Horizon 2020 Grant Agreement No 809965
Contract start date: July 1st 2018, Duration: 30 months

LAMBDA Deliverable 6.2
Foresight Exercise

Due date of deliverable: 31/12/2020
Actual submission date: 30/12/2020

Revision: Version 1.0

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme, H2020-WIDESPREAD-2016-2017 Spreading Excellence and Widening Participation under grant agreement No 809965.

Author(s)	Emanuel Sallinger, Sahar Vahdati (UOXF), Valentina Janev (PUPIN)
Contributor(s)	
Internal Reviewer(s)	Diego Collarana (Fraunhofer IAIS). Nikola Tomašević (PUPIN)
Approval Date	
Remarks	

Workpackage	WP 6 Monitoring, Evaluation and Impact Analysis
Responsible for WP	University of Oxford
Deliverable Lead	University of Oxford (Emanuel Sallinger)
Related Tasks	Task 6.1 Foresight Exercise

Document History and Contributions

Version	Date	Author(s)	Description
0.1	30.06.2020	Sahar Vahdati, Emanuel Sallinger, Valentina Janev	Deliverable 6.1
0.2	07.09	Emanuel Sallinger	Draft of Deliverable 6.2
0.3	1.12.2020	Emanuel Sallinger	KRRL Workshop
0.4	14.12.2020	Valentina Janev	Events in Serbia and West Balkan
0.5	28.12.2020	Diego Collarana, Nikola Tomašević	Review
0.6	29.12.2020	Valentina Janev	Final version

Executive Summary

In this deliverable we report about the LAMBDA networking and brainstorming events where different aspects of Big Data have been discussed and challenges identified. Results of discussions have been integrated in LAMBDA researchers' scientific publications, please check https://project-lambda.org/Publications.

A summary of Big Data challenges is presented in Section 2. The University of Oxford and University of Bonn organized two panel discussions on perspectives and the future of Knowledge Graphs as part of the International Workshop on Knowledge Representations and Representation Learning (KR4L) 2020, please check https://project-lambda.org/KRRL-2020

Section 3 and Section 4 reports the activities related to LAMBDA Foresight Exercise conducted with experts from Serbia and the West Balkan countries from February 2020 to June 2020. Experts were interviewed and they were invited to a panel discussion that occurred online during the Big Data Analytics summer school 2020.

Section 5 reports about related events organized by the Institute Mihajlo Pupin in Serbia and Montenegro, please check Research-Industry Forums, https://project-lambda.org/Research-Industry-Forums and past events https://project-lambda.org/Past-Events

Table of Contents

Abbreviations and Acronyms

AI	Artificial Intelligence
BDA	Big Data Analytics
BDVA	Big Data Value Association
ECAI	European Conference on Artificial Intelligence
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
KG	Knowledge Graphs
KRR	Knowledge Representation and Reasoning
KRRL	Knowledge Representations and Representation Learning
RATEL	Republic Agency for Electronic Communications and Postal Services machine learning
RL	Representation Learning
WP	Work Package

List of Figures

List of Tables

1. Introduction

The topics of Smart Data Management, Big Data, Linked Data, Knowledge Graphs, Open Data, Semantic technologies and Smart Analytics have spawned a tremendous amount of attention among scientists, industry leaders and decision makers in Europe in the last decade. With the emergence of Big Data, the last decade also witnessed a technology boost for artificial intelligence (AI)-driven technologies. While the expectation is that Big Data and AI will bring significant advances for business, science and society at large, there are concerns with different aspects of data use, processing and exchange that call for drafting policy regulations and recommendation. In the last three years, the LAMBDA project was an opportunity to study and analyse different EU policies related to data, see Table 1.

Table 1. EU policies and Recommendations

Policy	Goals
European Strategy for Data, COM(2020) 66 final, 19 February 2020[1]	Aims to make the EU a leader in a data-driven society.
Open Data Directive [2] 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information	Aims to improve the provision of real-time access to dynamic data through appropriate technical means, by increasing the supply of valuable public data for re-use, including that of public companies and research organizations.
Data Governance Act COM(2020) 767 final, 25 November 2020[3]	Support the set-up and development of common European data spaces in strategic domains, involving both private and public players: health, environment, energy, agriculture, mobility, finance, manufacturing, public administration and skills.
General Data Protection Regulation (GDPR, 2016/679)[4]	Lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. Protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.
FAIR data principles (2016)[5]	Enhance the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

The overall goal of policies related to date is to ensure that

- *data can flow easily through EU-wide and cross-sector value chains, for which a highly harmonised legislative environment is essential*

[1] https://ec.europa.eu/digital-single-market/en/european-strategy-data
[2] https://eur-lex.europa.eu/eli/dir/2019/1024/oj
[3] https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act
[4] https://gdpr-info.eu/
[5] https://www.force11.org/group/fairgroup/fairprinciples;

- *the European model of data sharing, with trusted data intermediaries for B2B data sharing and for personal data spaces, takes off, given the cross-border nature of data sharing and the importance of such data sharing.*

The activities conducted and results delivered in the LAMBDA framework are related to topics relevant for implementation of the above mentioned policies, i.e. to semantic technologies, vocabularies, open data and data sharing, data repositories, data sharing services, data integrity, transparency and trust.

In the last three years, LAMBDA researchers organized and participated at many events where the above mentioned policies were discussed. As members of the Big Data Value Association (BDVA/DAIRO), PUPIN and Fraunhofer IAIS researchers attended different presentations from the European Commission related to recent policy documents. The BDVA and the Industrial Data Space Association activities were also presented to the members of the LAMBDA network during the Big Data Analytics Summer School 2020, please see *Data for AI: The EC's Horizon Europe/Digital Europe Programmes and BDVA's standpoint*[6]. Table 2 gives a list of events where the above mentioned policies were discussed with stakeholders from Serbia and the Region, while Table 3 points to international events (panel discussions at workshops) organized by LAMBDA researchers at international conferences.

Table 2. Events in Serbia and Montenegro

Year	Event	Target participants (Serbia and the Region)
2020	Foresight Exercise 2020	Academia
2020	Open Data - Opportunities and Innovation, Podgorica, March 2020	Public Administration, Academia, Industry
2020	Interconnection of South- Eastern electricity markets, January 2020	Stakeholders from Energy Sector
2019	The Challenge and Opportunity of 5G in Serbia, November 2019	Academia, Public Administration, Telecommunications Industry
2019	Research Data and Open Science in South-eastern Europe, December 2019, Belgrade, Serbia	Academia, Public Administration
2018	Circular Serbia, December 2018	Public administration, Recycling Industry

Table 3. Events at international conferences

Year	Event	Target participants (international)
2020	Knowledge Representation & Representation Learning, September 2020, panel discussions: Emerging Topics in Academia and Industry Future Directions - Looking Ahead	Academia, Industry
2019	ECML PKDD 2019, September 2019, workshop 'New Trends in Representation Learning with Knowledge Graphs'	Academia, Industry

This Deliverable point to Big Data challenges identified by LAMBDA (see Section 2) and events organized by the LAMBDA discussions (see Section 3 to Section 5).

[6] https://project-lambda.org/sites/default/files/2020-06/Session_2_Data_and_AI.pdf

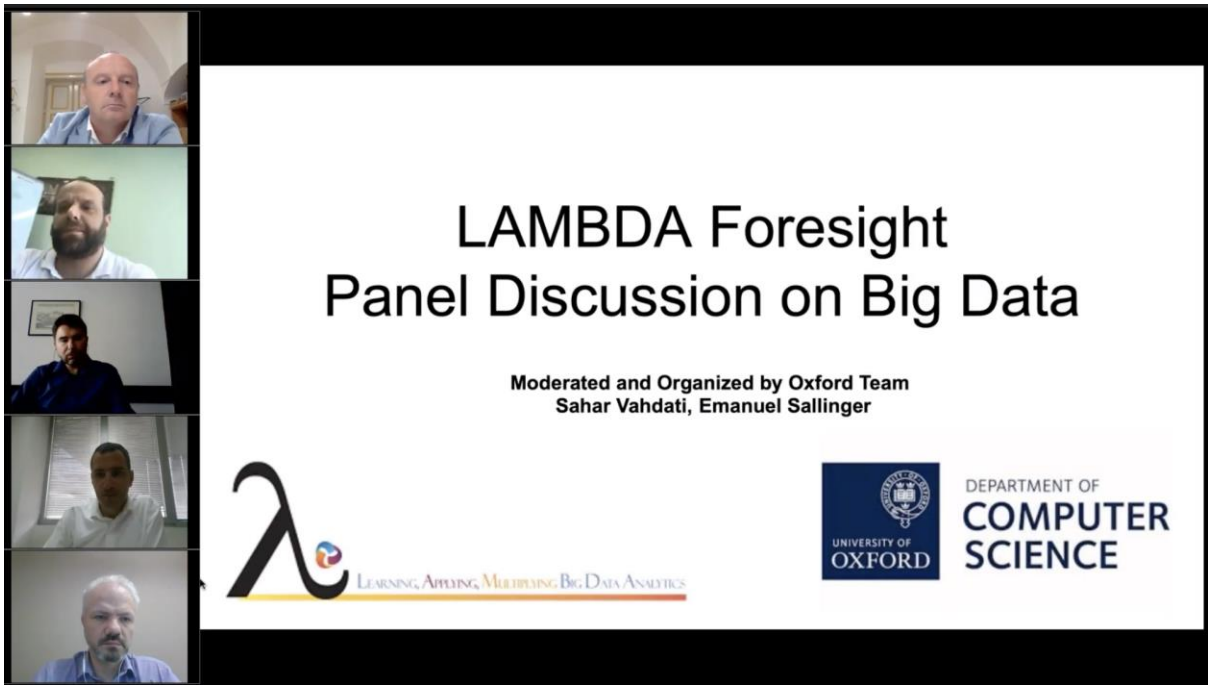


Figure 1. LAMBDA Foresight Panel

This Deliverable is also related to

1. Deliverable 2.1 Big Data Challenges and Analysis of Scientific and Technological Landscape that gives an overview of the Big Data concepts, outlines some of the relevant challenges in Big Data domain and reviews and describes the current state of the art tools relevant to Big Data applications.
2. Deliverable 6.1 Foresight Exercise that report about the Foresight Exercise of the LAMBDA project organized in June 2020 with leading experts from Serbia and the West Balkan region, see Figure 1.

2. Knowledge Graphs and Big Data Challenges (summaries)

2.1 Ecosystem of Big Data and Challenges

The rapid development of digital technologies, IoT products and connectivity platforms, social networking applications, video, audio and geolocation services has created opportunities for collecting/accumulating a large amount of data. **The term Ecosystem is defined in sciatic literature as a complex network or interconnected systems.** While in the past corporations used to deal with static, centrally stored data collected from various sources, with the birth of the web and cloud services, cloud computing is rapidly overtaking the traditional in-house system as a reliable, scalable and cost-effective IT solution.

Hence, in the LAMBDA project and the LAMBDA book, we have introduce definitions, methods, tools, frameworks and solutions for big data processing starting from the process of information extraction, via knowledge processing and knowledge representation to storing and visualization, sense-making, and practical applications.

For more info, please check *Chapter 1 Ecosystem of Big Data (Valentina Janev)* and *Chapter 3 Big Data Outlook, Tools, and Architectures (Hajira Jabeen)* in Janev, V., Graux, D., Jabeen, H., Sallinger, E. (Eds.) Knowledge Graphs and Big Data Processing. Lecture Notes in Computer Science vol. 12072, pp. 1-208. Springer International Publishing. ISBN 978-3-030-53198-0. DOI: https://doi.org/10.1007/978-3-030-53199-7

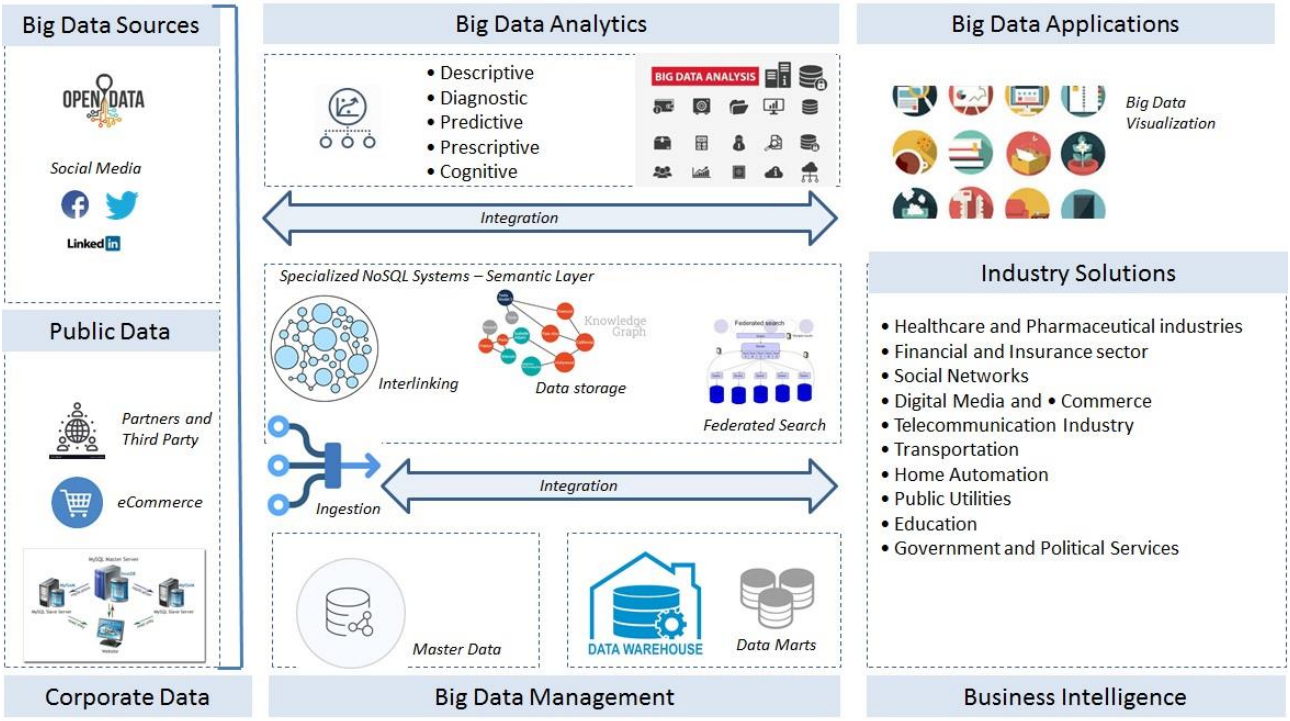


Figure 2. From Data to Applications

The high volumes of structures and unstructured data, stored in a distributed manner, and the wide variety of data sources pose problems related to data/knowledge representation and integration, data querying, business analysis and knowledge discovery. In order to exploit the full potential, big data professionals and researchers have to address different data and infrastructure management

challenges that cannot be resolved with traditional approaches (see Janev and Vrneš[7]; Dedić and Stanier[8]. Hence, in the last decade, different techniques have emerged for acquisition, storage, processing and information derivation in the big data value chains using knowledge graphs[9]. Three main categories of challenges are as follows:

- Data challenges related to the characteristics of the data itself (e.g. data volume, variety, velocity, veracity, volatility, quality, discovery and dogmatism);
- Process challenges related to techniques (how to capture data, how to integrate data, how to transform data, how to select the right model for analysis and how to provide the results);
- Management challenges related to organizational aspects such as privacy, security, governance and ethical aspects.

The characteristics of data combined with targeted business goals pose plenty of challenges while dealing with big data. In this section, we briefly cover the main challenges involved in using big data. Data, process and management challenges are interlinked and influence each other. Herein, we summarized some of the challenges (see also Chapter 1 and Chapter 3 of the LAMBDA Book):

- **Heterogeneity**

Heterogeneity is one of the major features of big data, also characterised as the variety. It is data of different types and formats. The heterogeneous data introduces the problems of data integration in big data analytics, making it difficult to obtain the desired value. The major cause of data heterogeneity is disparate sources of data that generate data in different forms. The data can be text data coming from emails, tweets or replies; log-data coming from web activities, sensing and event data coming from IoT; and other forms. It is an important challenge to integrate this data for value-added analytics and positive decision making.

- **Uncertainty of data**

The data gathered from heterogeneous sources like sensors, social media, web activities, and internal-records is inherently uncertain due to noise, incompleteness and inconsistency (e.g., there are 80% - 90% missing links in social networks and over 90% missing attribute values for a doctor diagnosis in clinic and health fields). Efficient analysis to discover value from these huge amounts of data demands tremendous effort and resources. However, as the volume, variety and velocity of the data increases, the uncertainty inherent in the data also increases, leading to doubtful confidence in the resulting analytics and predicted decisions.

- **Scalability**

The volume of data is drastically increasing and therefore an important challenge is to deal with the scalability of the data. It is also important to develop efficient analytics solutions and architectures that can scale up with the increasing data without compromising the accuracy or efficacy. Most of the existing learning algorithms cannot adapt themselves to the new big-data paradigms like dealing with missing data, working with partial data access or dealing with heterogeneous data sources. While the problem complexity of big data is increasing at a very fast rate, the

[7] Janev, V., Vranes, S.: The role of knowledge management solutions in enterprise business processes. J. Univ. Comput. Sci.11(4), 526–546 (2005)222.

[8] Dedić N., Stanier C. (2017) Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. In: Piazolo F., Geist V., Brehm L., Schmidt R. (eds) Innovations in Enterprise Information Systems Management and Engineering. ERP Future 2016. Lecture Notes in Business Information Processing, vol 285. Springer, Cham. https://doi.org/10.1007/978-3-319-58801-8_10

[9] Janev, V., Vraneš, S.: Applicability assessment of semantic web technologies. Inf. Process. Manag.47, 507–517 (2011)

computational ability and the solution capability is not increasing at a similar pace, posing a vital challenge.

- **Timeliness**

When looking for added business values, timing is of prime importance. It is related to capturing data, execution of analytics and making decisions at the right time. In a dynamic and rapidly evolving world, a slight delay (sometimes microseconds) could lead to incorrect analytics and predictions. In an example case of a bogus online bank transaction, the transaction must be disapproved in a timely manner to avoid possible money loss.

- **Data security**

Data storage and exchange in organizations has created challenges in data security and privacy. With the increasing sizes of data, it is important to protect e.g. transaction logs and data, real-time data, access control data, communication and encryption data. Also, it is important to keep track of data provenance, perform granular auditing of logs, and access control data to determine any misuse of data. Besides, the difference between legitimate use of data and customer privacy must be respected by organizations and they must have the right mechanisms in place to protect that data. Depending on the type of data being processed, security can sometimes be a crucial component that requires special attention.

- **Storage management**

Following the problem with processing, storage management is another unavoidable barrier regarding big data. Storing the huge quantity of data between its acquisition, processing and analysis requires gigantic memory capacity, thus rendering traditional solutions obsolete.

- **Data and system complexity**

The inherent complexity of big data (data complexity) makes its perception, representation, understanding and computation far more challenging and results in sharp increases in the computational complexity required compared to traditional computing models based on total data. The design of system architectures, computing frameworks, processing modes, and benchmarks for highly energy-efficient big data processing platforms is the key issue to be addressed in system complexity[10].

- **Responsiveness, reliability and correctness**

Contemporary cloud-based solutions are also considered to be on the edge of feasibility since responsiveness can be a critical issue, especially in real-time applications, where upload speeds are considered the main bottleneck. When simultaneously working with different data sources, the reliability of collected data will inevitably fluctuate with missed, partial and faulty measurements being unavoidable, resulting in serious potential trouble later on in the workflow, such as in the analytics stage. Hence, high-quality data management (i.e. data cleaning, filtering, transforming and other) actions are mandatory at the beginning of the process. Besides reliability, the correctness of the data is considered to be a key aspect of big data processing. High volumes, unstructured forms, the distributed nature of data in NoSQL data management systems and the necessity of near-to-real-time responses often lead to corrupted results with no method being able to guarantee their complete validity.

- **Other quality dimensions**

[10] Jin, X., Wah, B., Cheng, X., Wang, Y.: Significance and challenges of big data research. Big Data Res. **2**(2), 59–64 (2015)

Other quality dimensions that impact the design of a big data solution are completeness, consistency, credibility, timeliness and others. For instance, in real-time applications (e.g. stock market, financial fraud detection and transactions parsing, traffic management, energy optimization etc.), quick responses are required and expected immediately because the retrieved information can be completely useless if it is derived with high latency with respect to the collected data. An additional challenge from the human-computer perspective is the visualization of results. Although various ways in which the data can be displayed do not affect the data processing segment in any way, visualization is stated in the literature as a crucial factor because without adequate representation of the results, the derived knowledge is useless.

2.2 Knowledge Graphs-based Processing (Grand Challenges)

Knowledge Graphs (KGs) are one of the key trends among the next wave of technologies. Many definitions exist of what a Knowledge Graph is, and in *Chapter 2 Knowledge Graphs: The Layered Perspective (Luigi Bellomarini, Emanuel Sallinger, and Sahar Vahdati)* will choose a particular perspective, which we will call the layered perspective, and three views on Knowledge Graphs.

For more info, please see Chapter 2 in Janev, V., Graux, D., Jabeen, H., Sallinger, E. (Eds.) Knowledge Graphs and Big Data Processing. Lecture Notes in Computer Science vol. 12072, pp. 1-208. Springer International Publishing. ISBN 978-3-030-53198-0. DOI: https://doi.org/10.1007/978-3-030-53199-7

Additionally in the book the grand challenges are discussed related to processing big data with knowledge graphs-based tools. The UBO team has delivered the SANSA as a comprehensive system that addresses several challenges and provides libraries for the development of a knowledge graph value chain ranging from acquisition, distribution, and querying to complex analytics.
Grand challenges are discussed at the end of these chapters:

- *Chapter 5 Federated Query Processing (Kemele M. Endris, Maria-Esther Vidal, and Damien Graux)*
- *Chapter 7 Scalable Knowledge Graph Processing using SANSA (Hajira Jabeen, Damien Graux, and Gezim Sejdiu)*
- Chapter 8 Context-Based Entity Matching for Big Data (*Mayesha Tasnim, Diego Collarana, Damien Graux, and Maria-Esther Vidal)*

Grand challenges to face are:

- **Availability of data in RDF**

This challenge is to be linked to the research directions on federated queries (Chapter 5) and to the design of mappings to pave the road for datalake-oriented solutions such as the one presented by Mami et al.[11] . While the representation of data as knowledge graphs has gained lots of traction and large-scale knowledge graphs are being created, a majority of data being created and stored is not-RDF and therefore challenges such as the necessary efforts for data cleaning, and/or data maintenance should be taken into account.

- **RDF and query layer**

[11] Mami, M.N., Graux, D., Scerri, S., Jabeen, H., Auer, S., Lehmann, J.: Squerall: virtual ontology-based access to heterogeneous and large data sources. In: Ghidini, C., et al. (eds.) ISWC 2019. LNCS, vol. 11779, pp. 229–245. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30796-7 15

The distributed context requires smart partitioning methods (see Bernstein and Melnik[12], Kakas et al.[13] for detailed taxonomies) aligned with the querying strategies. One possibility would be to have dynamic partitioning paradigms which could be automatically selected based on data shape and/or query patterns, as envisioned in Alu¸c [14].

- **Responsiveness**

In a distributed context, processes often share resources with concurrent processes, and therefore the definition itself of what is a good query answer time may vary, as reviewed in the context of distributed RDF solutions by Graux et al.[15]. One could think of basing this performance evaluation on use-cases.

- **Machine Learning and partial access to data**

Most machine learning algorithms generally require access to all the training data and work by iterating over the training data to fit the desired loss function. This is challenging in the distributed setting where one might need to use multiple local learners or query processors (each working on a subset of the data) and optimize globally over (or collect) partial local results. For very large-scale distributed data, this working model may not be suitable[16]. Hence, there is a strong need to develop fundamentally new algorithms that can work with partial access to the data.

- **Challenge on the Semantic Web itself**

At the moment, using W3C standards, it is hard to be as expressive as with Property Graphs. This has led to the creation of RDF*[17],[18] in order to allow Semantic Web users to express statements of statements within an RDF extension. These new possibilities imply that the current landscape incorporates this extension while guaranteeing the same performances as before.

2.3 Knowledge Graphs (KGs) and Artificial Intelligence (AI) technologies: Perspectives and Future

Knowledge Graphs (KGs) in recent years have become a melting pot of Artificial Intelligence (AI) technologies. Two at first glance very different types of AI play a particularly important role here: Knowledge Representation and Reasoning (KRR) and Representation Learning (RL). The area of KRR is often associated with traditional AI techniques, while RL is associated with what is typically summarized as machine learning (ML). Yet, while in Knowledge Graphs both types of technologies come together, there currently is a perceived disconnect between the areas of RL and KRR. Most of the research is currently concentrated on one area or the other, yet arguably representation learning is central to making use of knowledge representation and reasoning techniques in modern, scalable AI applications.

[12] Bernstein, P.A., Melnik, S.: Model management 2.0: manipulating richer mappings. In: SIGMOD Conference, pp. 1–12. ACM (2007)
[13] Kakas, A.C., Kowalski, R.A., Toni, F.: Abductive logic programming. J. LogicComput.2(6), 719–770 (1992)
[14] Alu¸c, G., Ozsu, M.T., Daudjee, K., Hartig, O.: Chameleon-DB: a workload-aware robust RDF data management system. University of Waterloo, Technical reportCS-2013-10 (2013)
[15] Graux, D., Jachiet, L., Geneves, P., Layaıda, N.: A multi-criteria experimental ranking of distributed SPARQL evaluators. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 693–702. IEEE (2018)
[16] Papailiou, N., Konstantinou, I., Tsoumakos, D., Koziris, N.: H2RDF: adaptive query processing on RDF data in the cloud. In: Proceedings of the 21st International Conference on World Wide Web, pp. 397–400 (2012)
[17] Hartig, O.: Reconciliation of RDF* and property graphs. arXiv preprintarXiv:1409.3288(2014)185.
[18] Hartig, O., Thompson, B.: Foundations of an alternative approach to reification in RDF. arXiv preprintarXiv:1406.3399(2014)

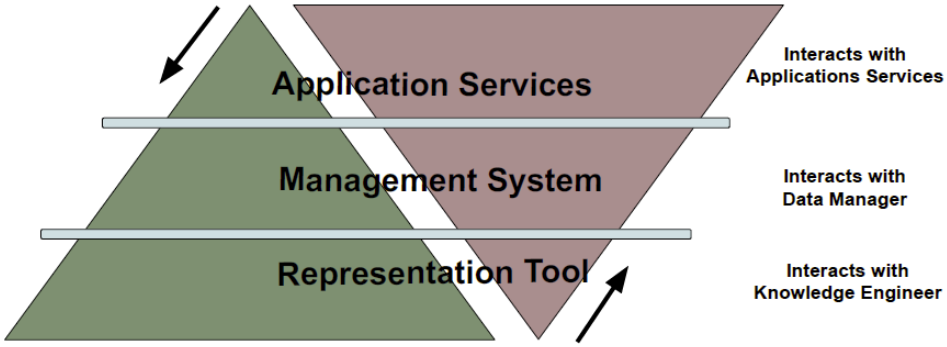


Figure 3. Ordered Pyramids of Views on KGs

LAMBDA researchers organized two foresight panels at the International Workshop on Knowledge Representations and Representation Learning (KR4L) 2020[19], part of the European Conference on Artificial Intelligence (ECAI) 2020.

Discussion topics:

The plenary talk (Dr. Emanuel Sallinger) was divided into four parts:

- **Modern Knowledge Graphs**
  - A motivation for the use of modern Knowledge Graphs.
  - Financial Knowledge Graphs as a particularly interesting area.
- **A Melting Pot of Technologies**
  - Types of technologies meeting in Knowledge Graphs (data exchange and integration, data wrangling, graph databases, business intelligence tools, reasoners, machine learning frameworks, etc.).
- **A Meeting Point of Research**
  - Areas of the narrower and wider research fields related to Knowledge Graphs, and in particular KRR and RL.
  - The evolution of these areas in the last few years.
- **Perspectives and Future**
  - We are going to consider multiple perspectives on Knowledge Graphs, including the *layered perspective* that looks at KGs as representation tools, management systems and application services.
  - The future of KRR and RL in KGs.

Many aspects of KRR and RL were discussed including:

- **The use of RL and ML techniques for KRR:**
  - *Traversing Knowledge Graphs with Good Old (and New) Joins*
  - *An Evolutionary Algorithm for Rule Learning over Knowledge Graphs*
- **The use of KRR in RL and ML:**
  - *Cluster Discovery from Sensor Data Incorporating Expert Knowledge*
  - *The Effect of Rule Injection in Leakage Free Datasets*
  - *A Performance Strategy: Multiple Slices of a KGE Model in Low Dimensions*

[19] https://project-lambda.org/KRRL-2020

- **Jointly use of KRR and RL/ML techniques:**
  - *Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs (Invited Paper)*
  - *Blockchains as Knowledge Graphs – Blockchains for Knowledge Graphs (Vision Paper)*

Two panels were organized as presented in Table 4.

Table 4. Panel discussions at KRRL, part of the European Conference on Artificial Intelligence 2020

Led by	Title
Luigi Bellomarini, Vice Director of IT Research, Central Bank of Italy	• **Emerging Topics in Academia and Industry**
Sahar Vahdati and Mojtaba Nayyeri (InfAI and University of Bonn)	• **Future Directions - Looking Ahead**

3. PUPIN Foresight Exercise 2020 - Preparatory Activities

3.1 Target Participants

The starting point for the foresight exercise was to design a list of scientists, leaders, decision makers from the regions. Initially a list of target participants has been created by brainstorming internally and reviewing the online profile of the selected people. After several internal discussions, the names have been shortlisted and the invitation for participating in the LAMBDA Foresight exercise has been sent out.

We have received very positive replies from leading scientists of the region to contribute.

3.2 Questionnaire

In the context of LAMBDA Foresight exercise, one of the early steps was to contact and collect information in different level about:

- The connection of research and industrial activities of the participants to Big Data
- The opinion of participants about aspects of Big Data in the past, current, and future

Each of these directions have been divided into several sections.
The questionnaire is still available online[20] and we discuss the individual questions and their replies below and in the following subsections.

[20] https://docs.google.com/forms/d/e/1FAIpQLSfIadFkOS8Q8SZhhFukcVT5Un8vWPCNH2tVlPjd5VznwoKpcA/viewform

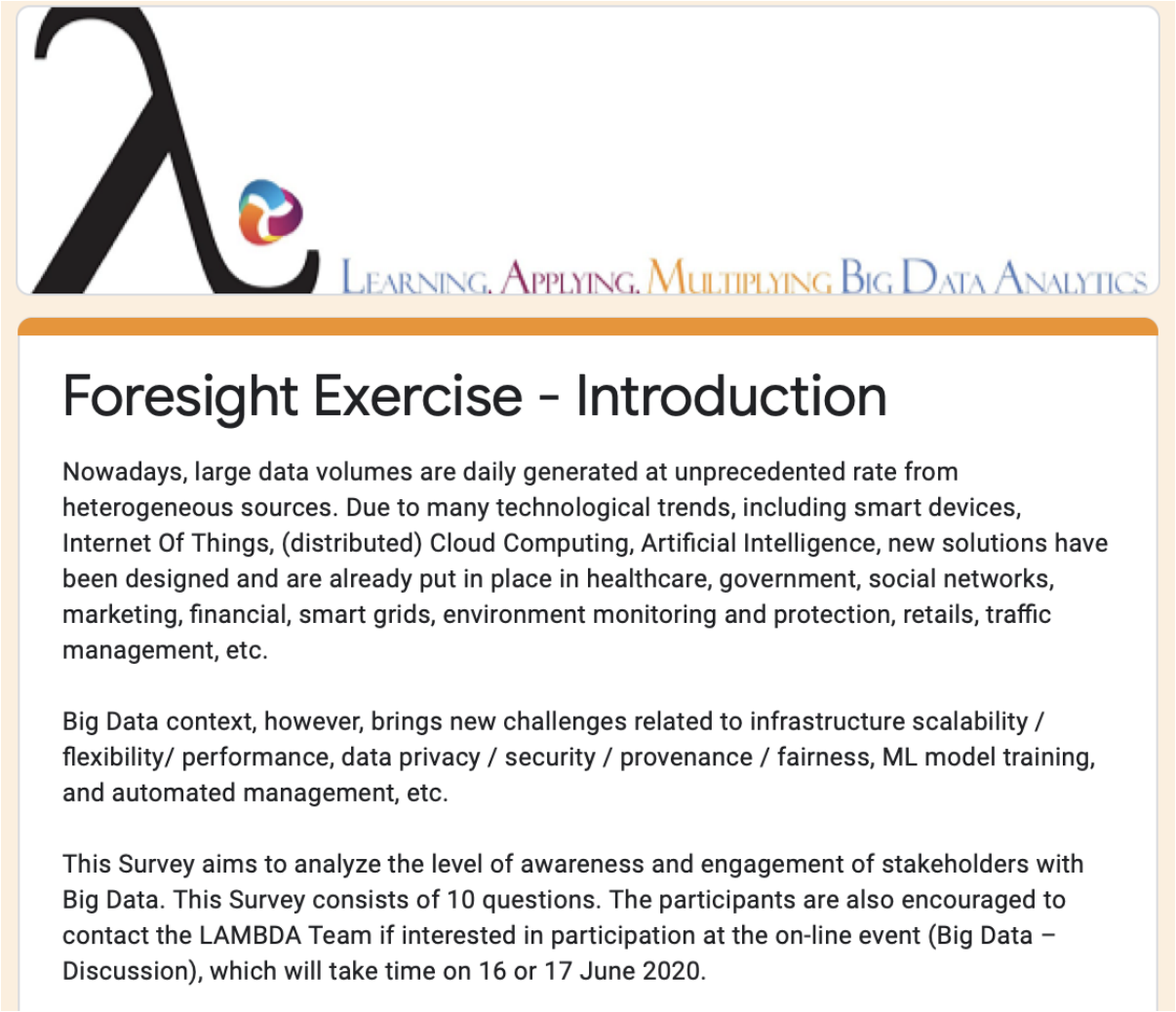


Figure 4. Foresight Questionnaire

3.3 Replies to the Questionnaire

Here we present the answers to the questionnaire from the domain experts. Overall, we had twelve replies and the results were collected and influenced the design of the foresight panel and future plans.

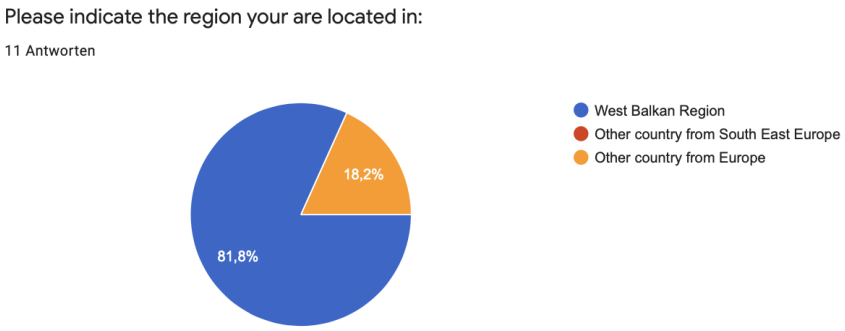


Figure 5. Structure and origin of participants

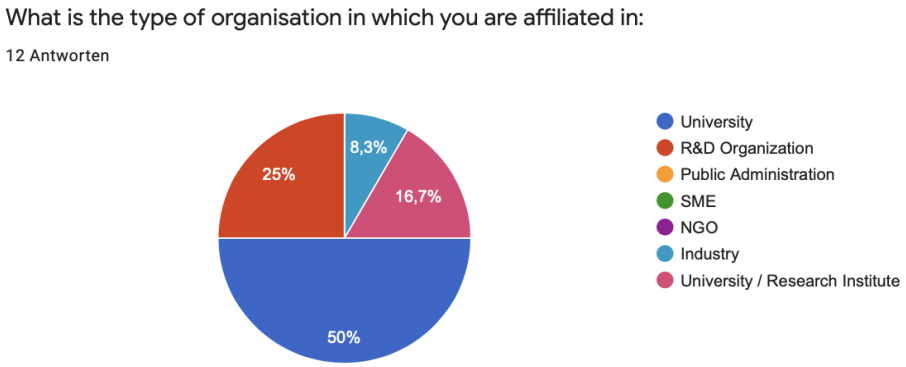


Figure 6. Results to Question - What is the type of organisation in which you are affiliated?

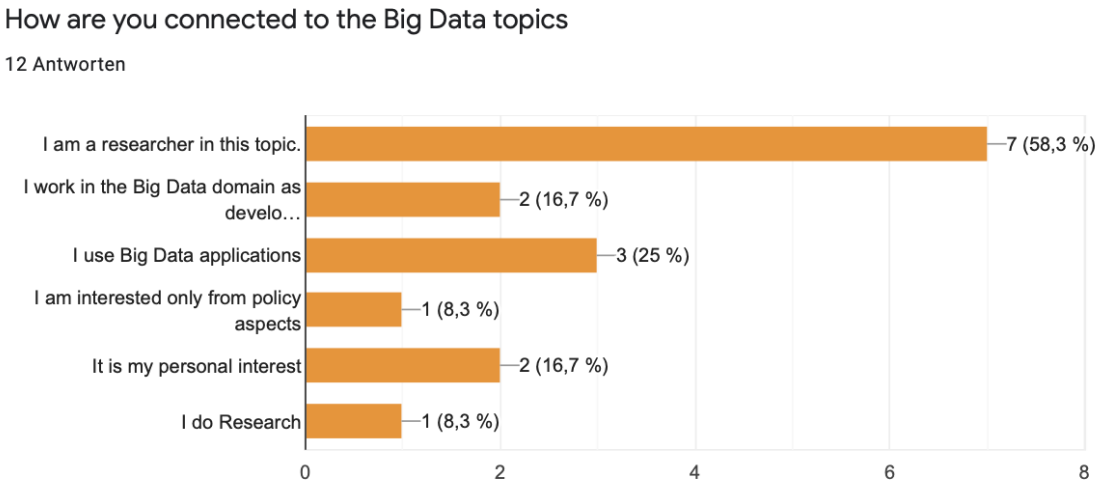


Figure 7. Results to Question - How are you connected to the Big Data topics?

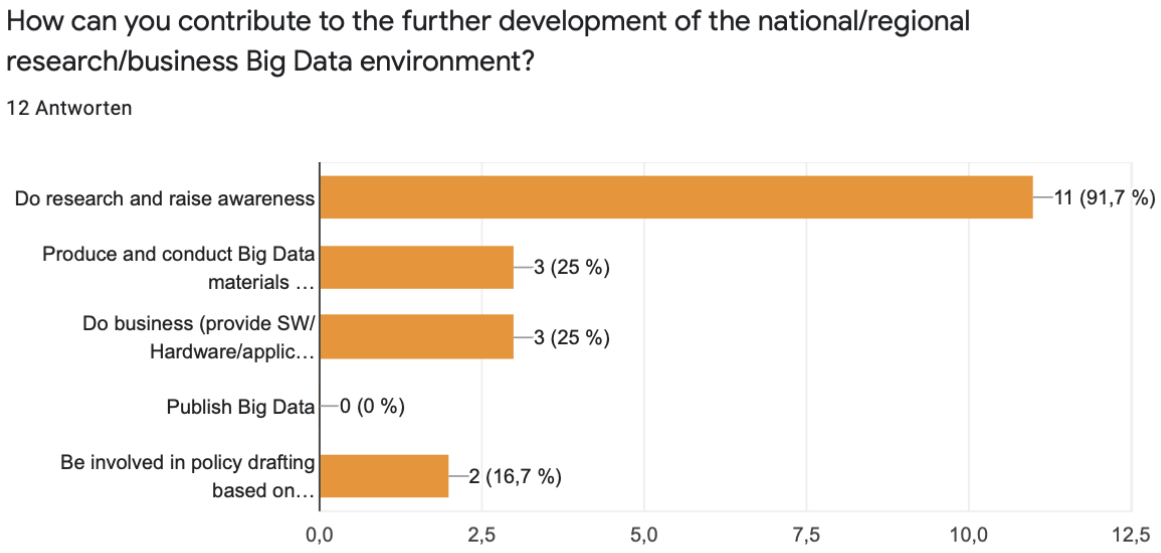


Figure 8. Results to Question - How can you contribute to the further development of the national/regional research/business Big Data environment?

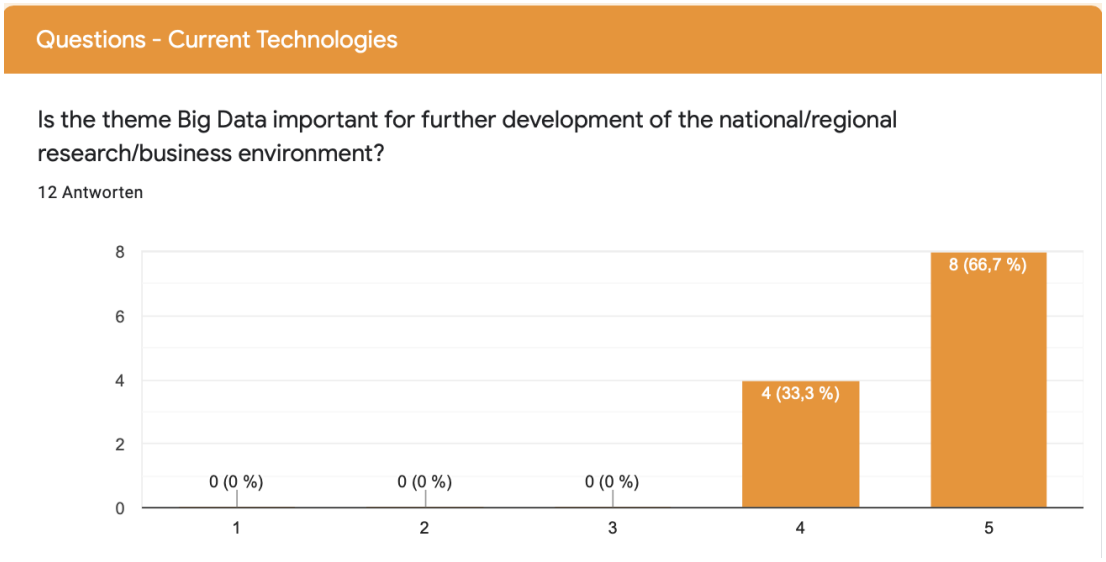


Figure 9. Results of inquiry - Further development of the national/regional research/development environment

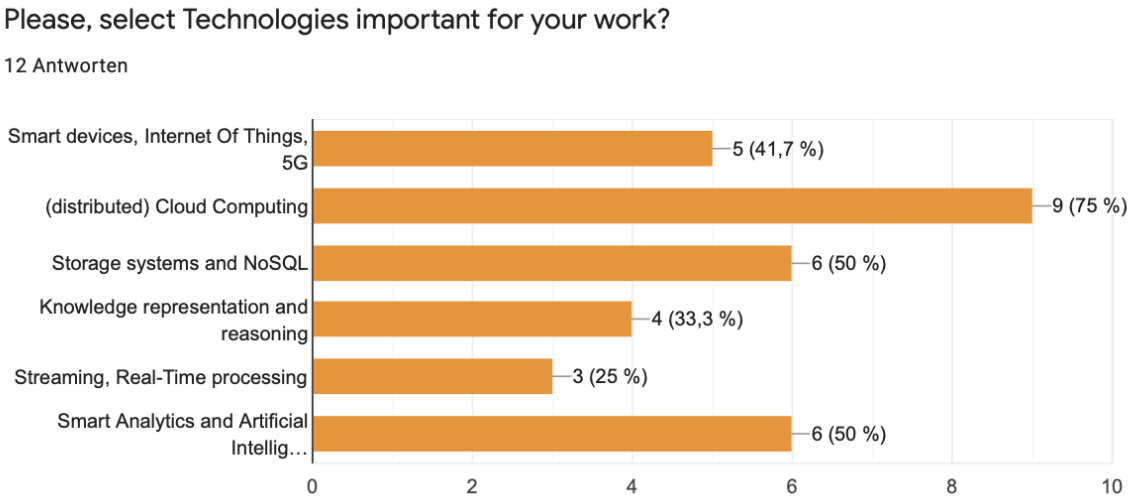


Figure 10. Results of inquiry - Technologies important for your work

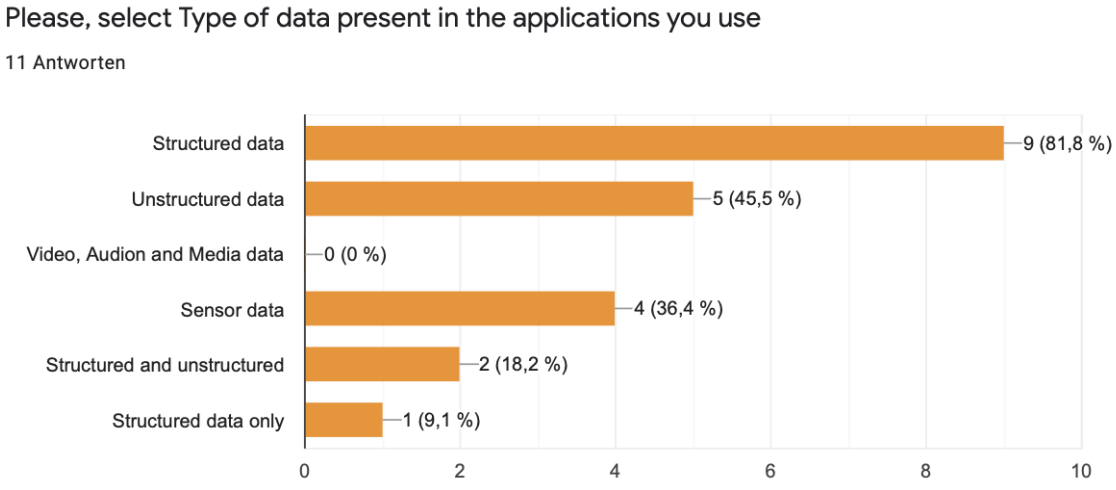


Figure 11. Results of inquiry - Type of data present in the applications you use

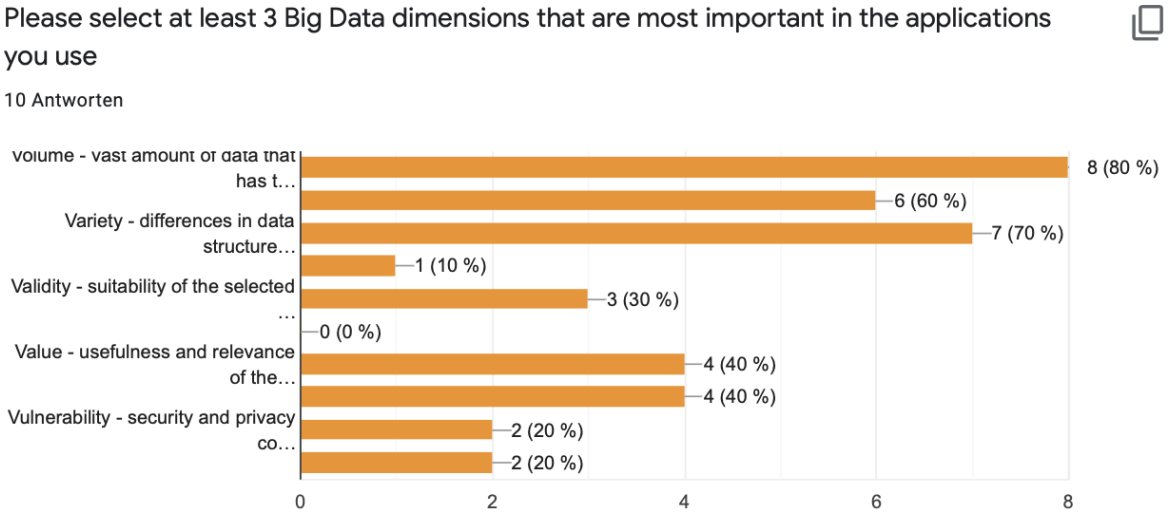


Figure 12. Results of inquiry - Big Data dimensions that are most important in the applications you use

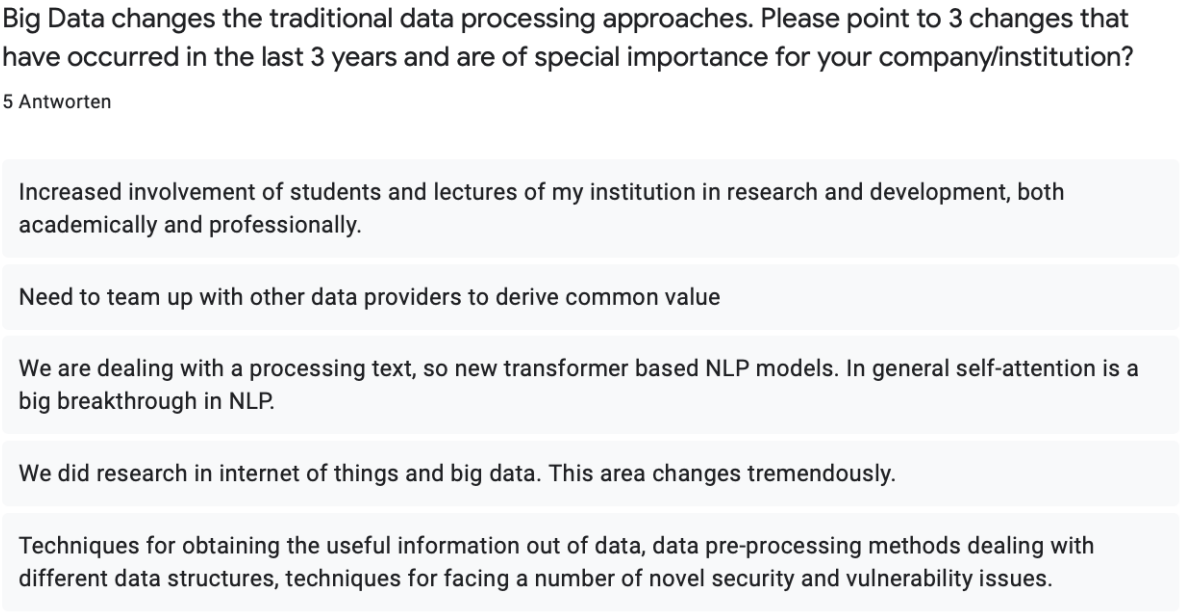


Figure 13. Results of inquiry - Changes that have occurred in the last 3 years

Questions - Emerging Technologies in your country

Please name at least 3 Big Data research trends that in your opinion will make the biggest breakthrough in the next 5 years

4 Antworten

Methods for improving veracity (provenance, blockchain, etc.); Methods to support distributed data processing (retaining data at source, security); Methods to support interoperability (easier visualisation and extension of common data models).

Big Data + AI

Research and development in the domains of: security and privacy protection, proper synchronization of the collected large data volumes, estimation of the data validity, the extraction of information that is most valuable for further analysis and processing of data in timely manner, solution customization for specific case needs, etc.

Machine learning and AI, Analytics, Visualization of data...

Figure 14. Results of inquiry - Big Data research trends that in your opinion will make the biggest breakthrough in the next 5 years

Please select 3 industry domains that in your opinion will have the biggest impact from emerging technologies in your country

5 Antworten

Energy, telecommunications and of course IT industry

Smart Manufacturing, Health, Transport

Robotics

Communication systems and network technologies (security, connection QoS, mobility, service availability); Energy and utility (smart meters, energy consumption control, active user involvement in service maintenance and enhancement) industry; Banking systems (Security - sophisticated intrusion detection and prevention systems, wide privacy concerns, advanced risk analysis and more accurate analytics techniques).

IT services in Healthcare, Energy sector, Marketing, Business ...

Figure 15. Results of inquiry - 3 industrial domains that will have the biggest impact from the emerging technologies

Please name at least 3 Big Data tools that you would like to examine in the next 3 years e.g. Hadoop, Spark, link, Storm, Kafka, Cassandra, Tableau, MongoDB, Apache Zookeeper, etc.

4 Antworten

TimeLion (Kibana), TensorFlow, Docker Swarm

Tableau, Apache Zookeeper

Storm, Spark, RapidMiner, R

Hadoop, Tableau, MongoDB...

Figure 16. Results of inquiry - Big Data tools worth testing

What is your opinion on required governmental actions for facilitating further development of Big Data research in your region?

5 Antworten

Bringing legislation that will enable educational and research institutions to have a much higher level of collaboration and synergy with industry.

Support datamarket setups to promote safe data exchange and sharing, incentivizing both research and industry to improve technology further

Opening more data

The necessary governmental action would be to provide more funds for the purchase of best possible hardware and software for the research and development in this area.

Government need to attract development using traditional and big data for our countries, but some of initiatives are only on paper.

Figure 17. Results of question - What is your opinion on required governmental actions for facilitating further development of Big Data research in your region?

Would you like to participate in the online LAMBDA Event that will be organized by the LAMBDA consortium on June 16 2020?

11 Antworten

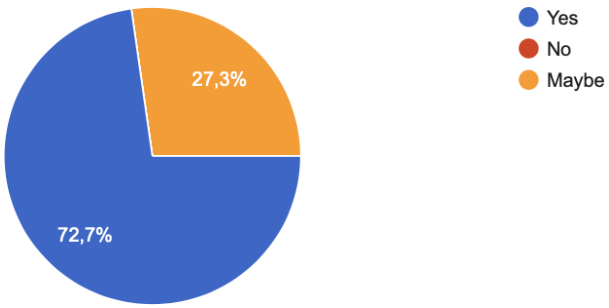


Figure 18. Results of inquiry about willingness to participate in the BDA School

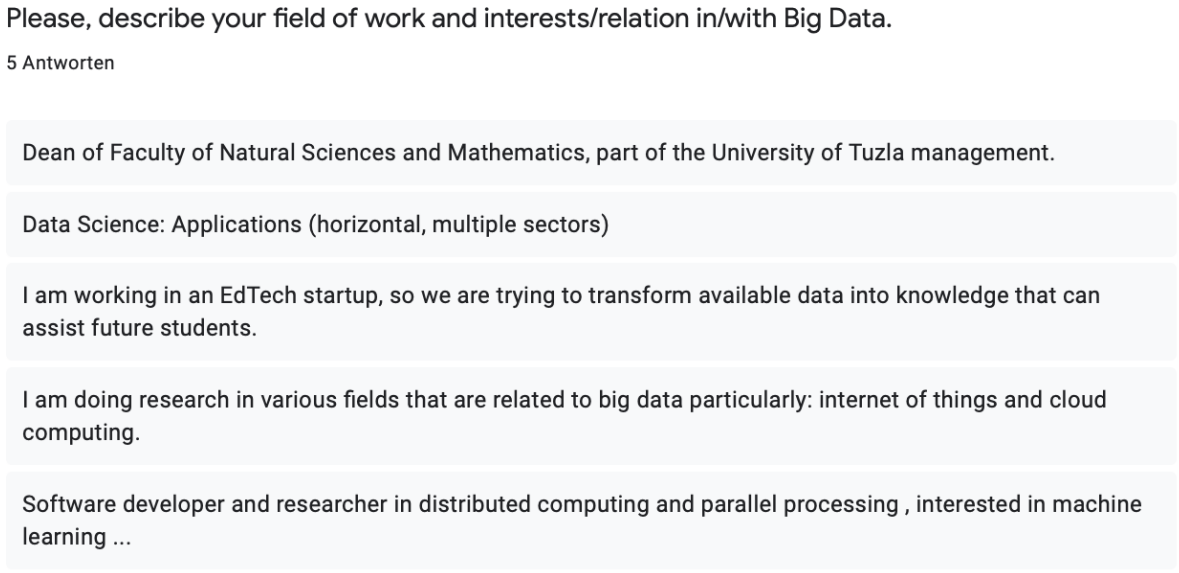
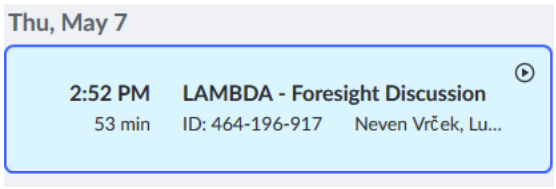


Figure 19. Results of inquiry about field of work

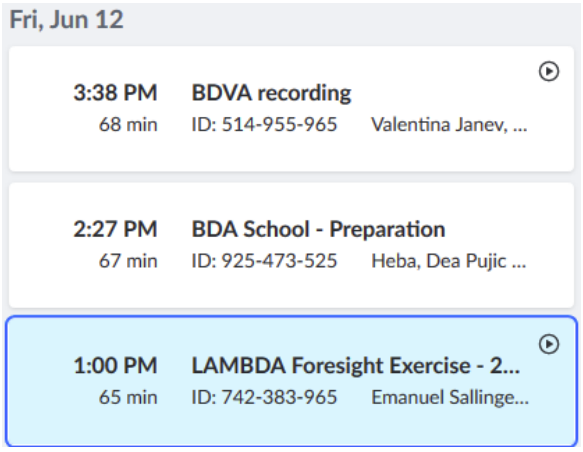
3.4 Meeting 1



In order to get to know the panel participants in person and to have an initial introduction on the ultimate objective of the foresight exercise, the Oxford team organized a meeting on 7 May 2020. A diverse selection of potential participants in the foresight exercise was invited.

In this meeting, we discussed the current status of Big Data in the Balkan area in general and the required needs for future developments. Then, we discussed the questionnaire and asked for feedback on that. Each of the participants were asked to prepare presentation slides including content to present themselves in the foresight exercise session and explain their connection to Big Data topics.

3.5 Meeting 2



In order to coordinate for the main foresight session, a second meeting has been organized on 12 June 2020. All the panel participants were present in this meeting and the Oxford team explained and simulated the foresight exercise with them.

The provided presentation slides have been discussed. The Oxford team made unification for the presentation content and everyone exercised how the actual foresight session will be.

4. PUPIN Foresight Forum, BDA School 2020

4.1 Target Participants

The foresight exercise has been organized by the University of Oxford, and moderated by Prof. Emanuel Sallinger and Dr. Sahar Vahdati.

Figure 20. Foresight panel – Opening

4.2 Panel Participants

Table 5 briefly introduces the panellists.

Table 5. Foresight Panel, June 2020 – List of participants

Participant	Short biography
Prof. Dr.sc. Neven Vrček	• Professor at University of Zagreb Faculty of Organization and Informatics (FOI). He is also the president of the Sectoral council for information sciences at Ministry of Science and Education, Croatia. He serves as a member of the Supervisory Board at the Ruđer Bošković Institute and is the former dean of FOI and Head of Department of IS Development. In his research profile, he co-authored more than 100 journal and conference papers and several books. He is also the experienced manager and member of supervisory board with a demonstrated history of working in the high education sector, information technology, project management and services industry. His research interests include: Smart Industry, Digital transformation, Open Data, Internet of Things, e-business, Entrepreneurship, Technology.
Prof. Dr. Dimitar	• Head of Information systems and network technologies Department at the faculty of Computer Science and Engineering, Cyril and Methodius

Trajanov	University –Skopje. He is the leader of Social Innovation Hub and the CEO of MindTRON Technologies. He has been co-authoring more than 150 journal and conference papers and he is the author of seven books. Involved in more than 60 research and industry projects. His research interests include Data Science, Machine Learning, NLP, FinTech, Semantic Web, Open Data, Sharing Economy, Social Innovation, e-commerce, Entrepreneurship, Technology for Development, Mobile Development, and Climate Change.
Dr. Luka Filipović	• Scientist at the IT center of the University of Montenegro, Software department. His main activities are at the faculty of electrical engineering, and he has several projects such as FP6/FP7/H2020 projects. His main topics are around Distributed and parallel computing.
Prof. Dr. Vedad Pašic	• Professor in the field of pure mathematics at the University of Tuzla since 2016. He teaches a variety of courses at the Department of Mathematics at the University of Tuzla. He is the Dean of the Faculty of Natural Sciences and Mathematics. His main areas of research are operator theory and theories of gravity, but he has also published work in functional analysis and numerical analysis.
Dr. Nikola Tomasević	• Project manager at Mihajlo Pupin Institute where he works since 2007. He received a Dipl. Ing. Degree in July 2007 at the School of Electrical Engineering, University of Belgrade, Serbia. In December 2013, he defended his PhD thesis at the Department of Communications and Information Technologies of the School of Electrical Engineering, University of Belgrade. He is involved in technical management and research activities of R&D projects in various domains. Currently, under the H2020 Work Programme, he is managing two H2020 projects (REACT and RESPOND) and taking active role in several other H2020 projects (such as InBetween, IDEAS and LAMBDA). So far, he took part in a number of EU H2020, FP7 and FP6 projects (H2020 SlideWiki, FP7 EPIC-HUB, CASCADE, EMILI, Reflect, and FP6 Web4Web) and also was actively involved in R&D projects financed by the Ministry of Science and Technological Development of Serbia (SOFIA and AMICA). In his scientific career, his research activities were focused on energy efficiency, emergency management, recommendation and support systems, semantic web technologies, mobile communication systems, learning analytics and natural language processing. He (co-)authored more than 40 scientific and technical papers as journal, conference and workshop contributions. He also serves as a reviewer for respectable journals (Applied Energy (Elsevier), Transactions on Wireless Communications (IEEE), International Journal of Neural Systems (World Scientific), Artificial Intelligence Review (Springer), etc.), as a PC member and session chair of international conferences (such as TELFOR and ICTERI).

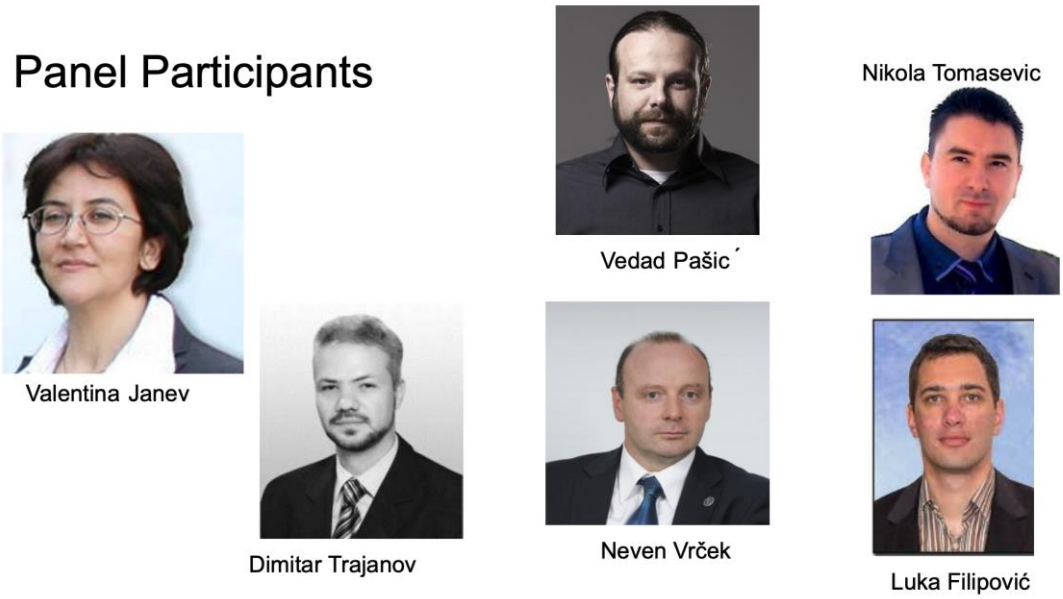


Figure 21. Foresight panel – Participants

Figure 22. Foresight panel – Dr. Luka Filipović

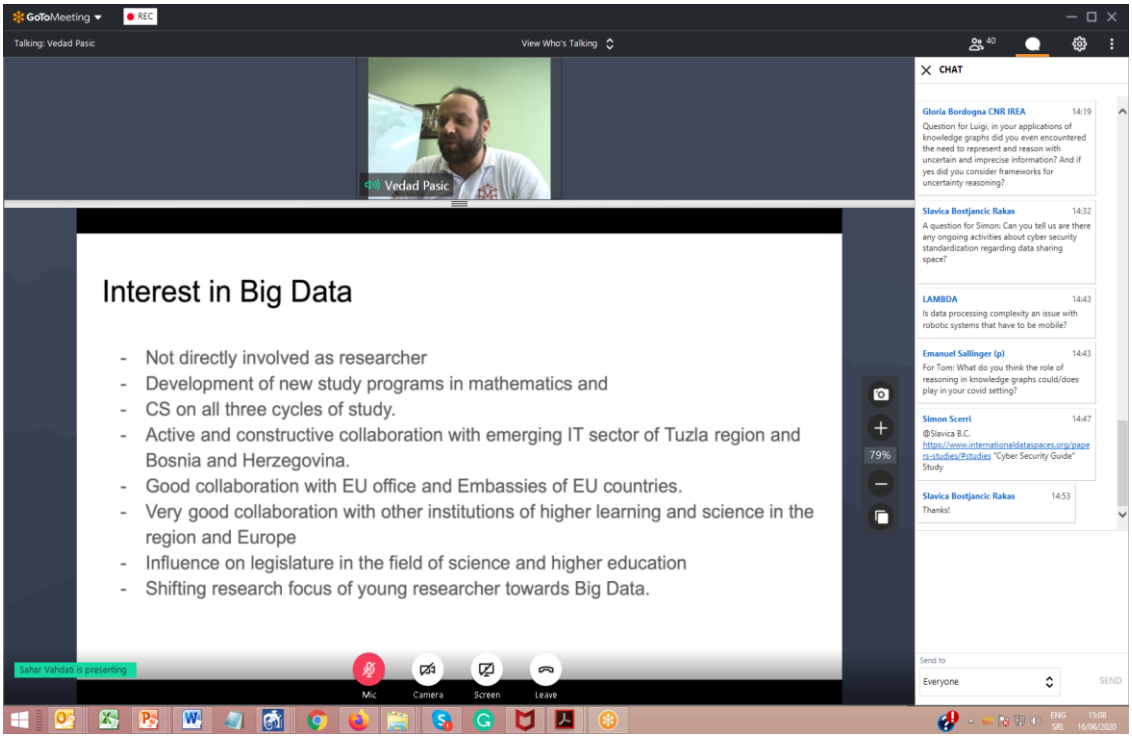


Figure 23. Foresight panel – Discussion

4.3 Panel Discussion

In preparation of the panel discussion, the Oxford team had designed a series of initial questions which could be asked to the participants. These questions have been discussed together with the other members of the consortium and finalized. A set of three main questions has been selected to be asked in three rounds and two back-up questions have been reserved for the case the discussion goes fast and time remains.

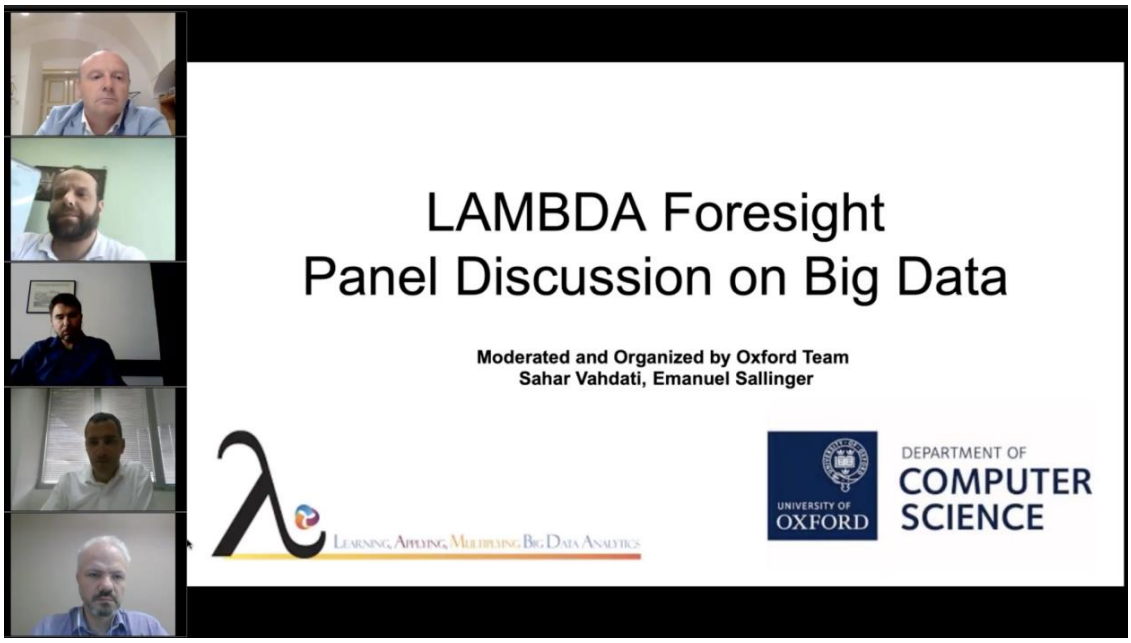


Figure 24. Foresight panel – Opening

Question 1. The first question had a focus on industry domains. From the panel participants, the ones who are very close to the industry and have strong network have been selected to be addressed. The discussion started by posing the question to the participants.

What are the industry domains that in your opinion will have the biggest impact from emerging technologies in your country/the region?

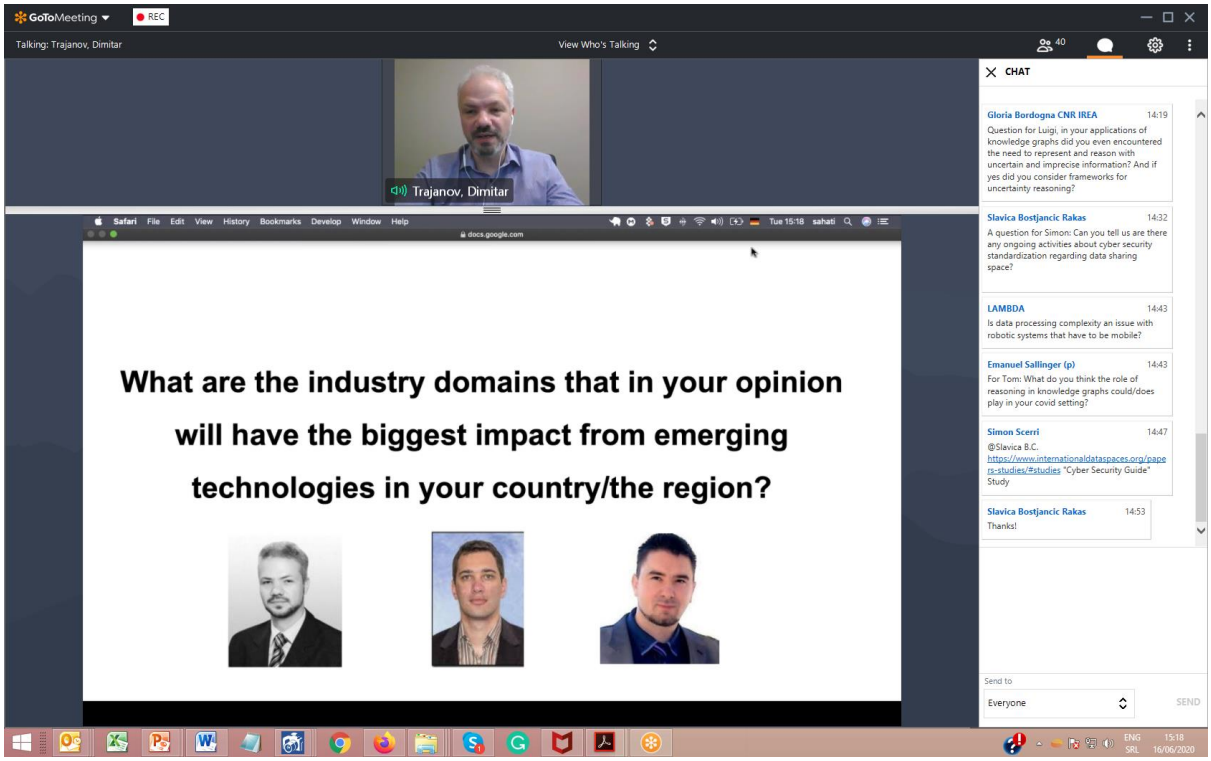


Figure 25. Foresight panel – Questions

- **Dimitar Trajanov:** Everyday, there is more and more data around us. Taking finances as an example, one can see how it influences our daily life from investment to risk assessments. Big data and NLP and ML are making a lot of changes and everyday there are new products that are closer and closer to the needs of the users. There are lots of chatbots for example that are replacing lots of human-based conversations and the ethical implications of that. Much more automated services are accessible by deep analysis of Big data.

- **Luka Filipović:** By emphasizing on life science which has global impact and the West Balkan region, he highlighted bioinformatics. The essence of Big Data has been discussed

by him in the topics of epidemy, biology, pharmacology, and he connected it to the currently ongoing global crises, COVID-19. He gave the example of software that was used to perform genetic analysis empowered by Big Data where they could combine multi-gene analysis with multiple options. In his opinion, biomedical and bioinformatics topics together with Big Data technologies are the future.

- **Nikola Tomasevic:** He pointed out about very many required technologies in several projects that are even running currently in many domains. The overview of trends in the energy domain was given by him such that unlocking flexibilities on the demand side, indoor energy efficiency and many of those require better policies. Nikola gave an extensive explanation about statistics of the results in total consumption and applied analytics using Big data. Coming from the domain of smart energy, any change in behavior and ultimate results require historical and stream data which becomes huge quantities of data points with different qualities. Therefore, it is highly needed to be considered in all regions.

Question 2. The second question has a focus on governmental actions.

- **Vedad Pašic:** As he mentioned, he is deeply involved in governmental actions in research and the situation about related topics to Big Data is not impressive. In preparation for the foresight panel, he consulted with several people. The situation is that science is not just something that one can through money to have only education but the actual impact with a concrete application. As the world is facing the crisis right now, the investigation in academia and connecting them to actual applications and companies who are actually doing the work in the market can be the biggest challenge that one can provide to legislators. Another goal is providing the possibility for young people to work hand in hand with companies and to also deepen their knowledge. As he pointed out, governments need to let the scientists take care of this.

- **Valentina Janev:** There is still a big need for more and more panels where people can gather and mastermind. Most of the conferences are not particularly focusing on these. She mentioned that on paper, many people talk about Big Data, however in practice there is a big gap with real possibilities. There are however, initiatives and data strategies and Pupin is involved. She mentioned that she is missing more discussions in these topics also involving governmental sectors.

Question 3. The third question has a focus on educational policies and required guidelines.

- **Neven Vrček:** He focused on the higher education point of view. He himself has been working on the development of higher qualification standards by providing support on many topics including Big Data and AI. In a broad project, they provided extensive educational services. In the course of this project, he together with the executive board tried to combine their own knowledge with knowledge from companies gathering their needs from the labor market. For example exploring the needs of skills when they hire experts in the Big Data area. To combine this into a multi-phase approach of life-long learning and multi-disciplinary education is what they learned through this project. It is hard to expect that lectures alone will “produce” experts in such new technologies. One needs to have a wider knowledge by providing educational services in such technologies for students from the law, medicine domain, etc. and interchange knowledge among them. Not only from research and implementation point of view but also from managerial point of view, combination of knowledge is needed nowadays and requires multi-phase learning including large set of techniques in many aspects, hard-core knowledge and anything that is applied in the market.

- **Vedad Pašić:** He pointed out that governmental and hiring sectors want to have quick

solutions for such problems and want quickly made experts with all the required skills. However in practice there is a long term education needed. By interviewing people in the IT sector with master or PhD level in the region, he concludes that a lot of basic knowledge is required as well as fresh education in all of the new technologies. He also pointed out that such panel discussions and more events are needed such that leading people in the region can brainstorm and provide solutions.

What is your opinion on required governmental actions for facilitating further development of Big Data research in your country/the region?

What is your opinion on required actions for educational policies and designing new curricula in your country/the region?

Figure 26. Foresight panel – Questions (cont.)

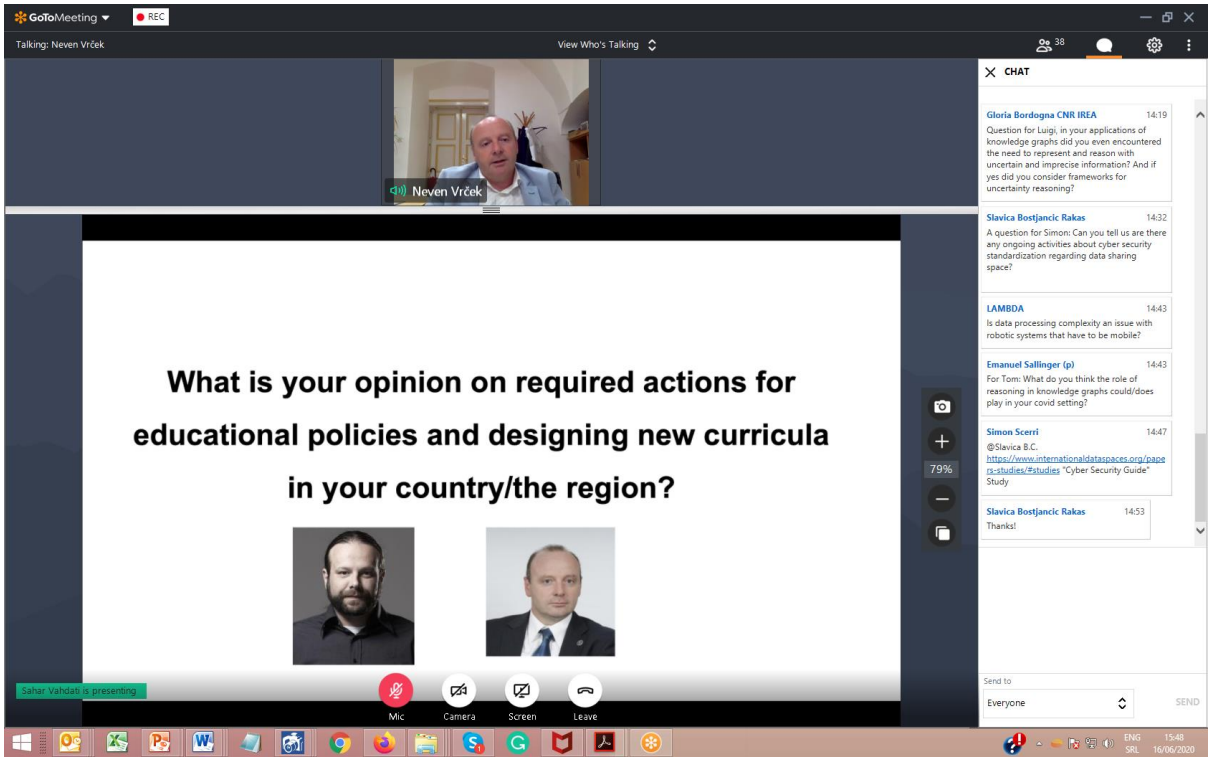


Figure 27. Foresight panel – Questions (cont.)

5. Other EU Policy-related Events

5.1 Research Data and Open Science in South-eastern Europe, December 2019, Belgrade, Serbia

Table 6. Roundtable "Research Data and Open Science in South-eastern Europe", December 2019, Belgrade, Serbia

Organizer	Participants
Data Center Serbia for the Social Sciences (Institute of Economic Sciences, Belgrade)	• CESSDA Consortium • Archives of Data in Social Sciences from West Balkan Countries • Representatives of Serbian government

Data Center Serbia for the Social Sciences (Institute of Economic Sciences, Belgrade), in collaboration with the CESSDA ERIC Training Group, the Croatian Center for Social Sciences (Faculty of Philosophy, University of Zagreb) and the Archives of Data in Social Sciences of Northern Macedonia (Institute for Sociological and Social Studies) Political and Legal Research, Skopje), organized a regional roundtable entitled "Legal and Ethical Constraints on the Management of Research Data and Open Science in Southeastern Europe" on December 11, 2019 at the Institute of Economic Sciences.

Dr. Valentina Janev on behalf of the LAMBDA consortium participated in the roundtable where very sensitive topics related to the implementation of GDPR (*General Data Protection Regulation* (EU) 2016/679 (*GDPR*) in Serbia and the Region were discussed.

Anne-Mette Somby, CESSDA ERIC gave an introductory talk on **GDPR and its implementation in data sharing in social sciences** and presented the **Data Management Expert Guide - an online tool to make research data FAIR**.

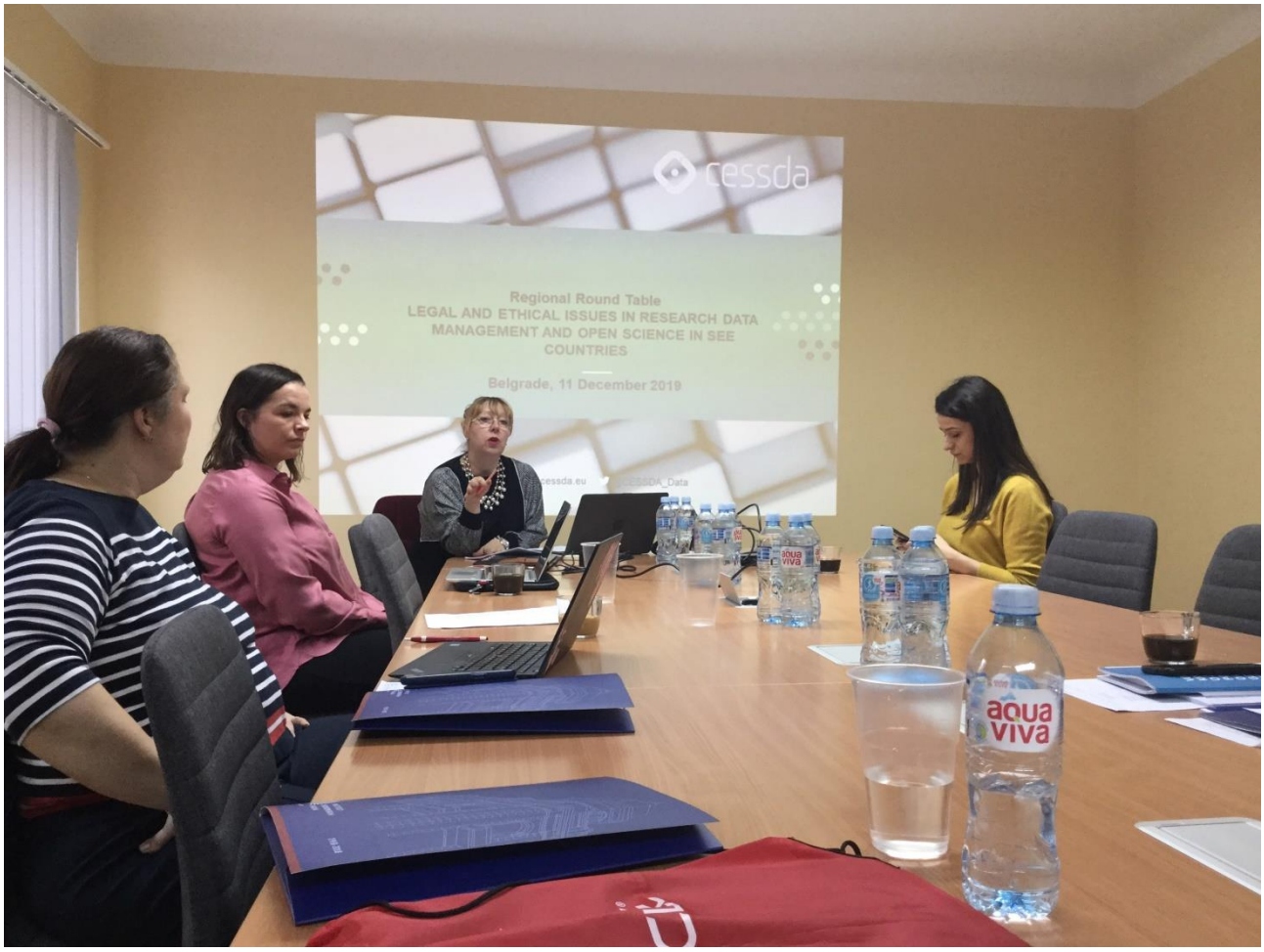


*Figure 28. Discussion on implementation of GDPR (*General Data Protection Regulation*) 2016/679 in Serbia and the Region*

5.2 Open Data - Opportunities and Innovation, Podgorica, March 2020

5.2.1 About

In cooperation with the ODEON project[21] and the Chamber of Economy and the Ministry of Public Administration of Montenegro, the PUPIN team organized an event 'Open Data - Business opportunities and innovation for SMEs and start-ups'. The event brought together experts from Serbia and Montenegro to discuss the European strategy for data, COM(2020) 66, 19.02.2020, and the Open Data Directive' (Directive (EU) 2019/1024).

Table 7. RoundTable "Open Data - Opportunities and Innovation", March 2020, Podgorica, Montenegro

Organizer	Participants
PUPIN Team in cooperation with the ODEON project	• Representatives of the public sector • Researchers • Start-ups

[21] https://odeon.interreg-med.eu/

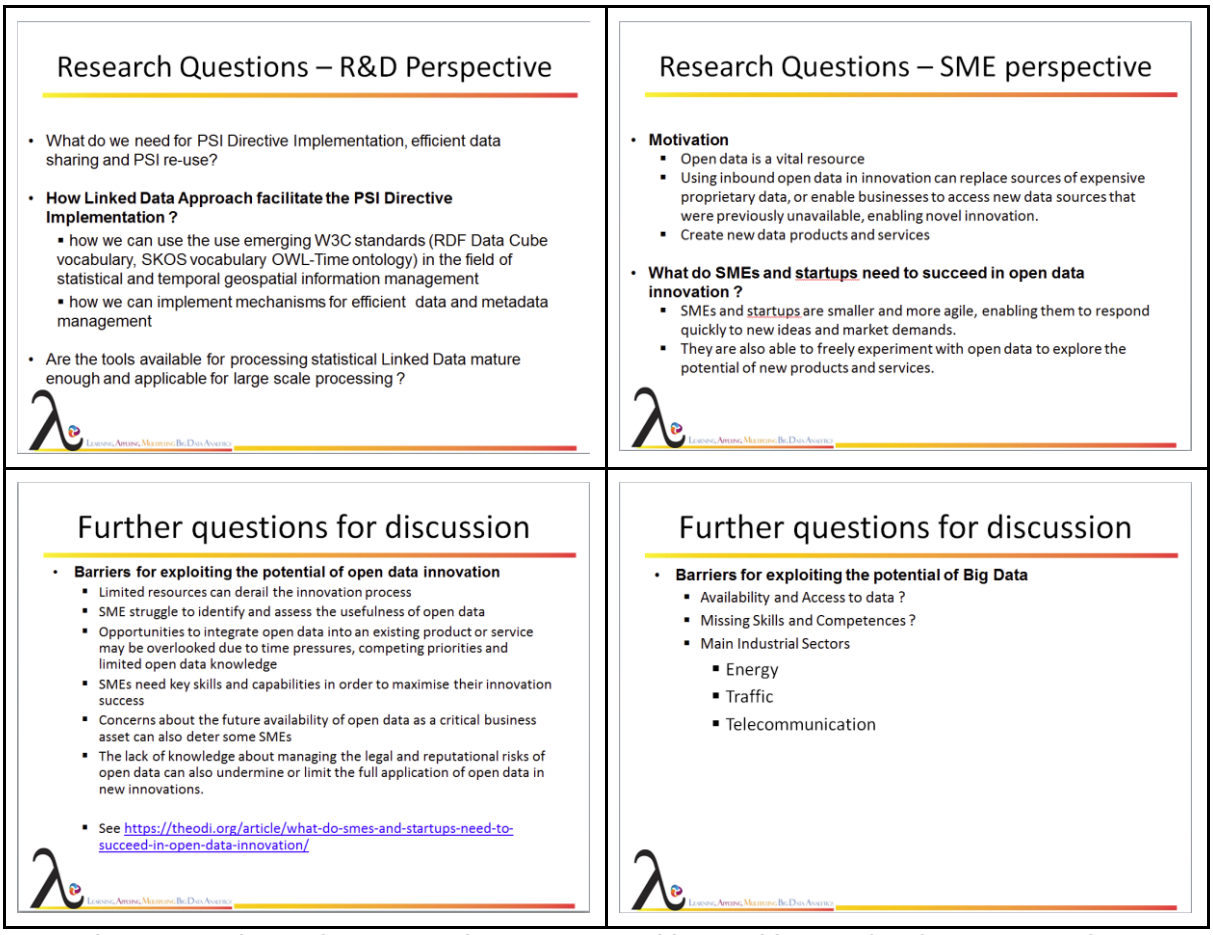


Figure 29. Discussion on 'Business opportunities and innovation for SMEs and start-ups'

5.2.2 Vision and Recommendations

The Government 3.0 paradigm for government operations aims at making the public administration systems more service-oriented and delivering customised public services by opening and sharing government-owned data to the public and businesses. In the last decade many European countries put forward Government 3.0 as a new paradigm, and as a result, improved efficiency in the provision of public services, increased transparency and interaction with citizens and society as a whole, but also created new businesses across Europe.

The Open Government Data Ecosystem is composed of

- Data providers, public organizations possessing datasets for which they include a description on one or more data portals, so that the datasets can be found more easily;
- Government Agencies, maintainers of national open data portals and/or data catalogues of the datasets made available by data publishers. Data portals make the description metadata of the datasets in their collection freely available to third parties. The national open data portals may also harvest collections of relevant datasets of other data portals(e.g., municipality data portals) and make them searchable via their user interface;
- Intermediary organizations, that provide data-sharing mechanisms across the EU e.g. metadata brokers, such as the European Data Portal, facilitate the collection and exchange of description metadata between data portals, as well as provide metadata harvesting, transformation, validation, harmonisation, publication services, translation of datasets and other services; and

- Public-Private partnership services that support the ecosystem, facilitate the collaboration between stakeholders and exploitation of open data/services;
- Data consumers who use data portals of their choice to search through various collections of datasets. The data portals allow the user to explore, find, identify and select the datasets coming from different data providers. Data consumers can also be systems (machines). Many different types of data consumers exist such as academia, media-journalists, NGOs or citizens willing, for example, to improve transparency or to add value to their services by combining data.

See also measures to boost data sharing and support European data spaces

- Proposal for a Regulation on European data governance (Data Governance Act)

or **EU PSI /Open Data Directives**

- Directive **2003/98/EC** of the European Parliament and of the Council of 17 November 2003 **on the re-use of public sector information,** https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32003L0098 - encouraged EU member states to make as much public sector information as possible for re-use, providing a common legislative framework for this area. The Directive was an attempt to remove barriers that hinder the re-use of public sector information throughout the Union.
- Directive **2013/37/EU** of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC **on the re-use of public sector information,** https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013L0037 - In 2013 it was amended to make it more aligned with open government data concepts („open by default“) and also to contemplate cultural heritage information, among other modifications.
- Consolidated text of the Directive 2003/98/EC on the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003L0098-20130717

Objectives of the recast in 2019 (*Directive* (*EU*) *2019/1024* of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information**) were** aimed to:

- **Reduce market entry barriers**, in particular for SMEs, by limiting the exceptions that allow public bodies to charge for the re-use of their data more than the marginal costs of dissemination;
- **Increase the availability of data** by bringing new types of public and publicly funded data into the scope of the Directive, such as data held by public undertakings in the utilities and transport sectors and research data resulting from public funding;
- **Minimise the risk of excessive first-mover advantage**, which benefits large companies and thereby limits the number of potential re-users of the data in question, by requiring a more transparent process for the establishment of public–private data arrangements;
- **Increase business opportunities** by encouraging the dissemination of dynamic data via application programming interfaces (APIs).

The PSI Directive *2019/1024* focuses on the economic aspects of the re-use of information rather than on access to information by citizens. It encourages the Member States to make as much information available for re-use as possible. The Directive introduces the concept of high value datasets, defined as documents the re-use of which is associated with important benefits for the society and economy.
They are subject to a separate set of rules ensuring their availability free of charge, in machine readable formats, provided via Application Programming Interfaces (APIs) and, where relevant, as bulk download, with the thematic scope provided in an Annex to the Directive.
The thematic categories of **high-value datasets** (Article 13 (1) of the Directive), are:

- ❑ Geospatial

- ❑ Earth observation and environment
- ❑ Meteorological
- ❑ Statistics
- ❑ Companies and company ownership
- ❑ Mobility

The Commission will adopt in 2021 a list of specific high value datasets by way of an implementing act, following an impact assessment.

Figure 30. Participants at Discussion on 'Business opportunities and innovation for SMEs and start-ups'

Analysis of Technological Challenges and recommendations for Transposing the PSI Directive in Serbia is given in book Chapter

- Valentina Janev (2019). Open Data: Challenges and Opportunities for Serbia. In I. Janev (Ed) Serbia: Current Political, Economic and Social Issues and Challenges. Nova Science Publishers, ISBN: 978-1-53615-060-5 (eBook), pp. 165-184. See also preprint

5.3 The Challenge and Opportunity of 5G in Serbia, November 2019

5.3.1 About

The event **5G in Serbia: Challenges and Opportunities** took place at the Institute Mihajlo Pupin on the 22nd of November 2019. The event was organized by the PUPIN Scientific Council (Dr. Vladimir Krstić).

Guest speaker: Goran Laovski, Republic Agency for Electronic Communications and Postal Services (RATEL).

Discussion topics:

- 5G changing the world - capabilities that differentiate 5G
- Special techniques and concepts used in 5G networks
- Potential 5G use cases
- 5G achievements and plans in Serbia
- Regulatory framework

Figure 31. Participants at event 5G in Serbia: Challenges and Opportunities

5.3.2 Vision and Recommendations

On the 22nd June 2019 Telenor has launched the first 5G base station in Serbia in Science Technology Park Belgrade, creating a 5G test environment that can be used by domestic and foreign companies, start-ups and students of technical faculties to develop technological solutions for the future.

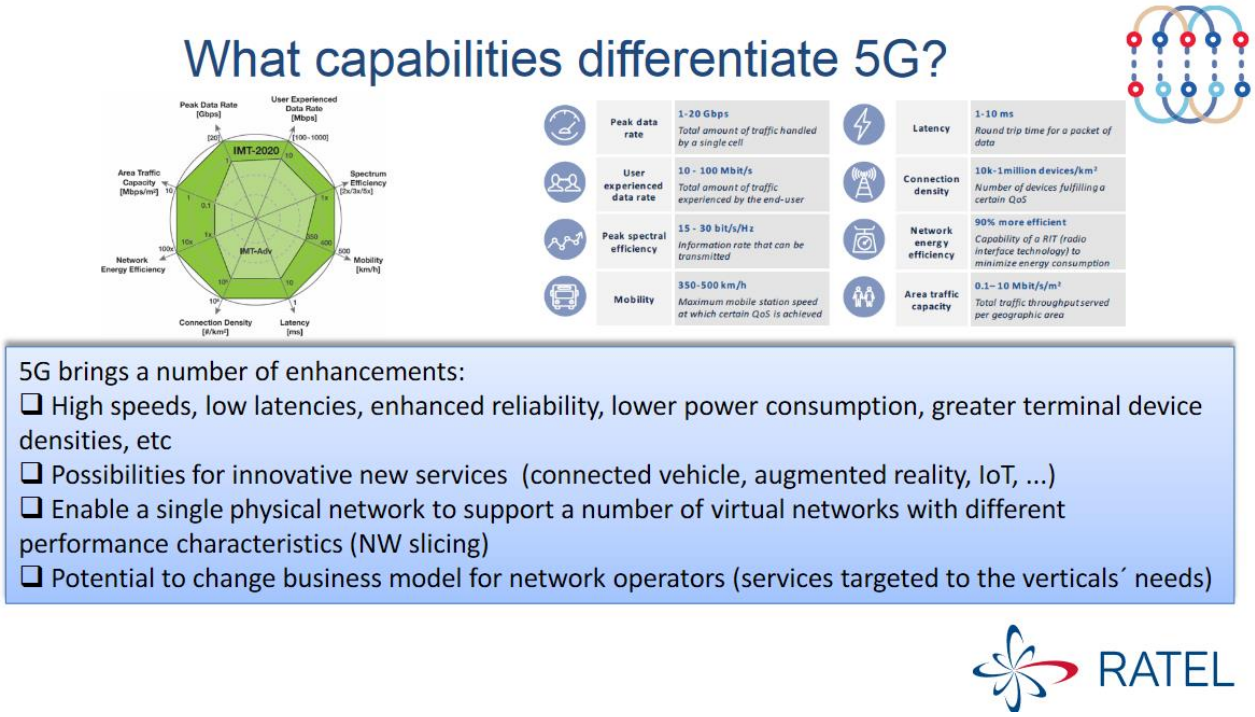


Figure 32. RATEL presentation at PUPIN (slide)

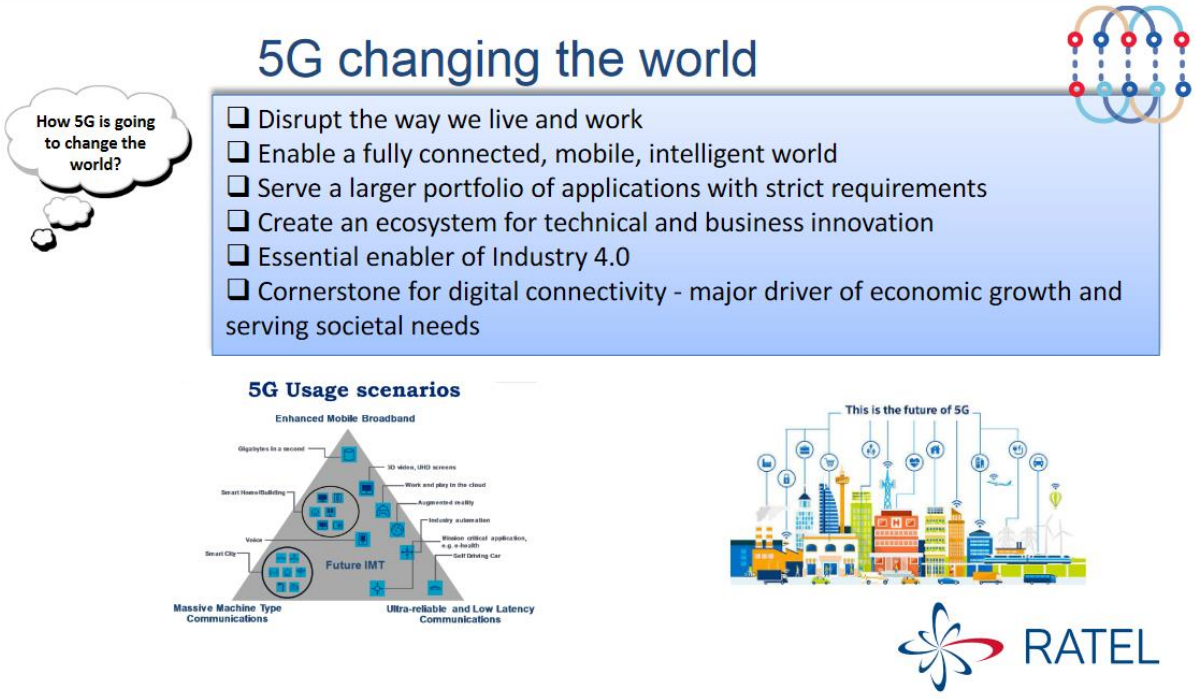
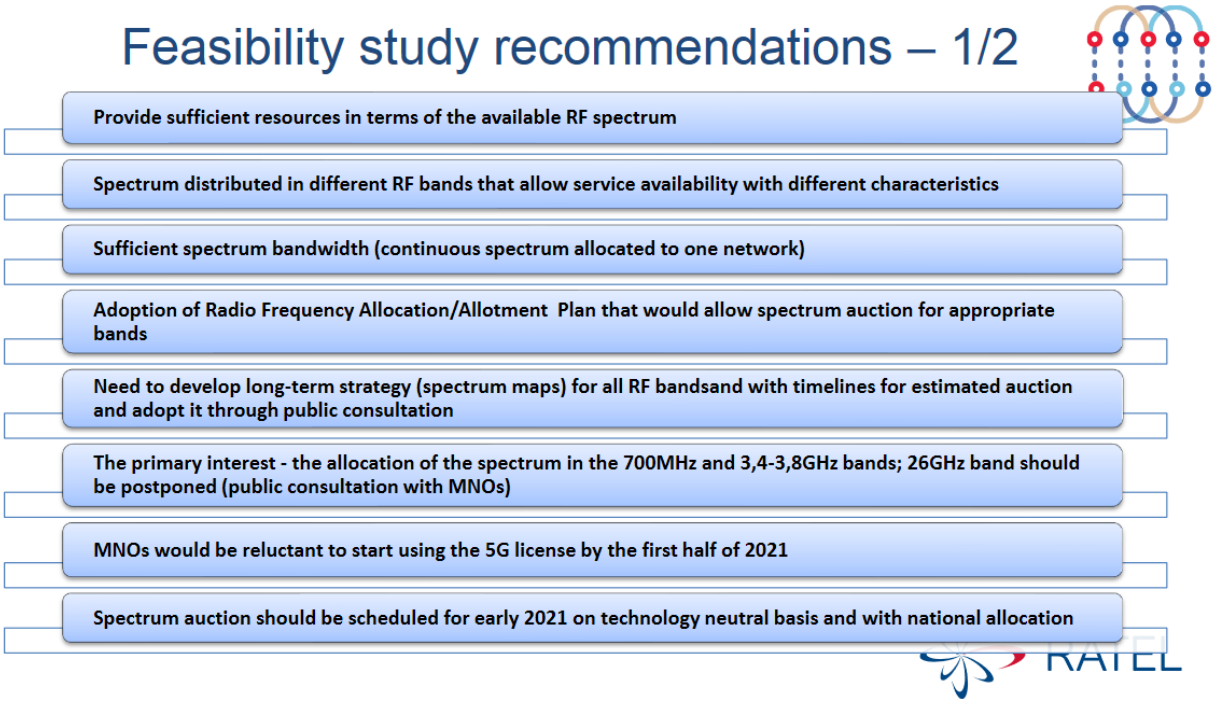


Figure 33. RATEL presentation at PUPIN (slide, cont.)

Figure 34. RATEL presentation at PUPIN (5G vision)



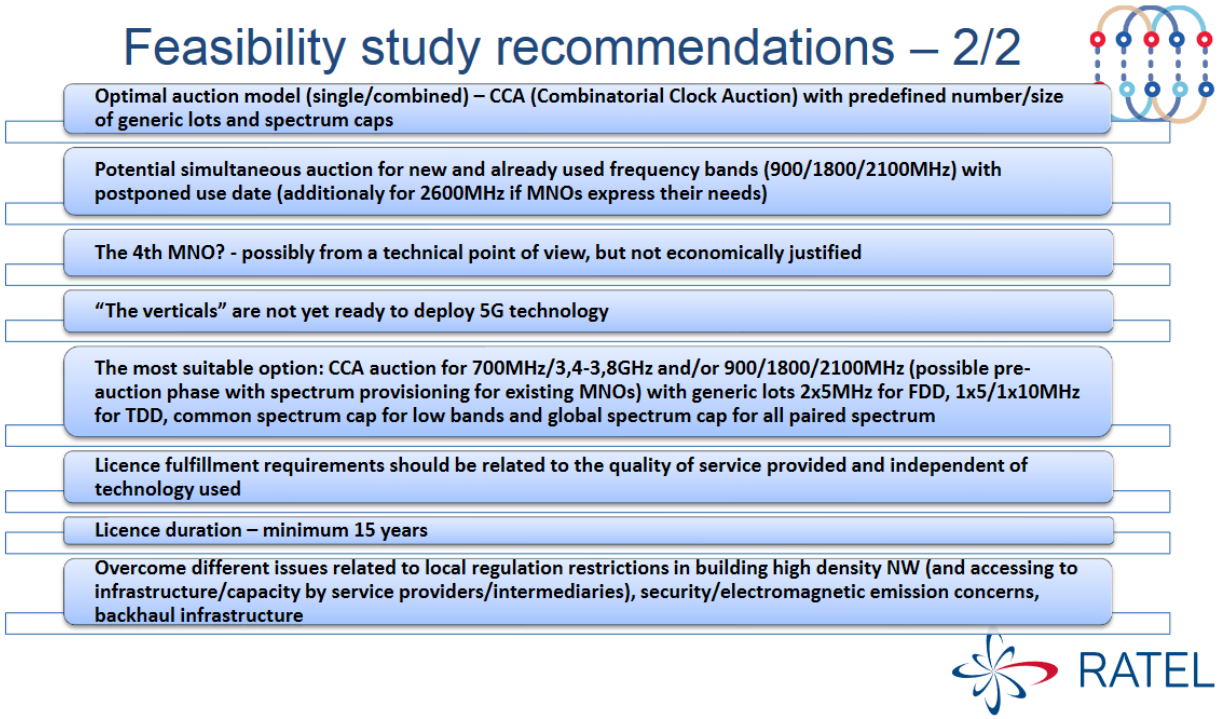


Figure 35. RATEL presentation at PUPIN (recommendations)

6. Conclusion

From the perspective of leading people in the region, we conclude that new policies and guidelines recommended by EU still need to be further implemented in Serbia and the West Balkan region. Providing specific guidelines for stakeholders requires more events bringing the leading people of the region in many sectors together in the form of panel discussions as were organized in LAMBDA or even bigger events.

It was also concluded that industry and enterprise sectors are having different needs than what the educational sectors offer at this time. This creates a big margin between the knowledge of the graduated people and the need in the market. More projects like LAMBDA are needed in the region to bring the best practice of research and industry from other parts of the Europe to the Balkan area.