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

Big Data course and Learning Model for Online education (LMO)
at the Laureate Online Education (University of Liverpool)

Yuri Demchenko (University of Amsterdam)
EDISON Project

EDISON Workshop @ EGI Community Forum 2015
12 November 2015, Bari, Italy

Outline

- Need for professional education in Big Data
- Big Data definition and Big Data Architecture Framework (BDAF)
- Common Body of Knowledge in Big Data
- Big Data and Data Analytics Course
- Collaborative Online Learning Model Principles at Laureate Online Education (LOE)
- Bloom's Taxonomy and Andragogy
- Summary and next steps

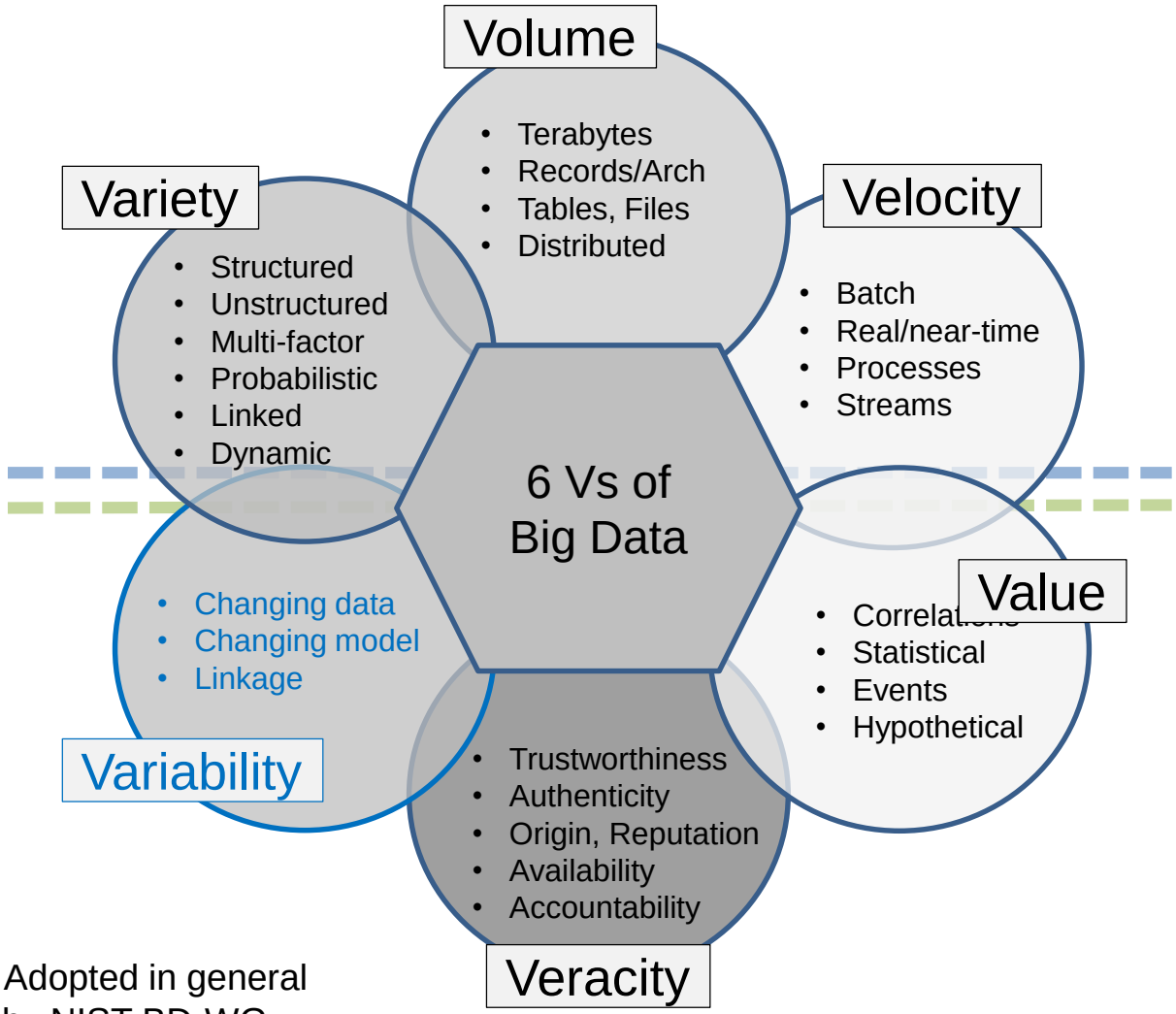
Professional Education Objectives

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

- Challenges due to the fact that Big Data is very wide technology domain
  - E.g., comparing to much narrower Cloud Computing
- New types of competences and knowledge in Big Data
  - Data analytics and Research methods

Improved: 6 V's of Big Data



Adopted in general by NIST BD-WG

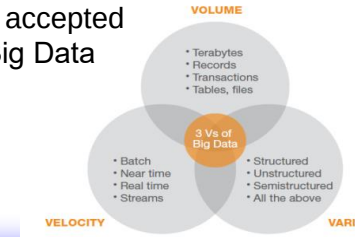
Generic Big Data Properties

- Volume
- Variety
- Velocity

Acquired Properties (after entering system)

- Value
- Veracity
- Variability

Commonly accepted 3V's of Big Data



Big Data Definition: From 6V to 5 Parts (1)

(1) Big Data Properties: 5V
- Volume, Variety, Velocity, Value, Veracity
- Additionally: Data Dynamicity (Variability)

(2) New Data Models
- Data linking, provenance and referral integrity
- Data Lifecycle and Variability/Evolution

(3) New Analytics
- Real-time/streaming analytics, interactive and machine learning analytics

(4) New Infrastructure and Tools
- High performance Computing, Storage, Network
- 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
- 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

Big Data Architecture Framework (BDAF)

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

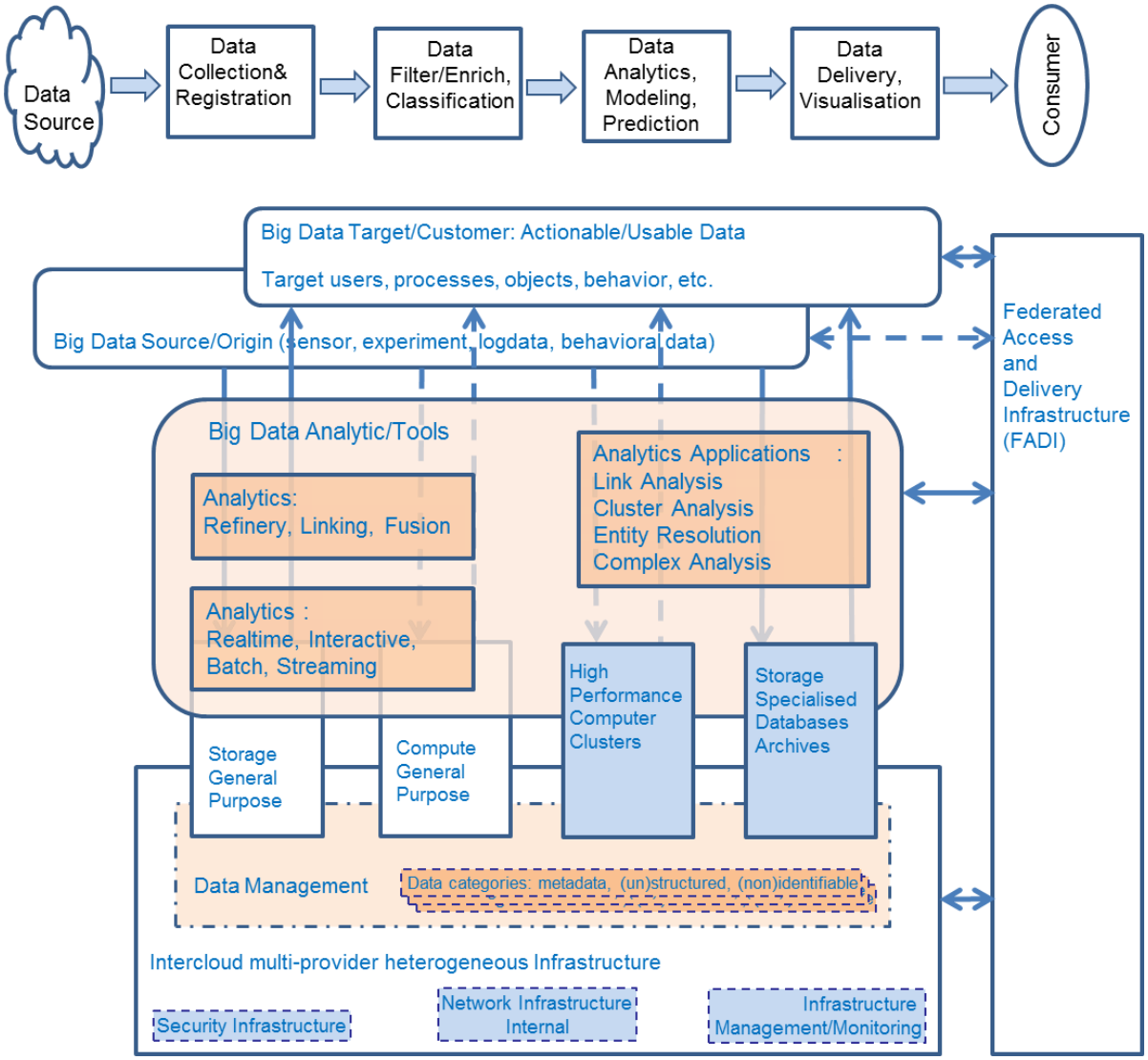
(2) Big Data Management
- Big Data Lifecycle (Management) Model
  - Big Data transformation/staging
- Provenance, Curation, Archiving

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

(4) Big Data Infrastructure (BDI)
- Storage, Compute, (High Performance Computing,) Network
- Sensor network, target/actionable devices
- Big Data Operational support

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

Big Data Infrastructure and Analytics Tools



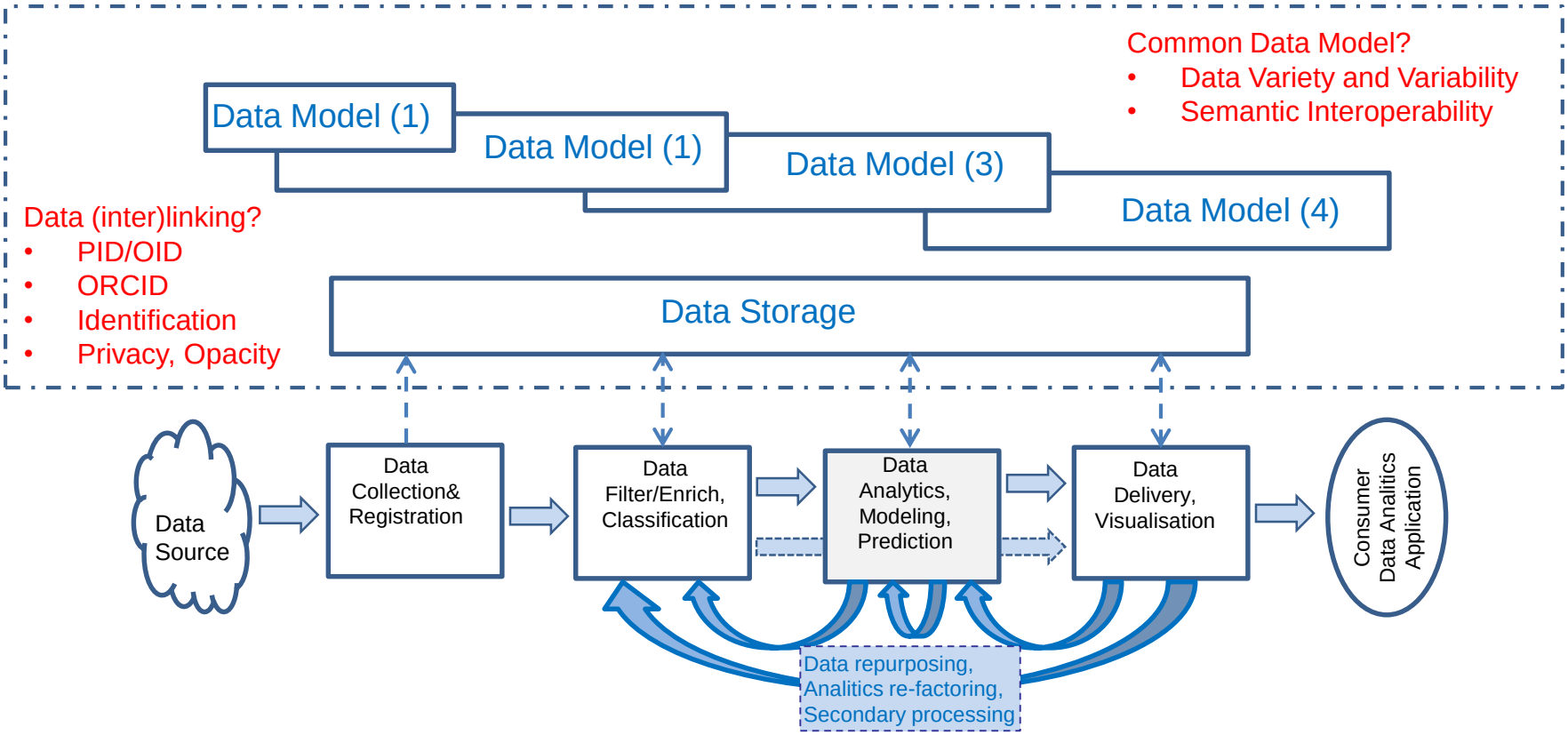
Big Data Infrastructure

- Heterogeneous multi-provider inter-cloud infrastructure
- Data management infrastructure
- Collaborative Environment (user/groups managements)
- Advanced high performance (programmable) network
- Security infrastructure

Big Data Analytics Platforms

- High Performance Computer Clusters (HPCC)
- Analytics/processing: Real-time, Interactive, Batch, Streaming
- Big Data Analytics tools and applications

Data Lifecycle/Transformation Model

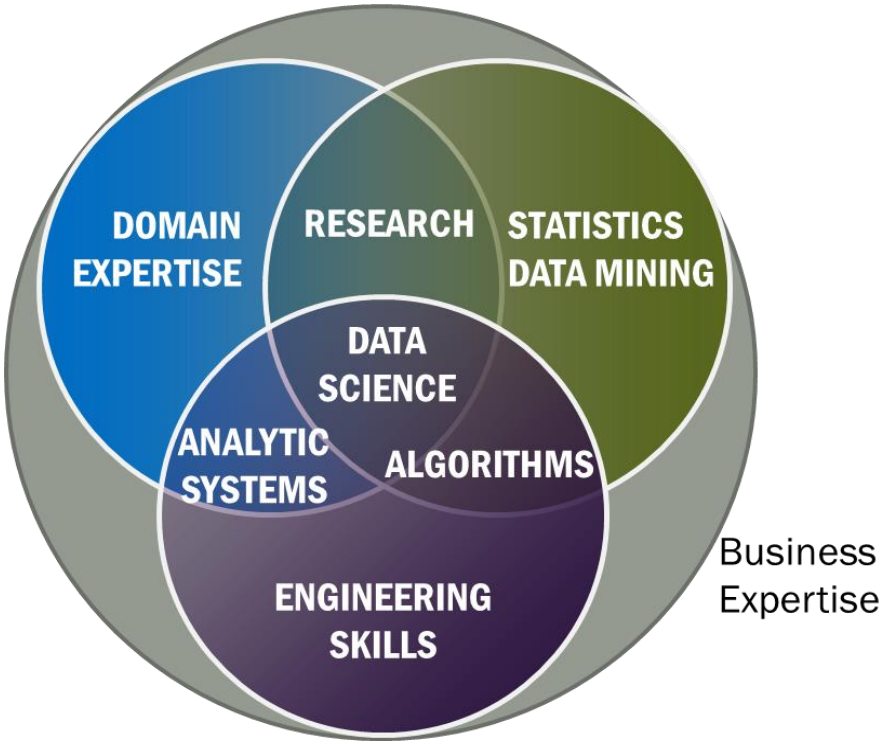


- Does Data Model changes along lifecycle or data evolution?
- Identifying and linking data
  - Persistent identifiers
  - Data ownership
  - Traceability vs Opacity
  - Referral integrity

Big Data and Data Science Competencies and Skills Taxonomy

- Analysing the Analysers. O'Reilly Strata Survey – Harris, Murphy & Vaisman, 2013
- NIST Big Data Working Group Data Science definition (2014)
- Data Science Competencies Taxonomy by HPC University (2014) http://hpcuniversity.org/educators/competencies/
  - Undergraduate Level Computational Science Competencies
  - Graduate Level Computational Science Competencies
  - Basic Data Driven Science Competencies
  - Advanced Data Driven Science Competencies
- A common European framework for ICT Professionals in all industry sectors (2012) http://www.ecompetences.eu/
  - Currently no Data Science but extensive industry oriented data management taxonomy
- The task of RDA IG on Education and Training to define the Data Science Competencies and Skills Taxonomy

NIST BD-WG Data Science Definition (2014-2015)



From "The Fourth Paradigm"

- ***Data Science*** *is extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and hypothesis testing.*

NIST Big Data WG definition

- *A* ***Data Scientist*** *is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming and systems engineering expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle.*

Common Body of Knowledge (CBK) in Big Data and Data Intensive Technologies

CBK refers to several domains or operational categories into which Big Data theory and practices breaks down

- The scope is very wide, need to combine few previously not connected domains
- This is one of attempts verified by practical course development

CBK Big Data and Data Intensive Technologies

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, data models, semantic web, SQL and NoSQL databases, Knowledge presentation
4. **Data Analytics Methods and Tools, Knowledge Presentation, Data Mining**
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

Big Data Course at LOE/UoL Structure

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, Enterprise Data Warehouses (EDW) and emerging *Agile Data Driven Enterprise (ADDE)*, 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.

Big Data Module Structure

- 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 (LN) and textbook reading assignment.
  - There are no synchronous lectures which makes possible the delivery of education to countries and to students with low Internet bandwidth, 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
  - Practical use of Bloom's Taxonomy
- Group project and hands on assignments

- Laureate International Universities owns 84 universities and colleges around the world with a population of 950,000 students.
- In addition Laureate partners with independent universities.
- Laureate Online Education (LOE) is the online education partner of the University of Liverpool.
- The University of Liverpool (UoL) is the original red brick university (1891) and a member of the Russell Group.

The Programmes at LOE/UoL

- LOE & UoL provide fully online teaching/education environment.
- Masters' Degrees in Information Technology, Management, Health, Law and Psychology
- IT degrees accredited by BCS
- Doctorates: DBA, EdD
- Laureate's programs are designed to push the boundaries of access to higher education from different countries (>175), cultural backgrounds and for students with varying educational background.
- Students are pushed beyond the boundaries of their customary thinking (i.e. pushed to think "outside of the box") and stimulate their self-motivated learning.
- Directed at Andragogy ("man-leading") rather than at Pedagogy ("child-leading")
- The key concepts are: constructivism, cooperative and peer learning, critical analysis, critical reflection and critical evaluation.

Andragogy in Adult Education

- Andragogy provides effective approach to online higher education. The following principles of andragogy 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 program knowledge domain and show relevance to professional or career needs.
- Rely on the internal motivation factors and provide such motivators as subject mastering satisfaction, knowledge opening their wider vision and general understanding.

Bloom's Taxonomy in Online Education

- 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 to facilitate deep and self-motivated learning.

This includes the following:

- Present learning tasks in terms of problem solving, not just 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.

Collaborative Online Learning Model Principles

a) Programs and courses are developed with input from nationally- and internationally-recognized Subject-Matter Experts (SME), leading practitioners, associations/professional groups.
   - Educational materials combine strong conceptual foundation, technology basis and applied mechanisms, standardization, best practices and industry implementation..

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

c) The course undergoes a quality review process that includes a Critical Reader (CR) review and recommendations and adoption to the common learning model. A final review is done by UoL.
   The quality reviews are continued all along the course “life span”.

Difference from campus education:

- Top down vs bottom up approach in course development

The Weekly Cycle

- Week starts on Thursday
- Material available on end day Wednesday:
  - Instructions, text lectures, videos, reading assignments, discussion questions (DQs), hand-in assignments (HIAs), projects
- Classes of around 18± students
- Presence in class: at least 3 days of the week
- Initial posting of answers to the DQs by Sunday
- 3-4 Follow–on postings by week end (Wednesday)
- Submission of HIA/Project by Wednesday
- A large number of assessment points (≥ 18)

The MSc degrees

- Fully online operation using a customized Blackboard LMS platform.
  - Asynchronous lectures which makes possible the delivery of education to countries and to students with limited Internet bandwidth, as well as bypassing time zone issues.
  - Recorded lectures and accompanying videos are planned for the future.
- Faculty members serve as moderators rather than lecturers
- 180 Credit Points programs.
- Eight courses of eight weeks
  - 15 Credit Points per course
  - Equivalent (according to the UK scheme) to 150 hours = 18.75 hours per week
- Final Dissertation
  - 60 Credit Points

Aims of Final Dissertation

- An important role belongs to the final dissertation module
  - Duration 9 months
  - Supervised by Dissertation Advisor (DA)

- Aims:
  - To give students the opportunity to work in a guided but independent fashion to explore a substantial problem in depth, making practical use of principles, techniques and methodologies acquired during the course of their program of study to date.
  - To give students experience of carrying out a large piece of individual work and producing a dissertation.
  - To enhance student communication skills.

- The students start by learning the basics of Research Methods and apply them to the dissertation development process following the **Dissertation 4-Step Model:**

- Hypothesis: Frame a hypothesis or hypotheses
- Research Methods: Use research methods to identify a mechanism to test the hypothesis
- IT Artefact: Implement the mechanism in the form of an IT artefact
- Evaluation: Evaluate the artefact in the context of the hypothesis

Big Data Module

- The Big Data Module is part of the Web Science and Big Data program
- It consists of 6 core modules, 2 electives (from a vast collection) and a dissertation:

  - Computer Structures
  - Computer Communication and Networks
  - Big Data
  - Professional Issues in Computing
  - Cloud Computing
  - Social Computing
  - Elective 1
  - Elective 2
  - Dissertation

Summary Plans

- Web Science and Big Data joins the other LOE/UoL IT programs:
  - MSc in Information Technology (IT)
  - MSc in Computer Security
  - MSc in Software Engineering
  - MSc in Information Systems and Technology
  - MSc in Information Systems Management (ISM)
  - MSc in Information Systems Project Management
  - MSc in Web Sciences and Big Data
- Which reaffirms that on-line education is more flexible and provides up-to-date coverage
- Possible next courses and programmes to be developed in 2016-2017
  - Business Intelligence
  - Data Science Engineering
  - Cloud Applications Engineering

Lessons Learnt

- No books to cover more than 1/3 of the course
  - Plenty of Data Analytics and Machine Learning
  - No books on Big Data or Scientific Data Infrastructure
    - Some books Data Warehouse, Scientific/HPC computing, NoSQL databases
- Potential use of Cloud Computing and Big Data platforms on clouds is promising but not for such wide students background like at UoL/LOE
- Instructor training for online courses guiding and moderation
- Use of MOOC or video lectures – suitable for pedagogy and less for andragogy
- Developing a good course for online education is a time consuming job