fig.4 Types of technology the users of language technologies are using.

Almost everyone is using CAT tools (91.15%) with various satellite technologies such as spell-checkers (79%), nearly the half of language technology users use Translation Management Systems (workflow), only 37% are using machine translation, and 7% are using more advanced and exotic tools such as Terminology management systems, content management systems, tools for linguistic QAs.

Clearly, tools are not yet so widely used in the language industry, and apart from CAT tools (which are now the requirements for the deliverables) even spell checkers are not used by all translation companies, and more advanced tools are even rarer.

We believe that this is a consequence of the vast array of tools, and here again standardisation based on open source, freely available tools can significantly improve the situation and increase tool adoption, as well as reduce interoperability and cost barriers.

The types of standards that developers of language products adhere to

The table below lists the most frequently used standards for software developers (sorted by the decreasing order of frequency):
Standards use among language technology developers.

Answer Choices	Responses	
TMX (Translation Memory Exchange – ETSI LIS),	70.59%	24
XLIFF (XML Localization Interchange File Format - OASIS),	67.65%	23
TBX (Term Base Exchange - ETSI LIS),	44.12%	15
XLIFF:doc (Subset of XLIFF 1.2 - Interoperability Now!/Linport),	29.41%	10
xml:tm (XML Text Memory – ETSI LIS),	23.53%	8
Our standards are proprietary and we provide only limited support of the above	23.53%	8
W3C ITS (Internationalization Tag Set – W3C),	20.59%	7
SRX (Segmentation Rules Exchange - ETSI LIS)	17.65%	6
GMX-V (Global information management Metrics eXchange – Volume – ETSI LIS),	5.88%	2
OAXAL (Open Architecture for XML Authoring and Localization - OASIS),	5.88%	2
TIPP (Translation Interoperability Protocol Package - Interoperability Now!/Linport),	5.88%	2

Total Respondents: 34

Fig.5 Most frequently used standards among tools developers.

Clearly, TMX and XLIFF (translation memory and transport data standards) are the most popular and important standards among tools and technology developers, closely followed by TBX. Such important, but perhaps not quite widely broadcast, standards as XLIFF:doc, xml:tm, ITS and SRX seem to be in the need of further improvement and popularization.

Standards use among language technology users.

Answer Choices	Responses	
TMX (Translation Memory Exchange – ETSI LIS),	74.77%	80
XLIFF (XML Localization Interchange File Format - OASIS),	67.29%	72
TBX (Term Base Exchange - ETSI LIS),	32.71%	35
XLIFF:doc (Subset of XLIFF 1.2 - Interoperability Now!/Linport),	28.04%	30
xml:tm (XML Text Memory – ETSI LIS),	20.56%	22
I am not aware of specific standards that I am using in the course of my work	14.95%	16
SRX (Segmentation Rules Exchange - ETSI LIS)	8.41%	9
W3C ITS (Internationalization Tag Set – W3C),	7.48%	8
TIPP (Translation Interoperability Protocol Package - Interoperability Now!/Linport),	4.67%	5
GMX-V (Global information management Metrics eXchange – Volume – ETSI LIS),	1.87%	2
OAXAL (Open Architecture for XML Authoring and Localization - OASIS),	1.87%	2

Total Respondents: 107

Fig.6 Most frequently used standards among users of language technology.

The situation among users of technology is somewhat different: the only standard that the users of language technology really use is TMX; even XLIFF is used by only 67% of users of language technology; all other standards adoption is lower than the adoption of the same standards among tools and technology developers.
The reason for this is, clearly, caused by the fact that users trail developers and tools; and this is why it is even more critically important to standardize the European language landscape so that the array of tools and standards and lack of interoperability would be less confusing and create less friction, duplication of effort and overhead.

The voice of the industry – what standards are missing?

It is especially interesting to see what standards are deemed missing by the respondents. Survey participants believe is essential to develop the following list:

1. There is no standard to promote and guide source content creation to enable translation needs (e.g. establishment of source terminology for use both in monolingual and multilingual projects).
2. Productivity measurement standards.
3. MT post-editing standards.
4. Language data model standards.
5. GRE[1]
6. A definition standard for XML data coming from content management systems. i.e. a standard analogous to a DTD with reference to language segmentation. This would save time adapting file format definitions.
7. Translation/Localization quality assessment standard (or adopt QT LaunchPad MQM internationally).
8. Domain ontology to classify text in uniform way and consequently to apply appropriate glossaries, TMs, MT engines.
9. Educational content descriptors.
10. Translation memory separation trigger.
11. Language-related geolocation data.
12. Project data standard – meaning, I'd like to see a standard for translation project data (i.e. a "pack" of translation memories, terminology databases and other reference material used to translate contents for a particular company or industry).
13. Standard way to score QA/match value in TUs (xliff and tmx).
14. Public data ontology.
15. Translation Web Services.
16. Sampling methodology.
17. MT QE.
18. Text annotation.
19. SIOC ontology.
20. Dublin Core.

As you can see, **quite a number of important areas for future standardisation have been identified**; some of the suggestions are already addressed by lesser-known new standards, some of them are currently in development.

[1] http://en.wikipedia.org/wiki/Graduate_Record_Examinations

The voice of the industry – what standards are underdeveloped?

Equally interesting are the standards that are believed by the respondents to be underdeveloped, seen in the following list of responses:

1. ISO 12616 Translation-oriented terminography is not serving its purpose at the moment.
2. Word count standard.
3. Speech recognition for sure.
4. GRE
5. tmx, tbx, and especially xliff – we experience high pre-translation losses when exchanging data on that basis between different translation memory systems.
6. Text/content quality assurance[2].
7. Translation/Localization quality management (project ISO 17100 is completely outdated, and GALA CRISP professional feedback on that matter made no difference).
8. Interoperability between CAT tools is still limping. Small differences in XLIFF implementations prevent from moving files between various TEnT safely.
9. DDF
10. Context recognition
11. Speech synthesis
12. Most except TBX
13. Concordance
14. Developing .XLIFF one step further AND actually using it as the one standard for everyone would be great. We see too many Technology vendors going their own way like SDL with .SDLXLIFF, for example.
15. Context related machine translation for diversion between medical, tech, financial, HR related etc.
16. Cohesive binding between technologies; interoperability.

As we can see, **there are quite a number of areas we can work on in the field of interoperability and standardisation**.

[2] Text and content quality assurance was cited as one of the most cited as missing and/or underdeveloped standard.

The voice of the industry – Which types of language resources, technologies or services that you use are mutually compatible / incompatible?

To see that we address the subject of imperfect standards even better, we have reformulated the question in yet another way and asked about incompatibility of resources and services – something that is closely related to the future European Language Cloud:

Here we have discovered some of the most interesting answers, as follows:

1. Management of SEO keywords and management of technical terminology is currently rather incompatible, as there is no established relation. Terminology assistance in content management platforms is not available. There are only QA processes. NLP for matching terms from a termbase to content either for the purpose of proactive terminology checks in monolingual content creation or in CAT tools.
2. Segmentation is incompatible.
3. As described under 5: in principle all systems are somewhat compatible, but only to a certain extent.
4. RC
5. Translation project files are often incompatible (e.g. Trados - DVX or Trados - MemoQ).
6. TMX did a good work for TMs.
7. Very little adoption of content interoperability standards in Language Learning.
8. CRM system with translation system.
9. TMX, XLIFF and TBX to some extend are mutually compatible and at the same time incompatible.
10. Apart from translation project data, many XLIFF files are not actually standard. Also, translation companies I work for require me to use specific tools for each translation. I guess they should be "educated" in open standards. Or maybe this is due to technical problems, as many tools use XLIFF in a non-standard way.
11. CAT-tool Wordfast is still incompatible with SDL Trados Studio.
12. I would really like to see a higher level of compatibility. Often various tools claim to be compatible, but at the end of the day, major revisions are required in order to be able to exchange files between two programs.
13. Liox TWS tools and Trados Studio.
14. Incompatible: various CAT tools.
15. TMX /XLIFF storage formats only store sentences. They have very limited functionality with regard to inter-sentential coherence (a text is more than a collection of sentences).

What are the language-related standards that you believe are important for Horizon 2020 program goals?

Finally, the question was directly asked about what standards are believed to be of critical importance for H2020 standards. The replies were:

1. Any of the terminology standards I have mentioned before (TBX, 12616, 26162, 12620, 704, 1087...).
2. Segmentation standards.
3. A standard like EN 15038 or 17100 would be useful, but not in the current version.
4. Translation quality assessment standard.
5. XLIFF, OAXAL
6. OpenExchange, TMX, TBX, SRX
7. Better tooling for e-commerce.
8. Website and software internationalization standards. It is essential for companies to get their websites translated into other languages.
9. Quality standards to avoid bad translations. File format standards in order to make file sharing possible between large ranges of CAT-tools.
10. Simplified language. I think the EU could demand that all documentation is written in Simplified English (for the English part of the documentation). This is important as English is often used as a pivot language, and as such, it would be very helpful in building better MT engines based on Simplified English.
11. External proofreading and QA with developed tools for consistency, conformity checks, and consistency checks.
12. IEC82079
13. Standards for text quality control and for exchange of glossaries, translation memories etc.
14. Terminology standards, parallel language corpora standards.
15. MT productivity measurements; term extraction validity measurements.

As we can see, **there are quite a number of gaps that remain**. A clear message of the different applications of standards is the first step towards addressing the different goals of the program.

Fixing the problem – self-certification or external certification?

It is interesting to see the replies to the issue of certification to the standards.
The vast majority of respondents would not even think about self-certification:

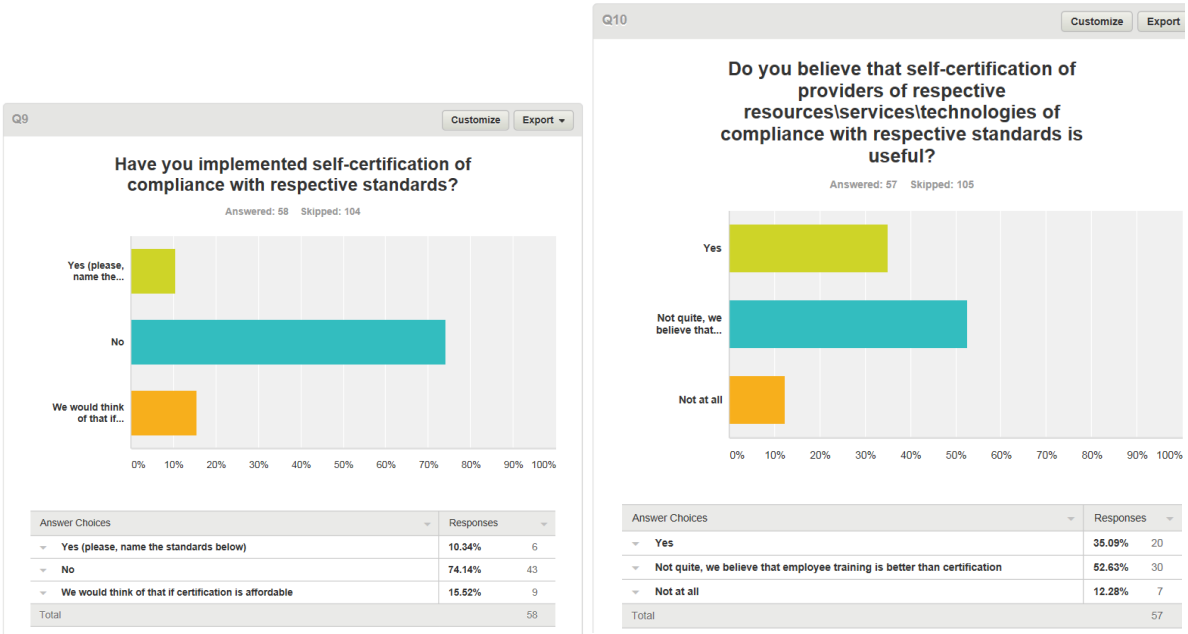


Fig.7 Opinion on Self-Certification.

The comments to this question are even more interesting:

1. “Certification is very useful. But it is the final step. There are many things missing before we can get to it. And it'll be good to even just have the knowledge out there and available for companies.”
2. “EN 15038 is outdated and non-flexible, ISO 9001 is too general.”
3. “EN 15038, I gain new clients when they are sure of my compliance with this certificate which my company (SALT Group) acquired since 2008 till today. I myself personally hold this certificate in two versions: 1- as One Person Enterprise freelancer and 2- as a Linguistic Auditor for Arabic Language licensed to grant OPE certificate to Arabic translators (personal EN 15038 standard Certificate)”.
4. “Certification does not prove the compliance with the exception of paperwork. If one is focused on quality, a certification is not mandatory.”

The majority of the respondents believe that **certification is something that third party certification bodies should do**, rather than self-certification:

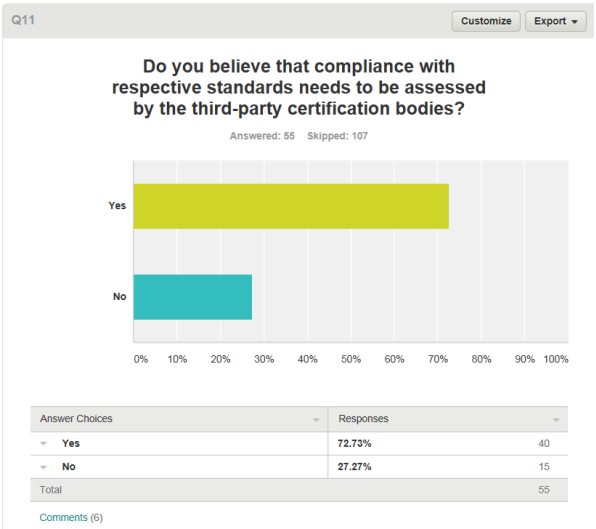


Fig.8 Opinion on certification by third party companies.

At the same time the concerns are:

1. “I believe that declaration of self-compliance plus monitoring by a national / international professional association is sufficient.”
2. “this would imply that third-party certification bodies all over the world use the same basis”
3. “I guess it depends on the standard and its real importance regarding interoperability.”

The conclusion is that certification is generally perceived as something that only professional third parties have to conduct. Consequently, if a set of standards is determined for the H2020 program, in theory LT-Innovate could act as a certification body for the users of the platform, and will be well perceived as such.

Fixing the problem – who will take the lead?

This is one of the most interesting set of questions to this research.

When asked “Would you agree that unified action would be required and useful to bring the variety and multitude of language-related standards together for the future of multilingual Europe without language barriers?” the following response has been obtained:

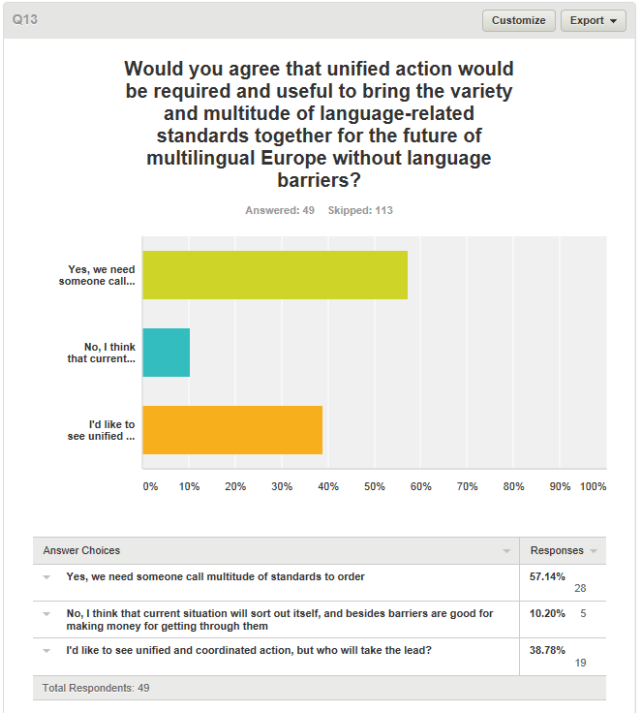


Fig.9 Would you agree that unified action is required? “Yes, but who will take the lead?”…

From this answer it’s clear that **the need for harmonization of standards for Horizon 2020 is evident**, but it is not clear who will take the lead in this effort.

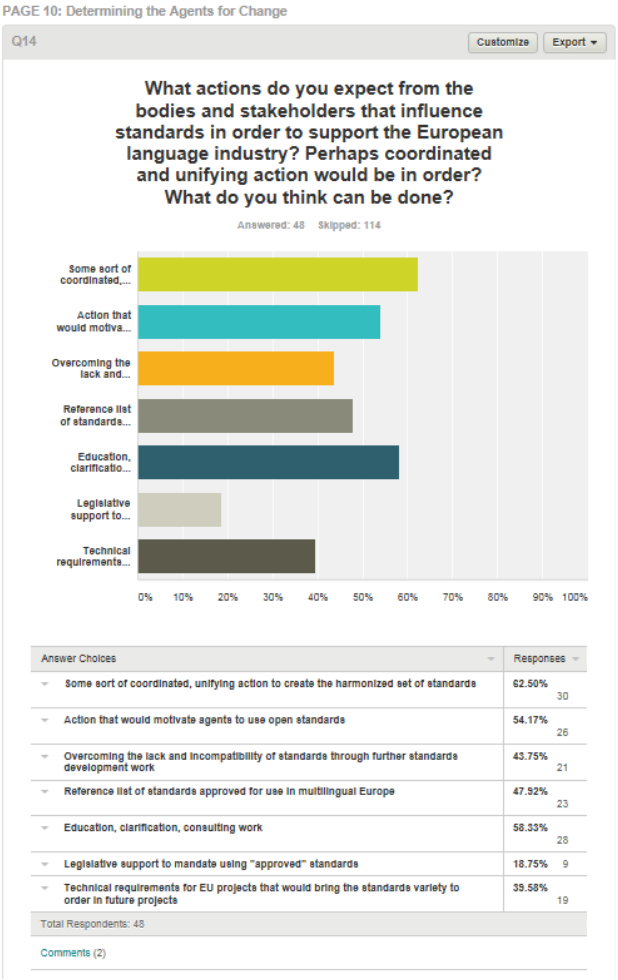


Fig.10 Types of actions that are expected from such unifying body.

In response to the question what specific action should be taken in this regard, the **publication, dissemination, education and consulting** work has been ranked sufficiently high, but also the active **action that would motivate agents to use open standards** has been ranked the highest. One of the most interesting conclusions from this survey was the fact that **legislative support ranked the lowest** among possible unifying and coordinating actions, with the following singular comment:

“Certainly NOT "legislative support to mandate using standards": there are already too many laws on the books. Enough already!”

Clearly, the market does not want *additional* regulation of its business, in any aspect that is connected to the language industry. Many standards are to an extent open and voluntary, not regulatory, and **clarification as to the use and certification of standards is seen to positively support the industry**.

When asked whether LT-Innovate would be the right agent, the answer was “yes” with good majority:

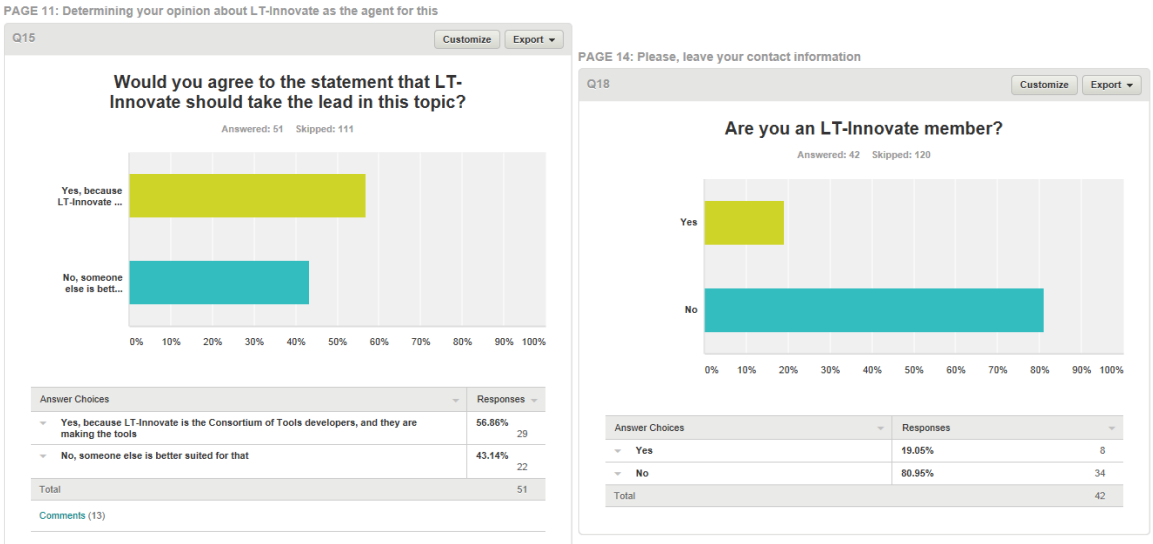


Fig.11 LT-Innovate is a very good candidate.

This opinion was especially interesting even though the majority of respondents were not LT-Innovate membership.

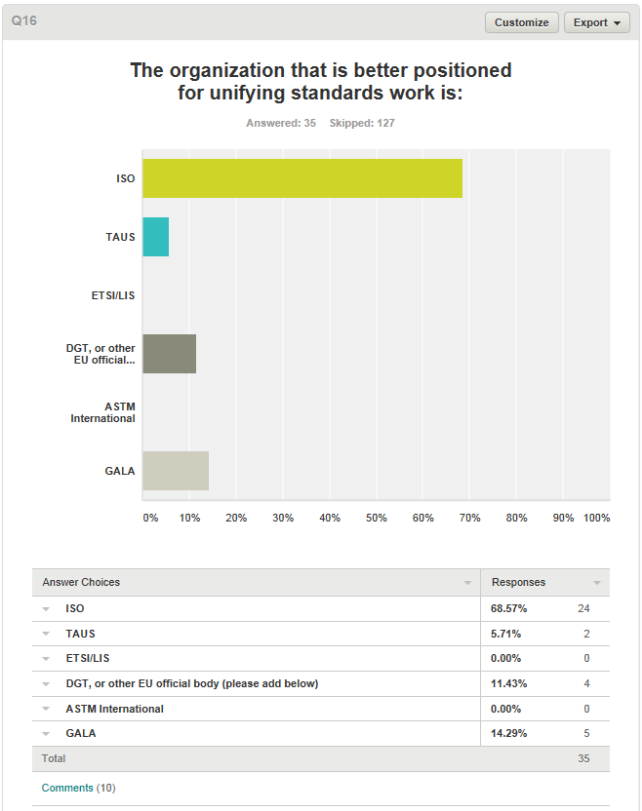


Fig.12 Among “others” ISO is most popular... the distant second is GALA CRISP program.

Not surprisingly, **ISO** as the most well-known European organization is also viewed as well positioned for unifying standards work. However, ISO is purely a global standards body and not the best vehicle to serve as connector between the industry, standards world and the European Commission for H2020 program, while LT-Innovate is.

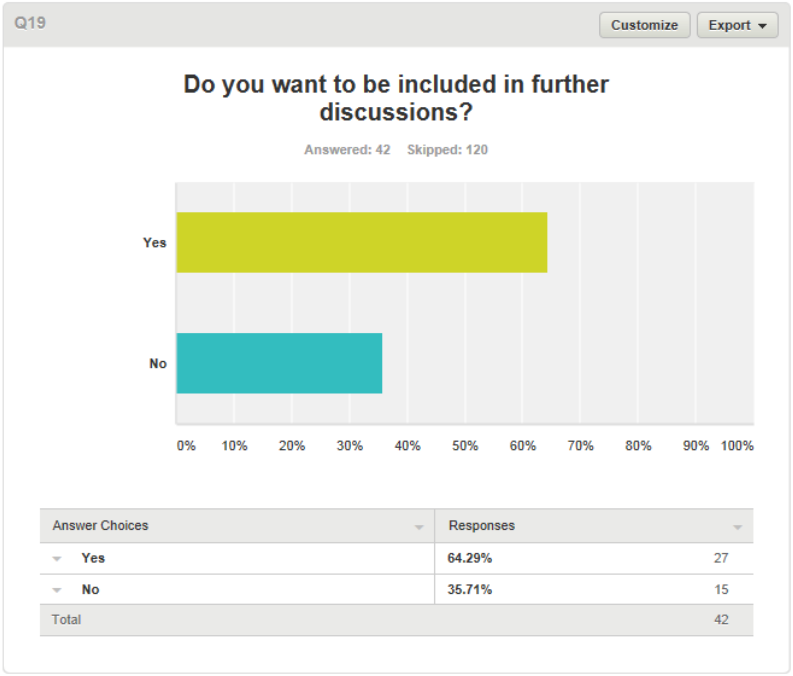
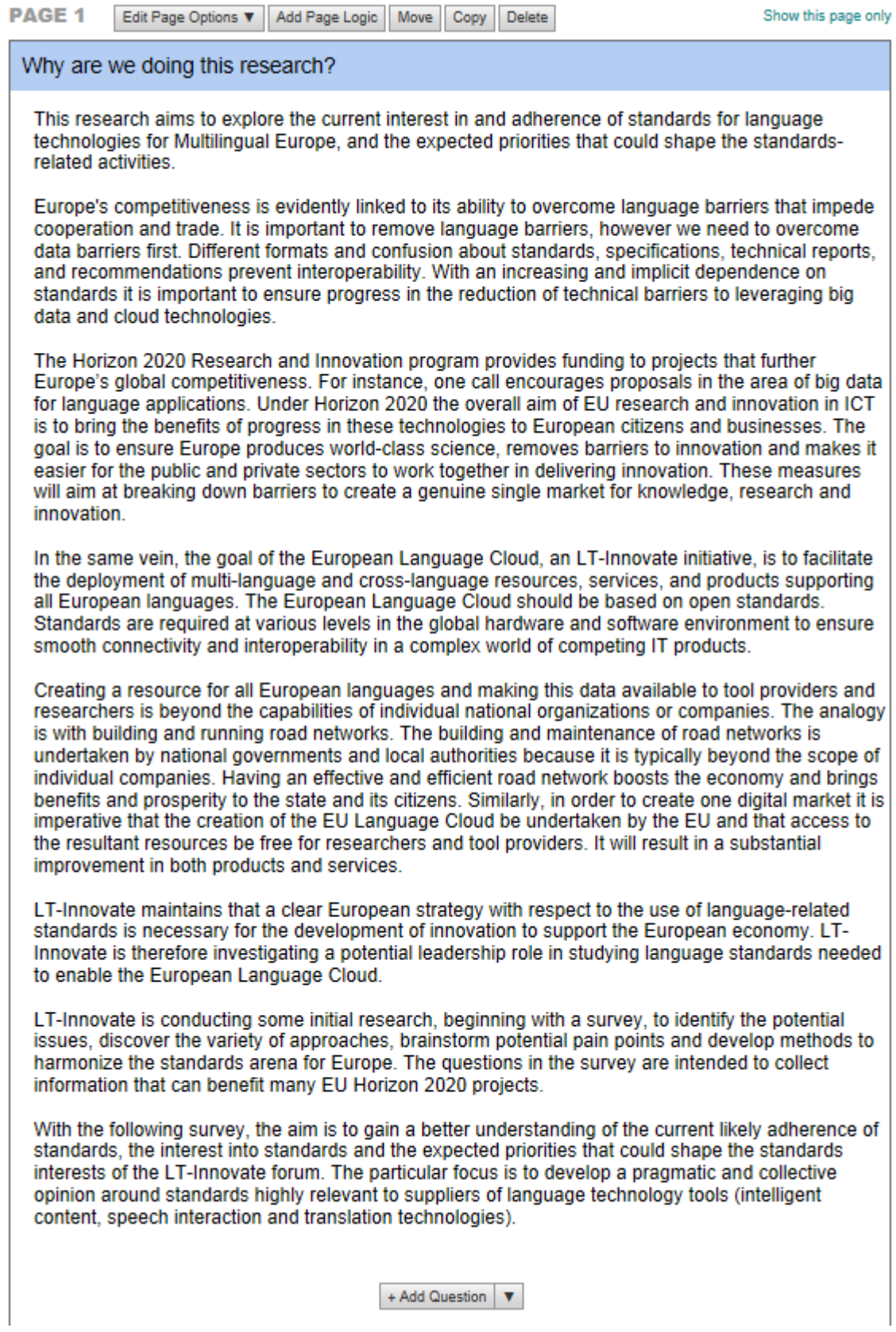
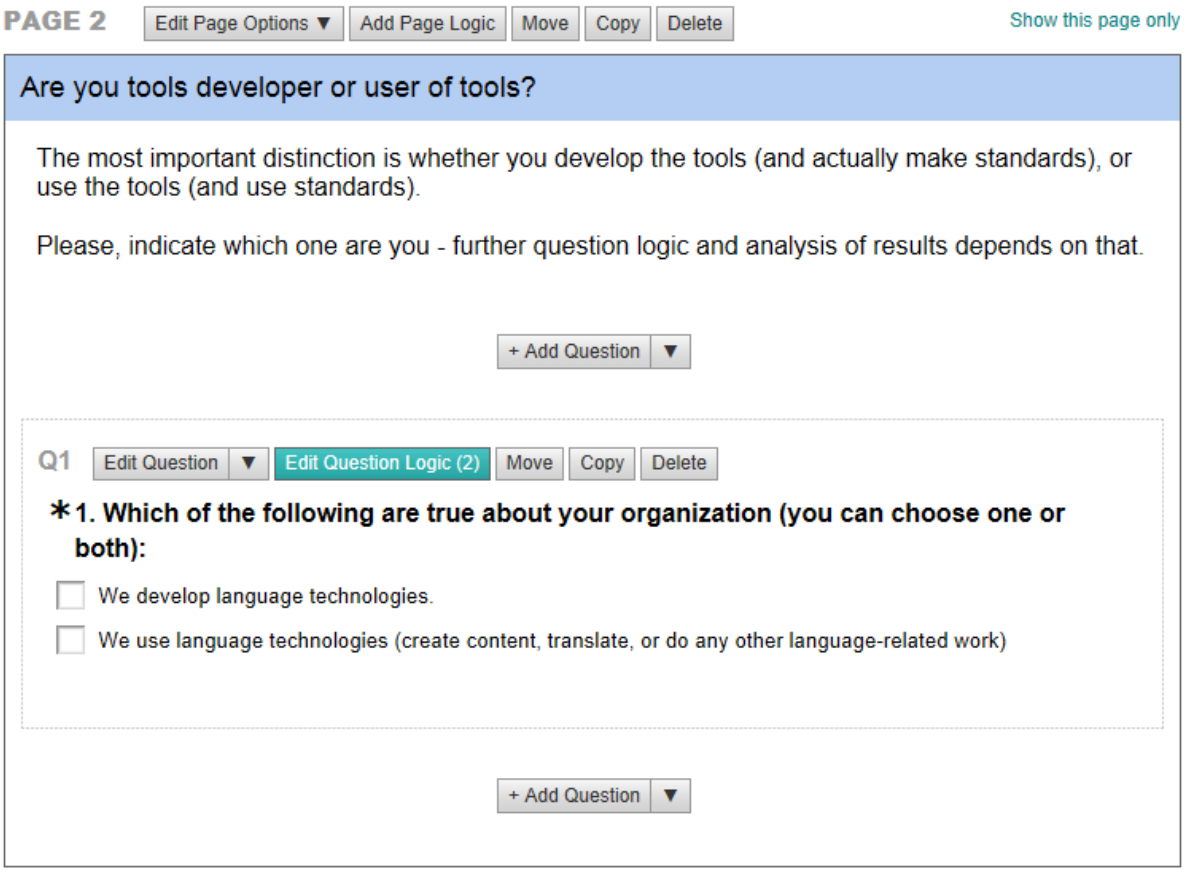


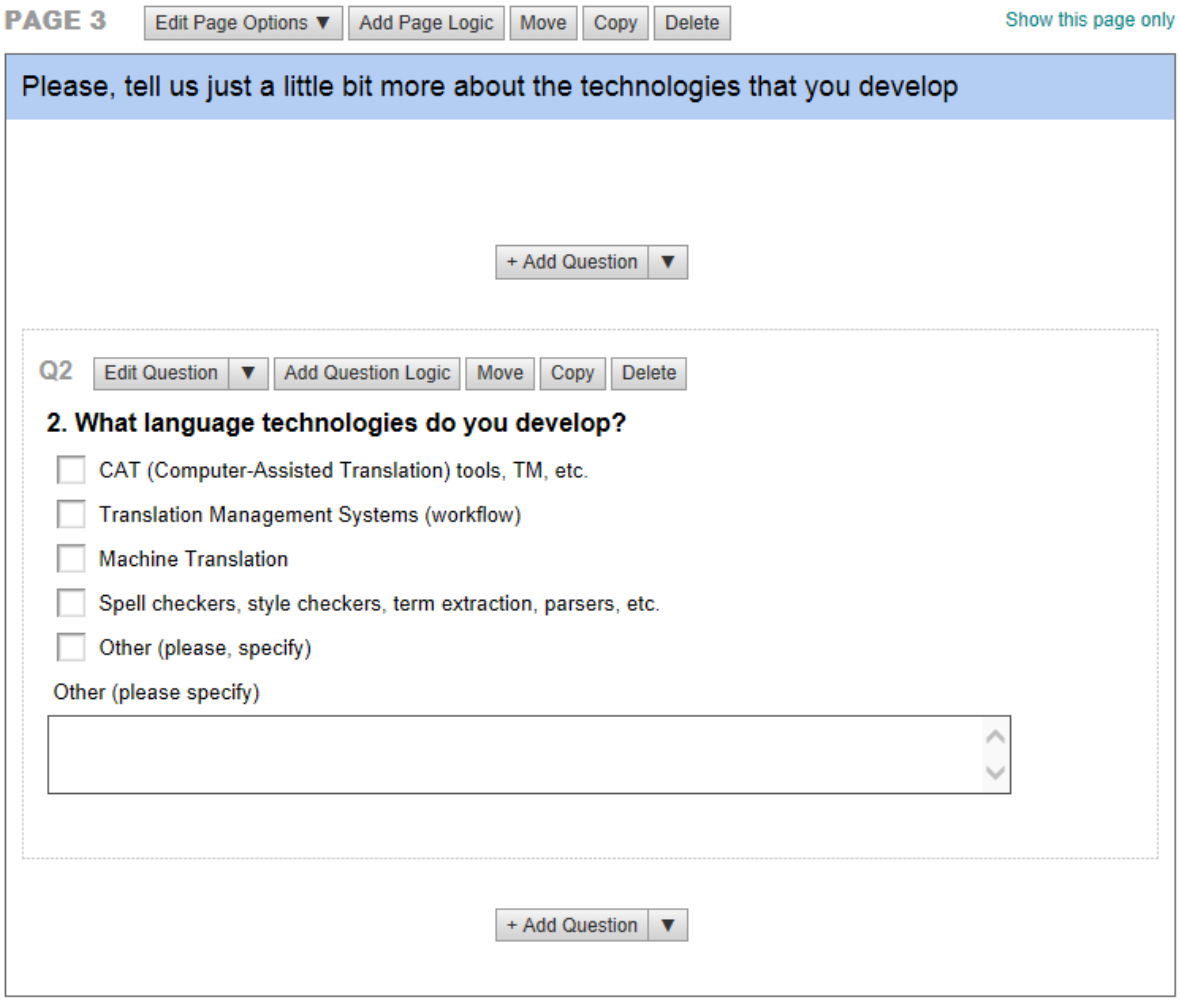
Fig.13 Expert volunteers are available and willing to participate

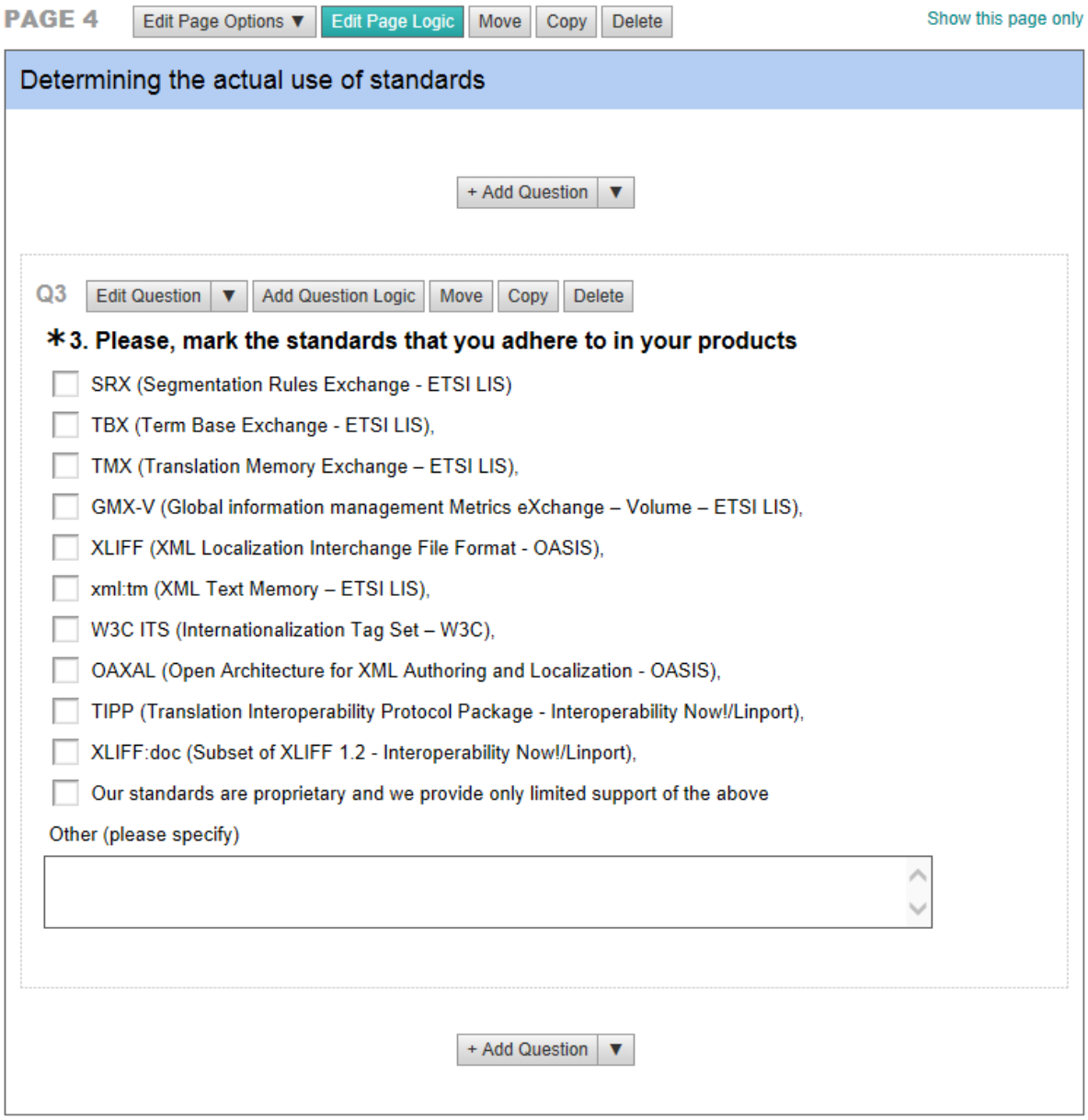
27 respondents expressed desire to be included in further discussions.

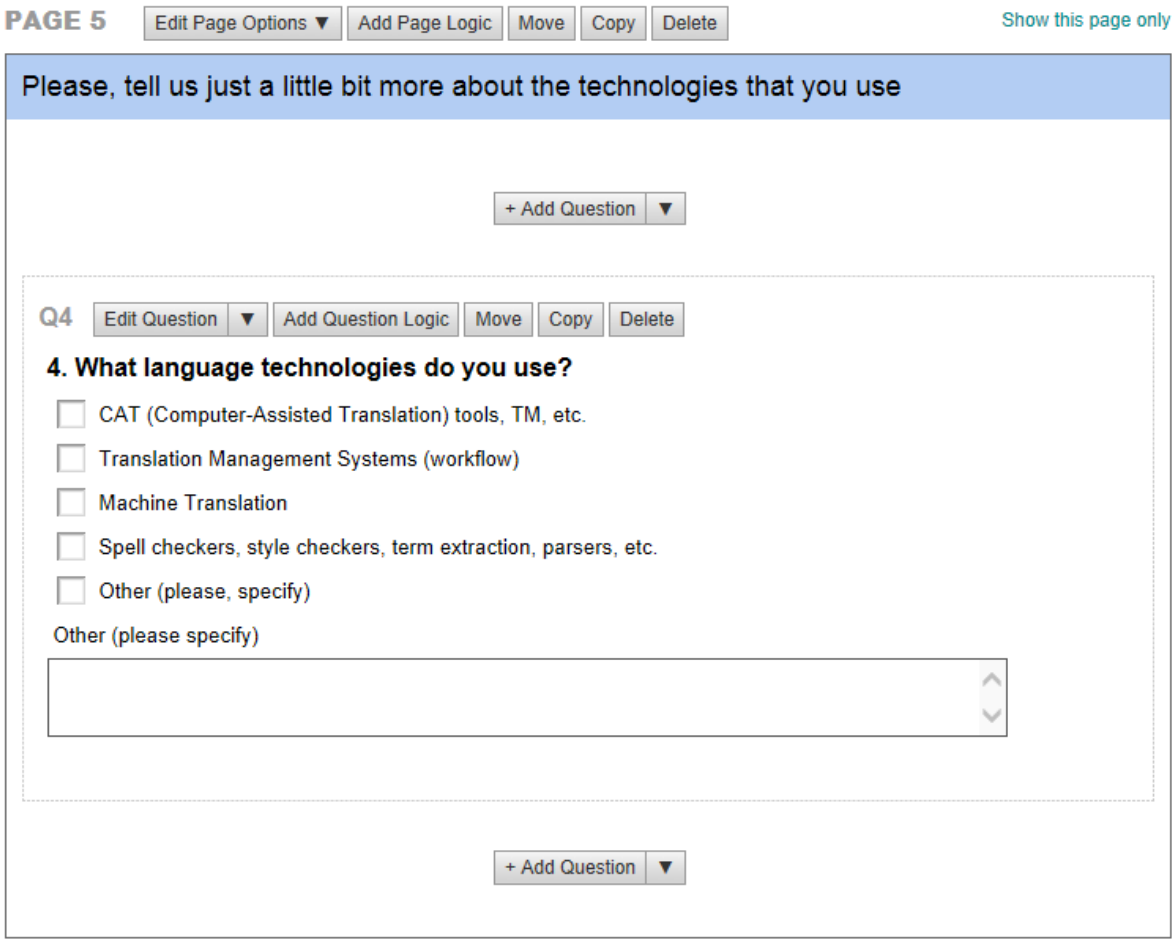
Annex: Survey questions




PAGE 2
Edit Page Options ▼
Add Page Logic
Move
Copy
Delete
Show this page only
Are you tools developer or user of tools?
The most important distinction is whether you develop the tools (and actually make standards), or use the tools (and use standards).
Please, indicate which one are you - further question logic and analysis of results depends on that.
+ Add Question ▼
Q1
Edit Question ▼
Edit Question Logic (2)
Move
Copy
Delete
*1. Which of the following are true about your organization (you can choose one or both):
We develop language technologies.
We use language technologies (create content, translate, or do any other language-related work)
+ Add Question ▼


PAGE 3
Edit Page Options ▼
Add Page Logic
Move
Copy
Delete
Show this page only
Please, tell us just a little bit more about the technologies that you develop
+ Add Question ▼
Q2
Edit Question ▼
Add Question Logic
Move
Copy
Delete
2. What language technologies do you develop?
CAT (Computer-Assisted Translation) tools, TM, etc.
Translation Management Systems (workflow)
Machine Translation
Spell checkers, style checkers, term extraction, parsers, etc.
Other (please, specify)
Other (please specify)
+ Add Question ▼


PAGE 4
Edit Page Options ▼
Edit Page Logic
Move
Copy
Delete
Show this page only
Determining the actual use of standards
+ Add Question ▼
Q3
Edit Question ▼
Add Question Logic
Move
Copy
Delete
*3. Please, mark the standards that you adhere to in your products
SRX (Segmentation Rules Exchange - ETSI LIS)
TBX (Term Base Exchange - ETSI LIS),
TMX (Translation Memory Exchange – ETSI LIS),
GMX-V (Global information management Metrics eXchange – Volume – ETSI LIS),
XLIFF (XML Localization Interchange File Format - OASIS),
xml:tm (XML Text Memory – ETSI LIS),
W3C ITS (Internationalization Tag Set – W3C),
OAXAL (Open Architecture for XML Authoring and Localization - OASIS),
TIPP (Translation Interoperability Protocol Package - Interoperability Now!/Linport),
XLIFF:doc (Subset of XLIFF 1.2 - Interoperability Now!/Linport),
Our standards are proprietary and we provide only limited support of the above
Other (please specify)
+ Add Question ▼


PAGE 5
Edit Page Options ▼
Add Page Logic
Move
Copy
Delete
Show this page only
Please, tell us just a little bit more about the technologies that you use
+ Add Question ▼
Q4
Edit Question ▼
Add Question Logic
Move
Copy
Delete
4. What language technologies do you use?
CAT (Computer-Assisted Translation) tools, TM, etc.
Translation Management Systems (workflow)
Machine Translation
Spell checkers, style checkers, term extraction, parsers, etc.
Other (please, specify)
Other (please specify)
+ Add Question ▼

PAGE 6 Edit Page Options ▼ Add Page Logic Move Copy Delete Show this page only

Determining the actual use of standards

+ Add Question ▼

Q5 Edit Question ▼ Add Question Logic Move Copy Delete

***5. Please, mark the standards that you use in your work**

- [ ] SRX (Segmentation Rules Exchange - ETSI LIS)
- [ ] TBX (Term Base Exchange - ETSI LIS),
- [ ] TMX (Translation Memory Exchange – ETSI LIS),
- [ ] GMX-V (Global information management Metrics eXchange – Volume – ETSI LIS),
- [ ] XLIFF (XML Localization Interchange File Format - OASIS),
- [ ] xml:tm (XML Text Memory – ETSI LIS),
- [ ] W3C ITS (Internationalization Tag Set – W3C),
- [ ] OAXAL (Open Architecture for XML Authoring and Localization - OASIS),
- [ ] TIPP (Translation Interoperability Protocol Package - Interoperability Now!/Linport),
- [ ] XLIFF:doc (Subset of XLIFF 1.2 - Interoperability Now!/Linport),
- [ ] I am not aware of specific standards that I am using in the course of my work

Other (please specify)

+ Add Question ▼

PAGE 7 Edit Page Options ▼ Add Page Logic Move Copy Delete Show this page only

What standards are missing, underdeveloped or disconnected?

Imagine that the future will bring all technologies together. What standards are missing or underdeveloped for this future to become the reality? What standards are disconnected?

+ Add Question ▼

Q6 Edit Question ▼ Move Copy Delete

***6. What standards are missing?**
(Language-related geolocation data? Public data ontology, etc.? Productivity measurements?)

+ Add Question ▼ Split Page Here

Q7 Edit Question ▼ Move Copy Delete

***7. What standards are underdeveloped?**
(Text/content quality assurance? Speech recognition and\or speech synthesis?)

+ Add Question ▼ Split Page Here

Q8 Edit Question ▼ Move Copy Delete

8. Which types of language resources, technologies or services that you use are mutually compatible / incompatible?

+ Add Question ▼

PAGE 8 Edit Page Options ▼ Add Page Logic Move Copy Delete Show this page only

Certification

Certification allows to make sure that organization is compliant. We often complain that language-related standards are not used properly, are we positive to the idea of certification?

+ Add Question ▼

Q9 Edit Question ▼ Add Question Logic Move Copy Delete

9. Have you implemented self-certification of compliance with respective standards?

- ○ Yes (please, name the standards below)
- ○ No
- ○ We would think of that if certification is affordable

Additional comment

+ Add Question ▼ Split Page Here

Q10 Edit Question ▼ Add Question Logic Move Copy Delete

10. Do you believe that self-certification of providers of respective resources\services\technologies of compliance with respective standards is useful?

- ○ Yes
- ○ Not quite, we believe that employee training is better than certification
- ○ Not at all

Other (please specify)

+ Add Question ▼ Split Page Here

Q11 Edit Question ▼ Add Question Logic Move Copy Delete

11. Do you believe that compliance with respective standards needs to be assessed by the third-party certification bodies?

- ○ Yes
- ○ No

Other (please specify)

PAGE 9 Edit Page Options ▼ Add Page Logic Move Copy Delete Show this page only

Discovery of the need

What language-related standards are important to furthering Europe's global competitiveness by fostering the development of technologies and resources that break down language barriers?

+ Add Question ▼

Q12 Edit Question ▼ Move Copy Delete

***12. What are the language-related standards that you believe are important for Horizon 2020 program goals?**

+ Add Question ▼ Split Page Here

Q13 Edit Question ▼ Add Question Logic Move Copy Delete

13. Would you agree that unified action would be required and useful to bring the variety and multitude of language-related standards together for the future of multilingual Europe without language barriers?

- [ ] Yes, we need someone call multitude of standards to order
- [ ] No, I think that current situation will sort out itself, and besides barriers are good for making money for getting through them
- [ ] I'd like to see unified and coordinated action, but who will take the lead?

In addition (or other than the above), I think that:

PAGE 10 Edit Page Options ▼ | Add Page Logic | Move | Copy | Delete Show this page only

Determining the Agents for Change

On this page we are trying to collect feedback on how do you think the things can be improved by taking some sort of coordinated unifying action.

+ Add Question ▼

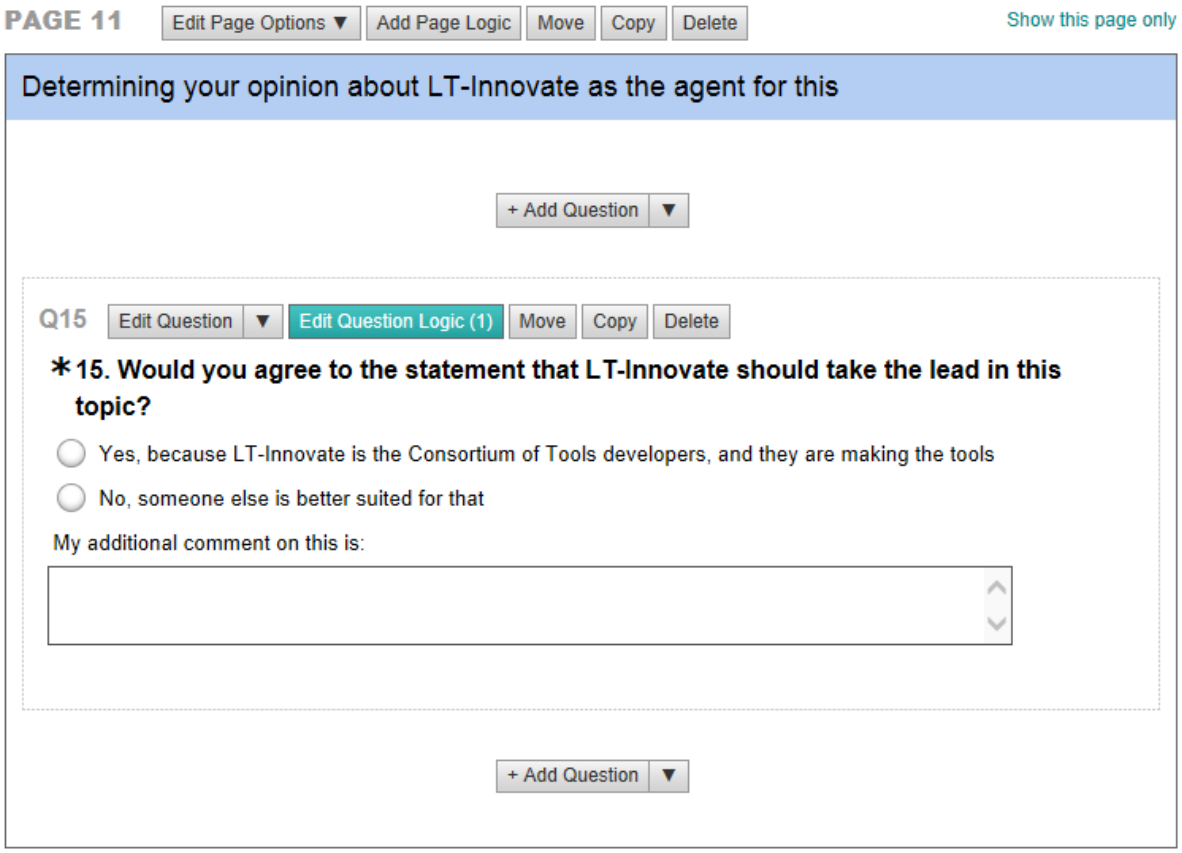
Q14 Edit Question ▼ | Add Question Logic | Move | Copy | Delete

14. What actions do you expect from the bodies and stakeholders that influence standards in order to support the European language industry? Perhaps coordinated and unifying action would be in order? What do you think can be done?

- [ ] Some sort of coordinated, unifying action to create the harmonized set of standards
- [ ] Action that would motivate agents to use open standards
- [ ] Overcoming the lack and incompatibility of standards through further standards development work
- [ ] Reference list of standards approved for use in multilingual Europe
- [ ] Education, clarification, consulting work
- [ ] Legislative support to mandate using "approved" standards
- [ ] Technical requirements for EU projects that would bring the standards variety to order in future projects

Something else (please elaborate)

+ Add Question ▼


PAGE 11
Edit Page Options ▼
Add Page Logic
Move
Copy
Delete
Show this page only
Determining your opinion about LT-Innovate as the agent for this
+ Add Question
▼
Q15
Edit Question
▼
Edit Question Logic (1)
Move
Copy
Delete
*15. Would you agree to the statement that LT-Innovate should take the lead in this topic?
Yes, because LT-Innovate is the Consortium of Tools developers, and they are making the tools
No, someone else is better suited for that
My additional comment on this is:
+ Add Question
▼

PAGE 12 Edit Page Options ▼ Edit Page Logic Move Copy Delete Show this page only

What other organization is suited better for unifying standards in language industry?

+ Add Question ▼

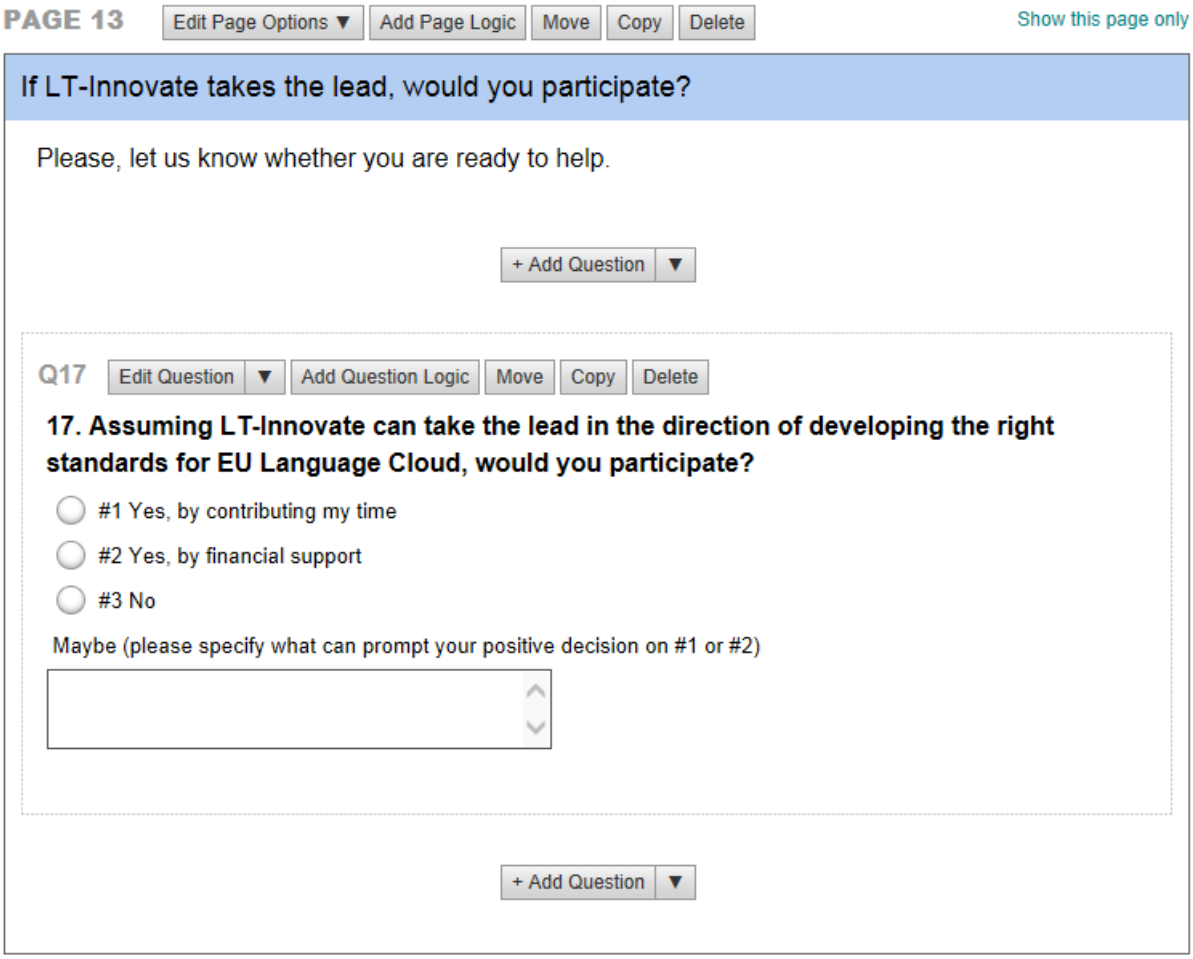
Q16 Edit Question ▼ Add Question Logic Move Copy Delete

16. The organization that is better positioned for unifying standards work is:

- ISO
- TAUS
- ETSI/LIS
- DGT, or other EU official body (please add below)
- ASTM International
- GALA

Other (please specify)

+ Add Question ▼


PAGE 13
Edit Page Options ▼
Add Page Logic
Move
Copy
Delete
Show this page only
If LT-Innovate takes the lead, would you participate?
Please, let us know whether you are ready to help.
+ Add Question ▼
Q17
Edit Question ▼
Add Question Logic
Move
Copy
Delete
17. Assuming LT-Innovate can take the lead in the direction of developing the right standards for EU Language Cloud, would you participate?
#1 Yes, by contributing my time
#2 Yes, by financial support
#3 No
Maybe (please specify what can prompt your positive decision on #1 or #2)
+ Add Question ▼

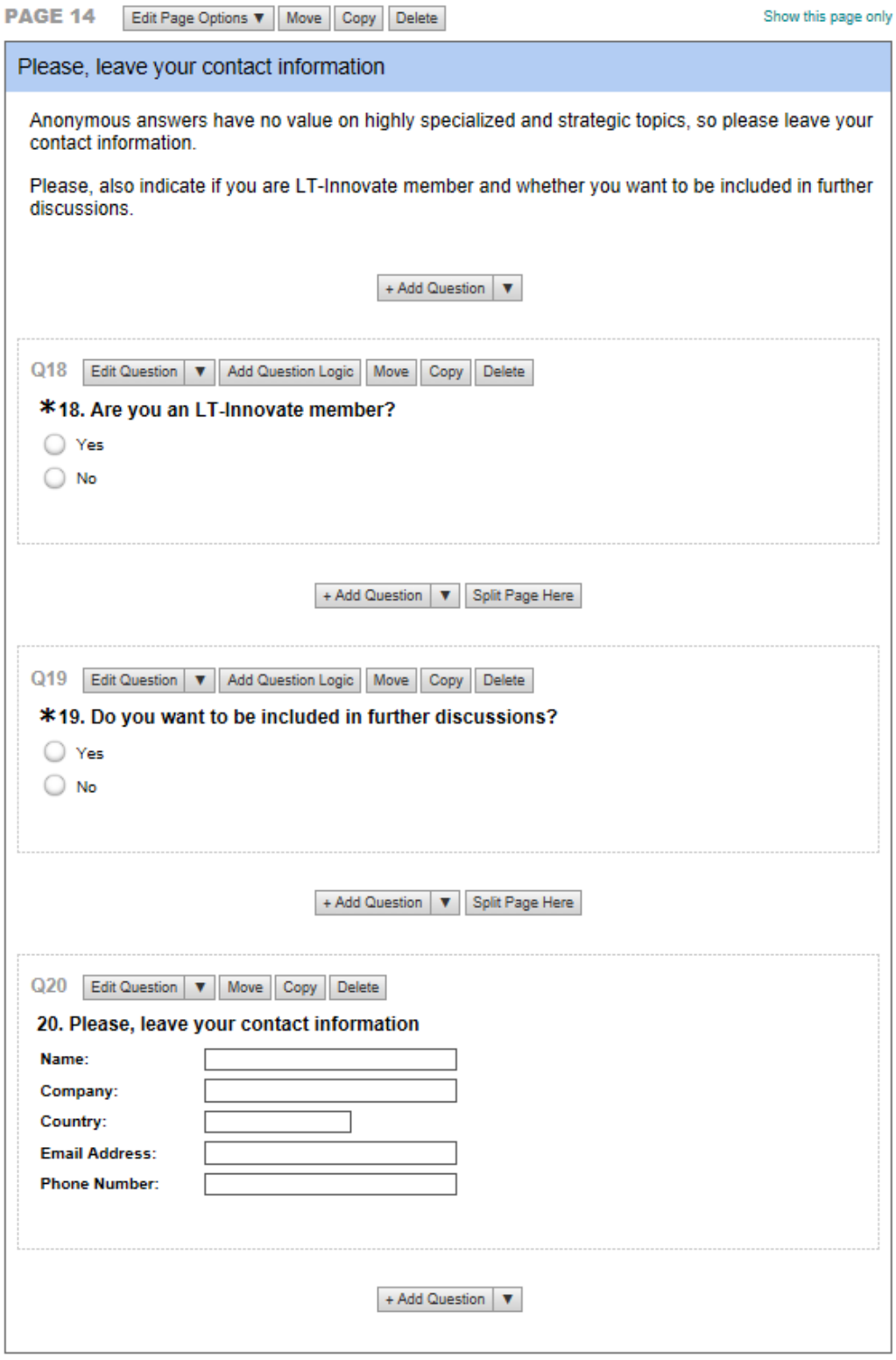


Fig.14 The survey screens.