Deploying a Strategy to Unlock Big Data Research and Teaching Activities in the West Balkan Region

Damien Graux
Inria, Université Côte d'Azur, CNRS, I3S, France
damien.graux@inria.fr

Valentina Janev
The Mihajlo Pupin Institute, University of Belgrade, Serbia
valentina.janev@pupin.rs

Hajira Jabeen
University of Cologne, Germany
hajira.jabeen@uni-koeln.de

Emanuel Sallinger
TU Wien, Austria and University of Oxford, UK
emanuel.sallinger@tuwien.ac.at|cs.ox.ac.uk

ABSTRACT

Big Data Analytics is a crucial component of the Big Data paradigm and deals with the extraction of knowledge from the enormous amount of data. As the number of Big Data related methods, tools, frameworks, and solutions is growing, there is a need to systematize the knowledge about the domain. Moreover, as this domain is rapidly involving, it is difficult to keep up to date with its latest and most efficient technologies; and this is thereby especially challenging for countries which have so far suffered from a lack of infrastructure in Big Data.

In this article, we report on the deployment of a strategy we design, fostering the knowledge and awareness around the challenges of Big Data analytics in the West Balkan region. In particular, we describe how we joined forces across multiple European institutions in order to design bespoke actions (from, e.g. classes, to student exchanges) to be applied in Serbia and the West Balkan region over several years.

CCS CONCEPTS

• **Information systems** → **Data management systems**; • **Applied computing** → **Education**.

KEYWORDS

Teaching Big Data Analytics; West Balkan Region; European Collaboration; Summer School

1 INTRODUCTION

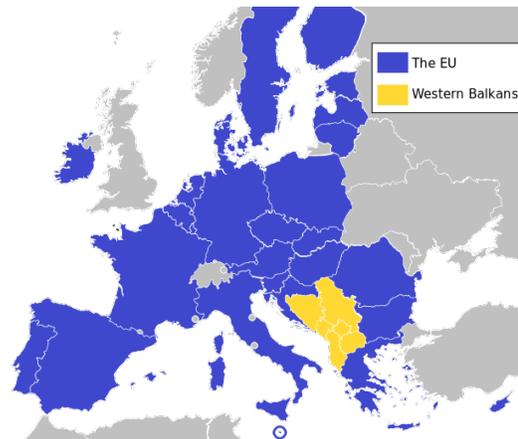Big Data refers to data sets which have large size and complex structure. The data size can range from dozens of terabytes to a

Figure 1: The Western Balkans and the European Union.

few Zettabytes and is still growing [22]. While more than 800,000 Petabyte ($1PB=10^{15}$ bytes) of data were stored in the year 2000, this volume will reach 35 Zettabytes ($1ZB=10^{21}$ bytes) by the end of 2019 [23], and is expected to grow 61% and exceeds 175 zettabytes by 2025 as per International Data Corporation expectations [16]. Big Data Analytics, hence, refers to the strategy of analysing large volumes of data that gathered from a wide variety of sources, including social networks, transaction records, videos, digital images and different kind of sensors.

Challenges [7] related to the European ability to exploit the potential of Big Data are: fragmentation of the data ecosystem, due to the different national policies, languages, and sectors involved; fragmentation of data research efforts and lack of effective exchange of results; the shortage of highly skilled persons for data-related jobs; and the complicated process of updating legislation. In an attempt to support the European data economy policy [5], in our project framework, our consortium proposed a training approach and established the infrastructure for collaborative work of teachers/trainers with PhD students and other interested parties.

This article describes the infrastructure that we established to reinforce organizational learning and capacity building in the West Balkan region (see Figure 1 for a precise location of the targeted region) and to facilitate teacher-trainee cooperation in the larger network of experts in the field of Enterprise Knowledge Graphs, Semantic Big Data Architectures, and Smart Data Analytics.

The rest of the article is structured as follows. Section 2 briefly reviews initiatives to teach Big Data. Then, Section 3 describes the high-level approach we adopted. In Section 4, we provide the reader with a complete overview on the designed lecture set, before detailing in Section 5, the specific actions we conducted. Finally, we present the created regional momentum around Big Data in Section 6 and conclude in Section 7.

2 RELATED WORK

To the best of our knowledge, our initiative is the first one which focused on bringing Big Data technologies through education to the West Balkan countries. Nevertheless, to date, multiple actions have already been deployed to teach Big Data technologies with curricula and specific studies have already explored the particular context of computer science education in the Balkans.

Teaching Big Data concepts. First of all, since the Big Data domain became (within the past decade) essential for most of the tech companies, teachers have set up several curricula in order to prepare their students. These programs are usually tailored for specific audiences as Big Data overlap multiple domains, or focus on specific technologies. For example, [20] designs the program to immunologists, [4] addresses key aspects of a Big Data unit for middle school, or [19] describes the state of the big data analysis education in Business information systems study in Slovakia. Other programs teach Big Data through specific tools *e.g.* [12] on Apache Spark, [14] using real-world datasets, or [15] thanks to open infrastructure.

In our project, we designed a program able to cover various aspects of the Big Data landscape (see Section 4 for a description of the lectures and modules) through the teaching of several state-of-the-art tools and technologies while showing the use of Big Data across multiple industrial use-cases.

Educational actions dedicated to the West Balkans. Due to its historical background, the West Balkan region has been receiving specific attention from computer science teachers within the past decades. In particular, several studies have been focused on the gender influences on studying computer science [9, 18]. More generally, several studies also describe transnational cooperation in higher education in the region [2, 3, 17]. Our project is therefore aligned with the aforementioned projects as it aims at building new cooperations between European Big Data experts and West Balkan students and industrial actors.

3 AN OPEN EDUCATION APPROACH

The overall objective is to stimulate scientific excellence and innovation capacity, to increase the research capacities and to unlock the research potential in the ICT area in the whole West Balkan region, turning the partner-institutes into regional points of reference when it comes to multidisciplinary ICT competence related to Big Data analytics. In early 2018, the consortium started activities for improving the skills and competences for smart data management through a set of actions including:

- the development of a Knowledge repository (Learning and Consulting Platform) that shall facilitate spreading excellence and exchange of learning materials and best practices between the international leading organizations and research institutions and industry from the West Balkan countries;
- organization of international events (training, workshops, webinars, conferences) in the West Balkan countries for raising awareness about future trends in Big Data, Semantic Tools and Technologies, standards and applications (or adoption) in the industry;
- forecasting exercises about future trends of data services in Europe.

The capacity building approach LAMBDA[1] introduced to the West Balkan partner institutes is based on the *Strategic Capacity Building Plan* that serves as a model and articulates how research institutions of disadvantaged Member States and regions can reach the level of internationally leading counterparts from Europe. Hence, in order to support the transfer of institutional knowledge and expertise to West Balkan staff, both from partner institutes and from relevant stakeholders in the region, a Learning and Consulting Platform was established using Drupal as a content management system. The platform facilitates collaboration including *e.g.* joined papers, information sharing, or stakeholders' database management. The learning materials that were produced in the project are free, stored in a public repository, and available online via an OpenCourseWare platform [8]. This repository aims at facilitating the exchange of learning materials, tools, project results, and best practices between the international leading organizations, research institutions and industry from the West Balkan countries.

4 BIG DATA LEARNING PROGRAM

In order to structure our Big Data learning program, we split it into **eight** modules which are tackling various angles of the knowledge we wanted the students to acquire. Following our open education approach, we release our lectures as different media types, meaning that in addition to the various actions we set up (see Section 5 for more details on them) this allows students to access them either though videos and slide decks or through conventional book chapters. Table 1 groups at a glance these figures about available media per module. Moreover, this openness and medium-variety helped us facing the COVID-19 crisis as it allows us to quickly deploy virtual classes and events since lockdowns and isolations forced us not to travel across Europe to meet physically.

Practically, these modules spread from very high-level and generic ones such as the *Foundations on Big Data* to specialized ones *e.g.* *Semantic Big Data Analytics* which target more advanced students. In the rest of the section, we describe briefly the content of multiple lectures we designed, grouped by modules, providing the reader with a panorama of the set-up knowledge transfer. We will finish by presenting the use-case oriented lectures that were made in order to tease Big Data technologies, paradigms and tools to specialized attendees who would be focusing on a specific industrial context. In addition, we indicate for each lecture whether it is intended for beginner attendees "(B)", for an advanced audience as technical details and theoretical concepts are discussed "(A)", or whether it also contains a hands-on session "(H)".

---

[1] https://project-lambda.org/

Module Name	Number of Lectures	Available as				
		Slides	Video	Paper	Chapter	Other
Surveys	3	1	–	2	1	–
Foundations on BD	3	–	–	–	3	–
Enterprise KG	5	–	3	2	2	–
BD & KG Tools	4	–	2	1	1	–
Semantic BD Archi.	5	4	2	–	1	–
AI & BD	4	2	3	–	–	–
Semantic BD Analytics	5	3	–	1	1	2
Case Studies	9	3	5	3	1	–

Table 1: Key figures of the 8 modules.

4.1 Surveys (3)

Survey on Big Data Tools (B). This introductory lecture discusses the Big Data processing pipeline and the Big Data Landscape from the following perspectives: Big Data Frameworks, NoSQL Platforms and Knowledge Graphs, Stream Processing Data Engines, Big Data Preprocessing, Big Data Analytics, and Big Data Visualization Tools.

Overview and Comparison of Machine Learning Algorithms (B). Big Data Analytics is a crucial component of the Big data paradigm and refers to the process of extracting useful knowledge from large datasets or streams of data. Due to enormity, high dimensionality, heterogeneous, and distributed nature of data, traditional techniques of data mining may be unsuitable to work with big data. In this lecture, different big data tools and machine learning algorithms are introduced, discussed, and analyzed.

Survey on Big Data Applications (B). The goal of this lecture is to shed light on different types of big data applications needed in various industries including healthcare, transportation, energy, banking and insurance, digital media and e-commerce, environment, safety and security, telecommunications, and manufacturing. For the selected industries, this lecture discusses challenges that can be addressed and overcome using semantic processing and knowledge reasoning approaches.

4.2 Foundations on Big Data (3)

Big Data Ecosystem (B). The rapid development of digital technologies, IoT products and connectivity platforms, social networking applications, video, audio, and geolocation services has created opportunities for collecting/accumulating a large amount of data. The high volumes of structures and unstructured data, stored in a distributed manner, and the wide variety of data sources pose problems related to data/knowledge representation and integration, data querying, business analysis and knowledge discovery. This lecture serves to characterize the relevant aspects of the Big Data Ecosystem with respect to big data characteristics, the components needed for implementing end-to-end big data processing and the need for using semantics for improving the data management, integration, processing, and analytical tasks.

Introduction to Knowledge Graphs (B). Knowledge Graphs (KGs) are one of the key trends among the next wave of technologies. This lecture actually presents why this multitude of definitions is one of the strengths of the area. In particular, it presents three views on Knowledge Graphs: KGs as Knowledge Representation Tools, KGs as Knowledge Management Systems, and KGs as Knowledge Application Services.

Big Data Outlook, Tools, and Architectures (A). This lecture covers the history of Big Data and discusses prominent related terminologies. The significant technologies including architectures and tools are reviewed. Finally, the lecture reviews big KGs that attempt to address the challenges (*e.g.*, heterogeneity, interoperability, variety) of big data through their specialised representation format.

4.3 Enterprise Knowledge Graphs (5)

What is a Knowledge Graph? (B). Knowledge Graphs are being used in a variety of applications including web search, answering questions, and for data integration. Knowledge graphs also target output for Natural Language Processing, computer vision algorithms, and Machine Learning algorithms more generally. This lecture summarizes KGs, how to create them, and how to use them with modern artificial intelligence algorithms.

Creation of Knowledge Graphs (A). This lecture introduces how Knowledge Graphs are generated. The goal is to gain: (1) an overview of the different approaches that were proposed for creating a Knowledge Graph and find out more details about the current prevalent ones; (2) an understanding of the different solutions to generate Knowledge Graphs; (3) knowledge to choose the mapping language that suits best a certain use case. After this lecture, the learner should have an understanding of the different solutions available to generate Knowledge Graphs and should be able to choose the mapping language that best suits a certain use case.

Extraction for Knowledge Graphs (A+H). This lecture discusses the topic of (web data) extraction for Knowledge Graphs. Web data extraction is essential to make information available on the web accessible and usable by Knowledge Graphs.

Swift Logic for Big Data and Knowledge Graphs (A). Many modern companies and organizations wish to maintain knowledge in the form of a corporate knowledge graph and to use and manage this knowledge via a knowledge graph management system. The lecture presents how complex reasoning tasks over KGs can be performed while guaranteeing performance both in terms of scalability and in terms of complexity.

Reasoning in Knowledge Graphs: An Embeddings Spotlight (A). In this lecture, reasoning in Knowledge Graphs is introduced. A broad overview focusing on the multitude of reasoning techniques is given: spanning logic-based reasoning, embedding-based reasoning, neural network-based reasoning, and their fascinating connections.

4.4 Big Data & Knowledge Graph Tools (4)

Spark using Scala (B+H). This lecture introduces Apache Spark (Architecture, Libraries), the underlying data structures (Resilient Distributed Dataset) and an example with Scala [21].

Data Science with Spark and Hadoop (B+H). This lecture technically introduces the use of Apache Spark [21] and MapReduce [6] via Hadoop for Data Science applications.

Context-Based Entity Matching for Big Data (A). In the Big Data era, where variety is the most dominant dimension, the RDF data model [13] enables the creation and integration of actionable knowledge from various data sources. This lecture presents how datasets

can be fused together thanks to the concept of semantic context: explaining the notion of similarity for RDF datasets and describing Formal Concept Analysis to map contextually equivalent entities.

Vadalog System (A). Over the past years, there has been a resurgence of Datalog-based systems in the database community as well as in industry. This lecture presents the Vadalog system: a state-of-the-art Datalog-based system for performing complex logic reasoning tasks, such as those required in advanced KG [1].

4.5 Semantic Big Data Architecture (5)

Reasoning in Knowledge Graphs (A). This lecture discusses reasoning in Knowledge Graphs. Reasoning is essential to gain value from Knowledge Graphs by deriving insights and making available new implicit data from existing data. The lecture both covers theoretical and practical aspects, and providing accessible examples based on Vadalog [1].

Introduction to Big Data Architecture (B). This technical lecture covers the existing advanced Big Data architectures following a bottom-up approach. The important knowledge to design and architect scalable solutions for challenging problems is introduced. The primary components in the architecture of such systems are also presented, such as distributed file systems or storage systems.

Big Data Solutions in Practical Use-cases (B). This lecture focuses on building Big Data solutions. The students are introduced to unique problem characteristics that drive Big Data and the unending technology options to solve them.

Distributed Big Data Frameworks (A+H). Processing frameworks are essential components of Big Data systems. There are three categories of such frameworks namely: Batch-only frameworks (Hadoop), Stream-only frameworks (Storm, Samza), and Hybrid frameworks (Spark, Hive, and Flink). This lecture introduces them and details Spark focusing on its data model.

Data Lakes and Federated Query Processing (A). Federated query processing techniques provide a solution to scale up to large volumes of data distributed across multiple data sources. This lecture summarizes the main characteristics of a federated query engine, reviews the current state of the field, and outlines the problems that still remain open and represent grand challenges for the area.

4.6 Artificial Intelligence & Big Data (4)

Data for AI: Foresight (B). This lecture presents the importance of data, the vision and actions from some industrial organizations point of view. The lecture lists some examples of technical solutions that have been suggested from secure and trusted data sharing.

AI and Knowledge Graphs (B). This lecture focuses on the importance of supporting different kinds of knowledge within a KG at the same time, especially when the KG is a component of large AI architectures where many flavours of AI and data processing come together. The focus here is on the connections.

Conversational AI (B). This lecture describes Speech-to-Text technologies and how they can be used to build Question Answering systems and pipelines over Knowledge Graphs in order to build complex AI systems.

The Revolution of AI (B). This lecture reviews major milestones in the history of intelligent systems and presents cutting-edge research projects that could shape our future with AI. Besides technical research objectives, it proposes economic and ethical initiatives.

4.7 Semantic Big Data Analytics in practice (4)

Distributed Big Data Libraries (A+H). Practically, the Big Data frameworks use different APIs for graph computation and graph processing. In this lecture, the important libraries built on top of Apache Spark are covered. These include SparkSQL, GraphX and MLlib. Students learn to build scalable algorithms in Spark.

Scalable Semantic Analytics Stack (B). This introductory lecture introduces SANSA [11] that provides support for: (1) efficient data distribution and semantics-aware computation of latent resource embeddings for knowledge graphs; (2) adaptive distributed querying; (3) efficient self-optimising inference execution plans; and (4) efficient distributed machine learning on semantic knowledge graphs of extremely large scale.

Distributed Semantic Analytics I (B+H). This hands-on lecture covers the needs and challenges of distributed analytics through practical examples and then dives into the practical details of SANSA [11] to perform scalable analytics for knowledge graphs.

Distributed Semantic Analytics II (A+H). This hands-on lecture covers the setup, APIs and several SANSA layers in details. At the end of it, the students can create programs that use SANSA APIs.

4.8 Case studies (9)

Semantic information infrastructures from business information delivery to water management (B). The lecture introduces semantic information infrastructures and shows a variety of examples where these infrastructures can provide competitive advantages to the organizations that have adapted them.

Soft computing for transparent synthesis of Geo Big Data (B). This lecture focuses on the strategies that can be set up in order to efficiently deal with large-scale geo-tagged data.

Spatio-temporal models for social and service robots (B). This lecture introduces the data challenges in the area of assistant robots and especially tackles the solutions deployed to deal with the large sensor-generated data flows.

Intelligent systems for road safety (B). This lecture presents how the Big Data technologies have been used so far in order to improve safe driving experience.

Reasoning on Financial Knowledge Graphs: The Case of Company Network (A). This lecture introduces the Enterprise Knowledge Graph of Italian companies of the Central Bank of Italy.

Embedding-based Recommendations on Scholarly KGs (B). One of the applications of Knowledge Graph Embedding models is to provide link predictions, which can also be viewed as a foundation for recommendation services, *e.g.* high confidence "co-author" links in a scholarly KG can be seen as suggested collaborations. In this lecture, KGEs are reconciled with a specific loss function (Soft

Margin) and examined with respect to their performance for the coauthorship link prediction task on scholarly KGs.

Open and Big Data – Utilization Perspective (B). Although each government in Europe with their public administration services can be treated as a big data ecosystem, the opportunities of interconnecting, integrating and processing data on the European level presents a real challenge nowadays. In this lecture, the potential and challenges of implementing the European data strategy in the West Balkan region are discussed, bearing in mind the needs and opportunities of SMEs and NGOs.

Data Analytics for Energy Sector (B). Big Data technologies are often used in domains where data is generated, stored, and processed with rates that cannot be efficiently processed by one computer. One of those domains is definitely the domain of energy. Here, the processes of energy generation, transmission, distribution, and use have to be concurrently monitored and analyzed in order to assure system stability without brownouts or blackouts. Some of these tools, deployed for monitoring and controlling energy, are presented in this lecture. This balances the end-user perspective (such as Non-Intrusive Load Monitoring, Energy Conservation Measures, and User Benchmarking) and the perspective of the grid (production, demand and price forecasting).

Predictive Analytics in Renewable Energy Systems (B). With the aim of improving ecological interest, the share of renewable energy sources in energy production has to be increased. Nonetheless, that growth adversely influences the grid's instability, as a result of dependencies of renewable solutions on weather conditions. Therefore, to provide stable energy delivery, planning the consumption in advance with respect to the availability of renewable production is necessary. This lecture is focused on comparing current state-of-the-art approaches for different renewable energy sources.

5 KNOWLEDGE TRANSFERRING ACTIONS

As presented in Section 3, the overall goal of our initiative is to transfer knowledge on Big Data technologies in West Balkan so that students are familiar with these tools and paradigms. Therefore, we did not only focus on creating a set of lectures (see their detailed presentations in Section 4) but rather built momentum thanks to the participation of multiple top researchers from various institutes and universities across Europe. Each expert was then in charge of advertising their own domain of expertise, allowing thereby students to be at the forefront of these various research domains.

Moreover, we also set up several actions in order to strengthen the connections between states and people and in order to create means of collaboration and knowledge exchange. In particular, we promoted our initiative in multiple international conferences, organized summer schools and invited Master students from Balkan to Western European states.

5.1 Advertising initiatives

In order to promote the project and to connect with experts from Big Data related domains, part of our efforts was orientated towards international research events. Participating there allowed us to involve new researchers and institutions in the project. In two years, the different members of the consortium attended more than 30

Lecturers		Attendees from West Balkan	
Country	Number	Country	Number
Germany	5	Serbia	39
Bulgaria	1	Croatia	3
Hungary	1	Bosnia & Herzeg.	2
Romania	2	Montenegro	2
Serbia	4	North Macedonia	2
UK	1		
Austria	1		
Italy	1		
Total:	**16**	**Total:**	**48**

Table 2: Lecturers and attendees by country (school 2019).

distinct conferences and also (co-)organized more than 15 events[2] dedicated to networking around our curriculum. And as a consequence, the project has now involved more than 200 experts from more than 100 organizations in its knowledge transfer initiative.

5.2 Summer Schools

So far, as of January 2021, we conducted two summer schools on site, in Serbia. This gave the experts the opportunity of sharing directly their knowledge with students from the West Balkan states.

5.2.1 Physical 2019 edition. In 2019, we organized a summer school in Belgrade (Serbia) in order to share knowledge with students. It was a 3-day event, scheduled as follows:

1st Research Industry Forum composed of keynotes and presentations from companies;
2nd Lectures of experts from partner institutions;
3rd Lectures of experts from partner institutions.

The Research-Industry Forum was a one-day event organized together with the summer school so that PhD students (participants of the school) have an opportunity to learn about the newest technologies and trends in this and related fields, as well as about use cases from the following companies: OntoText, SAS Institute, CISCO, Banca d'Italia, Meltwater, etc.

The school, for a first edition, was overall well attended, as we received attendance from almost 50 students from 5 West Balkan countries. Moreover, the scientific program was ensured by more than 15 experts coming from 8 different countries. The detailed countries and figures are presented in Table 2.

5.2.2 Virtual 2020 edition. Preparation of lectures is an activity that goes on continuously from the very beginning of the project. The preparation of the 2020 edition of the summer school started at the end of 2019. The date was fixed at the beginning of 2020, while the Program was discussed in February 2020. Originally, it was planned to be a live event like the school organized in 2019. However, due to the COVID-19 pandemic, the decision was made to hold an online 2-day event, while the 3rd day activities (hands-on sessions) were postponed for few months.

Nevertheless, the virtual event attracted actually more participants (*ca.* 70) than the 2019 edition. Moreover, as presented in Table 3, the number of distinct countries increased, from 12 to 17. Such an improvement of participation might be explained by two

---

[2]Networking events: https://project-lambda.org/Past-Events

Represented Countries	Number of	
	Lecturers	Attendees
Serbia	2	33
Italy	2	5
Germany	4	3
UK	2	2
Czech Rep.	1	1
Belgium	1	1
Austria	1	1
Croatia	1	1
Bosnia & Herzeg.	1	1
India	1	
Greece	1	
Bulgaria	1	
Romania		1
North Macedonia	1	
Montenegro	1	
Albania		1
Ireland	1	
Total:	**21**	**50**

Table 3: Lecturers and attendees by country (school 2020).

Faculty, University	Country
Sch. of Electrical Eng., Uni. of Belgrade	Serbia
Fac. of Agriculture, Uni. of Novi Sad	Serbia
Fac. of Sciences, Uni. of Novi Sad	Serbia
Fac. of Electronic Eng., Uni. of Niš	Serbia
Fac. of Mechanical Eng., Uni. of Niš	Serbia
Fac. of Transport and Traffic Eng., Uni. of Belgrade	Serbia
Sch. of Eng. Management, Union Nikola Tesla Uni.	Serbia
Fac. for Information Technologies, Belgrade Metropolitan Uni.	Serbia
Fac. of Organization and Informatics, Uni. of Zagreb	Croatia
Fac. of Computer Science & Eng., St. Cyril and Methodius Uni.	North Macedonia
Fac. of Electrical Eng., Uni. of Montenegro	Montenegro
Fac. of Natural Sciences and Mathematics, Uni. of Tuzla	Bosnia & Herzeg.

Table 4: Adoption of lectures in West Balkan universities.

factors: first; the second edition was built up on the success of the first one, and second; the fact that the event went online allows an easier way of attendance as people do not have to travel. And the laster possibility explains the fact that, for instance some people were joining from very far *e.g.* from India.

5.3 Staff Visits and Exchanges

Lastly, we set up staff exchange activities. Indeed, we consider that establishing deeper connections with key people allows to create a stronger momentum within the knowledge transfer process to the whole region as these people will then promote the initiative and advertise it for the next round of lectures. In practice, these actions were targeting two types of people.

First, in order to build strong connections, we aim to strengthen the links between research group leaders. To do so, the consortium organized working groups where participants were able to brainstorm to build bridges between their own domain of expertise and Big Data. Such groups led to extensive discussions with possible future industrial partners and therefore pave the road to partnerships between Western Europe countries and West Balkan countries.

Second, as the project aims at unlocking Big Data research activities, we also set up student exchanges from West Balkans to Western Europe. These stays of several days inside partner laboratories gave the opportunity to students to directly follow Big Data experts in their own premises. They were then able to present their current research topics and to ask how Big Data technologies could be used to help solving some of their technical challenges.

6 FEEDBACK & FUTURE ADOPTION

During the two summer schools, we collected feedback from the audience. A feedback questionnaire was distributed to participants of the school and in-person interviews were conducted with selected participants. Twenty-three participants answered the questionnaire,

e.g., "I would like to see the lectures presented at the School integrated in one of the master courses". The feedback will be used to better plan our next summer school. Overall, the collected feedback was positive and emphasised the quality and diversity of the learning material; in parallel, participants also suggested that more interactions could take place between the teachers and participants.

Moreover, we already received multiple interest expressions to adopt our lecture set. Additionally, as presented in Table 4, twelve schools and faculties (from the region) have already used and integrated the teaching material we shared into their own curricula. We therefore acknowledge that our initiative spreads across the region and various experts who build the curriculum are now involved in exchange programs, fostering Big Data adoption.

7 CONCLUSION

In this article, we presented the program we started to deploy three years ago in Europe in order to develop the teaching and research activities in the domain of Big Data in the West Balkan region. To do so, we set up a consortium of domain experts who collaborated together to build a comprehensive set of lectures (exploiting multiple media) on the considered domain while taking care of advertising Big Data through various technical case studies and published them both online[3] and in an open-access book [10] already downloaded more than 40 000 times as of April 2021. In parallel, we organized several events to promote our initiatives, prepared staff exchanges and finally initiated a summer school series to directly teach in Serbia.

Our goal is now to continue our efforts of knowledge transfer with the West Balkan region. To do so, we are planning to extend the set of lectures, providing for instance, more hands-on sessions and to enrich the medium-diversity of each lecture (*i.e.* text-material, slide decks and videos). Moreover, we are currently starting the process to organize the third edition of our summer school. Finally, because our initiative has been well-received in the West Balkan region, we would be happy to start also applying our strategy in another region: since all the lectures are openly accessible, the current curricula could be easily extended by using parts of our lecture set.

---

[3]https://project-lambda.org/Learning

REFERENCES

[1] Luigi Bellomarini, Emanuel Sallinger, and Georg Gottlob. 2018. The Vadalog System: Datalog-based Reasoning for Knowledge Graphs. *Proc. VLDB Endow.* 11, 9 (2018), 975–987.

[2] Klaus Bothe and Zoran Putnik. 2015. Master crash course on "software engineering" in a multi-cultural environment. *Computer Applications in Engineering Education* 23, 6 (2015), 854–864.

[3] Zoran Budimac, Zoran Putnik, Mirjana Ivanović, and Klaus Bothe. 2013. Transnational cooperation in higher education in Balkan countries. *Novi Sad Journal of Mathematics* 43, 1 (2013), 167–177.

[4] Philip Sheridan Buffum, Allison G Martinez-Arocho, Megan Hardy Frankosky, Fernando J Rodriguez, Eric N Wiebe, and Kristy Elizabeth Boyer. 2014. CS principles goes to middle school: learning how to teach "Big Data". In *Proceedings of the 45th ACM technical symposium on Computer science education*. 151–156.

[5] European Commission. 2017. Building a European data economy. https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy.

[6] Jeffrey Dean and Sanjay Ghemawat. 2008. MapReduce: simplified data processing on large clusters. *Commun. ACM* 51, 1 (2008), 107–113.

[7] Florin Gheorghe Filip and Enrique Herrera-Viedma. 2014. Big Data in the European Union. *The Bridge Vol. 44, No. 4* (2014). https://www.nae.edu/Publications/Bridge/128772/129172.aspx.

[8] Damien Graux, Valentina Janev, Hajira Jabeen, and Emanuel Sallinger. 2021. A Big Data Learning Platform for the West Balkans and Beyond. In *26th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE)*.

[9] Mirjana Ivanović, Zoran Putnik, Zoran Budimac, Klaus Bothe, and Katerina Zdravkova. 2013. Gender influences on studying computer science: non-EU Balkan case. In *Proceedings of the 6th Balkan Conference in Informatics*. 171–178.

[10] Valentina Janev, Damien Graux, Hajira Jabeen, and Emanuel Sallinger. 2020. *Knowledge graphs and big data processing*. Springer Nature. https://link.springer.com/book/10.1007/978-3-030-53199-7.

[11] Jens Lehmann, Gezim Sejdiu, Lorenz Bühmann, Patrick Westphal, Claus Stadler, Ivan Ermilov, Simon Bin, Nilesh Chakraborty, Muhammad Saleem, Axel-Cyrille Ngomo Ngonga, and Hajira Jabeen. 2017. Distributed Semantic Analytics using the SANSA Stack. In *Proceedings of 16th International Semantic Web Conference - Resources Track (ISWC'2017)*. http://svn.aksw.org/papers/2017/ISWC_SANSA_SoftwareFramework/public.pdf

[12] Mark C Lewis. 2019. Big Data Analytics with Spark. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 1242–1242.

[13] Frank Manola, Eric Miller, Brian McBride, et al. 2004. RDF primer. *W3C recommendation* 10, 1-107 (2004), 6.

[14] Anurag Nagar. 2016. Enhancing Teaching of Big Data by Using Real World Datasets. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. 496–496.

[15] Anurag Nagar. 2017. Developing big data curriculum with open source infrastructure. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. 700–701.

[16] Andy Patrizio. 2018. IDC: Expect 175 zettabytes of data worldwide by 2025. *Network World* (2018).

[17] Zoran Putnik, Mirjana Ivanović, Zoran Budimac, Klaus Bothe, and Hans-Dieter Burkhard. 2014. Teaching in balkan countries: multilingual and multicultural approach. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*. 393–400.

[18] Nevena Ranković, Elinda Kajo Mece, Mirjana Ivanović, Asya Stoyanova-Doycheva, Miloš Savić, and Dragica Ranković. 2019. Female Students' Attitude Towards Studying Informatics and Expectations for Future Career-Balkan Case. In *Proceedings of the 9th Balkan Conference on Informatics*. 1–7.

[19] M Sarnovsky and J Paralic. 2015. Teaching big data analysis at Technical University in Kosice in business information systems study program. In *2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 1–6.

[20] Joachim L Schultze. 2015. Teaching 'big data' analysis to young immunologists. *Nature immunology* 16, 9 (2015), 902–905.

[21] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauly, Michael J Franklin, Scott Shenker, and Ion Stoica. 2012. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In *Presented as part of the 9th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12)*. 15–28.

[22] Amelie Chi Zhou and Bingsheng He. 2019. Big Data and Exascale Computing. *Sherif Sakr, Albert Y. Zomaya (eds) Encyclopedia of Big Data Technologies, Springer, Cham.* (2019).

[23] Paul Zikopoulos, Chris Eaton, et al. 2011. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.