and not in meaning. The survey analysis allowed us to obtain the list of hard and soft skills that are required by the labor market.

REFERENCES

1. GRIGORESCU A., ZAMFIR A.-M., SIGURDARSON H.T., LAZARCZYK CARLSON E.: Skill Needs among European Workers in Knowledge Production and Transfer Occupations (2022) Electronics (Switzerland), 11 (18), art. no. 2927, DOI: 10.3390/electronics11182927.
2. JONY S.S.R., KANO T., HAYASHI R., MATSUDA N., RAHMAN M.S.: An Exploratory Study of Online Job Portal Data of the ICT Sector in Bangladesh: Analysis, Recommendations and Preliminary Implications for ICT Curriculum Reform (2022) Education Sciences, 12 (7), art. no. 423, DOI: 10.3390/educsci12070423.
3. TOMMASI F., PERINI M., SARTORI R.: Multilevel comprehension for labor market inclusion: a qualitative study on experts' perspectives on Industry 4.0 competences (2022) Education and Training, 64 (2), pp. 177-189. DOI: 10.1108/ET-04-2021-0146.
4. TAVAKOLI, M., FARAJI, A., VROLIJK, J., MOLAVI, M., MOL, S.T., KISMIHÓK, G.: An AI-based open recommender system for personalized labor market driven education (2022) Advanced Engineering Informatics, 52, art. no. 101508, DOI: 10.1016/j.aei.2021.101508.
5. SCHLEGEL T., PFISTER C., HARHOFF D., BACKES-GELLNER U.: Innovation effects of universities of applied sciences: an assessment of regional heterogeneity (2022) Journal of Technology Transfer, 47 (1), pp. 63-118, DOI: 10.1007/s10961-020-09839-w.
6. KIENER F., GNEHM A.-S., CLEMATIDE S., BACKES-GELLNER U.: IT skills in vocational training curricula and labour market outcomes (2022) Journal of Education and Work, 35 (6-7), pp. 614-640. DOI: 10.1080/13639080.2022.2126968.
7. WYSKWARSKI M. Requirements for project managers—what do job advertisements say? (2021) Sustainability (Switzerland), 13 (23), art. no. 12999, DOI: 10.3390/su132312999.
8. SALAS-VELASCO M.: Mapping the (mis)match of university degrees in the graduate labor market (2021) Journal for Labour Market Research, 55 (1), art. no. 14, DOI: 10.1186/s12651-021-00297-x.
9. LI G., YUAN C., KAMARTHI S., MOGHADDAM,M., JIN X. Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis (2021) Journal of Manufacturing Systems, 60, pp. 692-706. DOI: 10.1016/j.jmsy.2021.07.007.
10. ALTONJI J.G., ZHONG L.: The labor market returns to advanced degrees (2021) Journal of Labor Economics, 39 (2), pp. 303-360. DOI: 10.1086/710959.

DOI: https://doi.org/10.53052/9788367652018.04

Kateryna MERKULOVA[1], Tetiana KOVALIUK[1], Liudmyla ZUBYK[1],
Yelyzaveta ZHABSKA[1], Sergii DOTSENKO[1],
Aleksandra KŁOS-WITKOWSKA[2], Krzysztof WITOS[2], Marcin BERNAS[2],
Vasyl MARTSENYUK[2], Georgi DIMITROV[3], Dejan RANCIC[4],
Oleksii BYCHKOV[1]

BADANIE ISTNIEJĄCYCH PROGRAMÓW NAUKOWYCH W DZIEDZINIE BIG DATA

Streszczenie: Celem badań jest zbadanie potencjału innowacyjnego projektów naukowych i komercyjnych z zakresu Big Data. Zidentyfikowane zostały główne kierunki rozwoju analityki Big Data, cele i zadania rozwoju Big Data, rezultaty projektów przyczyniające się do rozwoju regionów oraz korzyści płynące z projektów dla użytkowników. Uzyskane informacje mogą posłużyć do określenia treści programów edukacyjnych dla szkoleń Big Data.

Słowa kluczowe: Big Data, projekty naukowe, innowacje

RESEARCH OF EXISTING SCIENTIFIC PROGRAMS IN THE FIELD OF BIG DATA

Summary: The purpose of the research is to study the innovative potential of scientific and commercial projects in the field of Big Data. It identified the main directions of Big Data analytics development, the goals and objectives of Big Data development, the project results in a contribution to the development of regions, and the benefits of projects for users. The information obtained can be used to determine the content of educational programs for Big Data training courses.

Keywords: Big Data, scientific projects, innovations

1. Introduction

Technology based on Big Data are widely applied in a series of scientific projects [1 - 6]. The given research was carried out in the context of project no. 2020-1-PL01-KA203-082197 entitled “Innovations for Big Data in a Real World”. The objective

[1] Taras Shevchenko National University of Kyiv, Kiev, Ukraine: oleksiibychkov@knu.ua
[2] Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: (awitkowska, mbernas, kwitos, vmartsenyuk)@ath.bielsko.pl
[3] University of Library Studies and Information Technologies, Sofia, Bulgaria: g.dimitrov@unibit.bg
[4] University of Niš, Nis, Serbia: dejan.rancic@elfak.ni.ac.rs

was to conducted a survey of the academic community involved in the teaching of university students from four partner countries in order to determine the competencies of teachers in the field of Big Data and Data Science. Namely, here we are focusing to the application of Big Data in scientific projects.
The survey was performed on-line using Google Forms tools. The survey was performed during a period from the 1'th of September 2020 to 28'th of February 2021. To obtain wide range of data, multiple choice question fields, with additional open question field were offered to overcome the shortcomings of narrowed answers suggestions. The survey contains both open and closed questions. To make a process of data collection un-biased no additional recommendation was added. No events were reported during that time that could influence the result.

Target

This survey is a part of the research within IO1 in connection with the objectives of project 2020-1-PL01-KA203-082197 "Innovations for Big Data in a Real World" (IBigWorld) under the Erasmus+ program. This project aims to join Universities, business and provide innovative solutions to develop Big Data experts. The questions of this research were aimed at studying the innovative potential of scientific and commercial projects in the field of Big Data. It is necessary to identify the main directions of Big Data analytics development, goals and objectives of Big Data development, the projects results contribution to the development of regions, the benefits of projects for users. The information obtained can be used to determine the content of educational programs Big Data courses.

Collection and analysis of data

To collect and analysis data on the potential of science and commercial projects respondents from different universities of the four project partner countries were interviewed 30 responses were received. The FORM 3: EXISTING SCIENTIFIC PROJECTS IN DATA SCIENCE was offered to researches of different universities. Partner Organisations are presented on Fig. 1 and Table 1.

Data description

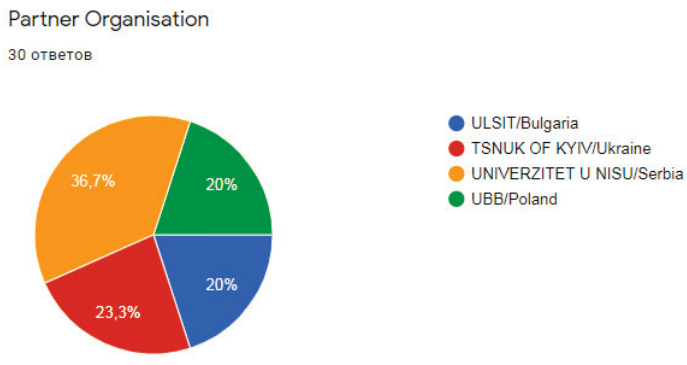

Figure 1. Percentage of researches surveyed in comparison with other project partners

Table 1. Percentage of researches surveyed in comparison with other project partners

Country	Total	Percentage
Serbia	11	36,7
Ukraine	7	23,3
Bulgaria	6	20
Poland	6	20

2. Research of Existing Big Data Projects in the World and Their Characteristics

2.1. Name and family of the researcher

The goal of the question was to determine the activity of respondents and project participants.

Data description

The survey showed that 12 researchers participated in the analysis of existing projects in the field of Big Data and their characteristics in the world. The number of reviewed projects, distributed by researchers, is shown in Fig. 2. The 30 projects was considered.

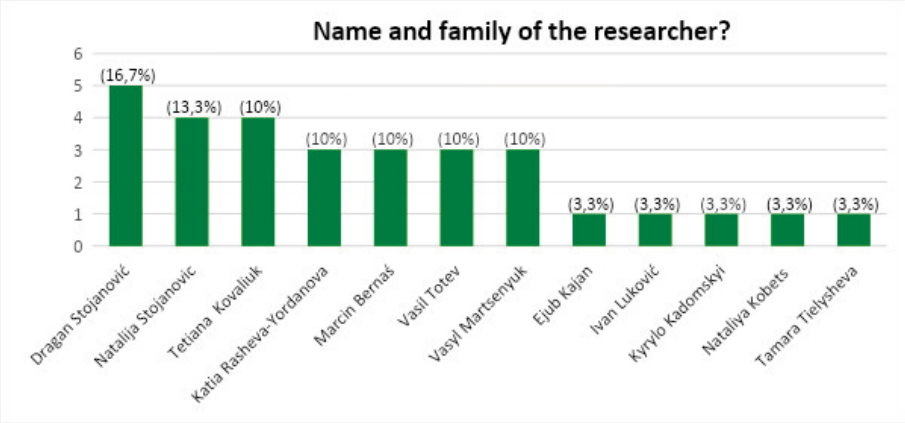

Figure 2. Name and family of the researcher and number of found projects

Discussion

Researchers have been actively studying Big Data projects in different countries. The limited number of reviewed projects is justified by the lack of free access to information about the project's details.

Conclusion

The difficulty of obtaining detailed information about scientific and commercial projects in the field of Big Data suggests that the field of Big Data is in a state of development. Projects are trade secrets.

2.2. What existing scientific program in the field of BIG DATA was it found?

The goal of the question was to get information about areas of application of Big Data.

Data description

The results of the survey are given in Table 2. The names of the projects and the scope of application are very diverse. Unable to quantify project areas of application automatically.

Table 2. Name of Big Data projects

#	Name of Big Data project	Areas of application
1	Intelligent Big Data Analytics for Industry 4.0 Systems	Industry 4.0
2	Big Data Europe	Big Data Integrator
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	Cloud Computation
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	Data Mining
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	Decision-making
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	Energy Industry
7	AnalyticOps	Enterprise automation
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	Grapevine-powered industries
9	Forming of Data Science Competence for bridging the Digital Divide	Higher Education
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	Higher Education
11	Agile Analytics on Big Data Cubes	Horizon 2020
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	Horizon 2020
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	Horizon 2020
14	Diagnosis and energy optimizing in production systems of Industrial IoT	IoT
15	NB-IoT Vodafone	IoT
16	BigMedilytics - Big Data for Medical Analytics	Medical Analytics
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	Scientific Research
18	Innovative Big Data solutions for smart cities (Big4Smart)	Smart cities
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	Smart cities
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	Smart cities

21	Kyiv Smart Sity,	Smart cities
22	SoBigData Research Infrastructure	Social data mining
23	SoBigData	Social data mining
24	SoBigData++	Social data mining
25	Application of clustering methods in some problems of speech signal processing	Speech signal processing
26	1991 Open Data Incubator	Startups support
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	Supercomputer technologies
28	Leveraging Big Data to Manage Transport Operations	Transport industry
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	
30	N/A	

Discussion

Found projects on different topics and focused on different areas of application. Typical applications: smart city, internet of things, higher education and industry.

Conclusion

The analysis of the survey results showed that the researches covered different areas of Big Data applications. There are projects supported by the European Union within the framework of the program Horizon 2020.

2.3. Who is the main partner in this scientific program? Name of company, community, state organisation or scientific organisation, University, country, using BIG DATA

The goal of the question is to identify project partners to obtain information on communications within projects.

Data description

The survey results are shown in Table 3. Universities, government organizations, commercial companies, European Commission represent partners of Big Data projects. Figure 3 shows the percentage of partner types. Universities (54%) are the main partners of the Big Data projects. Companies (20%) present commercial projects to the market. Research organizations (23%) conduct fundamental research.

Table 3. Big Data project partners

#	Name of Big Data project	Name of partner
1	Intelligent Big Data Analytics for Industry 4.0 Systems	University of Bradford
2	Big Data Europe	European Commission
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	Barcelona Supercomputing Center (BSC)

4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	University of Reading
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	Heriot-Watt University
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	Warsaw University of Technology, Faculty of Electronics and Information Technology, Poland
7	AnalyticOps	Buran.Gpoup
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	Agroknow - Greece
9	Forming of Data Science Competence for bridging the Digital Divide	University of Library Studies and Information Technologies (ULSIT), Bulgaria
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	University of Belgrade; University of Novi Sad; University of Niš, etc.
11	Agile Analytics on Big Data Cubes	Jacobs University Bremen Ggmbh
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	Foundation For Research And Technology Hellas (Forth)
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	University Of Piraeus Research Center (Uprc)
14	Diagnosis and energy optimizing in production systems of Industrial IoT	SK.AI
15	NB-IoT Vodafone	Vodafone Ukraine
16	BigMedilytics - Big Data for Medical Analytics	Philips Electronics Nederland B.V. - Netherlands
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	Technical University of Sofia, Bulgaria; Association “Innovation Center for Information and In-silico Technology and Expert Knowledge Transfer – In-silico Intellect”
18	Innovative Big Data solutions for smart cities (Big4Smart)	Sofia University "St. Kliment Ohridski"
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	University of Nis, Serbia
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	Faculty of Electronic Engineering, University of Nis, Serbia
21	Kyiv Smart Sity,	Department of Information and Communication Technologies of the Kyiv City State Administration,
22	SoBigData Research Infrastructure	Consiglio Nazionale Delle Ricerche (Italy)
23	SoBigData	Coordinator: CNR - Consiglio Nazionale delle Ricerche - (PI: Fosca Giannotti) - Italy

24	SoBigData++	Consiglio Nazionale delle Ricerche (CNR) - Italy
25	Application of clustering methods in some problems of speech signal processing	Kiev Academic University
26	1991 Open Data Incubator	SocialBoost
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	V.M.Glushkov Institute of Cybernetics of the NAS of Ukraine
28	Leveraging Big Data to Manage Transport Operations	Stiftinga Vestlandsforsking (Norway)
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	Wroclaw University of Science and Technology, Faculty of Electronics
30	N/A	University Lion 1 France, Zayed University, State University of Novi Pazar Serbia

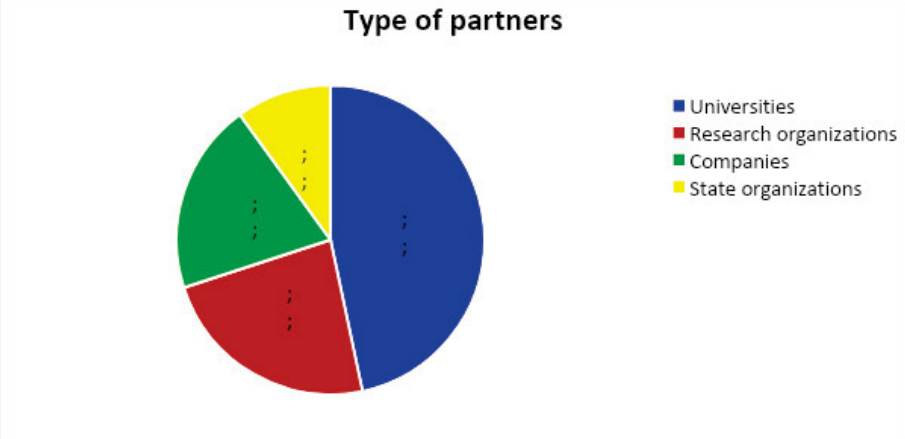

Figure3. Big Data project partners

Discussion

Universities are carriers of innovation and think tanks, so researches is concentrated there mainly. Research organizations need funding and, in the absence of it, do not conduct research.

Conclusion

It is necessary to encourage universities to do research. Universities should become centres of technology transfer to industry.

2.4. Goals of the projects

The purpose of the survey is to obtain information, what goals define Big Data projects.

Data description

The survey results are shown in Table 4. The goals of the projects are very diverse. There are projects aimed at developing algorithms, analytical data models, methods for obtaining high-quality statistical information (5 projects).
Other projects solve the problem of building smart infrastructure in cities, including threats from a large number of moving objects (5 projects). Some of the projects are exploring the possibilities of medical analytics (2 projects). The problems of social mining and Big Data are considered in a number of projects (2 projects). Technology projects explore the issues of building heterogeneous databases, data lake storages, data pipelining, data cubes and new software frameworks. (9 projects). Evaluation of the system and programs of higher education, bridging the digital Big Data divide is considered in projects (2 projects). Industrial projects are aimed at creating applications for the mobile, transport, energy industries (4 projects).
The grouping of projects by goals is shown in Fig. 4.

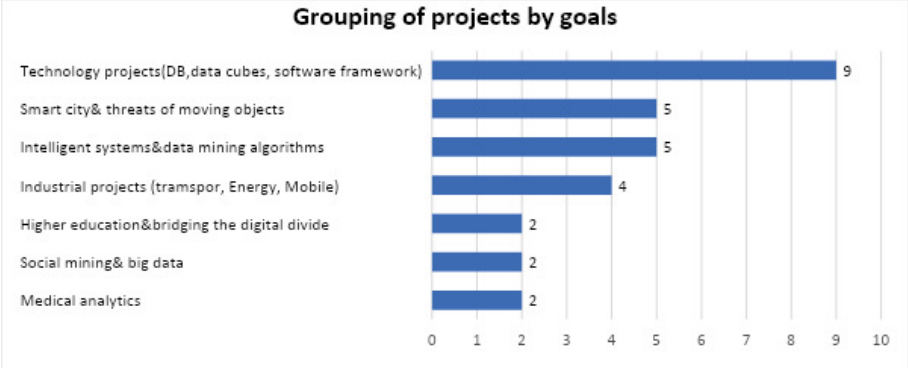

Figure 4. Grouping of Big Data projects by goals

Table 4. Goal of the Big Data project

#	Name of Big Data project	Goal of Big Data project
1	Intelligent Big Data Analytics for Industry 4.0 Systems	This project aims to develo+E2:E23p new intelligent systems capable of online decision making by analysing large-scale data at real-time in highly complex engineering/manufacturing environments.
2	Big Data Europe	Create a easy to use Big data platform with good practice pipelines
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	CLASS aims to develop a novel software architecture framework to help Big Data developers to efficiently distributing data analytics workloads along the compute continuum (from edge to cloud) in a complete and transparent way, while providing sound real-time guarantees. This ability opens the door to the use of Big Data into critical real-time systems, providing to them superior data analytics capabilities to implement more intelligent and autonomous control applications.
4	Big Data Analytics and Mining: investigating and testing	The project will investigate and test distributed formulations of data mining algorithm that are suitable

	distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	for the MapReduce paradigm and for other distributed computing approaches
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	Therefore, this research seeks to develop a project data analytic model using Big Data and Machine Learning for front-end decision-making in transportation megaprojects.
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	The aim of this project is to extend the possibilities of the data quality monitoring system with methods machine learning and creating new visualization systems for collected data. Artificial method intelligence will allow to relieve the currently existing data quality analysis infrastructure, based on people work to a large extent by taking into account historical data and external factors (e.g. ambient temperature)
7	AnalyticOps	Data science life cycle management, Design and Deploy Data Lake Storage on an Open Hadoop Platform, Big Data Management and Data Science for the Retail Industry
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	BigDataGrapes aims to help European companies in the wine and natural cosmetics industries become more competitive in the international markets. It specifically tries to help companies across the grapevine-powered value chain ride the Big Data wave, supporting business decisions with real time and cross-stream analysis of very large, diverse and multimodal data sources. Two main goals: To develop and demonstrate powerful data processing technologies that will increase the efficiency of companies that need to take important business decisions dependent on access to vast and complex amounts of data. To catalyse the creation of a data ecosystem and economy that will increase the competitive advantage of companies that serve with IT solutions these sectors.
9	Forming of Data Science Competence for bridging the Digital Divide	The main aim of the project is to create a methodological basis for overcoming the digital divide in the field of Data Science.
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	Evaluation of Higher Education System and programs in Serbia
11	Agile Analytics on Big Data Cubes	EarthServer-2 makes Agile Analytics on Big Earth Data Cubes of sensor, image, simulation, and statistics data a commodity for non-experts and experts alike through • navigation, extraction, aggregation, and recombining of any-size space/time data cubes; • easy to install & maintain value-adding services extending the existing portfolio of data and compute centers; • based on open standards, in particular: the OGC Big Data standards and the forthcoming ISO SQL/MDA (“Multi-Dimensional Arrays”) standard.

12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	I-BiDaaS aims to empower users to easily utilize and interact with Big Data technologies, by designing, building, and demonstrating, a unified solution that: significantly increases the speed of data analysis while coping with the rate of data asset growth, and facilitates cross-domain data-flow towards a thriving data-driven EU economy.
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	datAcron is a research and innovation collaborative project targeting at introducing novel methods to detect threats and abnormal activity of very large numbers of moving entities in large geographic areas. Towards this target, datAcron aims to advance the management and integrated exploitation of voluminous and heterogeneous data-at-rest (archival data) and data-in-motion (streaming data) sources, so as to significantly advance the capacities of systems to promote safety and effectiveness of critical operations for large numbers of moving entities in large geographical areas. Technological developments in datAcron will be validated and evaluated in user-defined challenges that aim at increasing the safety, efficiency and economy of operations concerning moving entities in the air-traffic management (ATM) and maritime domains.
14	Diagnosis and energy optimizing in production systems of Industrial IoT	Developing solutions for detecting anomalous behavior, malfunctioning or wear; provide analytics for energy optimization
15	NB-IoT Vodafone	To make NB-IoT the best technology for those applications and devices that will benefit from the kind of interconnect and roaming capabilities that the mobile industry has provided since its inception
16	BigMedilytics - Big Data for Medical Analytics	BigMedilytics (Big Data for Medical Analytics) is the largest EU-funded initiative to transform the region's healthcare sector by using state-of-the-art Big Data technologies to achieve breakthrough productivity in the sector by reducing cost, improving patient outcomes and delivering better access to healthcare facilities simultaneously. The project is composed of 12 pilots that address three themes with the greatest impact on the sector—Population Health and Chronic Disease Management, Oncology and Industrialization of Healthcare Services— and covers the entire Healthcare Continuum from Prevention to Diagnosis, Treatment and Home Care.
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	(1) Genetic regulatory elements identification for unknown genes detection in sequenced genomes and for the aim of genomic mapping. (2) To predict the type and malignancy of breast cancer based on information for mutations in genes associated with it, the level of expression and the associated epigenetic information. The driving force of the project is an interdisciplinary team, combining expertise in information science and technology, engineering foundations and technical deployment of software methods and tools, as well as established scholars in the areas of molecular biology and medical genetics.

18	Innovative Big Data solutions for smart cities (Big4Smart)	Development of methodology, implemented by an open technological platform, that support making informed and timely decisions on Big Data for building smart cities.
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	The goals of the CitySensing project are focused on research and development of methods, tools and software systems for efficient and effective processing, analysis and mining of Big mobility data collected leveraging mobile crowd sensing and Internet of Things paradigms in Smart Cities.
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	The goal of the CitySensing project is focused on research and development of methods, tools and software systems for efficient and effective processing, analysis and mining of Big mobility data collected leveraging mobile crowd sensing and Internet of Things paradigms in Smart Cities.
21	Kyiv Smart Sity,	Kyiv Smart Sity: unite city authorities, business representatives and activists for the joint development of smart infrastructure in the capital of Ukraine
22	SoBigData Research Infrastructure	SoBigData serves a wide community of data scientists carrying out studies of all aspects of society. It supports policy making by novel ways to produce high-quality statistical information; empowers citizens with self-awareness tools; and promotes ethical uses of Big Data.
23	SoBigData	SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by “Big Data”. SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research.
24	SoBigData++	SoBigData++ strives to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and Big Data to understand the complexity of our contemporary, globally-interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015.
25	Application of clustering methods in some problems of speech signal processing	To derive cost functions for spectral clustering based on measures of error between a given partition and a solution of the spectral relaxation of a minimum normalized cut problem.
26	1991 Open Data Incubator	To transform big volumes of open government data into a resource for creating startups that provide services to Ukrainian citizens, companies and government agencies.
27	Methods of highly productive calculation of mathematical	Application of Big data analytics technology to heterogeneous environments and databases

	models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	
28	Leveraging Big Data to Manage Transport Operations	Leveraging Big Data to Manage Transport Operations (LeMO) project will explore the implications of the utilisation of Big Data to enhance the economic sustainability and competitiveness of European transport sector. The project will study and analyse Big Data in the European transport domain in particular with respect to five transport dimensions: mode, sector, technology, policy and evaluation. LeMO will accomplish this by conducting a series of case studies, in order to provide recommendations on the prerequisites of effective Big Data implementation in the transport field.
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	The project aims to make a significant contribution to the development of advanced analysis and processing tools BD type data with specific properties. These are the so-called non-negative data, which accounts for a large proportion of all data digital. These include, for example, collections of images and films, spectral signals, text documents, various biomedical data, etc. To search and extract the desired information from such huge data, it becomes necessary to reduce it volumes, as well as extraction of semantic features.
30	N/A	N/A

Discussion

Accelerating data growth is an objective reality. Social networks, mobile devices, data from measuring devices, business information and more are several types of sources that can generate gigantic amounts of information. In this regard, there is a growing need for the efficient use of this data. The goals of the projects analyzed by the researchers meet the requirements of the Big Data market.

Conclusion

Many business representatives note that difficulties in implementing Big Data projects are associated with a shortage of specialists - data scientist, data engineer, data analyst. The rate of return on investments in Big Data directly depends on the quality of work of employees engaged in deep and predictive analytics. The variety of goals and therefore the content of Big Data projects indicates the need to develop educational programs and train specialists who are able to solve problems in the field of Big Data.

2.5. Type of data used

The goal of the question is to determine the sources and nature of data that are processed in the Big Data projects

Data description

The survey results showed (Table 5, Fig. 5) that the largest amount of data comes from sensors and researchers (30%). Unstructured data is 16.7%. Other data types account for 3.3%.

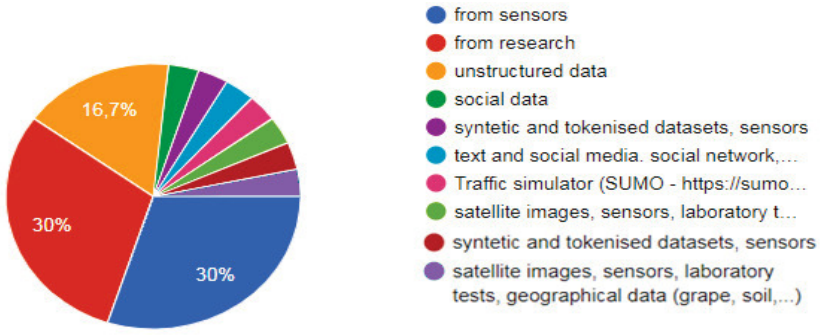

Figure 5. Type of data used

Table 5. Type of data used

#	Type of data used	Number of Projects	Percentage
1.	from sensors	9	30
2.	from research	9	30
3.	unstructured data	5	16.7
4.	social data	1	3.3
5.	sensors, data from social media	1	3.3
6.	traffic simulator (SUMO)	1	3.3
7.	text and social media. social network, human mobility, web data	1	3.3
8.	text and social media data, social network data, human mobility, web content	1	3.3
9.	synthetic and tokenised datasets, sensors	1	3.3
10.	satellite images, sensors, laboratories, geographical data (grape, soil,...)	1	3.3

Discussion

The dissemination of IoT projects requires processing data from sensors. Research projects use data that researchers generate. Many social projects process social data.

Conclusion

The variety of data sources requires the development of ways to collect and store them. These issues should be focused on the developers of educational programs in the field of Big Data.

2.6. Start of the project and finish of the project

The purpose of the survey is to determine the duration of Big Data projects and their relevance

Data description

The survey results showed (Table 6, Fig. 6) that respondents indicated project start dates from 2012 to 2021. The project completion dates are set in 2018 - 2023. The maximum project duration is 9 years; the minimum duration is 0. The average project duration is 3 years.

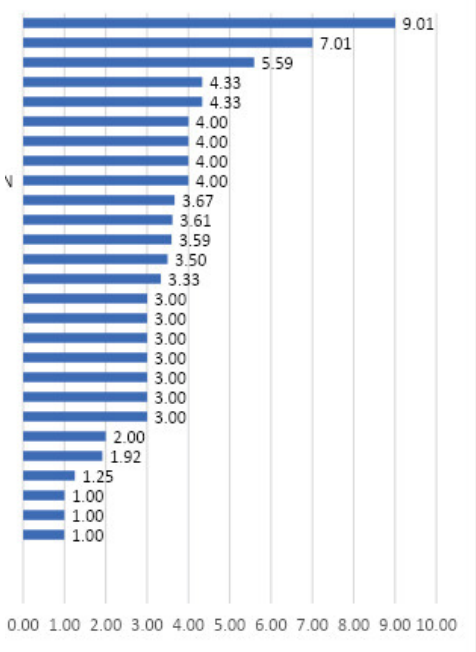

Figure 6. Durations of the projects

Table 6. Duration of the Big Data projects

#	Name of Big Data project	Start of the project	End of the project	Duration of the projects / year
1	Intelligent Big Data Analytics for Industry 4.0 Systems	19.01.2021	19.01.2021	0,00
2	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	19.01.2021	19.01.2021	0,00
3	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	19.01.2021	19.01.2021	0,00

4	Diagnosis and energy optimizing in production systems of Industrial IoT	01.05.2020	30.04.2021	1,00
5	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	15.11.2018	15.11.2019	1,00
6	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	18.11.2019	18.11.2020	1,00
7	Application of clustering methods in some problems of speech signal processing	01.10.2018	01.01.2020	1,25
8	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	01.01.2020	01.12.2021	1,92
9	Forming of Data Science Competence for bridging the Digital Divide	20.12.2017	20.12.2019	2,00
10	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	20.12.2016	20.12.2019	3,00
11	Leveraging Big Data to Manage Transport Operations	01.11.2017	31.10.2020	3,00
12	Agile Analytics on Big Data Cubes	01.05.2015	30.04.2018	3,00
13	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	01.01.2018	31.12.2020	3,00
14	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	01.01.2018	31.12.2020	3,00
15	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	01.01.2016	31.12.2018	3,00
16	Innovative Big Data solutions for smart cities (Big4Smart)	20.12.2017	20.12.2020	3,00
17	Big Data Europe	27.02.2015	28.06.2018	3,33
18	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	01.01.2018	30.06.2021	3,50
19	1991 Open Data Incubator	20.07.2014	20.02.2018	3,59
20	NB-IoT Vodafone	20.02.2018	01.10.2021	3,61
21	BigMedilytics - Big Data for Medical Analytics	01.01.2018	31.08.2021	3,67
22	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	20.01.2017	19.01.2021	4,00
23	Extraction of low-dimensional hidden factors from large sets of non-negative data	01.03.2016	29.02.2020	4,00
24	SoBigData++	01.01.2020	31.12.2023	4,00
25	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	01.10.2014	01.10.2018	4,00
26	SoBigData Research Infrastructure	01.09.2015	31.12.2019	4,33

27	SoBigData	01.09.2015	31.12.2019	4,33
28	Kyiv Smart Sity,	01.03.2015	01.10.2020	5,59
29	AnalyticOps	01.01.2015	01.01.2022	7,01
30	N/A	01.01.2012	31.12.2020	9,01

Discussion

The indicated start and end dates of projects mean that the projects are up-to-date. Business needs for Big Data solutions have intensified in the last decade, as evidenced by project dates. The variety of projects, their complexity is confirmed by their duration.

Conclusion

Meeting the business needs for Big Data solutions requires expertise, technology, tools, infrastructure, and new challenges. It is obvious that the training of specialists in the field of Big Data is an urgent task

2.7. Hyperlink to the organization

The purpose of the question is to determine the hyperlink to the organizations, who are project developers, get information about its activities, type of organizations and other characteristics.

Data description

The survey results are showed in Table 7. Organizations are classified by type: education, research, government, commercial. 47% (14) are universities, 23% (7) are research organizations, 20% (6) are commercial and 10% (3) are state organizations (Fig. 7).

Table 7. Hyperlink to the organization, who are developing the Big Data project

#	Name of Big Data project	Hyperlink to the organization	Type of organization
1	Intelligent Big Data Analytics for Industry 4.0 Systems	*https://www.findaphd.com/phds/university-of-bradford/?40g710*	Educational
2	Big Data Europe	*https://ec.europa.eu/info/index_en*	State
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	*https://www.bsc.es/*	Research
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	*http://www.reading.ac.uk/*	Educational
5	Big Data for Big Decisions – Application of Big Data Analytics and	*https://www.hw.ac.uk/*	Educational

	Machine Learning for Front-end Decision-making in Transportation Megaprojects		
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	*http://www.elka.pw.edu.pl/eng#*	Educational
7	AnalyticOps	*http://buran.group/*	Commercial
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	*https://agroknow.com/*	Commercial
9	Forming of Data Science Competence for bridging the Digital Divide	*https://unibit.bg/*	Educational
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	*http://www.bg.ac.rs/*	Educational
11	Agile Analytics on Big Data Cubes	*http://www.jacobs-university.de/*	Educational
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	*https://www.ics.forth.gr/*	Research
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	*https://www.unipi.gr/unipi/en/*	Educational
14	Diagnosis and energy optimizing in production systems of Industrial IoT	*https://skai.tech/*	Commercial
15	NB-IoT Vodafone	*https://www.vodafone.ua/uk*	Commercial
16	BigMedilytics - Big Data for Medical Analytics	*http://www.philips.com/*	Commercial
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	*https://insilicokdd.com/network-partners/*	Educational
18	Innovative Big Data solutions for smart cities (Big4Smart)	*https://www.uni-sofia.bg/*	Educational
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	*http://www.elfak.ni.ac.rs*	Educational
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	*http://www.elfak.ni.ac.rs*	Educational
21	Kyiv Smart Sity,	*https://project.liga.net/projects/smart_city/*	Satate
22	SoBigData Research Infrastructure	*http://www.cnr.it/*	State
23	SoBigData	*https://www.cnr.it/*	Research
24	SoBigData++	*https://www.cnr.it/*	Research
25	Application of clustering methods in some problems of speech signal processing	*https://sites.google.com/view/datasciencekau*	Research
26	1991 Open Data Incubator	*https://socialboost.com.ua*	Commercial
27	Methods of highly productive calculation of mathematical models of	*http://incyb.kiev.ua/proekt*	Research

	heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies		
28	Leveraging Big Data to Manage Transport Operations	*http://www.vestforsk.no/*	Research
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	*https://weka.pwr.edu.pl/en/*	Educational
30	N/A	*N/A*	Educational

Discussion

Most of the organizations analysed (47%) are educational institutions. This suggests that the main research and design work is concentrated in universities. The smallest number of reviewed organizations (10%) are state-owned. This means that states do not give sufficient attention to scientific and project work.

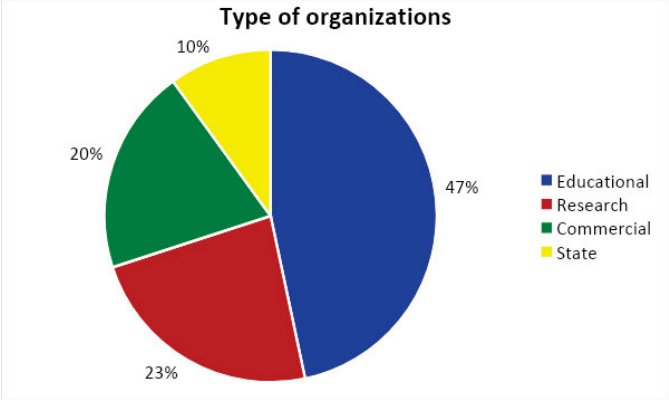

Figure 7. Classification of organizations by type

Conclusion

To carry out scientific, educational and commercial projects, universities must have the appropriate resources: personnel, equipment, finances and time. Universities will be able to achieve the desired results in the structure of educational ecosystems, research and production clusters, organized according to the quadruple helix model.

2.8. Hyperlink to the project

The purpose of the question is to determine the hyperlink to the project for detailed information about its.

Data description

The survey results are showed in Table 8. Figure 8 illustrates the following sources of project financing: 33.3% (10) of project financed by program Horizon 2020; Government finance 30% (9) projects; 33.3% (10) of the reviewed projects are funded by the Horizon 2020 program; 30% (9) projects have public funding. Commercial

companies support 23.3% (7) of projects reviewed. Respondents identified by 10% (3) international projects. One project is a Public-Private Partnership project.

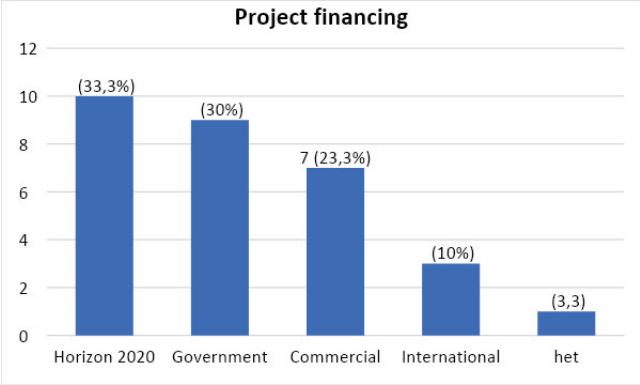

Figure 8. Sources of project financing

Table 8. Hyperlink to the Big Data project

#	Name of Big Data project	Hyperlink to the Big Data project	Financing
1	Intelligent Big Data Analytics for Industry 4.0 Systems	*https://www.findaphd.com/phds/project/intelligent-big-data-analytics-for-industry-4-0-systems/?p122152*	commercial
2	Big Data Europe	*https://www.big-data-europe.eu/*	Horizon 2020
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	*https://class-project.eu/*	Horizon 2020
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	*https://www.findaphd.com/phds/project/big-data-analytics-and-mining-investigating-and-testing-distributed-formulations-of-data-mining-algorithms-that-are-suitable-for-the-mapreduce-paradigm-and-for-other-distributed-computing-approaches/?p81074*	commercial
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	*https://www.findaphd.com/phds/project/big-data-for-big-decisions-application-of-big-data-analytics-and-machine-learning-for-front-end-decision-making-in-transportation-megaprojects/?p128378*	commercial
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE	*https://projekty.ncn.gov.pl/index.php?projekt_id=336268*	government

	experiment at the Large Hadron Collider at CERN		
7	AnalyticOps	*http://buran.group/solutions/*	commercial
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	*https://www.bigdatagrapes.eu*	Horizon 2020
9	Forming of Data Science Competence for bridging the Digital Divide	*http://data-science.unibit.bg/*	government
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	*http://peshes.ius.bg.ac.rs/*	government
11	Agile Analytics on Big Data Cubes	*https://cordis.europa.eu/project/id/654367*	Horizon 2020
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	*https://www.ibidaas.eu/*	Horizon 2020
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	*http://datacron1.ds.unipi.gr:9082/*	Horizon 2020
14	Diagnosis and energy optimizing in production systems of Industrial IoT	*https://www.facebook.com/search/top?q=sk%20ai*	commercial
15	NB-IoT Vodafone	*https://business.vodafone.ua/iot/nb-iot*	commercial
16	BigMedilytics - Big Data for Medical Analytics	*https://www.bigmedilytics.eu/*	Public-Private Partnership
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	*https://insilicokdd.com/*	government
18	Innovative Big Data solutions for smart cities (Big4Smart)	*http://big4smart.eu/*	government
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	*https://hpi.de/en/research/future-soc-lab.html*	international
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	*https://hpi.de/en/research/future-soc-lab.html*	international
21	Kyiv Smart Sity,	*https://www.kyivsmartcity.com/projects/*	government
22	SoBigData Research Infrastructure	*https://cordis.europa.eu/project/id/654024*	Horizon 2020
23	SoBigData	*http://project.sobigdata.eu/ http://www.sobigdata.eu/*	Horizon 2020
24	SoBigData++	*http://plusplus.sobigdata.eu/*	Horizon 2020
25	Application of clustering methods in some problems of speech signal processing	*https://events.imath.kiev.ua/event/653/*	government
26	1991 Open Data Incubator	*http://1991.vc/en/about-incubator/*	commercial

27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	*N/A*	government
28	Leveraging Big Data to Manage Transport Operations	*https://cordis.europa.eu/project/id/770038*	Horizon 2020
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	*https://projekty.ncn.gov.pl/index.php?projekt_id=290661*	government
30	N/A	*N/A*	

Discussion

An analysis of the project sites showed that universities conduct high-quality scientific research, which is supported by Horizon 2020 grants. This indicates that universities, being knowledge clusters, carry out relevant and breakthrough projects. Another source of funding for projects that solve urgent problems in accordance with the list of priority research areas in countries is government funding. Commercial companies support some projects.

Conclusion

The development of knowledge-intensive industries is necessary for a high-quality innovative breakthrough in the creation of a competitive high-tech economy. It is universities that can become points of growth that ensure the innovative development of the economies of countries. Indeed, the specificity of the activities of universities makes it possible to effectively integrate personnel training, research, and commercialization of the results of intellectual activity. Analysis of Big Data projects has proven the ability of universities to generate knowledge in the form of disruptive scientific research.

2.9. Hyperlink to the found Big Data set

The purpose of the question is to obtain information about the nature of Big Data sets, storage methods, mapping Big Data and other.

Data description

The survey results are showed in Table 9. Fig. 9 presents the percentage of datasets open for access. The respondents researched the sites of the projects and stated the following. Most projects (77%) do not reflect information about the structure and composition of datasets on project sites. Some sites (6.6%) have a description of the data structure but do not have access to the datasets themselves. A number of projects (12.8%) post project reports with links to datasets on their websites. You can get access after preliminary registration.

Table 9. Hyperlink to the found Big Data set

#	Name of Big Data project	Hyperlink to the found BIG DATA set	Is data available freely?
1	Intelligent Big Data Analytics for Industry 4.0 Systems	N/A	No
2	Big Data Europe	N/A - platform+good practice	No
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	N/A	No
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	N/A	No
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	N/A	No
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	N/A	No
7	AnalyticOps	*http://buran.group/solutions/big_data_analysis/*	No Description Data Lake solution, Hadoop cluster performance optimization, heterogeneous storage construction
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	BigDataGrapes D2.2 - Data Management Plan & Support Pack *https://zenodo.org/record/4546011#.YDgOJHko_Qx*	BigDataGrapes doesn't have any public repositories yet in https://github.com/BigDataGrapes
9	Forming of Data Science Competence for bridging the Digital Divide	not stated	No
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	N/A	No
11	Agile Analytics on Big Data Cubes	*https://ec.europa.eu/research/participants/documents/downloadPublic?doc*	Yes (links are in Data Management Plan – Final)

		umentIds=080166e5ba572792&appId=PPGMS	
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	*https://www.ibidaas.eu/tools/*	Yes Open Source Components (COMPSS, Hecuba, and other)
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	*http://datacron1.ds.unipi.gr:9082/services/open_datasets/*	Yes The dataset contains four categories of data: Navigation data, vessel-oriented data, geographic data, and environmental data
14	Diagnosis and energy optimizing in production systems of Industrial IoT	not publicly available	No
15	NB-IoT Vodafone	*https://business.vodafone.ua/produkty/bigdata-scoring*	No, according to the tariff plan
16	BigMedilytics - Big Data for Medical Analytics	N/A	No
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	not stated	No
18	Innovative Big Data solutions for smart cities (Big4Smart)	not stated	No
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	*ExtraSensory dataset - http://extrasensory.ucsd.edu/; CrowdSignals.io - http://crowdsignals.io/*	Yes Yes
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	*https://sumo.dlr.de/docs/index.html*	Yes
21	Kyiv Smart Sity,	N/A	No
22	SoBigData Research Infrastructure	*https://cordis.europa.eu/project/id/654024/results*	Yes XML and Project report
23	SoBigData	*https://sobigdata.d4science.org/home*	Yes Access to the SoBigData e-infrastructure methods and datasets after register.
24	SoBigData++	N/A	No
25	Application of clustering methods in some problems of speech signal processing	N/A	No
26	1991 Open Data Incubator	N/A	No
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big	N/A	No

	Data) based on supercomputer technologies		
28	Leveraging Big Data to Manage Transport Operations	*https://lemo-h2020.eu/interaction-maps*	No, only report on the project
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	data from Google+, Facebook and Twitter	No
30	N/A	N/A	No

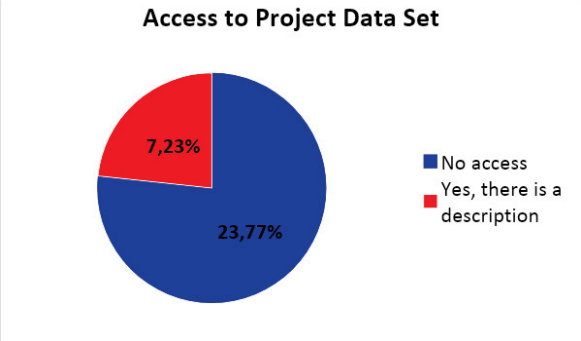

Figure 9. Open data of the projects

Discussion

An analysis of the project sites shows that the datasets are closed for free use in most commercial and research projects. Information about projects funded by Horizon 2020 grants is posted on project sites in the form of reports describing the structures of the datasets.

Conclusion

It was not possible to obtain reliable information about the structures, composition, data types of the project datasets analyzed as a result of the review of project sites. As a rule, data is a commercial secret of projects.

2.10. Achieved results of the project

The main result of the project is the achievement of the stated main goal of the project. The final result of the project is compared with the stated goal of the project. Therefore, the purpose of the question is to estimate achievable goals.

Data description

The survey results are showed in Table 10.

Table 10. Achieved results of the project

#	Name of Big Data project	Achieved results of the project
1	Intelligent Big Data Analytics for Industry 4.0 Systems	In progress - expected -->The project will fully utilise Artificial Intelligence, Data Analytics, Big Data and Industrial IoT technologies to offer 'smart' solutions for real-time monitoring and analysis of data gathered from the production/engineering processes and extracting the information needed for making the right decisions in time to predict certain events in advance.
2	Big Data Europe	The Big Data Europe (BDE) platform taht makes Big Data simpler, cheaper and more flexible than ever before. Considers basic building blocks to get started with common Big Data technologies and make integration with other technologies or applications easy.
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	*https://class-project.eu/publications, https://class-project.eu/results/deliverables*
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	in progress -->expected: providing effective and efficient algorithms and tools for Big Data Analytics and Mining is a fundamental aspect.
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	in progress - expected -->The deliverables of this research may include a global cost database of past transportation megaprojects. Identifying what input data is required to develop predictive analytic models from the dataset. A predictive analytic model to support front-end decision-making with realistic project cost forecasts on future megaprojects.
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	Data Quality Assurance plays an important role in many high-energy physics experiments. To ease the burden of human quality label assignment, authors investigated several state-of-the-art machine learning methods that can automate this process. The evaluated selection of the machine learning methods include artificial neural networks, support vector machines (with linear and non-linear kernels), as well as Random Forests and Naive Bayes Classifier. Obtained results indicate, that over 75% of all data points classified by human experts, can be correctly evaluated without any human interaction using Random Forest classifier with over 95% precision. Experiments was conducted Time Projection Chamber (TPC).
7	AnalyticOps	Optimizing hadoop cluster performance, construction of heterogeneous data storage, unification of access to hadoop data and databases
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	BigDataGrapes is targeting technology challenges of the grapevine-powered data economy as its business problems and decisions requires processing, analysis and visualisation of data with rapidly increasing volume, velocity and variety: satellite and weather data,

		environmental and geological data, phenotypic and genetic plant data, food supply chain data, economic and financial data and more. *https://www.bigdatagrapes.eu/deliverables*
9	Forming of Data Science Competence for bridging the Digital Divide	To conduct an empirical study of the degree of digital skills and Data Science literacy of the population in order to reveal different categories of social groups. To identify the competencies, knowledge and skills that make up Data Science literacy. To develop an educational model, curriculum and learning aids targeted at some social groups on the development of their Data Science skills.
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	PESHES Applciation.
11	Agile Analytics on Big Data Cubes	The project has advanced the existing, world-leading rasdaman Array Database technology wrt. query functionality, inter-federation data processing with automatic data and query distribution, tape archive integration, and 3D/4D visualization based on NASA's virtual globe technology. Large data centers (ECMWF, PML, MEEO/ESA, GeoScience Australia, JacobsUni) were set up water, air, weather, and planetary services on 3D & 4D data cubes up to Petabyte-size with user-tailored clients for both visual and textual ad-hoc mix&match. The project has advanced open Big Data standards in OGC, RDA, and ISO (in particular: write ISO SQL/MDA).
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	*https://www.ibidaas.eu/deliverables/*
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	Real-time detection and prediction of trajectories, Detection and prediction of important events related to moving entities, Advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from moving entities,Real-time in-situ processing of multiple data streams,Provision of integrated views of streaming data with archival data, Provision of advanced solutions for managing spatio-temporal data. *http://datacron1.ds.unipi.gr:9082/systems-tools-resources/*
14	Diagnosis and energy optimizing in production systems of Industrial IoT	Ready solutions for industrial automation: production monitoring, malfunctioning detection
15	NB-IoT Vodafone	Big Data Scoring, Vodafone Analytics, Vodafone Smart City, Narrow Band Internet of Things (NB-IoT) 2020: Income UAH 18 billion (+ 14%), 19 million customers, UAH 3.8 billion of investments
16	BigMedilytics - Big Data for Medical Analytics	*https://www.bigmedilytics.eu/deliverables/*
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision	(1) Genetic regulatory elements identification for unknown genes detection in sequenced genomes and for the aim of genomic mapping.

	Making Based on Analysis of Big Data Streams for Scientific Research	(2) To predict the type and malignancy of breast cancer based on information for mutations in genes associated with it, the level of expression and the associated epigenetic information. The driving force of the project is an interdisciplinary team, combining expertise in information science and technology, engineering foundations and technical deployment of software methods and tools, as well as established scholars in the areas of molecular biology and medical genetics.
18	Innovative Big Data solutions for smart cities (Big4Smart)	Quantitative progress assessment of implementation of smart city strategy; Feedback on efficiency of current policies; Timely and informed decision making; Increased understanding of future challenges; Identification of best practices to be followed in other cities.
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	The Big Data application provides processing and analytics of Big sensor data originated from smart personal devices with the aim to detect and recognize human activities and behaviour.
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	The TrafficSense, traffic monitoring, control and adaptation platform, was developed based on Big Mobility Data processing and analytics. It provides continuous monitoring of traffic situation and detection of important traffic parameters, conditions and events, such as travel times along the street segments and traffic congestions in a real time. Upon detecting the traffic congestion on an intersection, the TrafficSense application leverages the feedback control loop mechanism to provide traffic adaptation based on dynamic configuration of traffic lights duration in order to increase the traffic flows in critical directions at the intersections. We tested and evaluated developed application on the distributed cloud computing infrastructure. By varying the streaming workload and the cluster parameters we show the feasibility and applicability of our approach and the platform. http://casopisi.junis.ni.ac.rs/index.php/FUAutContRob/article/view/6720
21	Kyiv Smart Sity,	Public data portal, Public budget, Public procurement, E-petitions to the Kyiv City Council, CALL-center, GPS-trackers, Doctor's appointment, Kindergarten registration, Single account of a Kyiv resident
22	SoBigData Research Infrastructure	SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by “Big Data”. Building on several established national infrastructures, SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. It will not only strengthen the existing clusters of excellence in social

		data mining research, but also create a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities. In addition, as an open research infrastucture, SoBigData will promote repeatable and open science.
23	SoBigData	SoBigData open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art research data, methods, and services, into new research. It will contribute to the creation of a new community of researchers – data scientists – ready to exploit the opportunities of Big Data and to incorporate it in data-driven science and innovation. In addition, as an open research infrastructure, SoBigData will promote repeatable and open science.
24	SoBigData++	Becoming an advanced community, SoBigData++ is empowering its tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments. It will be open to users with diverse background, accessible on project cloud - aligned with EOSC - and also exploiting supercomputing facilities.
25	Application of clustering methods in some problems of speech signal processing	N/A
26	1991 Open Data Incubator	Digital decentralization, the data is collected digitally, in cloud, free via API.
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	Methods of highly productive calculation of mathematical models
28	Leveraging Big Data to Manage Transport Operations	The LeMO project performed seven case studies in transport related areas through the course of this project. These case studies involved organisations actively using Big Data for specific purposes and enable LeMO to understand strategies, actions and changes in behaviour associated with Big Data and identify their resultant merits and demerits. These case studies have produced evidence-based, clear and precise questions based on rigorous knowledge that illuminate opportunities, problems and viable solutions to be further investigated in the LeMO roadmap. The identification of these issues was complemented by a horizontal analysis to identify challenges, opportunities, limitations and other consequences of cross-disciplinary nature, and thus relevant to Big Data in transport sector.
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	As a result of the project, new algorithms will be created that will certainly contribute to the further development of learning methods machine and artificial intelligence.
30	N/A	Several scientific papers

Discussion

The results of the analyzed projects indicate the scale of these projects and correspond to the goals.

Conclusion

To achieve the set goals and obtain the results declared in the project, resources are needed: people, time, finances. An educational project aimed at developing educational courses on Big Data should contribute to understanding what content should be included in programs and what competencies should be formed in students.

2.11. Used Big Data tools

The purpose of the question is to collect information about the tools used in Big Data projects for determine the tools most used in the Big Data projects.

Data description

The results of the survey (Fig. 10) showed that most projects used Hadoop (19 projects, 63.3%). Cassandra is also a popular tool (10 projects, 33.3%). The rest of the projects used many other tools.

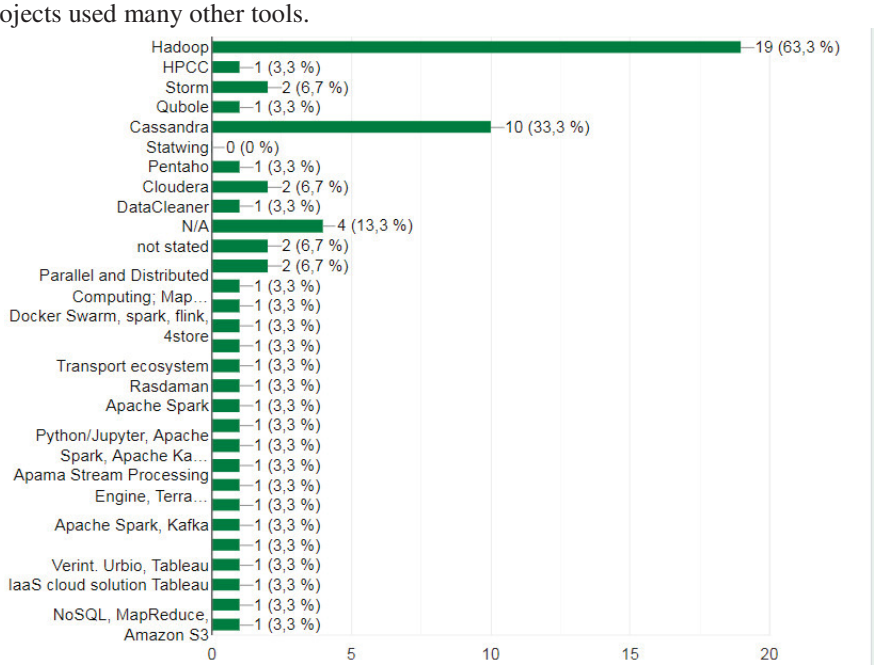

Figure 10. Used Big Data tools in Big Data projects

Discussion

One of the problems of introducing Big Data courses in universities is the high cost of equipment, tools, which must be used to teach students. The diagram (figure 10) gives an idea of which tool should be mastered first.

Conclusion

The competencies of students with a Big Data education should include the ability to use tools to handle Big Data sets. In this case, Hadoop can be one of the first tools that students will master.

2.12. Number of participants in the project

The purpose of the question is to get information about the complexity of projects in the form of the number of project participants.

Data description

The result of the survey (Table 11, Fig. 11) showed that on average, project development teams include 12 members. Individual projects have teams of 30 or more participants (3 projects, 10%). The number of participants in 8 (26.6%) projects could not be found. The respondent indicated the value of 600,000 in the project "Kyiv Smart Sity" incorrectly. This value means the number of users of the system "Kyiv Smart Sity".

Table 11. Number of participants in the project

#	Name of Big Data project	Number of participants in the project
1	Intelligent Big Data Analytics for Industry 4.0 Systems	N/A
2	Big Data Europe	16
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	6
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	N/A
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	N/A
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	1
7	AnalyticOps	N/A
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	9
9	Forming of Data Science Competence for bridging the Digital Divide	8
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	30
11	Agile Analytics on Big Data Cubes	6
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	13
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	8
14	Diagnosis and energy optimizing in production systems of Industrial IoT	5
15	NB-IoT Vodafone	N/A

16	BigMedilytics - Big Data for Medical Analytics	35
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	11
18	Innovative Big Data solutions for smart cities (Big4Smart)	18
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	2
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	2
21	Kyiv Smart Sity,	15
22	SoBigData Research Infrastructure	11
23	SoBigData	12
24	SoBigData++	31
25	Application of clustering methods in some problems of speech signal processing	N/A
26	1991 Open Data Incubator	N/A
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	N/A
28	Leveraging Big Data to Manage Transport Operations	5
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	1
30	N/A	15

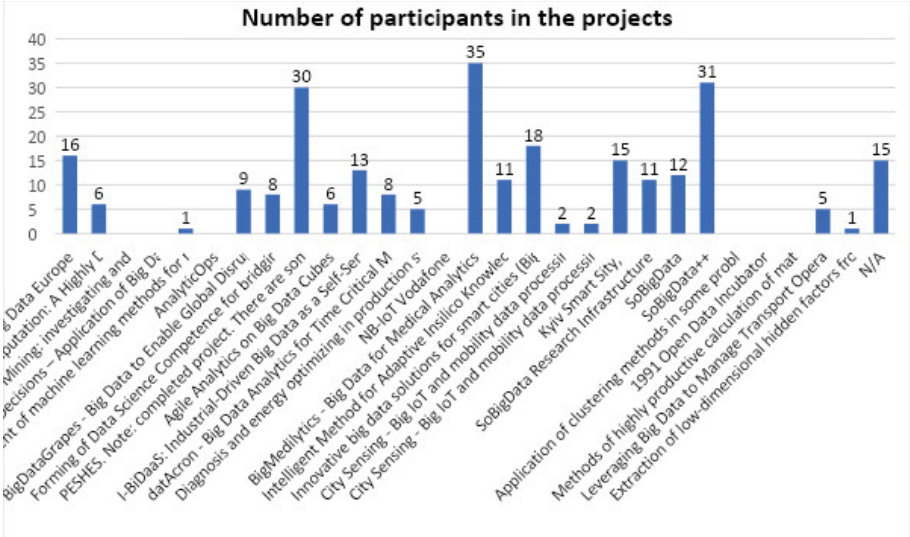

Figure 11. Number of participants of the projects

Discussion

Big data project teams are the resource that ensures a successful project. The number of project participants is determined depending on the complexity and financing of the project.

Conclusion

Teamwork, the ability to communicate, knowledge of conflict management are the skills that students need to develop, including the relevant disciplines in educational programs based on Big Data

2.13. Names and link for each partner in the project

The purpose of the question is to identify each partner in the Big Data project and determine the average number of participants for projects.

Data description

The result of the survey are in the Table 12 and Fig. 12. Some projects did not display information about their partners (N/A label). The number of partners indicated by the researchers is in the range from 0 to 34. As a result, the average number of partners in the analyzed projects is five. A value of zero means no information about partners on project sites.

Table 12. Names and link for each partner in the project

#	Name of Big Data project	Name and link for each partner in the project
1	Intelligent Big Data Analytics for Industry 4.0 Systems	N/A
2	Big Data Europe	Partners' list *https://www.big-data-europe.eu/partners/*
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	1. Barcelona Supercomputing Center (BSC) - *https://www.bsc.es/* 2. Atos - *https://atos.net/en/* 3. Comune di Modena - *https://www.comune.modena.it/* 4. IBM Research - Haifa - *http://www.research.ibm.com/labs/haifa/* 5. Università degli studi di Modena e Reggio Emilia (UNIMORE) - *https://www.unimore.it/* 6. Maserati - *https://www.maserati.com/*
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce paradigm and for other distributed computing approaches	N/A
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	N/A
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their	WEITI - Wydział Elektroniki i Technik Informacyjnych *www.elka.pw.edu.pl*

	visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	
7	AnalyticOps	N/A
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	1. Agroknow, Greece - Nikos Manouselis 2. SRIMA AI, Bulgaria - Todor Primov 3. Consiglio Nazionale Delle Richerche (CNR, Italy) - Raffaele Perego 4. Katholieke Universiteit Leuven (KULeuven, Belgium) - Katrien Verbert 5. Geocledian GmbH (GEOCLEDIAN Germany) - Stefan Scherer 6. Institut National de la Recherché Agronomique (INRA, France) - Pascal Neveu 7. Agricultural University of Athens (AUA, Greece) - Katerina Biniari 8. Abaco SpA (ABACO, Italy) - Simone Parisi 9. SYMBEEOSIS EY ZHN S.A. - (Symbeeosis, Greece) Konstantinos Rodopoulos
9	Forming of Data Science Competence for bridging the Digital Divide	N/A
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	There are many partners. We recommend to look them through the project's site
11	Agile Analytics on Big Data Cubes	1. Rasdaman GMBH, Plymouth Marine Laboratory Limited, 2. European Centre For Medium-Range Weather Forecasts, Meteorological And Environmental Earth Observation Srl, 3. Communication & Information Technologies Experts Anonymos Etaireia Symvouleftikon Kai Anaptyxiakon Ypiresion, 4. The Australian National University
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	1. Foundation For Research And Technology Hellas (FORTH) 2. Barcelona Supercomputing Center (BSC) 3. Ibm Israel - Science And Technology Ltd (IBM) 4. Centro Ricerche Fiat (CRF) 5. Software Ag (SAG) 6. Caixabank (CAIXA) 7. The University Of Manchester (UNIMAN) 8. Ecole Nationale Des Ponts Et Chaussees (ENPC) 9. Atos Spain S.A. (ATOS) 10. Aegis It Research Ltd (AEGIS) 11. Information Technology For Market Leadership (ITML) 12. University Of Novi Sad Faculty Of Sciences (UNSPMF) 13. Telefonica Investigacion Y Desarrollo Sa (TID)
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	1. University Of Piraeus Research Center - Greece 2. Frhf-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V (Fraunhofer) Germany

		3. National Center For Scientific Research ``Demokritos`` Greece 4. Naval Academy Reseach Institute - France 5. Nato Sto- Centre For Maritime Research And Experimentation - Italy 6. Boeing Research And Technology Europe - Spain 7. Reference Center For Research, Development And Innovation In Atm - Spain 8. Imisg Global - United Kingdom
14	Diagnosis and energy optimizing in production systems of Industrial IoT	N/A
15	NB-IoT Vodafone	N/A
16	BigMedilytics - Big Data for Medical Analytics	1. Fundacion Para La Investigacion Del Hospital Clinico De La Comunitat Valenciana, Fundacion Incliva Spain 2. 2. Instituto Tecnologico De Informatica Spain 3. 3erasmus Universitair Medisch Centrum Rotterdam Netherlands 4. Achmea Bv Netherlands 5. Gie Axa France 6. Optimedis Ag Germany 7. Atos Spain Sa Spain 8. Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek Tno Netherlands 9. Technische Universiteit Eindhoven Netherlands 10. Contextflow Gmbh Austria 11. Huawei Technologies Duesseldorf Gmbh Germany 12. Royal College Of Surgeons In Ireland Ireland 13. Region Stockholm Sweden 14. Hasso-Plattner-Institut Fur Digital Engineering Ggmbh Germany 15. "National Center For Scientific Research ""Demokritos""" Greece 16. Athens Technology Center Anonymi Biomichaniki Emporiki Kai Techniki Etaireia Efarmogon Ypsilis Technologias Greece 17. Gottfried Wilhelm Leibniz Universitaet Hannover Germany 18. Universidad Politecnica De Madrid Spain 19. Servicio Madrileno De Salud Spain 20. Medizinische Universitaet Wien Austria 21. Ibm Israel - Science And Technology Ltd Israel 22. Institut Curie France 23. Teknologian Tutkimuskeskus Vtt Oy Finland 24. Deutsches Forschungszentrum Fur Kunstliche Intelligenz Gmbh Germany 25. Charite - Universitaetsmedizin Berlin Germany 26. Aok Nordost - Die Gesundheitskasse Germany 27. Universitaetsklinikum Essen Germany 28. University Of Southampton United Kingdom 29. My Mhealth Limited United Kingdom 30. Astrazeneca Uk Limited United Kingdom 31. Stichting Olvg Netherlands

		32. Stichting Elisabeth-Tweesteden Ziekenhuis Netherlands 33. Erasmus Universiteit Rotterdam Netherlands 34. Privredno Drustvo Za Pruzanje Usluga Istrazivanje I Razvoj Nissatech Innovation Centre Doo Serbia
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	*https://insilicokdd.com/network-partners/*
18	Innovative Big Data solutions for smart cities (Big4Smart)	ULSIT: https://unibit.bg/
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	1. Faculty of Electronic Engineering, University of Nis, Serbia, http://www.elfak.ni.ac.rs 2. Future SOC Lab, Hasso Plattner Institute, Potsdam, Germany
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	1. Faculty of Electronic Engineering, University of Nis, Serbia, http://www.elfak.ni.ac.rs 2. Future SOC Lab, Hasso Plattner Institute, Potsdam, Germany
21	Kyiv Smart Sity,	N/A
22	SoBigData Research Infrastructure	1. The University Of Sheffield Http://Www.Shef.Ac.Uk/, 2. Universita Di Pisa, 3. Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V., 4. Tartu Ulikool, 5. Scuola Imt (Istituzioni, Mercati, 6. Tecnologie) Alti Studi Di Lucca, 7. Gottfried Wilhelm Leibniz Universitaet Hannover, 8. King's College London, 9. Scuola Normale Superiore, 10. Aalto Korkeakoulusaatio Sr, 11. Eidgenoessische Technische Hochschule Zuerich, 12. Technische Universiteit Delft
23	SoBigData	1. USFD - The University of Sheffield- (PI: Hamish Cunningham) - UK 2. UNIPI - Università di Pisa - (PI: Dino Pedreschi) - Italy 3. FRH - Fraunhofer IAIS and IGD- (PI: Gennady Andrienko) - Germany 4. UT - Tartu Ulikool- (PI: Marlon Dumas) - Estonia 5. IMT - Scuola IMT (Istituzioni, Mercati, Tecnologie) Lucca - (PI: Guido Caldarelli) - Italy 6. LUH - Gottfried Wilhelm Leibniz Universitaet Hannover - (PI: Wolfgang Nejdl) - Germany 7. KCL - King's College London- (PI: Tobias Blanke) - UK 8. SNS - Scuola Normale Superiore- (PI: Fabrizio Lillo) - Italy 9. AALTO - Aalto University- (PI: Aristides Gionis) - Finland 10. ETHZ - ETH Zurich- (PI: Dirk Helbing) - Switzerland

		11. TUDelft - Technische Universiteit Delft- (PI: Jeroen Van Den Hoven) - Netherlands
24	SoBigData++	1. Consiglio Nazionale delle Ricerche (CNR) 2. The University of Sheffield (USFD) 3. Università di Pisa (UNIPI) 4. Fraunhofer Gesellschaft zur Foerderung der Angewandten Forschung E.V. (FRH) 5. Tartu Ulikool (UT) 6. Scuola IMT (Istituzioni, Mercarti, Tecnologie) Alti Studi di Lucca (IMT) 7. Gottfried Wilhelm Leibniz Universitaet Hannover (LUH) 8. King's College London (KCL) 9. Scuola Normale Superiore (SNS) 10. Aalto Korkeakoulusaatio SR KORKEAKOULUSAATIO SR (AALTO) 11. Eidgenoessische Technische Hochschule Zuerich (ETH Zürich) 12. Technische Universiteit Delft (TUDelft) 13. Stichting EGI (EGI) 14. Ecole d'Economie de Paris (PSE) 15. Università degli Studi di Roma La Sapienza (UNIROMA1) 16. OpenAIRE AMKE (OpenAIRE) 17. Centre National de la Recherche Scientifique CNRS (CNRS) 18. Re-Imagine Europa (RIE) 19. Kozep-Europai Egyetem (CEU) 20. Universitat Rovira I Virgili (URV) 21. Center for the Study of Democracy (CSD) 22. Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (BSC) 23. Universidad Pompeu Fabra (UPF) 24. Kunliga Tekniska Hoegskolan (KTH) 25. Università degli Studi dell'Aquila (UAQ) 26. Società Cooperativa Sociale Eliante ONLUS (Eli) 27. CarloRattiAssociati S.R.L. (CRA) 28. Nubisware SRL (Nubisware) 29. Universiteit Van Amsterdam (UvA) 30. Stacc OU (STACC) 31. Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna (SSSA)
25	Application of clustering methods in some problems of speech signal processing	N/A
26	1991 Open Data Incubator	N/A
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	N/A
28	Leveraging Big Data to Manage Transport Operations	1. Johann Wolfgang Goethe-Universitaet Frankfurt Am Main ,

		2. Confederation of Organisations In Road Transport Enforcement AISBL, 3. Bird & Bird LLP , 4. Panteia BV, 5. Bird & Bird (Belgium) LLP
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	*https://weka.pwr.edu.pl/en/*
30	N/A	N/A

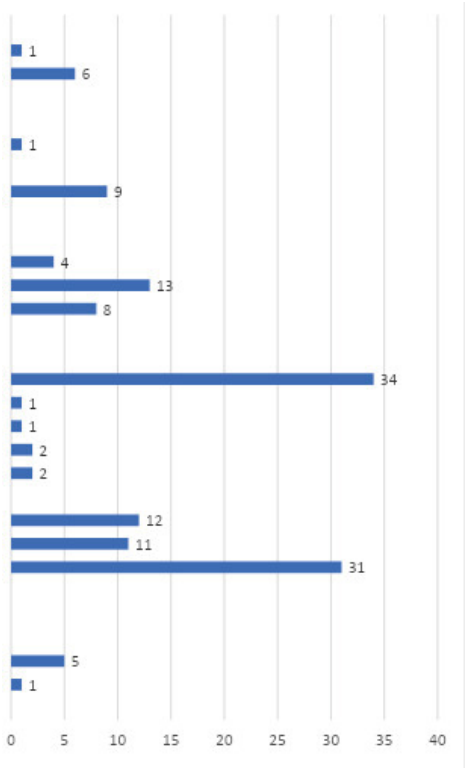

Figure 12. Number of partners in the Big Data projects

Discussion

The partners in the project are the organizations with whose help the project can be implemented. Many organizations may lack resources (human, material, etc.), experience, expertise in certain areas related to the project. In this case, joining forces can help the successful implementation of the project. International projects and projects funded by EU programs (Horizon 2020) require mandatory participation of partners. Role of project partners: elaboration of the project idea, cooperation during the implementation of the project, there may be joint financing, examination and assessment of project results, etc. The success of a project may depend on the

number and quality of project partners. The analyzed projects, which ended successfully, had five or more partners.
Projects with the number of participants was not specified or exceeded 3 people turned out to be more popular. This is probably due to the agreements on non-proliferation of proprietary information, which are concluded with employees at the enterprises where the researches were carried out. There are also a few projects that were implemented by one performer. This may indicate the fulfilment of orders by individual researchers, possibly in the presence of limitations in their funding. More popular are projects carried out by more than 3 participants. Attracting a large number of participants may be due to the timing of the project or the complexity of the tasks.

Conclusion

The participation of partners in projects contributes to the success of projects. Projects funded by EU programs (Horizon 2020) require partners to participate in projects. As a rule, the staff of the companies carry out commercial projects. Sub-contractors may be employed to carry out highly specialized project tasks. Commercial and grant-financed projects can be considered more successful, probably due using of more flexible personnel management schemes. The increase in the number of project participants should be justified and determined by the complexity of the task being performed, as well as the duration of the project. The project partners partially act as the guarantor of the expected success.

2.14. Project managers

The purpose of the question is to identify the Project managers of Big Data projects. This information can be used to build the organizational structure of the project, project consultations and etc.

Data description

The result of the survey as showed in Table 13 and Fig.13, 14, 15. The Fig. 13 shows the number of project managers in each project. The Fig. 14 presents names and numbers (as a percentage) of project managers in projects. The Fig. 15 presents the number of Prof, Dr and no academic titles

Table 13. Name of project managers

#	Name of Big Data project	Name of project managers
1	Intelligent Big Data Analytics for Industry 4.0 Systems	Dr Savas Konur, Dr Dhaval Thakker, Dr Geev Mokryani, Dr Sohag Kabir
2	Big Data Europe	Kimmo Rossi
3	CLASS - Edge and Cloud Computation: A Highly Distributed Software for Big Data Analytics	Eduardo Quinones Moreno (Barcelona Supercomputing Center - BSC)
4	Big Data Analytics and Mining: investigating and testing distributed formulations of data mining algorithms that are suitable for the MapReduce	Dr G DiFatta

	paradigm and for other distributed computing approaches	
5	Big Data for Big Decisions – Application of Big Data Analytics and Machine Learning for Front-end Decision-making in Transportation Megaprojects	Prof Stephen Ogunlana
6	Development of machine learning methods for monitoring the quality of large-volume data and interactive methods of their visualization on the example of the ALICE experiment at the Large Hadron Collider at CERN	Dr Tomasz Trzciński
7	AnalyticOps	N/A
8	BigDataGrapes - Big Data to Enable Global Disruption of the Grapevine-powered industries	Dr. Nikos Manouselis, Dr. Raffaele Perego
9	Forming of Data Science Competence for bridging the Digital Divide	Assist. Prof. Katia Rasheva, phd
10	PESHES. Note: completed project. There are some other project initiatives (e.g. through IDEJE program)	Prof. Dr Ivanka Popović, University of Belgrade; Prof. Dr Ivan Luković - coordinator of University of Novi Sad, etc...
11	Agile Analytics on Big Data Cubes	Peter Baumann, p.baumann@jacobs-university.de
12	I-BiDaaS: Industrial-Driven Big Data as a Self-Service Solution	Dr. Sotiris Ioannidis
13	datAcron - Big Data Analytics for Time Critical Mobility Forecasting	Prof. George Vouros Dept of Digital Systems, University of Piraeus, Greece Email: georgev@unipi.gr
14	Diagnosis and energy optimizing in production systems of Industrial IoT	Michael Dvornichenko
15	NB-IoT Vodafone	O. Pidgirna
16	BigMedilytics - Big Data for Medical Analytics	Supriyo Chatterjea
17	Intelligent Method for Adaptive Insilico Knowledge Discovery and Decision Making Based on Analysis of Big Data Streams for Scientific Research	Prof. PhD Plamenka Borovska
18	Innovative Big Data solutions for smart cities (Big4Smart)	Assoc. Prof. Dr. Dessislava Georgieva Petrova-Antonova
19	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities	Prof. Dragan Stojanovic
20	City Sensing - Big IoT and mobility data processing and analytics in Smart Cities (Extension)	Prof. Dragan Stojanovic
21	Kyiv Smart Sity	Yurii Nazarov, Director of the Department of Information and Communication Technologies of the Kyiv City State Administration.
22	SoBigData Research Infrastructure	THE PROJECT COORDINATOR Fosca Giannotti, Ph.D, Two Deputy coordinators assist the PC: Kalina Bontcheva (USFD) and Dino Pedreschi (UNIPI).

		The Project Manager is Roberto Trasarti (CNR). Project Administrative Manager (PAM), Chiara Falchi (CNR).
23	SoBigData	THE PROJECT COORDINATOR Fosca Giannotti, Ph.D, director of research at CNR, chair of the Project Management Board, the Project Steering Board and the Project Management Team
24	SoBigData++	Fosca Giannotti, Ph.D, director of research at CNR, chair of the Project Management Board, the Project Steering Board and the Project Management Team
25	Application of clustering methods in some problems of speech signal processing	N/A
26	1991 Open Data Incubator	Denis Gursky, cofounder
27	Methods of highly productive calculation of mathematical models of heterogeneous middle ranges and processing of great tributes (Big Data) based on supercomputer technologies	Dr. Prof. Olexander Khimich
28	Leveraging Big Data to Manage Transport Operations	Anders-Johan Almås
29	Extraction of low-dimensional hidden factors from large sets of non-negative data	Dr hab. Rafał Zdunek
30	N/A	Zakaria Maamar

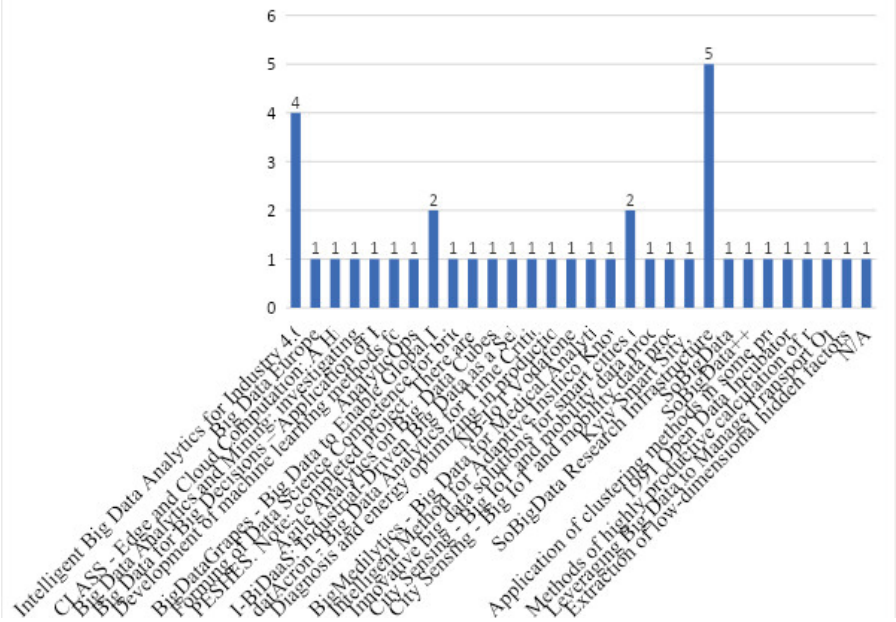

Figure 13. Number of project managers in each project

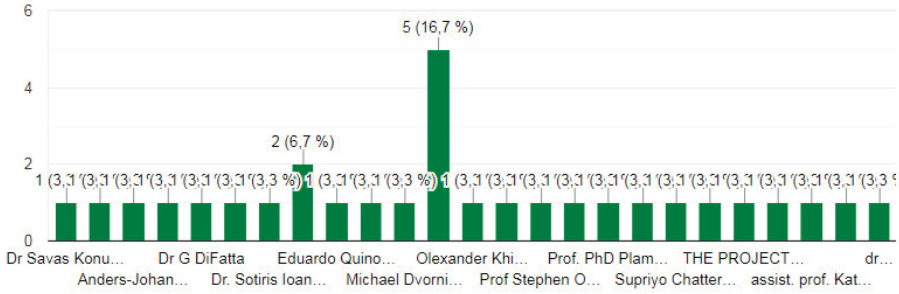

Figure 14. Names and numbers of project manager in each projects

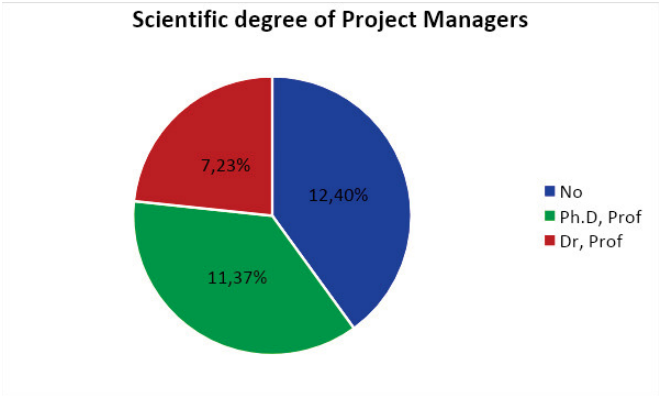

Figure 15. Number of Prof, Dr, and no academic titles

Discussion

The manager's science degree acts as a guarantee of high quality of the project implementation without breaking deadlines. About half of project managers are professors or have a Doctor Sciences degree.

Conclusion

An academic degree is not required for a project leader. But its presence is desirable, since it acts as a guarantee of high-quality project implementation without violating deadlines. About half of the project managers are professors or have a Doctor Sciences degree. Commercial and grant-financed projects can be considered more successful, probably due using of more flexible personnel management schemes. Project managers with sciences degrees have certain advantages since they have fundamental and technical knowledge, team building and management skills already. High professionalism and organizational skills are needed a manager to lead to the success of the project.

3. Conclusions

For the reasons given, there was studied the innovative potential of scientific and commercial projects in the field of Big Data. It was identified the main directions of Big Data analytics development, the goals and objectives of Big Data development, the project results in a contribution to the development of regions, and the benefits of projects for users. The information obtained can be used to determine the content of educational programs for Big Data training courses.

REFERENCES

1. DEEPA N., PHAM Q.-V., NGUYEN D.C., BHATTACHARYA S., PRABADEVI B., GADEKALLU T.R., MADDIKUNTA P.K.R., FANG F., PATHIRANA P.N.: A survey on blockchain for Big Data: Approaches, opportunities, and future directions (2022) Future Generation Computer Systems, 131, pp. 209-226. DOI: 10.1016/j.future.2022.01.017.
2. XIANG C.: Study on the Role of Big Data Age in Making Fitness Programs for Individual Science (2022) Lecture Notes on Data Engineering and Communications Technologies, 130, pp. 162-166. DOI: 10.1007/978-3-030-99581-2_22.
3. REGGIO G., ASTESIANO E. Big-Data/Analytics Projects Failure: A Literature Review (2020) Proceedings - 46th Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2020, art. no. 9226362, pp. 246-255. DOI: 10.1109/SEAA51224.2020.00050.
4. DOUBAL F.N., ALI M., BATTY G.D., CHARIDIMOU A., ERIKSDOTTER M., HOFMANN-APITIUS M., KIM Y.-H., LEVINE D.A., MEAD G., MUCKE H.A.M., RITCHIE C.W., ROBERTS C.J., RUSS T.C., STEWART R., WHITELEY W., QUINN T.J.: Big data and data repurposing - using existing data to answer new questions in vascular dementia research (2017) BMC Neurology, 17 (1), art. no. 72, . DOI: 10.1186/s12883-017-0841-2.
5. BARTLETT A., LEWIS J., WILLIAMS M.L.: Generations of interdisciplinarity in bioinformatics (2016) New Genetics and Society, 35 (2), pp. 186-209. DOI: 10.1080/14636778.2016.1184965.
6. DELIC K.A., RILEY J.A.: On big science: A survey (2015) ACM International Conference Proceeding Series, 21-22-October-2015, art. no. a34,. DOI: 10.1145/2809563.2809622