


STOA Workshop

Language equality in the digital age

Participants' booklet

WORKSHOP
STOA | SCIENCE AND TECHNOLOGY OPTIONS ASSESSMENT

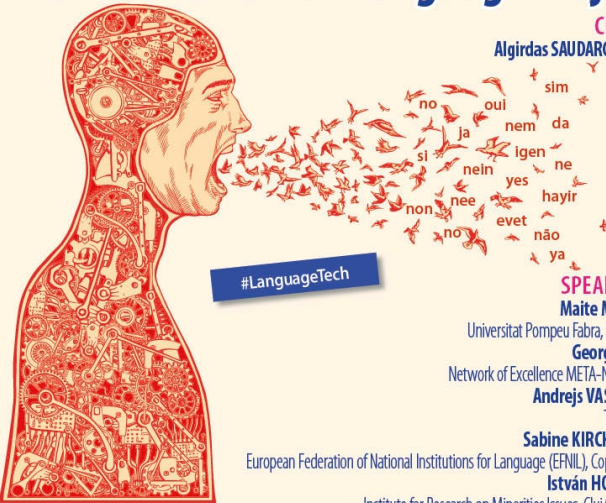
Tuesday 10.01.2017 – **14:00-17:00**
EUROPEAN PARLIAMENT, BRUSSELS,
PAUL-HENRY SPAAK BUILDING – ROOM **P7C050**



European Parliament

Language equality in the digital age - Towards a Human Language Project

CHAIR:
Algirdas SAUDARGAS, MEP



SPEAKERS:
Maite MELERO,
Universitat Pompeu Fabra, Barcelona
Georg REHM,
Network of Excellence META-NET, Berlin
Andrejs VASILJEVS,
Tilde, Riga
Sabine KIRCHMEIER,
European Federation of National Institutions for Language (EFNIL), Copenhagen
István HORVÁTH,
Institute for Research on Minorities Issues, Cluj/Kolozsvár
Hans USZKOREIT,
German Research Centre for Artificial Intelligence (DFKI), Berlin

With the participation of:
JILL EVANS (MEP), **Ádám KÓSA** (MEP), **Csaba SÓGOR** (MEP) and **Evžen TOŠENOVSKÝ** (MEP and STOA Panel Vice-Chair)

#LanguageTech

Registration before 6th January at
<http://www.europarl.europa.eu/stoa/>

EPRS | European Parliamentary Research Service

STOA workshop

Language equality in the digital age

Participants' booklet

**10 January 2016, 14:00 - 17:00
European Parliament, Brussels
Paul-Henri Spaak building, room 7C050**

Prepared by Zsolt G. Pataki & James Tarlton, Scientific Foresight Unit (STOA)

Available at <http://www.europarl.europa.eu/stoa/cms/home/workshops/language>

Join the conversation on Twitter by using the hashtag **#LanguageTech** and by tweeting at **@EP_ThinkTank**

CONTENTS

1. PROGRAMME	5
2. SUBJECT OF THE WORKSHOP	7
3. CHAIR.....	12
ALGIRDAS SAUDARGAS, MEP	12
4. MODERATOR.....	13
MAITE MELERO	13
5. SPEAKERS	14
5.1 GEORG REHM.....	14
5.2 ANDREJS VASILJEVS.....	14
5.3 SABINE KIRCHMEIER.....	15
5.4 ISTVÁN HORVÁTH	15
5.5 HANS USZKOREIT	16
5.6 JILL EVANS	17
5.7 ADÁM KÓSA	17
5.8 CSABA SÓGOR	18
5.9 EVŽEN TOŠENOVSKÝ	18
6. ABOUT STOA	19
6.1 MISSION.....	19
6.2 ADMINISTRATION	20

1. PROGRAMME

Chair: **Algirdas Saudargas**, MEP, LT

Moderator: **Maite Melero**, Senior Researcher, Group of Computational Linguistics, Universitat Pompeu Fabra, ES

- | | |
|---------------|--|
| 14:00 – 14:10 | Welcome of the Chair
Algirdas Saudargas, MEP, (EPP, LT), ITRE & CULT Committee |
| 14:10 – 14:20 | Setting the stage: The potential of Human Language Technologies
Maite Melero, Senior Researcher, Group of Computational Linguistics, Universitat Pompeu Fabra, ES |
| 14:20 – 14:40 | Human Language Technologies in a multilingual Europe
Georg Rehm, Senior Researcher, German Research Centre for Artificial Intelligence and General Secretary, Network of Excellence META-NET, DE |
| 14:40 – 15:00 | Challenges and opportunities posed by Multilingualism in the Digital Single Market
Andrejs Vasiljevs, CEO, Tilde, LV |
| 15:00 – 15:20 | Social implications of unbalanced level of technology support for European languages
Sabine Kirchmeier, Vice-President of the executive committee, European Federation of National Institutions of Languages, DK |
| 15:20 – 15:40 | Culturally sensitive marketing, an alternative way to improve public status of minority languages
István Horváth, President, President of the Romanian Institute for Researching the Problems of National Minorities, RO |
| 15:40 – 16:00 | Human Language Technologies and public policies
Hans Uszkoreit, Scientific Director, German Research Centre for Artificial Intelligence and Network of Excellence META-NET, DE |
| 16:00 – 16:45 | Panel: Human Language Technologies: Potential impact on European economy and society |

Moderator: Maite Melero, Group of Computational Linguistics,
Universitat Pompeu Fabra, ES

Panellists:

Georg Rehm, German Research Centre for Artificial Intelligence
and Network of Excellence META-NET, DE

Andrejs Vasiljevs, Tilde, LV

Sabine Kirchmeier-Andersen, European Federation of National
Institutions of Languages, DK

István Horváth, Romanian Institute for Researching the
Problems of National Minorities, RO

Hans Uskoreit, German Research Centre for Artificial
Intelligence and Network of Excellence META-NET, DE

Jill Evans, MEP (EFA, UK), CULT Committee

Ádám Kósa, MEP (EPP, HU), EMPL Committee

Csaba Sógor, MEP (EPP, RO) LIBE & EMPL Committee

Marco Marsella, European Commission, DG Connect, Learning,
Multilingualism and Accessibility

16:45 - 16:55

Conclusions

Algirdas Saudargas, MEP, (EPP, LT), ITRE & CULT Committee

16:55 - 17:00

Final Remarks

Evžen Tošenovský, MEP (ECR, CZ), STOA Panel Vice-Chair

2. SUBJECT OF THE WORKSHOP

This STOA workshop will provide a detailed overview of the current status and trends of Human Language Technologies (HLT) and the specificities and challenges of the European HLT sector. The workshop will also cover the economic impact of language barriers in Europe and the social implications of not having a balanced level of technological support for all European languages. Finally the workshop will discuss policies to effectively tapping the potential of these technologies in order to achieve a fully integrated European Union (EU) in the digital age.

A crucial topic for Europe

With 24 official languages and more than 60 national and regional minority languages, this is a very relevant topic for the EU. Multilingualism represents, at the same time, the richness of the cultural diversity in Europe, and one of the most substantial challenges for the creation of a truly integrated EU.

Citizens need to communicate in their own languages across the borders of Europe in order to foster a common identity and increase citizens' engagement in the European construction. Lowering language barriers will also benefit the European economy and society by leveraging the opportunities of the Digital Single Market (DSM).

This is a formidable challenge which HLT can undoubtedly contribute to address, thus becoming one of the most crucial technologies for the EU.

HLT are key to overcome language barriers

HLTs are software systems designed to handle human language in all of its forms (spoken, written or signed) which are found behind many everyday digital products such as intelligent assistants, speech-based interfaces and automatic translators.

The emergence of new approaches, based on increased computational power and access to huge amounts of data, such as deep-learning neural networks, are making HLT a real solution to overcome language barriers. The increasing availability of high quality translation and real-time speech recognition services are creating the tools for businesses and public bodies to seize the opportunity of providing their services, contents, and support in any language.

Specific challenges in Europe for minority languages

Improvements in HLT rely mainly on the ability to access and maintain ever larger and more finely tuned linguistic data and resources. Although minority languages are the ones to gain most from language technologies, tools and resources for them are scarce – in some cases almost non-existent – fostering a widening technology gap between English and the other official, co-official or non-official EU languages, some of which might already be facing digital extinction.

In order to bridge this technology gap, policies should focus on fostering technology development for European languages other than English, particularly the smaller and less well-resourced ones.

A fragmented European HLT industry with low market impact

Europe has a strong scientific base in language engineering and technology, and there is no shortage of innovative new entrants. European R&D has produced a steady stream of small

companies working on language technologies. However, a fragmented European industry is not able to effectively respond to the current technology challenges, while top players in language technologies, such as Google or Microsoft, are not European and may be unfit to address the specific needs of a multilingual Europe.

With notable exceptions, rather than building on the important results and success stories generated by publicly-funded HLT projects, Europe has tended to pursue isolated research activities with little pervasive impact on the market. Recent initiatives such as META-NET, Cracking the Language Barrier federation, LT-Innovate and the Connecting Europe Facility programme have attempted to shift this trend by starting to bring the fragmented community together.

Furthermore, the European HLT community acknowledges a lack of coordination between research efforts and the market of HLT applications and services. Recent proposals, originating from this community, attempt to address these concerns by recommending a concerted European effort of administration, research and industry, allowing European HLT SMEs to compete in the global landscape while at the same time contributing to reducing the technology gap for all European languages.

Language barriers are likely to have significant social and economic consequences

In Europe, language barriers are expected to have profound and intertwined social and economic consequences such as: (1) fostering a language divide, (2) hampering workers' mobility, (3) hindering the access to cross-border public-services, (4) reducing citizens' engagement and participation in the political process, and (5) creating fragmented markets of cross-border trade and e-commerce, particularly for SMEs.

Promoting an English monolingual policy in the EU would leave behind 60% of the European population, with high disparities among the countries. In countries such as Hungary, Spain, Portugal, and Bulgaria, less than 20% of the population is able to speak English, compared to 80% in the Netherlands. Moving towards a multilingual scenario of 6 languages (English, German, French, Italian, Spanish, and Polish) would improve the situation, although still half of the population of 11 European countries would be unable to properly communicate using any of those languages. Moreover, policies based on a few languages would create a profound language divide, leaving behind less educated and older population, and would be unfair to speakers of minority languages.

In the new global environment, language becomes a strong barrier for mobility, hindering the creation of an integrated EU. The percentage of EU citizens who have ever moved to a different EU country to live and work is only 5.8% and language differences are likely to be one of the main reasons. In 2014, the internal migration rate in the EU was 6.4 times lower compared to the US. By lowering language barriers in the EU, the working mobility rate could increase up to three fold.

One of the reasons that could partially explain why language barriers are so important for EU mobility is the lack of public services in the destination country which are provided in the language of the EU migrant. In a DSM, with free movement of citizens and goods, public administrations should provide efficient and cross-border citizen services. However, the European Union's internal market is also fragmented regarding e-Government services. Out

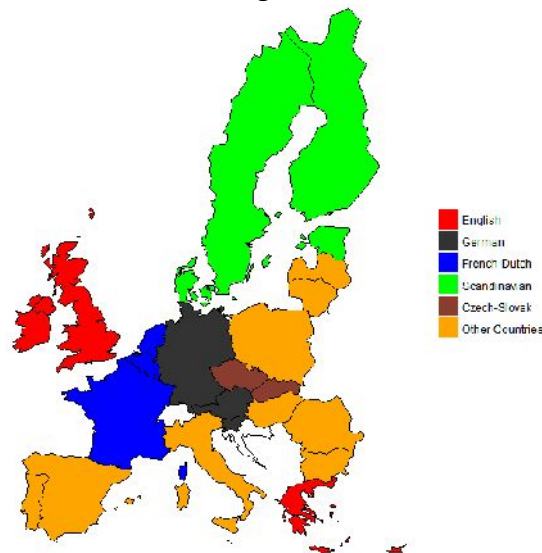
of the 66% of public administrations portals that offer information in languages different from the country's official language(s), only 39% offer information in a language other than English.

Language barriers also affect the construction of a European identity. Citizens and interest groups only engage in a meaningful conversation with decision-makers if it is done in their native language, so an important part of the social intelligence is lost. Participation and collaboration in Europe requires information and data to be fully provided in all official languages and that citizens, businesses and civil society can address European, national, and regional authorities and other stakeholders of all the member states in their native languages. This is particularly important in the current situation of there being a growing sense of disengagement and disillusionment about the political process.

The DSM is fostering a global European area where firms, workers, and citizens can share knowledge, services, products, and labour force. However, language barriers are hindering the achievement of this goal. Only 16% of European citizens have purchased online from other EU countries in 2015, accounting for 30% of total e-commerce users in that year, with symptoms of stagnation. About 50% of European consumers are not prepared to buy in another EU language, while 30% of the companies that sell on-line to consumers consider that language is an important barrier.

The average number of cross-border e-shoppers between countries with low language barriers is four-fold higher compared to countries with high language barriers¹. As a result, e-commerce in Europe is fragmented into 6 groups shaped by language similarities.

Figure 1: E-commerce fragmentation of the EU DSM



Source: Compiled by the authors based on (Civic Consulting, 2011)

For countries with smaller languages this has profound consequences. Take Hungary, a country with an isolated language and a low percentage of population speaking foreign languages. Currently, the people from the rest of the EU shopping on line to Hungarian websites is negligible, while overcoming language barriers would increase the on-line

¹ Language barriers are calculated by using the percentage of population in the origin country speaking, at least good, one of the majority official languages of the destination country

transactions to Hungary from other countries by up to 5.9 million. That is more than twice the current population buying on-line in Hungary (2.2 million).

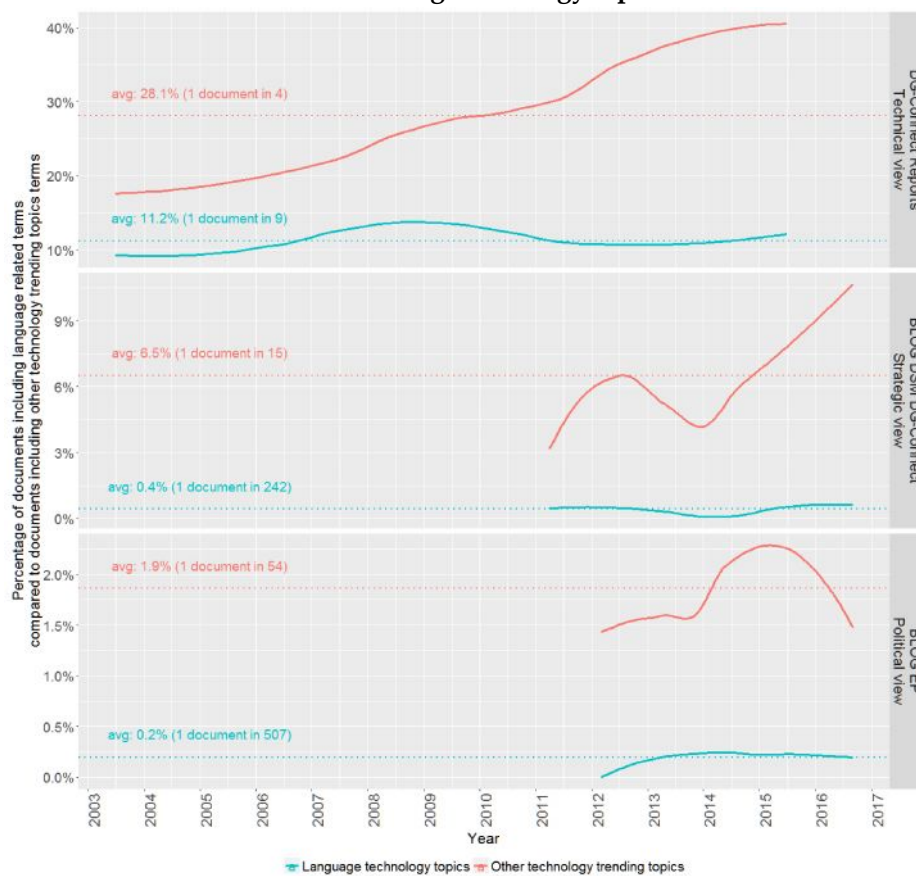
The effect is particularly negative for SMEs, a crucial pillar of the European economy. Up to 41% of large companies sell through e-commerce channels domestically and 23% to other EU countries, compared to 16% and 7% of small companies respectively. One of the main reasons of this gap is the language barrier.

All in all, language barriers are likely to have a very profound negative effect on multiple cross-border social and economic activities in Europe, seriously challenging the European construction and the creation of a truly integrated DSM.

HLT are not properly considered in current policies of the EU

HLT seems the only feasible way to overcome language barriers while preserving cultural diversity. Therefore, HLT should have a very relevant standing in the policy-agenda of the EU. However, our analysis of over 3.000 technical, political, and strategic documents and posts of the EU institutions, by using text-mining techniques to compare language technology terms to other trending technology terms, shows disappointing results. Language technology is an irrelevant topic compared to other trending technologies, and the gap is increasing.

Figure 2: Average percentage of documents including language technology topics compared to other trending technology topics



Source: Compiled by the authors based on (DG-CONNECT, 2014; European Commission, 2015f; European Parliament, 2015)

Neither multilingualism nor HLT are properly reflected in current Information and Communication Technologies (ICT) policies of the EU. In fact, the DSM Strategy of 2015 only makes a brief mention of multilingual services. There is no mention of the role of HLT in providing these services nor of language as one of the most important barriers for the EU DSM. The good news is that within the DSM strategy there are two actions that are oriented to promote research on HLT and to provide these technologies as a service, although the funding is too scarce to make a difference.

Some EU countries, such as the UK, Ireland, and particularly Spain, and countries outside the EU, such as India and South Africa, have developed interesting plans regarding multilingualism and technology that provide useful policy ideas. Industry and research groups have also developed strategic plans that provide valuable insights.

Policy recommendations to effectively draw upon HLT

The following policy options are proposed and assessed.

- **Institutional policies** to evolve the current institutional framework to draw upon emerging technology trends that better fit the challenges of Europe, while properly assessing the results.
 - Reinforce the role of HLT within the institutional framework of multilingualism related bodies
 - Create tools to properly evaluate HLT policies
- **Research policies** to integrate research and industry, providing Europe with the ability to compete with other markets, while contributing to the equality of all European citizens in their everyday digital experience, regardless of their language.
 - Refocus and strengthen research in HLT
 - Create a European language technology platform of data and services
 - Bridge the technology gap between European languages
- **Market policies** to improve the HLT sector in Europe by raising awareness among European stakeholders of the relevance of these technologies to further increasing the demand of services and by fostering the creation and growth of competitive European firms while increasing the availability of highly qualified workers.
 - Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services and products in multiple languages
 - Create of European public or mixed investment instruments and accelerator programs targeting HLT companies
 - Increase the availability of personnel qualified on HLT
- **Public service policies** to create multilingual public services in the European, national, regional and local administrations, while contributing to developing a healthy innovative HLT sector by using public procurement tools.
 - Foster the translation of national and regional public webs and documents to other EU languages by using HLT
 - Facilitate public procurement of innovative technology and pre-commercial public procurement

3. CHAIR

Algirdas Saudargas, MEP



Algirdas Saudargas was re-elected a Member of the European Parliament in May 2014 (Group of the European People's Party - Christian Democrats) after serving as a MEP on the previous parliamentary term. He currently serves on the Industry, Research and Energy Committee. Previously, he was Member of the Lithuanian Parliament (1992-2003) and Lithuanian Ambassador to the Holy See (2004-2008). He has also served as Minister of Foreign Affairs (1990-1992 and 1996-2000).

He holds a University Degree in biophysics by the Kaunas Medical Institute (now Lithuanian University of Health Sciences).

4. MODERATOR

Maite Melero



Maite Melero, PhD is a senior researcher and associate professor at GLiCom, the Computational Linguistics Research Group at Universidad Pompeu Fabra. She has a long experience in the field of Natural Language Processing and Data Analysis, and has participated in a number of EU and nationally funded projects in the computational linguistics field over the past two decades. She has been a member of the Expert Advisory Group on Language Engineering Standards (EAGLES), which is a worldwide reference in standards for linguistic information encoding. As senior researcher in the NLP group at Microsoft Research, she was in charge of the Spanish Grammar Checker still used by MS Office, and of the Spanish generation of the Microsoft MT system. She is the author of the volume dedicated to Spanish of the META-NET White Paper series (The Spanish Language in the Digital Age).

5. SPEAKERS

5.1 Georg Rehm

Georg Rehm works as a Senior Researcher in the Language Technology Lab at the German Research Center for Artificial Intelligence (DFKI), in Berlin. He is the General Secretary of META-NET. Furthermore, he is the Coordinator of the EU/EC-funded project CRACKER which initiated, among others, the emerging European Federation Cracking the Language Barrier. Georg is also the coordinator of the BMBF-funded project Digitale Kuratierungstechnologien.

Additionally, Georg Rehm is the Manager of the German/Austrian Office of the World Wide Web Consortium (W3C), hosted at DFKI in Berlin.



Georg Rehm holds an M.A. in Computational Linguistics and Artificial Intelligence, Linguistics and Computer Science from the University of Osnabrück. Georg Rehm has authored, co-authored or edited more than 130 research publications and co-edited, together with Hans Uszkoreit, the META-NET White Paper Series Europe's Languages in the Digital Age as well as the META-NET Strategic Research Agenda for Multilingual 2020.

5.2 Andrejs Vasiljevs

Dr. Andrejs Vasiljevs is the co-founder and CEO of a leading European language technology company Tilde specializing in machine translation and other technologies for smaller languages. Andrejs is deputy chair of the Multilingual Europe Technology Alliance (META-NET), a board member of the European Language Resource Association (ELRA), LT-Innovate forum and the Latvia State Language Commission. He has also served as an elected member of the Bureau of the UNESCO Intergovernmental Council for IFAP (2010-2014). Andrejs has received a Ph.D. in Computer Sciences from the University of Latvia and has authored more than 50 research papers.



5.3 Sabine Kirchmeier



Sabine Kirchmeier has a Ph.D. in linguistics and computational linguistics and 30 years of experience with research on language technology and machine translation. She has been head of the Department of Language Technology at Copenhagen Business School for 10 years and director of the Danish Language Council since 2006. Since 2008 she has served as deputy president of the executive committee of EFNIL (European Federation of National Institutions for Language) and as project manager of the European Language Monitor (ELM).

5.4 István Horváth



István Horváth is Full Professor of the Sociology and Social Work Faculty of Universitatea Babeş-Bolyai since 2013. He was the Deputy Dean of the aforementioned faculty between 2002 and 2008, in charge of assessing the scientific activity of the faculty.

István Horváth is a reputed expert in migration and interethnic relations, as he has worked in several projects related to these issues, for both Romanian and international institutions. He holds a Ph.D. in Philosophy of History and a B.A. in Sociology and Philosophy from the Universitatea Babeş-Bolyai.

5.5 Hans Uszkoreit

Hans Uszkoreit is Scientific Director and Head of the DFKI Language Technology Division. He also serves as site coordinator of DFKI Berlin and as Honorary Professor at Technical U. Berlin.

Uszkoreit studied linguistics and computer science at Technical U. Berlin and U. of Texas at Austin, where he received his Ph.D. in 1984. He then worked at the Artificial Intelligence Center of SRI International, the Center for the Study of Language and Information of Stanford U. and the Science Division of IBM Germany. In 1988 he accepted a Full Professor position at Saarland University where he taught until 2015. From 2010-2015 he served as Coordinator of the



European Network of Excellence META-NET (Multilingual Europe Technology Alliance) with 60 research centers in 34 countries. Uszkoreit is Member of the European Academy of Sciences and of the International Committee of Computational Linguistics as well as past president of the Association of Logic, Language and Information. He has co-founded several start-up companies and serves on various industrial and scientific advisory boards. He leads several of the most prominent European MT projects such as EuroMatrix, EuroMatrixPlus and QTLauchPad. His current work centers around the analysis of large-scale language data, information and dialog applications and automatic translation. He recently received two Google research awards for his work on large-scale text data. His research results are documented in more than 200 international publications.

5.6 Jill Evans



Jill Evans is Member of the European Parliament since 1999 (Group of the Greens – European Free Alliance) after her re-elections in 2004, 2009 and 2014. She is member of the Committee on Culture and Education.

5.7 Adám KÓSA



Ádám Kósa was born in Budapest, Hungary. He is deaf and his mother tongue is Hungarian Sign Language. He graduated from the University of Pázmány Péter, in Hungary as a lawyer in 2005. He is president of the Hungarian Association for the Deaf and Hard of Hearing (SINOSZ) since 2005 and a member of the European Parliament since 2009. Between 2009 and 2014 he was the president of Disability Intergroup in the European Parliament; currently he is co-chair.

Mr Kósa got important reports and opinions on people with disabilities such as European Disability Strategy (2011), general regulation on European funds ("CPR regulation") and instruments in terms of employment and social affairs (2012) as well as the implementation of the anti-discrimination directive in the field of employment and training (2013). Ádám

Kósa has been elected as a MEP of Year in 2013 of the European Parliament Magazine. Currently he works on dossiers on marginalised communities and other equal treatment-related topics. Finally, as a newly appointed rapporteur, he deals with the emerging issue of robotics with a view to civil law.

5.8 Csaba Sógor

Csaba Sógor is Member of the European Parliament since 2007 (Group of the European People's Party – Christian Democrats), after his re-election in 2009 and 2014. He is member of the Committee on Civil Liberties, Justice and Home Affairs and the Delegation for relations with Japan. Previously, he was Senator of the Romanian Parliament (2000-2007).



5.9 Evžen Tošenovský

Evžen Tošenovský is Member of the European Parliament since 2009 (Group of the European Conservatives and Reformists), after his re-election in 2014. He is member of the Committee on Industry, Research and Energy and he also serves as Vice-Chair of the STOA Panel. Previously he was Member of the Presidency of the Committee of the Regions (2003-2008), President of the Moravian-Silesian Region (2000-2008) and Mayor of Ostrava (1993-2000).

Evžen Tošenovský is Vice-Chair of the Science and Technology Options Assessment Panel (STOA). He is also a substitute member of the Committee on Transport and Tourism, the Committee of Inquiry into Emission Measurements in the Automotive Sector, and the Delegation for relations with Switzerland and Norway.



6. ABOUT STOA

6.1 Mission

The Science and Technology Options Assessment (STOA) Panel forms an integral part of the structure of the European Parliament (EP). Launched in 1987, STOA is tasked with identifying and independently assessing the impact of new and emerging science and technologies. The goal of its work is to assist, with independent information, the Members of the European Parliament (MEPs) in developing options for long-term, strategic policy-making.

The STOA Panel

The STOA Panel consists of 24 MEPs nominated from the eight permanent parliamentary committees: AGRI (Agriculture & Rural Development), CULT (Culture & Education), EMPL (Employment & Social Affairs), ENVI (Environment, Public Health & Food Safety), IMCO (Internal Market & Consumer Protection), ITRE (Industry, Research & Energy), JURI (Legal Affairs) and TRAN (Transport & Tourism). Mairéad McGuinness MEP is the EP Vice-President responsible for STOA and member of the Panel. The STOA Chair for the first half of the 8th legislature is Paul Rübig, with Eva Kaili and Evžen Tošenovský elected as 1st and 2nd Vice-Chairs.

The STOA Approach

STOA fulfils its mission primarily by carrying out science-based projects. Whilst undertaking these projects, STOA assesses the widest possible range of options to support evidence-based policy decisions. A typical project investigates the impacts of both existing and emerging technology options and presents these in the form of studies and options briefs. These are publicly available for download via the STOA website: www.europarl.europa.eu/stoa/.

Some of STOA's projects explore the long-term impacts of future techno-scientific trends, with the aim to support MEPs in anticipating the consequences of developments in science. Alongside its production of 'hard information', STOA communicates its findings to the EP by organising public events throughout the year.

Focus areas

STOA activities and products are varied and are designed to cover as wide a range of scientific and technological topics as possible, such as nano-safety, e-Democracy, bio-engineering, assistive technologies for people with disabilities, waste management, cybersecurity, smart energy grids, responsible research & innovation, sustainable agriculture and health. They are grouped in five broad focus areas: eco-efficient transport and modern energy solutions; sustainable management of natural resources; potential and challenges of the Internet; health and life sciences; science policy, communication and global networking.

6.2 Administration

Director-General, Directorate-General for Parliamentary Research Services (EPRS)

Anthony Teasdale

Director, Directorate C, Impact Assessment & European Added Value

Wolfgang Hiller

Head of Unit - Scientific Foresight Unit (STOA)

Theo Karapiperis

Head of Service - STOA Secretariat

Zsolt G. Pataki

Head of Service - Scientific Foresight

Lieve Van Woensel

Administrators

Philip Boucher

Mihalis Kritikos

Nera Kuljanic

Christian Kurrer

Gianluca Quaglio – Seconded National Expert

Assistants

Serge Evrard

Rachel Manirambona

Damir Plese

Trainee

James Tarlton

This is a publication of the STOA Secretariat
Directorate for Impact Assessment and European Added Value
Directorate-General for Parliamentary Research Services, European Parliament



PE 581.937