A Big Data Learning Platform for the West Balkans and Beyond

Damien Graux
ADAPT Centre, Trinity College Dublin, Ireland
damien.graux@adaptcentre.ie

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

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

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

ABSTRACT

As the number of Big Data related methods, tools, frameworks and solutions is growing, there is a need to classify and make available the knowledge related to this domain. This is especially useful for countries which, so far, have suffered from some lack of infrastructure in Big Data. In this article, we describe the deployment of an open online platform gathering lectures and resources, from multiple partnered European institutions, tailored for West Balkan students through focuses on local Big Data challenges.

CCS CONCEPTS

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

KEYWORDS

Teaching Big Data Analytics, West Balkans, European collaboration

1 INTRODUCTION

Big Data refers to data sets which have large sizes and complex structures. The data size can range from dozens of terabytes to a few zettabytes and is still growing [7]. 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. In an attempt to support the European data economy policy [1], our consortium proposed a training approach [2] and established the infrastructure for collaborative work of teachers/trainers with PhD students and other interested parties such as industries.

This article describes the infrastructure that we deployed to reinforce organizational learning and capacity building in the West Balkan region (see aside for a precise location of the targeted region) and to facilitate teachers-trainees cooperation in the larger network of experts[1] in the field of Knowledge Graphs, Semantic Big Data Architectures, and Smart Data Analytics. Moreover, our consortium spreads over Europe: Germany, Serbia, the UK; and gathers experts from 15+ different countries [2].
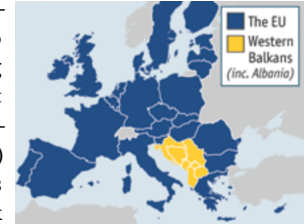
2 AN OPEN EDUCATION PLATFORM

Our overall objectives are 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 such as: the organization of international events (training, workshops, webinars, conferences), the development of a specific learning kit about Big Data Analytics [2], or the publication of an open-access book[2] [3].

In particular, to increase the audience and the availability of our initiative, we developed an online Knowledge repository to help students acquiring new skills:

<https://project-lambda.org/Learning>

The learning materials that were produced are free, stored in a public repository and available online via the OpenCourseWare platform SlideWiki [4]. This project repository aims at facilitating the exchange of learning materials, tools, project results and best practice between the international leading organizations and research institutions and industry from the West Balkan countries.

2.1 Technical Considerations

Practically, the Learning and Consulting Platform was established using Drupal[3] as a content management system. Some of the advantages of Drupal are related to the availability of different features, its easy customization and its security, *i.e.*, Drupal is less resource-intensive and its pages typically load quicker and have

---

[1] https://www.linkedin.com/groups/12129621/
[2] Downloaded more than 40 000 times as of January 2021.
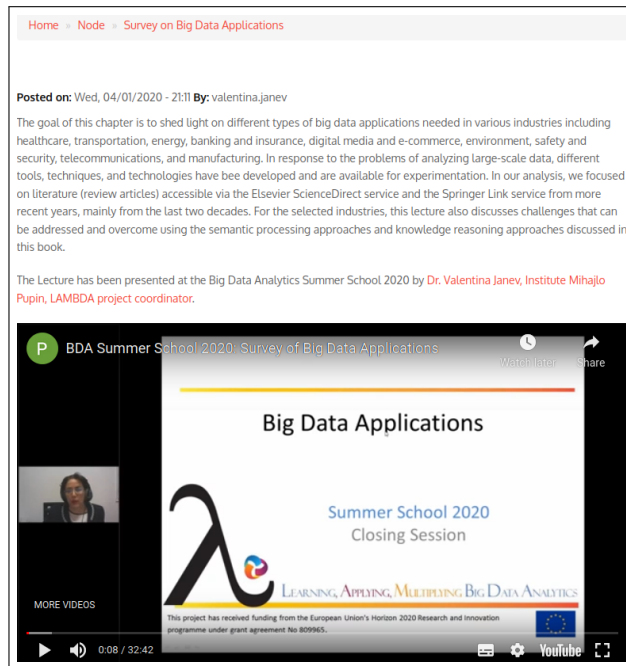[3] https://www.drupal.org/

Figure 1: Screenshot of a lecture on "Big Data Applications".

faster response times than WordPress and Joomla [5]. In order to support the needs, many different types of contents were defined, where each content type is described with a set of attributes. For instance, lecture is described with category, contributor, format (video, slides, paper, book chapter).

Hierarchically, a Drupal-based platform offers three groups of users: anonymous users; authenticated users ; administrators. Currently, two different roles have been defined for authenticated users: *partner* having full access to the private pages of the portal and *associated partner* having full access to the Stakeholder database (restricted) and contents in the Knowledge Repository.

2.2 A Comprehensive Set of Lectures

In order to structure our Big Data learning programme, we split it into eight modules which are tackling various angles of the knowledge we want the students to acquire. We release our lectures in different media types: slide decks, videos, scientific articles, chapters, hands-on sessions. In total, the platform offers more than 40 lectures[4]. Practically, these modules range 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. The interface (see Figure 1) provides at once various information about the lecture: from an abstract to a video recording of the teacher aligned with the relevant slide.

2.3 A Big Data tool classification

In addition to the large set of lectures, the platform also presents a *tool repository*[5]. By analysing the most popular frameworks used to

handle Big Data, we proposed a characterization of the landscapes in the topics related to Big Data. Indeed, we propose to categorise the Big Data landscape as follows: Cloud Marketplaces; Hadoop as a Web Service/Platform; Operational Database Management Systems; NoSQL/Graph databases; Stream Processing Engines; Analytics Software/System/Platform; Data Analytics Languages; Optimization Library for Big Data; Library/API for Big Data; ML Library/API for Big Data; Visualization Software/System; Distributed Messaging System. Overall, more than 80 tools have been identified and classified so that the students can be guided by their use-case when they need to select a tool for a specific application.

2.4 An OpenCourseWare Integration

The platform provides access to the SlideWiki OpenCourseWare system[6]. In a nutshell, SlideWiki is an open, web-based OpenCourse-Ware authoring platform that aims to provide an open and accessible platform to create and share qualitative, rich and engaging educational content following the 5R principles of OERs [6]. The platform allows educators to create, edit, translate and reuse HTML slide presentations complemented with comments, links to sources and supporting materials as well as questions to help learners.

The SlideWiki tool has the ability to import and export data from/into different data formats, thus SlideWiki users that use the SlideWiki presentation mode can merge the content we are sharing directly into their presentations. Moreover, SlideWiki allows to create self-assessment exam sheets; therefore, teachers are able to set up exams where users can test their knowledge about a lecture.

3 CONCLUSIONS

In order to foster knowledge about Big Data analytics in the West Balkan area, we set up various actions, from staff-exchange to summer schools. Among these actions, we also deployed an online and open platform gathering all the material presented to students plus additional analyses. So far, our platform counts more than 150 registered users and shares 40+ lectures in different formats. Moreover, the initiative of opening such a platform allows users to learn at anytime virtually (which proved to be very useful during the COVID-19 outbreak) and from anywhere. The latter thereby broadens the possible audience of our lectures which can now be taken outside the West Balkan region.

REFERENCES

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

[2] Damien Graux, Valentina Janev, Hajira Jabeen, and Emanuel Sallinger. 2021. Deploying a strategy to unlock Big Data research activities in the West Balkan region. In *ACM Conference on Innovation and Technology in Computer Science Education.*

[3] 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.

[4] Ali Khalili, Sören Auer, Darya Tarasowa, and Ivan Ermilov. 2012. SlideWiki: elicitation and sharing of corporate knowledge using presentations. In *Int. Conference on Knowledge Engineering and Knowledge Management.* Springer, 302–316.

[5] Savan K Patel, VR Rathod, and Jigna B Prajapati. 2011. Performance analysis of content management systems-joomla, drupal and wordpress. *International Journal of Computer Applications* 21, 4 (2011), 39–43.

[6] David Wiley et al. 2014. David Wiley,"The Access Compromise and the 5th R". *An Open Education Reader* (2014). https://opencontent.org/blog/archives/3221.

[7] Amelie Chi Zhou and Bingsheng He. 2019. Big Data and Exascale Computing.. In *S. Sakr, A. Y. Zomaya (eds) Encyclopedia of Big Data Technologies, Springer, Cham.*

---

[4]https://project-lambda.org/Knowledge-repository/Lectures
[5]https://project-lambda.org/tools-for-experimentation

---

[6]https://slidewiki.org/