Instructional Model for Building effective Big Data Curricula for Online and Campus Education

[1]Yuri Demchenko, [2]Emanuel Gruengard, [1,3]Sander Klous,

[1]University of Amsterdam
System and Network Engineering Group
Amsterdam, The Netherlands
e-mail: y.demchenko@uva.nl

[2]Laureate Online Education
e-mail: emanuel.gruengard@online.liverpool.ac.uk
[3]KPMG
Amsterdam, Netherlands
e-mail: Klous.Sander@kpmg.nl

Abstract— **This paper presents current results and ongoing work to develop effective educational courses on the Big Data (BD) and Data Intensive Technologies (DIT) that is been done at the University of Amsterdam in cooperation with KPMG and by the Laureate Online Education (online partner of the University of Liverpool). The paper introduces the main Big Data concepts: multicomponent Big Data definition and Big Data Architecture Framework that provide the basis for defining the course structure and Common Body of Knowledge for Data Science and Big Data technology domains. The paper presents details on approach, learning model, and course content for two courses at the Laureate Online Education/University of Liverpool and at the University of Amsterdam. The paper provides also background information about existing initiatives and activities related to information exchange and coordination on developing educational materials and programs on Big Data, Data Science, and Research Data Management.**

Keywords- Education and Training on Big Data Technologies; Online Education; Instructional methodology; Common Body of Knowledge; Bloom's Taxonomy; Andragogy; Big Data Architecture Framework.

I. INTRODUCTION

Big Data, also referred to as Data Intensive Technologies, are becoming a new technology trend in science, industry and business [1, 2, 3]. Big Data are becoming related to almost all aspects of human activity from just recording events to research, design, production and digital services or products delivery to the final consumer. Current technologies such as Cloud Computing and ubiquitous network connectivity provide a platform for automation of all processes in data collection, storing, processing and visualization.

Emergence of Big Data technologies indicates the beginning of a new form of the continuous technology advancement that is characterized by overlapping technology waves related to different components of the modern digital economy from production and consumption to collaboration and general social activity. This creates demand for continuous and dynamically adopted education for preparing new specialists and training for new skills.

There is a growing demand for new type of specialists with strong technical background and deep knowledge of the Big Data technologies what actually can be identified a new profession of the Data Scientist [4]. However, there is no widely available the Data Science professional education

and general Big Data training programs as well as there is no common approach to how to effectively build professional level Big Data curricula.

The presented paper and the current development of the Big Data courses follows successful example of developing advanced instructional model and educational framework for Cloud Computing presented in the earlier authors' paper [5].

The remainder of the paper is organized as follows. Section II provides information about needs for new Big Data related professions and challenges in Big Data education. Sections III provide basic information about the Big Data definition and Big Data Architecture Framework. Section IV introduces the Common Body of Knowledge for Big Data (CBK-BD) and refers to required competencies. Section V discusses two examples of the education programs and courses development in which the authors are involved, and the paper concludes with the summary and future developments in section VI.

II. DEMAND FOR NEW PROFESSION AND PECEIVED CHALLENGES

Successful adoption of Big Data technologies is a result of mutual enrichment between science and industry. With a long tradition of working with constantly increasing volume of data, modern e-Science can offer industry the scientific analysis methods, while industry can bring advanced and fast developing Big Data technologies and tools to science and wider public.

Besides extensive use in modern research and science, the Big Data technologies demonstrated their power and competitive advantage in many recent campaigns (such as recent US elections), personalised service offerings and effective risk management where are characterised by extensive use of statistical methods, machine learning and predictions. To benefit from such technologies, a new type of specialists is needed that could combine statistical methods, computational platform, and scientific approach.

A. Horizon2020: Education and skills development for Research e-Infrastructure

The new European H2020 Framework research and technology development program [6] includes measures to strengthen the human capital of research infrastructures and support skills and expertise development specifically needed to construct, operate and use research infrastructures that are becoming increasingly complex and expensive. The

special priority item "INFRASUPP-4-2015: New professions and skills for e-infrastructures" includes the following activities:

(1) Defining or updating university curricula for the e-infrastructure competences mentioned above, and promoting their adoption.
(2) Developing and executing training programmes (including for lifelong learning)
(3) Create a reference model which defines new profession competencies
(4) Support networking and information sharing among practicing e-infrastructure experts and data scientists.

B. Challenges in Big Data Education

The discussed curriculum seeks to provide knowledge in three inter-related technology domain comprising modern Big Data and Data Intensive Technologies: Data Structures and Data Management, Big Data Infrastructure and computing platforms, Data Analytics and Machine Learning. Our objective is to empower future professionals with the ability to develop new knowledge and build stronger expertise in practical technology application.

An effective professional education needs to provide for the professional level of knowledge to achieve the following:
1) Master basic concepts and major application areas
2) Compare similar concepts (and concepts inter-relation) and alternatives, as well as application specific areas
3) Appraise basic technologies and their relation to the basic concepts.

In contrary to more focused technology domain such as Cloud Computing, Big Data fuses computer technology, data analytics and research methods – the three domains that earlier have been considered as rather independent. How their fusion will evolve is still not clear but the leading Cloud Service Providers demonstrate a trend to add Big Data Analytics services to their cloud offerings building on their ability to integrate around available computing and infrastructure facilities and easy support for distributed cooperative groups.

At the same time, many not technical domains, and first of all non-computer sciences, are talking about Big Data literacy that should include knowledge of the data analytics methods and tools, research methods, and existing platform, primarily cloud based to effectively use data intensive technologies in research.

In current continuous technology development and evolution, the usual educational approaches do not serve well in providing advanced curriculum to professionals on new emerging technologies. In the following sections we will explain the new instructional methods and educational approaches we found effective for such new emerging technology domains as Cloud Computing and Big Data.

III. BIG DATA TECHNOLOGY DOMAIN DEFINITION

A. 6 Vs of Big Data

We refer to our recent papers [7, 8] where we summarized the existing at that time discussions and proposed the Big Data definition as having the following 6V properties: Volume, Velocity, Variety that constitute native/original Big Data properties, and Value, Veracity, Variability as acquired features which are considered in the context of a target data use/application and depending on a specific process or model.

The Big Data 6Vs properties provide a basis for the Big Data definition but not sufficient for effective Big Data definition as a technology domain that could provide a roadmap for the Big Data technology taxonomy and roadmap definition.

B. Structural Big Data Definition

In our research and in our contribution to the NIST Big Data Working Group (NBD-WG) [9] we proposed a structural Big Data definition that consists of five parts as listed below [8]:
(1) Big Data Properties: 6V
- Volume, Variety, Velocity, Value, Veracity, Variability
(2) New Data Models
- Data types, data structure, data linking, provenance
- Data Lifecycle and Variability/Evolution
(3) New Analytics
- Real-time/streaming analytics, new statistical methods, interactive and machine learning analytics
(4) New Infrastructure and Tools
- Cloud based Big Data infrastructure, storage, network, high performance computing
- Heterogeneous multi-provider services integration
- New Data Centric (multi-stakeholder) service models
- New Data Centric security models for trusted infrastructure and data processing and storage
(5) Source and Target that are important aspect defining data types and data structures, e.g. raw data, data streams, correlated data, and required processing methods
- High velocity/speed data capture from variety of sensors and data sources
- Data delivery to different visualisation and actionable systems and consumers
- Full digitised input and output, (ubiquitous) sensor networks, full digital control

To reflect the major Big Data features and Big Data Ecosystem components and based on the widely accepted Gartner definition [10], we propose an improved Big Data definition [8]:

"Big Data (Data Intensive) Technologies are targeting to process high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure high-veracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices."

C. Big Data Architecture Framework

We refer to the definition of the Big Data Architecture Framework (BDAF) proposed by the authors in [8] that summarises the majority of currently ongoing research activities and standardisation processes at NIST, ISO/IEC, as well as by industry leaders.

The proposed BDAF comprises of the following 5 components that address different aspects of the Big Data

Ecosystem and Big Data definition aspects which we consider to some extent orthogonal and complementary:

(1) Data Models, Structures, Types
- Data formats, non/relational, file systems, etc.

(2) Big Data Management
- Big Data Lifecycle Management
- Big Data transformation/staging
- Provenance, Curation, Archiving

(3) Big Data Analytics and Tools
- Big Data Applications and Analytics Methods
- Target use, presentation, visualisation

(4) Big Data Infrastructure (BDI)
- General Compute and Storage infrastructure (cloud based), High Performance Computing (HPC), Network
- Sensor network, target/actionable devices
- Big Data services delivery and Operational support

(5) Big Data Security and Privacy
- Data security and privacy in-rest, in-move, trusted processing environments

The proposed BDAF definition is rather technical and infrastructure focused; it actually reflects the major technology oriented stakeholders in Big Data. However the proposed BDAF together with the Big Data definition allow structured definition of the main knowledge and technical domains for Big Data curriculum.

The required new approach to data management and processing in Big Data industry is reflected in the Big Data Lifecycle Management (BDLM) model (see Figure 1) proposed as a result of analysis of the existing practices in different scientific communities and industry technology domains.
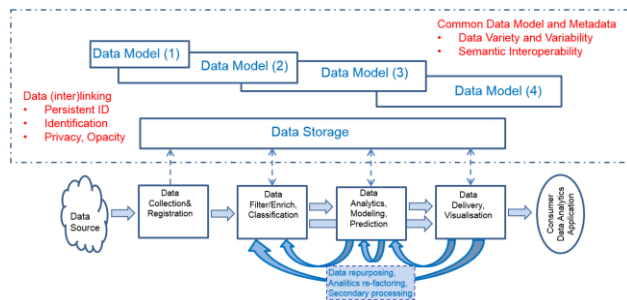

Figure 1. Big Data Lifecycle in Big Data Ecosystem.

New BDLM requires data storage and preservation at all stages what should allow data re-use/re-purposing and secondary research/analytics on the processed data and published results. However, this is possible only if the full data identification, cross-reference and linkage are implemented in BDI. Data integrity, access control and accountability must be supported during the whole data lifecycle. Data provenance is an important component of the discussed BDLM and must also be done in a secure and trustworthy way.

IV. COMMON BODY OF KNOWLEDGE IN BIG DATA (CBK-BD) AND SKILLS TAXONOMY

CBK refers to several domains or operational categories into which Big Data theory, technology and practices breaks down. It is like a blueprint or framework which students can see how things fit together. Based on the proposed above Big Data definition and Big Data Architecture Framework, we identify the following CBK-BD elements:

1. Big Data Definition and Big Data Architecture Framework, Data driven and data centric applications model, Stakeholders and Roles
2. Big Data use cases and application domains taxonomy and requirements, Big Data in industry and science
3. Data structures, SQL and NoSQL databases
4. Data Analytics Methods and Tools, Knowledge Presentation
5. Big Data Management and curation, Big Data Lifecycle, Data Preservation and Sharing, Enterprise Data Warehouses, Agile Data Driven Enterprise
6. Cloud based Big Data infrastructure and computing platforms, Data Analytics application and new Data Scientist skills required
7. Computing models: High Performance Computing (HPC), Massively Parallel Computing (MPP), Grid, Cluster Computing
8. Big Data Security and Privacy, Certification and Compliance

The analytical part of the CBK-BD must be based on the strong mathematical foundation for Data Analytics focused curriculum.

The CBK should be mapped to the skills and competencies model to define the curriculum profile that responds to the needs of the specific community or stakeholder groups. The O'Reilly Strata industry research [4] defines the four Data Scientist profession profiles and their mapping to the basic set of technology domains and competencies as shown in Figure 2.
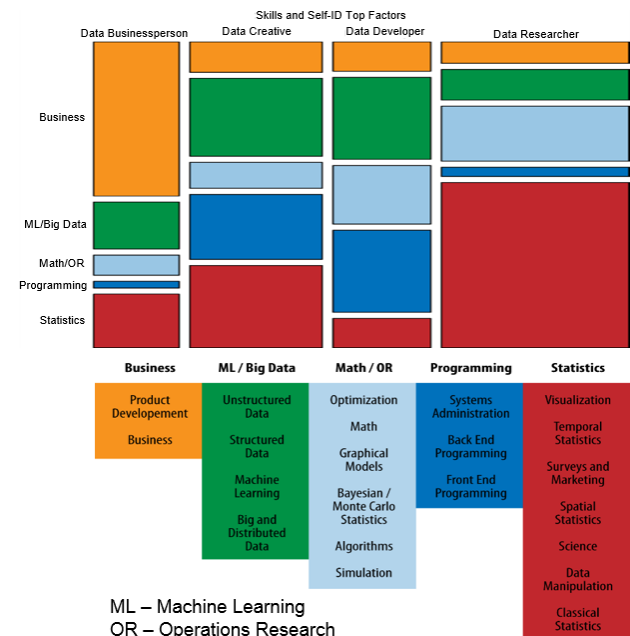

Figure 2. Data Scientist skills and profiles according to O'Reilly Strata survey [8]

There are few known projects and community activities attempting to develop a consistent competency taxonomy for Data Science and data preservation community such as HPC University in US [11, 12], APARSEN (Association Permanent Archives) [13]. The goal to create a taxonomy of the required skills for research community is a part of the activity by the RDA Interest Group on Education and Training on handling of research data [14].

The HPC University competencies and skills taxonomy [11] provides comprehensive list of the Computational and Data Science education competencies for the undergraduate

and graduate level, and for basic and advanced Data Driven Science competencies. We can take those as a basis for the science and research domain.

The common component of the possible curriculum profiles the Big Data Fundamentals should include such topics as Big Data definition and architectures, Big Data Lifecycle management; Big Data analytics methods and platforms. They should provide sufficient basis for performing the main professional tasks and further independent professional development. Different groups of the Data Science professionals and specialists can extend their education and training with the specific knowledge and skills according to the CBK-BD.

V. IMPLEMENTATION EXAMPLES AND USED APPROACHES

This section will discuss two examples of implementing the proposed approach in the Big Data courses developed by the authors for different target groups of students and stakeholders.

A. Big Data course at Laureate Online Education

Laureate Online Higher Education (LOHE) [15, 16], the online education partner of the University of Liverpool, provides fully online teaching/education environment based on customized Blackboard platform.

Laureate's courses are designed to push the boundaries of access to higher education from different countries, cultural backgrounds, and for students with varying educational background. The common method here is to push students beyond the boundaries of their customary thinking (i.e., to push them to think "outside of the box") and stimulate their self-motivated learning.

1) Collaborative Online Learning Model Princiles

The following are the main principles that are applied to achieve effective learning in online programs:

a) Programs and courses are developed with input from nationally- and internationally-recognized Subject-Matter Experts (SME), leading practitioners, associations/professional groups, and international representatives.

- Educational materials combine strong conceptual foundation, technology basis and applied mechanisms, standardization, best practices and industry implementation.
- Programs and courses fully leverage technological and media resources to optimize collaboration and communication.

b) Programs and courses are designed to create an inspiring and transforming student experience and promote collaborative student experiences

- Programs are future-oriented and forward thinking, both in providing course materials that reflects current status and trends in the technology domain, and in facilitating critical and analytical students' thinking.
- Students are responsible for their learning and they exercise elements of control over their learning environment. They are inspired through opportunities to engage in reflection and critical thinking, to connect theory to practice, their own experience and educational group experience in the weekly classroom discussions.

- Work on individual and group projects and hands on assignment.

d) Laureate's programs and courses are designed to expose students to diverse ideas, opinions, perspectives, and experiences - both brought by instructors and based on knowledge and experience exchange in the classroom.

e) The course undergoes a quality review process that includes a critical reader review and recommendations, and adoption to the common learning model. The quality reviews are continued all along the course "life span".

2) Bloom's Taxonomy and Andragogy in Online Education

The Laureate courses are using such effective learning models as Bloom's taxonomy to facilitate self-learning in online environment and Andragogy to improve effectiveness of learning for adult students.

The courses are developed using best practices for online education and applying Bloom's taxonomy with strong emphasis not only on the Cognitive Domain but also on the Affective Domain [17, 18] to facilitate deep and self-motivated learning. This includes the following:

- Present learning tasks in terms of problem solving, not only as demonstration of accumulated knowledge, and encourage multiple approaches to problem solving.
- Provide opportunities for collaboration with others, including: discussions; sharing of experience, perceptions, and alternate viewpoints; and group activities.
- Allow students to draw on their own experience as part of their learning and to incorporate their own goals into the work of the course.

Andragogy provides effective approach to online higher education. The following principles of andragogy (adult learning) [19, 20] are incorporated:

- Define a rationale for learning and make a case for the value of doing the work.
- Create environments where self-directed skills are nurtured.
- Have different experiences, background, learning styles, motivation, interests, and goals.
- Have a life-centered orientation to learning; motivate to learn the whole course knowledge domain and show relevance to their professional or career needs.
- Instruction should help the students perform tasks, deal with problems, and thrive in real-life situations.
- Rely on the internal motivation factors and provide such motivators as subject mastering satisfaction, knowledge opening their wider vision and general understanding.

The important role belongs to the final dissertation module where the formation of the future specialist is finalized. The students learn the basics of the research methods and apply them to the dissertation development process that includes hypothesis, research questions, scholarly contribution, solution development, and hypothesis verifications.

Comparing to campus based programs, the online programs need to be constructed rather top-down providing the whole complex of knowledge from the general concepts to application and practical knowledge. While campus based courses often use specific and sometimes unique research and scientific environment to motivate the

students' learning through engagement into practical projects supervised by research staff and teachers.

3) Module structure

The following activities and education forms are used in the Big Data module:

- The module consists of 8 weekly seminars that includes 2 Discussion Questions (DQ), Hands-in Assignment (HA, or homework) and project assignment.
- Each seminar is provided with the Lecture Notes and textbook reading assignment. There are no synchronous lectures which makes also possible education delivery to countries and to students with low Internet connectivity, as well as bypassing time zone issues. Recorded lectures and accompanying videos are planned for the future.
- Discussion questions and asynchronous discussion are the main form of educational activity. DQ answers are submitted to the discussion forum and the students are required to contribute to the discussion. The students benefit from the knowledge and experience sharing during discussion and learn how to defend own answer. Instructor plays a role of moderator and the students' knowledge and activities assessor.
- Discussion questions are designed in such a way that to stimulate the students' higher cognitive activities starting from the basic literature search to analyzing and evaluating collected and their application to problem solving.

Criteria for assessing the students' knowledge and learning outcome are based on the Bloom's Taxonomy and require some experience from the instructor.

4) Module Content

This module provides students with a clear picture of the emerging Big Data and Data Intensive Technologies and Data Analytics techniques that power the practical use of Big Data. The Big Data module consolidates recent developments in Cloud Computing, Data Analytics, Web/Internet technologies, mobile and sensor networks and provides a basis for further concise knowledge synthesis by the students. It gives students a chance to experience a variety of Big Data technologies and related analytical techniques.

The weekly seminars have the following topics:

Seminar 1: Introduction. Big Data technology domain definition, Big Data Architecture Framework

Seminar 2: Big Data use cases from science, industry and business

Seminar 3: Big Data Infrastructure components and platforms, Enterprise Data Warehouses, MapReduce and Hadoop, distributed file systems and database architectures, data structures, NoSQL databases.

Seminar 4: Big Data analytic techniques, introduction to RapidMiner. Statistical techniques for modeling data.

Seminar 5: Processes behind Big Data Analytics: Rule Extraction Algorithms and Cluster Analysis, Decision tree induction.

Seminar 6: Classification and forecasting techniques: Machine Learning, Neural Networks and Support Vector Machines/ Measurement techniques: Receiver Operating Curves and Gains Charts.

Seminar 7: Big Data Management and Emerging Agile Data Driven Enterprise, Big Data Service and platform Providers.

Seminar 8: Big Data Security and Privacy, data centric security models. Big Data privacy issues and regulations, Privacy Enhancement Techniques.

The Big Data course is part of the Web Sciences and Big Data program [16] that includes other complementary courses such as Cloud Computing and Programming the Internet.

B. Big Data, Data Science and Cloud Computing courses at the University of Amsterdam

The Faculty of Science at the University of Amsterdam (UvA) is in the process of developing the full set of Computer Science based courses on the Cloud Computing and Big Data. Having a strong basis in both Computer and Network related courses, on one hand, and Informatics, Machine Learning and Data Analytics, on the other hand, UvA gradually migrates the focus of Master program to adopt current trends in using clouds and adopting Data Science and Data Analytics methods in research and industry.

Recently developed and taught the Web Services and Cloud-Based Systems course (April-May 2014) [21] provides a wide spectrum of knowledge on Cloud Computing that creates a strong basis for the Big Data and Data Science related courses which development is under way.

1) Data Science and Data Intensive Technologies Course structure

The Data Science program is being jointly supported by few departments that has traditional focus on the different aspects of the Big Data technologies. Consequently, the Data Science program implemented by different departments has the following specialization: (1) Data Models and Data Management; (2) Machine Learning, Data Analytics; (3) Big Data Infrastructure (cloud based), Computer Platforms, Big Data tools. Departments may also offer more specialized courses based on cooperation with research or industry partners as one described below.

The aim of the UvA Data Science program is to cover both the conceptual and architectural issues in defining the Big Data infrastructure, and Data Analytics applications, as well as more specific information about Big Data ecosystem components, tools, and applications. The courses will provide detailed analysis of the Big Data use cases in science, industry and business what is used to motivated better understanding of the Big Data technologies and required solutions. The course design uses the two major components around which the whole curriculum is built: conceptual Big Data definition and Architecture Framework for Big Data Ecosystem described above.

2) Data Science and Applied Mechanisms Design course

This course is proposed jointly by UvA and KPMG and aims to teach students a combination of Mechanism Design and Big Data techniques to better understand, predict and influence the behavior of large scale information systems. The course will cover the application of mechanism design principles to large scale information systems and the measurability of these systems with the Big Data technologies, allowing for example analysis of phase transitions (such as the global transition from carbon based energy to renewables). After the course, students will be able to understand how to model large scale information

systems with mechanism design techniques. They will be able to determine how key parameters in these systems can be measured and analyzed with Big Data technologies and they will be able to participate in the scientific discussion in this emerging field.

The course will include the following key topics:
- Introduction to Mechanism Design Theory;
- Introduction to Big Data Analytics;
- Modelling large scale information systems;
- Measuring and analyzing large scale information systems;
- Understanding, predicting and influencing phase transitions;
- Applying Mechanism Design and Big Data Analytics to real world systems.

Hands on labs and the project will use available at the Faculty of Science HPC and cloud facilities, as well as using Big Data and Data Analytics facilities from leading cloud providers such as Amazon Web Services, Microsoft Azure and local SURFsara academic supercomputing datacenter.

VI. CONLUSION AND FUTURE DEVELOPMENTS

This paper presents an integrated approach to developing the effective education and training courses on Big Data and Data Science technologies that can be profiled both for full time university students and for online education and IT professionals. It is also important that Big Data or Data Science courses are complemented by the Cloud Computing course to provide the better vision and experience for students.

The ongoing research and development work is primarily focused on improving the definition of the Common Knowledge Base for Big Data and Data Science and required skills taxonomy. The authors are involved into wide cooperation with other universities and research community to exchange experience and coordinated development of new courses.

Based on the first full implementation of the Big Data and Cloud Computing courses in the Web Sciences and Big Data program at the Laureate Online Education [16], the further adoption of the proposed model and approach will done for the University of Amsterdam Data Science program.

When successfully tested, the experience will be shared with the interested research community via Research Data Alliance (RDA) Interest Group on Educations and Skills Development [14, 22]. Ongoing coordination and discussion on the Data Science curriculum development is done in the framework of the RDA EDISON (Education for Data Intensive Science to Open New science frontiers) initiative which has a goal to support establishment of new professions by creating skills taxonomy and knowledge map for a new science and technology domain, share existing experience and disseminate best practices; support experts and educators visiting and exchange

REFERENCES

[1] NIST Big Data Working Group (NBD-WG). [Online]. Available: http://bigdatawg.nist.gov/home.php

[2] Riding the wave: How Europe can gain from the rising tide of scientific data. *Final report of the High Level Expert Group on Scientific Data. October 2010*. [Online]. Available at http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[3] The Fourth Paradigm: Data-Intensive Scientific Discovery. Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Microsoft Corporation, October 2009. ISBN 978-0-9825442-0-4 [Online]. Available: http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[4] Demchenko, Yuri, David Bernstein, Adam Belloum, Ana Oprescu, Tomasz W. Wlodarczyk, Cees de Laat, New Instructional Models for Building Effective Curricula on Cloud Computing Technologies and Engineering. Proc. The 5th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2013), 2-5 December 2013, Bristol, UK.

[5] Harris, Murphy, Vaisman, Analysing the Analysers. O'Reilly Strata Survey, 2013 [online] http://cdn.oreillystatic.com/oreilly/radarreport/0636920029014/Analyzing_the_Analyzers.pdf

[6] HORIZON 2020 Work Programme 2014–2015, 4. European research infrastructures (including e-Infrastructures) [online] http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-infrastructures_en.pdf

[7] Demchenko, Y., P.Membrey, P.Grosso, C. de Laat, Addressing Big Data Issues in Scientific Data Infrastructure. First International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2013). Part of The 2013 International Conference on Collaboration Technologies and Systems (CTS 2013), May 20-24, 2013, San Diego, California, USA.

[8] Demchenko, Yuri, Peter Membrey, Cees de Laat, Defining Architecture Components of the Big Data Ecosystem. Second International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2014). Part of The 2014 International Conference on Collaboration Technologies and Systems (CTS 2014), May 19-23, 2014, Minneapolis, USA.

[9] Output Documents: NBD-PWG V1.0 Pre-release Verions, Volume 1-7 [online] http://bigdatawg.nist.gov/V1_output_docs.php

[10] Big Data defintion, Gartner, Inc. [Online]. Available: http://www.gartner.com/it-glossary/big-data/

[11] HPC University [online] http://hpcuniversity.org/

[12] The HPC University competencies and skills taxonomy [online] http://hpcuniversity.org/educators/competencies/

[13] APARSEN (Association Permanent Archives) [online] http://www.alliancepermanentaccess.org/index.php/aparsen/aparsen-deliverables/

[14] RDA Interest Group on Education and Training on handling of research data [online] https://www.rd-alliance.org/group/education-and-training-research-data.html

[15] Laureate Online Higher Education. [online] http://www.university-liverpool-online.com/online-learning

[16] Laureate Online Education: Online Degree: MSc in Web Sciences and Big Data [online] http://www.university-liverpool-online.com/programmes/informationtechnology/msc-in-web-sciences-and-big-data

[17] Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman

[18] Learning Domains or Bloom's Taxonomy, by Don Clark. [online] http://www.nwlink.com/~donclark/hrd/bloom.html

[19] Knowles, M.S., The Modern Practice of Adult Education: From Pedagogy to Andragogy, March 1988, ISBN-10: 0842822135

[20] Merriam, S.B., Andragogy and Self-Directed Learning: Pillars of Adult Learning Theory [online] http://umsl.edu/~wilmarthp/modla-links-2011/Merriam_pillars%20of%20anrdagogy.pdf

[21] Computer Science: Web Services and Cloud-Based Systems [online] http://gss.uva.nl/future-msc-students/information-sciences/information-sciences/information-sciences/content/folder-3/computer-science.html

[22] Research Data Alliance. [onlie] https://rd-alliance.org/node