Project acronym:	**EDSA**
Project full name:	**European Data Science Academy**
Grant agreement no:	**643937**

D3.2 Report on the delivery of video-lectures, webinars and face-to-face trainings

Deliverable Editor:	**A. Voss (Fraunhofer)**
Other contributors:	**OU, SOTON, JSI, ODI, TU/e, Persontyle, KTH**
	C. Phethean (SOTON)
Deliverable Reviewers:	**I. Novalija (JSI)**
Deliverable due date:	**31/07/2016**
Submission date:	**29/07/2016**
Distribution level:	**Public**
Version:	**1.0**

This document is part of a research project funded
by the Horizon 2020 Framework Programme of the European Union

Change Log

Version	Date	Amended by	Changes
0.1	Jan 8, 2016	A. Voss, Fraunhofer	Initial version
0.2	Jan 11, 2016	A. Voss, Fraunhofer	Updated 3.3.1 in order to clarify the key questions we are trying to address with the feedback around F2F delivery
0.3	May17, 2016	A. Voss, Fraunhofer	Update for new courses
0.4	June 21, 2016	A. Voss, Fraunhofer	Description of partners' deliveries
0.5	June 23, 2016	Angi Voss	Adaptations as discussed during EDSA plenary meeting
0.6	July 7, 2016	Angi Voss	Conclusions
0.7	July 18, 2016	Angi Voss	Changes due to internal reviews
0.8	July 28, 2016	Elena Simperl	Scientific Review
1.0	July 29, 2016	Aneta Tumilowicz	Final QA

Table of Contents

List of Tables

List of Figures

1. Executive Summary

This deliverable reports on the courses delivered in WP3. In total, 700 videolectures, 2 MOOCs and 32 different face-to-face courses related to the project have been provided by the partners in Tasks 3.1-3.3 since the start of the project and we are making good progress towards reaching the target figures of audience.

All these courses constitute potential topics and a source of learning material for the core curriculum in WP2. The learning material produced in WP2 – as it becomes available as MOOCS or for self-study in the EDSA portal – can be reused to enhance the original courses. Feedback obtained from their deliveries can then help to improve EDSA's learning material.

The feedback has been obtained either through forms from the participants of face-to-face courses or through automated learning analytics of online courses. The recommendations based on learning analytics can be found in the parallel deliverable D3.3, while this deliverable focuses on recommendations based on delivering F2F courses.

In the last year, after the release of the first version of the core curriculum and EDSA material, the spectrum of delivered courses has led us to extend the core curriculum in D2.2 to the topic of linked data and semantic web. Feedback on the F2F course "Visual Analytics" has led us to replace this topic by "Social Media Analytics". Moreover, feedback on the delivery is guiding the production of EDSA courses at ODI and has led to revisions of the EDSA modules on "Distributed Computing", "Foundations of Data Science" and "Machine Learning". At Soton, ODI and Fraunhofer IAIS first pieces of learning paths are emerging which interlink EDSA modules and F2F courses.

1. Introduction

The objective of WP 3 is to

1. Deploy the materials developed in WP2 in courses for different target groups and in different environments: webinars, videolectures and face-to-face training.
2. Gather data about the effectiveness of learning in the courses.
3. Analyse this data and obtain indications of how to improve the content or form of deployment.

WP3 produces two series of deliverables. While the first series refers to objective 1, the second one addresses objectives 2. Objective 3 is addressed by both series of deliverables.

1. D3.1 (month 6), D3.2 (month 18) and D3.4 (month 36) give an overview of all courses delivered, while recommendations for the learning resources and future delivery focus on face-to-face courses.
2. D3.3 (month 18) and D3.5 (month 36) contain the results of learning analytics for the online courses and corresponding recommendations.

Recommendations on the core curriculum as a whole is also obtained from the industrial board in WP1.

This deliverable D3.2 is the second of the first series. It reports on data science courses that have been delivered by the partners in Task 3.1 (videolectures), Task 3.2 (webinars and MOOCs) and Task 3.3 (face-to-face training) since D3.1 in month7.

All these courses constitute potential topics and a source of learning material for the core curriculum in WP2. The learning material then produced in WP2 – as it becomes available as MOOCS or for self-study

in the EDSA portal – can also be used to enhance the original courses. Feedback obtained from their deliveries can then help to improve EDSA's learning material.

The feedback has been obtained either through forms from the participants of face-to-face courses or through automated learning analytics of online courses. The recommendations based on learning analytics can be found in the parallel deliverable D3.3, while this deliverable focuses on recommendations based on delivering F2F courses.

2. Delivered Courses

The technical annex of the project contains target figures for the audience to be eventually reached by the delivery of different formats: eLearning (passive) corresponds to the videolectures of T3.1, eLearning (active) corresponds to the MOOCs of T3.2, while face-to-face trainings are covered by Task 3.3. The next table shows that we already have reached our target in the MOOC format and can be confident to reach it in the F2F format.

Table 1: Target audience figures

Activity	eLearning (passive) / Videolectures[1] (views counted)	eLearning (active) / MOOCs (registered students counted)	Face-to-face-training (participants counted)
Overall target	200,000	50,000	2,500
Audience reached until month 6, c.f. D3.1	1,005 for ESCW 2015	24,558	356
Total audience reached so far	31,733	50,089	2,018

The subsections below are devoted to the courses given in the different formats of Table 1. Where possible we provide the course characteristics suggested in WP2. This can be: title of the course, length, stage, sector, target group, experience, level of the course:

Stages: The core curriculum in WP2 distinguishes four so-called stages, where the courses reported here can address multiple stages:

- Foundations
- Storage & processing
- Analysis
- Interpretation & use

Target groups: We distinguish four roles of data scientists. A single course can address multiple roles in its target group:

- BE: managers, product developers and business experts

[1] The "Data Science" category was introduced after month 6, But the corresponding number of views cannot be reconstructed from one year ago. Therefore the figures for the two periods are not comparable. For months 1-6 the figure gives the views of videolectures from ESCW.

- DM: data-skilled persons: data managers, curators and data engineers
- DA: data analysts
- IT: system architects and application developers

Experience: We consider two levels of experience in the target group.

- S: Students
- P: Practitioners

This essentially corresponds to the distinction between commercial and academic courses.

Level: We consider courses at three levels:

- 1: Basic
- 2: Advanced
- 3: Expert

Sectors: We use the same sectors as in the online survey in WP1 (see column 3 in Table 8 in the appendix).

Data on the impact can be delivery dates, number of participants and further data which depends on the format of the course (online or offline) and the elicitation method (feedback forms or learning analytics). This deliverables focuses on F2F courses, while D3.3 focuses on online courses.

2.1 Videolectures at JSI

Task 3.1 of WP3 is dedicated to content delivery through Videolectures. VideoLectures.NET is a free and open access educational videolectures repository providing lectures from known scholars and scientists at various events, such as conferences, summer schools, workshops and science promotional events.

Below we present work performed in the context of EDSA training delivery through VideoLectures.Net portal.

Recommended Lectures at VideoLectures.NET portal

The EDSA project started to collate videos on topics related to data science in order to further data science learners' knowledge in the respective fields. Lectures are collated on a periodic basis and distributed through **http://edsa-project.eu/video-lectures**. The lectures are selected by the EDSA project partners based on their relevance to data science education and training around the European Union, and to the level of expertise of the speakers. EDSA focuses on gathering content from world-renowned experts in topics related to data science.

Introducing a Data Science Category at VideoLectures.NET portal

As a part of Task 3.1 (Content delivery through Videolectures), JSI introduced a new category at VideoLectures.NET portal. The "Data Science" category incorporates a number of relevant topics. The number of videos per category is given in parentheses below.

Econometrics (4), Statistics (85), Social_Sciences_Methodology_and_Statistics (6), Artificial Intelligence (417), Big Data (348), Bioinformatics (227), Compressed Sensing (27), Computational Linguistics (113), Computer Vision (573), Crowdsourcing (29), Data Mining (772), Data Visualisation (73), Decision Support (60), Image Analysis (103), Information Extraction (86), Information Retrieval (216), Internet, World Wide Web (177), Knowledge Extraction (270), Machine Learning (3544), Multilingual Information Access (168), Multimedia Search (10), Natural Language Processing (219), Network Analysis (351), Pattern Recognition

(107), Semantic Computing (14), Semantic Search (56), Semantic Web (1371), Sensor Networks (10), Social Computing (27), Social Media (211), Streaming Data (3).

These categories are matched to the stages of our data science curriculum as shown in the next table.

Table 2; Matching categories to the stages of data science

Stage	Categories
Foundations	Econometrics, Statistics, Social_Sciences_Methodology_and_Statistics, Data Mining, Information Extraction, Knowledge Extraction, Machine Learning, Semantic Web, Text Mining
Storage and processing	Big Data, Computer Vision, Crowdsourcing, Information Extraction, Information Retrieval, Knowledge Extraction, Machine Learning, Natural Language Processing, Semantic Computing, Visual Computing, Streaming Data, Text Mining, Web Mining
Analysis	Compressed Sensing, Computer Vision, Data Mining, Decision Support, Image Analysis, Information Extraction, Information Retrieval, Knowledge Extraction, Machine Learning, Natural Language Processing, Network Analysis, Pattern Recognition, Semantic Computing, Semantic Search, Text Mining, Visual Computing, Web Mining, Web Search, Semantic Search, Multimedia Search, Sensor Networks
Interpretation and use	Bioinformatics, Computational Linguistics, Data Visualisation, Image Analysis, Internet, World Wide Web, Multilingual Information Access, Multimedia Search, Sensor Networks, Semantic Search, Social Media, Web Search

In total over 10,000 lectures and tutorials related to data science can be found at Videolectures.NET which were recorded before the start of the project. Statistical information about the top data science related videos from years 2007-2016 can be found in Appendix 1: Videolectures.

The number of videos is quite balanced across three of the four stages, as Figure 1 shows. But note that a video may fall into multiple categories and a category into multiple stages. The stage "Interpretation and Use" is less generic and maybe therefore less filled. However, videos in all categories tend to follow the same pattern. There are a couple of very popular videos and then videos of medium and small popularity. Bioinformatics, data visualisation, image analysis contain popular videos, so it might be worth considering these fields for possible extension of our curriculum.

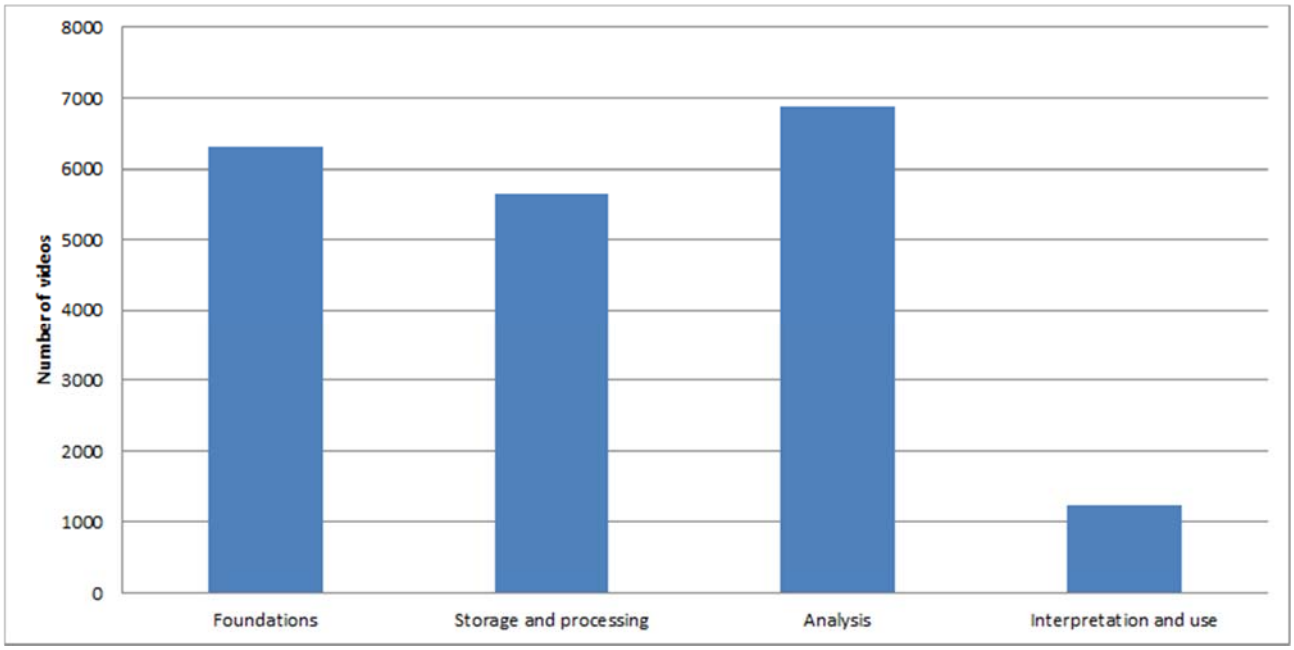

Figure 1: Number of videos per stage in the data science category

Training Delivery through Data Science Lectures at VideoLectures.NET portal

Among the videos in the data science category, over 700 have been published since the beginning of the EDSA project (in years 2015-2016).

The number of views of all videos in the data science category is 2 497 880, with 31 733 views in the years 2015-2016. Table 15 in Appendix 1: Videolectures presents a list of top data science videos (by views). Table 3 presents the top-ten of this list.

Table 3: Top-ten Data Science Videolectures Viewing Statistics (Data Science Videolectures published in 2015-2016)

Publishing Date	Lecture Title	Lecture URL	Views
28.07.2015	Deep Reinforcement Learning	http:// videoLectures.net/rldm2015_silver_reinforcement_learning	3072
5.12.2015	Two high stakes challenges in machine learning	http:// videoLectures.net/icml2015_bottou_machine_learning	848
28.07.2015	Basics of Computational Reinforcement Learning	http:// videoLectures.net/rldm2015_littman_computational_reinforcement	740

5.12.2015	Natural Language Understanding: Foundations and State-of-the-Art	http:// videoLectures.net/icml2015_liang_language_understanding	393
24.02.2016	It's Learning All the Way Down	http:// videoLectures.net/iccv2015_lecun_learning	329
10.02.2016	Multi-Task Recurrent Neural Network for Immediacy Prediction	http:// videoLectures.net/iccv2015_chu_neural_network	319
28.07.2015	Quickly Learning to Make Good Decisions	http:// videoLectures.net/rldm2015_brunskill_good_decisions	258
5.12.2015	Bayesian Time Series Modeling: Structured Representations for Scalability	http:// videoLectures.net/icml2015_fox_structured_representations	237
23.02.2016	Convex Optimization with Abstract Linear Operators	http:// videoLectures.net/iccv2015_boyd_convex_optimization	232
5.12.2015	Advances in Structured Prediction	http:// videoLectures.net/icml2015_daume_structured_prediction	215

2.2 Webinars, MOOCs, Self-Learning Courses

Next to the MOOC on process mining, which TU/ released before D3.1 on Coursera, Soton has published a MOOC on linked data and the semantic web on the FutureLearn platform, as discussed in D2.2. In total, 50,089 users have registered to these MOOCS so that we have already surpassed our planned target of 50,000 participants.

2.2.1 MOOCs at TU/e

The MOOC format responds very well to the purpose of improving the visibility of TU/e and offers an engaging online learning experience. TU/e has hence focused on delivering MOOCs for EDSA. 24,558 people registered on the "Process Mining: Data science in Action" course delivered by TU/e before the time period reported in this deliverable, for a total of 45,374 registered users as of June 7th.

The MOOC explains the key analysis techniques in process mining. Participants learn various process discovery algorithms. These can be used to automatically learn process models from raw event data. Various other process analysis techniques that use event data are presented. Moreover, the course provides easy-to-use software, real-life data sets, and practical skills to directly apply the theory in a variety of application domains.

Table 4: MOOCs by TU/e

Title	**Process Mining: Data Science in Action**	
Course characteristics		
Stage	Analysis	
Level	Basic	
Length	8 weeks	8 weeks
Delivery		
Start date of delivery	07/10/2015	Started 28/04/2015, remains open for registration permanently.
Number of registered participants since M7	14,015	6,801 as of 07/07/2016

Note that the self-study module on "Process Mining" hosted on the EDSA portal is only a small part of the Coursera MOOC reported here. The self-study course insists on abstract ideas, introducing the purposes and explaining in an abstract way the methods and the theory of process mining.

A new MOOC "Process Mining with ProM", also especially prepared for EDSA and published on FututeLearn, is quite a practical course with practical advices, focused on the tool ProM. It starts starts July 11th, lasts 4 weeks and, as of July 7th, has about 3,500 registered participants. This MOOC will be included in the next deliverable.

2.2.2 MOOCs at Soton

In line with the revisions to the EDSA curriculum made in D2.2, the University of Southampton have implemented the first FutureLearn MOOC within the EDSA project, releasing the "Introduction to Linked

Data and the Semantic Web" course in April 2016 as the delivery mechanism for the relevant module in Version 2 of the EDSA curriculum presented in D2.2 (M18). A FutureLearn MOOC was chosen due to the availability of existing resources from the EUCLID project, and as test case for the FutureLearn platform in the project - this has helped us to establish a number of guidelines for developing a MOOC on this platform and allowed us to prepare learning analytics for future MOOCs based on the data received.

Table 5: MOOCs by Soton

Title	**Introduction to Linked Data and the Semantic Web**
Course characteristics	Online course, interactive exercises, quizzes, community discussions, lecturer feedback
Stage	Storage & processing
Target group	DM: Database managers/administrators
Experience	Practitioners
Level	Basic
Length	3 weeks
Delivery	Online (Futurelearn MOOC)
Start date of delivery	11/04/2016
Number of registered participants since M7	4 715

2.3 Face-to-Face Trainings

The face-to-face training courses delivered so far have reached 2018 persons against the planned target of 2 500 by the end of the project. Part of the courses are offered to paying customers, typically practitioners, the others are given to students at the university.

Data on impact and feedback for commercial face-to-face courses are obtained from registration and feedback forms. As the partners use different forms, the data presented here varies with the partner. For academic courses such data is typically not collected.

2.3.1 Courses at Fraunhofer IAIS

At Fraunhofer IAIS courses for data scientists have been offered since 2014. The modules have been extended step-by-step to a more comprehensive data scientist training programme, with contributions from several other institutes in Fraunhofer's Big Data Alliance, which is coordinated by our institute.

- In 2015, funded by EIT Digital (www.eitdigital.eu/), two sector-specific blended learning modules were developed for smart energy systems and smart buildings, and a third course on security and privacy for big data. The online part consisted of annotated slides, quizzes and exercises for self-study after 2-3 days of face-to-face training.
- In 2016 a new sector-independent course started for data managers and two new courses started for data scientists in life sciences and health care.
- Our Big Data Alliance has set up a personal certification programme for data scientists at three levels – basic, specialist and expert – targeting persons who seek formal qualifications for their career building. A first building block is a comprehensive course called "data scientist basic level" of 5 days with a written exam at the next day. It was coordinated by our institute with contributions to data management by the Fraunhofer Institut für Angewandte Informationstechnik FIT, to all business related aspects by the Fraunhofer Institut für Experimentellles Software Engineering IESE, and to security and privacy by the Fraunhofer Institut für Sichere Informationstechnologie SIT.

Table 3 summarizes the more popular courses held at Fraunhofer IAIS including the new certificate course "Data Science Basic Level".

In addition to the courses in Table 6 we have conducted courses on:

- Business potentials of big data analytics: delivered 14/12/15, 16 participants
- Multimedia Analytics: delivered 7/12/15 and 27/4/16, 13 participants
- Linked enterprise data integration: delivered 14/4/16, 4 participants

Table 6: F2F-courses at Fraunhofer

Title	**Basic Data Analytics**	**Big Data Architecture**	**Big Data Analytics**	**Visual Analytics**	**Social Media Analytics**	**Data Scientist Basic Level**
Course characteristics						
Stage	Analysis	Storage & processing	Storage & processing, Analysis	Interpretation and use	Analysis	
Sector					Scientific and market research	
Target group	Data analysts	System engineers	System engineers	Data analysts	Data analysts	
Experience	Practitioners	Practitioners	Practitioners	Practitioners	Practitioners	Practitioners
Level	Advanced	Advanced	Advanced	Basic	Advanced	Basic
Length (in days)	2+ (1 opt.)	2	2	2	2	4.5

Delivery since M7						
Start date of delivery	1/09/15 20/10/15 17/11/15 25/11/15 22/02/16 22/02/16 18/05/16 31/05/16	25/09/15 21/08/15 25/08/15 12/11/15 02/12/15 17/12/15 20/04/16 16/03/16	30/11/15 19/08/15 29/09/15 20/10/15 10/11/15 18/04/16	29/9/16 25/2/17	9/12/15	6/6/16
Number of participants since M7	98	81	73	14	11	15

Apart from "Social Media Analytics", all courses in Table 3 are sector-independent. Most courses are at an advanced level and target two different groups of practitioners, data analysts and engineers of big data systems. The courses address different stages except foundations. They are either open and take place at our campus or they are commissioned by a company and delivered at their premises. The number of participants is limited to 10, 12 or 15, depending on the degree of interaction permitted in the course.

In general, we offer a course twice a year, adding offers as demand increases and capacities permit. By far the most popular courses are "Basic Data Analytics", "Big Data Analytics" and "Big Data Architecture". The latter two focus on big data, and we regularly offer them together in one week. These two courses are the precursors of the corresponding EDSA courses.

The percentage of female participants lies around 16%, with peaks of 30% in "Visual Analytics" and 25% in the new certificate course.

Only recently we have started to ask for the role of the participants, as shown in Figure 2. The lower number of software developers participating in "Visual Analytics" might explain the higher percentage of females in this course. In the new certificate course we are attracting business developers as a new target group.

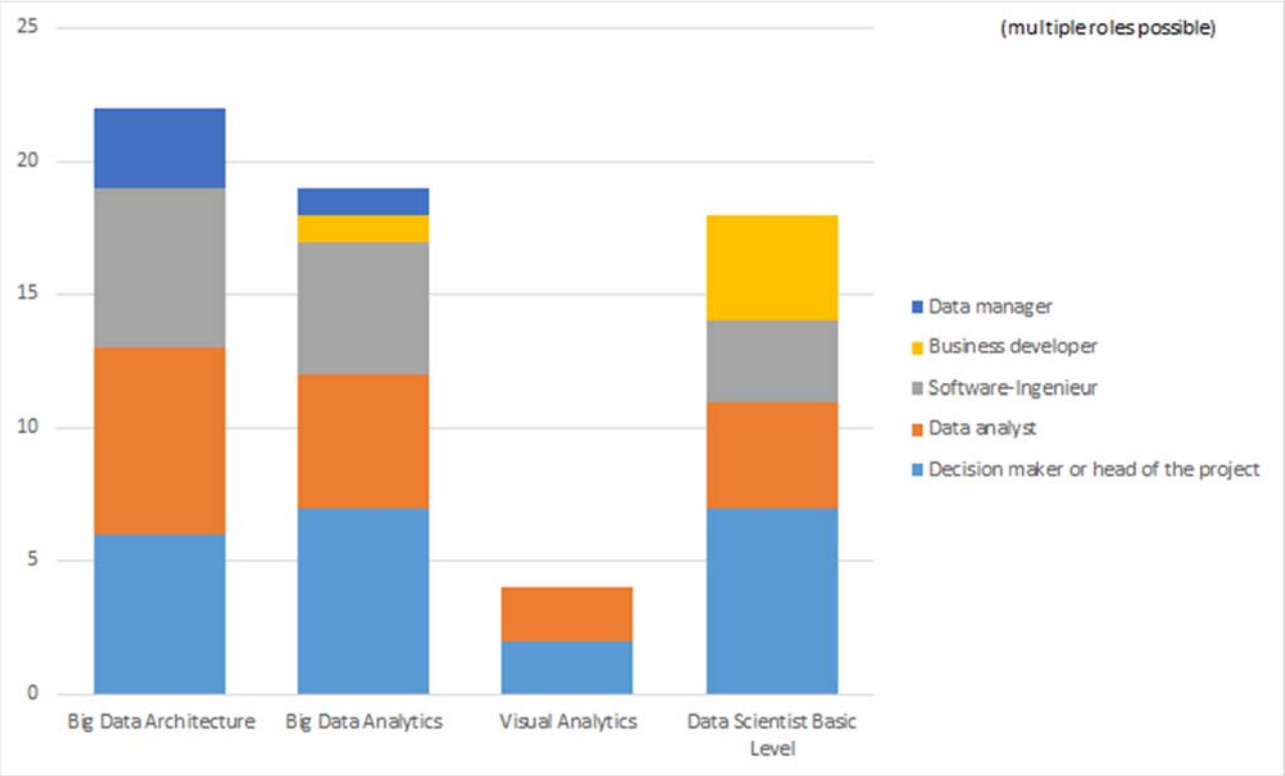

Figure 2: Roles of the participants per course

The 325 participants mostly come from the automotive industry (84), followed by consulting (55), telecom (43), manufacturing (40), the public sector (32), energy (18) and data and information systems (17). It is interesting that there is high demand for basic analytics in manufacturing but low demand in consultancy, but vice versa high demand for big data technology and analytics in consultancy, and low demand in manufacturing. In the automotive industry such a focus cannot be observed.

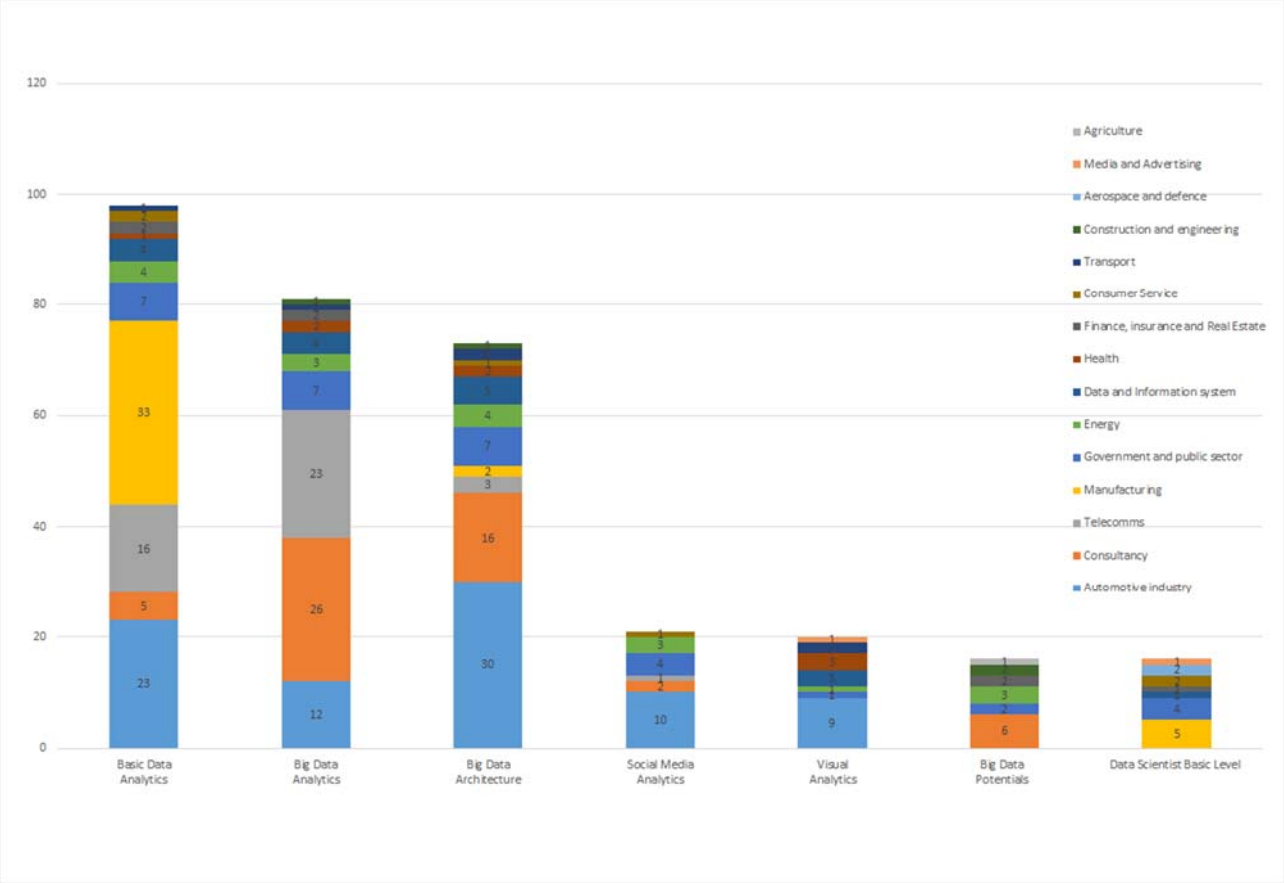

Figure 3: Sectors of the participants per course

The next diagram compares satisfaction with the content with practical relevance and with the strength of recommending the course to a friend or colleague. In general, practical relevance obtains the lowest scores. Our trainers believe that we would have to include more examples from real projects, which could be done in tailored inhouse deliveries of the courses. However, less practical relevance usually hardly affects participants' positive attitude towards recommending the course to their colleagues. The situation is different with the new certificate course. Here we need to identify more clearly the material that is relevant for the exam.

"Visual analytics" has turned out to be a difficult topic, suffering from more realistic data, from popular tools not being open or available for free, and a difficult balance between information visualization and visual data exploration. As discussed in D2.2, similar overlaps with other courses on visualization were encountered in the EDSA core curriculum. We therefore quit our original plan to turn this course into another EDSA module. Instead, we will develop an EDSA course on social media analytics, with an outlook on deep learning as a new hot topic.

Our participants are mostly content with the face-to-face format because their companies do not give them much time for preparation or further exercises. However, they do appreciate a limited amount of preparatory information which helps them to get attuned to the topic of the course. For our courses on "Big Data Architecture" and "Big Data Analytics" we therefore reuse the corresponding self-study modules of EDSA. In the future we might also use them for a kind of self-assessment, when potential customers need help in judging the relevance of a particular course for their career.

The situation is again different with the new certificate course. Almost half of the participants would have preferred a blended format, essentially in order to facilitate learning for the exam. The material we developed for EDSA could be reused for this purpose.

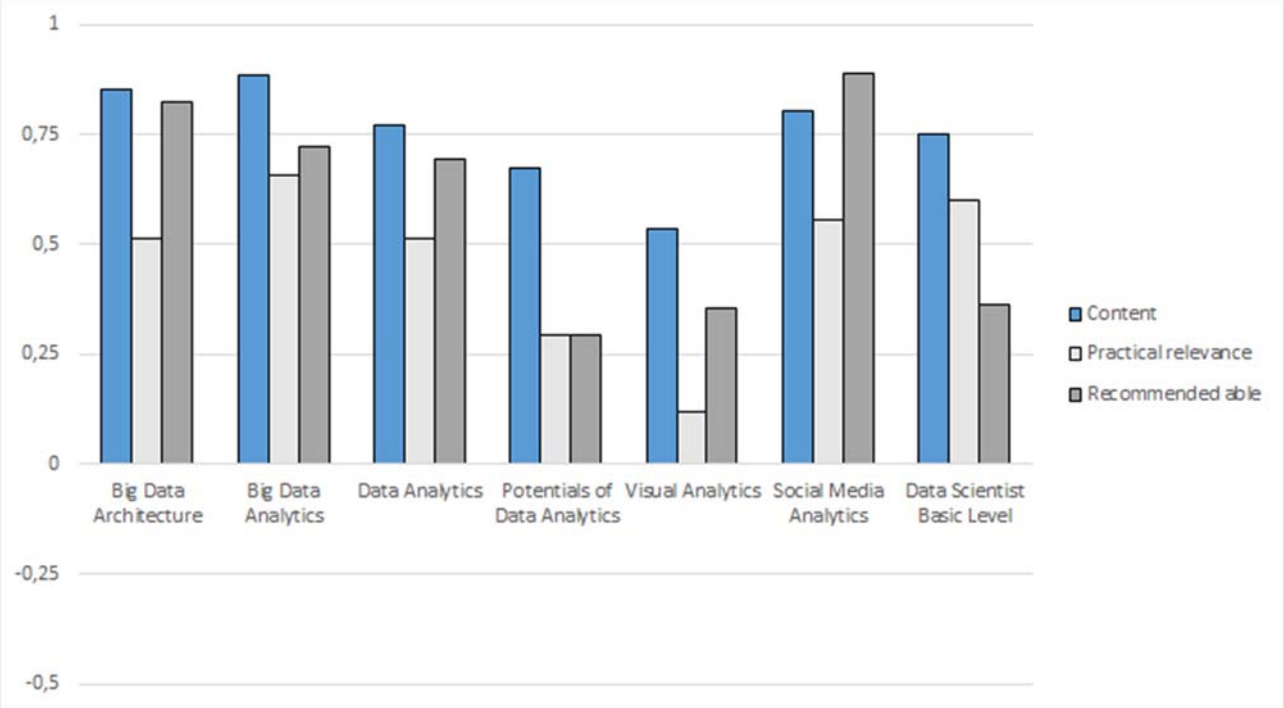

Figure 4: Feedback per course

We hoped to learn more about potential learning paths by asking the participants which topics they would like to learn about next, but hardly received any feedback.

2.3.2 Courses at Persontyle

So far in 2016 we have conducted 4 face-to-face training programs. Both of the workshops are designed for people who have some prior programing experience. Total of 150 participants attended these workshops.

We have designed both learning programs so that participants can learn enough to start practicing and experimenting with machine learning and deep learning, as there are great open source machine learning libraries (such as scikit-learn in Python) and cloud platforms that make it easy to create machine learning models from data.

Most developers these days have heard of machine learning, but when trying to find an 'easy' way into this technique, most people find themselves getting scared off by the abstractness of the concept of machine learning and terms as regression, unsupervised learning, probability density function and many other definitions.

Machine learning teachers like to explain how different learning algorithms work and spend tons of time on that. For a programmer/developer who wants to start using machine learning, being able to choose an algorithm and set parameters looks like the #1 barrier to entry, and knowing how the different techniques work seems to be a key requirement to remove that barrier.

The “Operational Machine Learning” workshop provides an agnostic introduction to operational ML with open source and cloud platforms. It is the first ML workshop to go all the way from data preparation to the integration of predictive models in real-world applications and their deployment in production.

Participants who attended the workshop learned how to use Python open source libraries scikit-learn, Pandas and SKLL, and cloud platforms Microsoft Azure ML, Amazon ML, BigML and Indico (along with their APIs).

The "Deep Learning Workshop" is designed to practically learn everything a practitioner need to design, train, and integrate neural network-powered artificial intelligence into applications with widely used open-source frameworks. The workshop is targeted for researchers, developers, hackers, postgraduate students, data scientists, quants, or data analysts that already know about machine learning and have experience in programing.

Table 7: Courses at Persontyle

Title	Deep Learning Workshops	Operational Machine Learning Workshop
Course characteristics		
Stage	Analysis	Analysis
Target group	Developers, Data Scientists, Analysts and Technical Architects	Developers
Experience	Practitioners	Practitioners
Level	Advanced	Advanced
Length	2-3 days	2 days
Delivery since M7		
Start date of delivery	03-04 Feb, London, UK 04-06 July, London, UK	11-12 May Madrid, Spain 16-17 June London, UK
Number of participants	90	60

Both of the workshops and feedback from attendees has helped us in designing and making the EDSA Study Guide "The Essentials of Data Analytics and Machine Learning" practical and relevant by covering the concepts, technology and applied practices required throughout the entire lifecycle, from asking the relevant questions to developing machine learning models and visualizing results.

As discussed and outlined in D2.2, we are adding four new modules in the Study Guide and also updating the modules listed to cover advanced topics suggested by the participants of the above workshops.

- Module 16 - Gaussian Processes
- Module 17 - Tree based methods
- Module 18 - Support Vector Machines (SVMs)
- Module 19 - Neural Networks and Deep Learning

Also, these F2F instructor-led workshops and feedback of the attendees will help us designing the new EDSA Machine Learning MOOC. This will be an eLearning program of 4 practical courses on machine learning concepts, practices, models and tools. The aim of this programme is to provide both a deep

understanding of the techniques and practices of machine learning and to expose a wide set of resources capable of being wielded by the data scientist and analyst in their work. Participants will encounter explanations of the theory behind the algorithms and models they are exposed to, giving them an understanding of the strengths and weaknesses of each which they should be able to use to reason about suitable approaches to real life problem – and to communicate such reasoning to other stakeholders in such problems.

2.3.3 Courses at ODI

The Open Data Institute (ODI) runs a suite of data and open data training courses in order to equip people with the knowledge and skills necessary to become data proficient. "Finding Stories in Open Data", "Open Data in Practice" and "Open Data Science" are all ODI courses which provide technical training to non-technical people, and provide an entry level introduction to data science.

The ODI chooses to run its courses in face-to-face classroom format due its combination of theoretical and practical education. Allowing the participants to put their learning into practice converts newly acquired knowledge into skills. The classroom format also fosters peer learning and gives participants the opportunity to ask questions directly to a subject matter expert. The ODI offers face to face courses of different lengths, along with online and blended training. This provides prospective participants with the option to choose a format that suits their learning style best. Participants who wish to learn more and learn at their own pace are encouraged to go to the EDSA courses portal to try the technical MOOCs available.

The ODI have been monitoring the feedback of our practical courses in order to inform how we might adapt the content to better serve user needs, and adapt the materials we plan to develop for EDSA. The feedback has informed us where we need to make changes and how useful the materials are. The forthcoming demand analysis will help to further evaluate these courses. Below you can see a breakdown of the feedback and recommendations by course.

Table 8: Courses at ODI

Title	Finding Stories in Open Data	Open Data in Practice	Open Data Science
Course characteristics			
Stage	Interpretation & use		
Sector	Media	Applicable across all public, private and third sectors.	Applicable across all public, private and third sectors.
Target group	BE: Journalists, project managers, developers, writers, artists, producers, communications managers, press relations, presenters and consultants.	All roles: Managers, technologists, all data roles, directors and those working in knowledge and transparency.	All roles: Business and technology professionals, developers, managers, executives, data analysts and teaching staff.

Experience	Students or practitioners.	Practitioners	Students or practitioners.
Level	Basic	Advanced	Basic
Length	1 day	3 days	0.5 days
Delivery since M7			
Start date of delivery	27/1/15, 10/3/15 22/4/15 23/6/15 16/9/15 20/10/15 9/2/16 15/4/16	3/2/15, 17/3/15 21/7/15 25/8/15 22/9/15 27/10/15 1/12/15 2/2/16 12/4/16 18/4/16	14/7/15 3/8/15 2/9/15 6/10/15 5/11/15 2/12/15 7/12/15 15/12/15 10/2/16 27/4/16
Number of participants	42	99	134

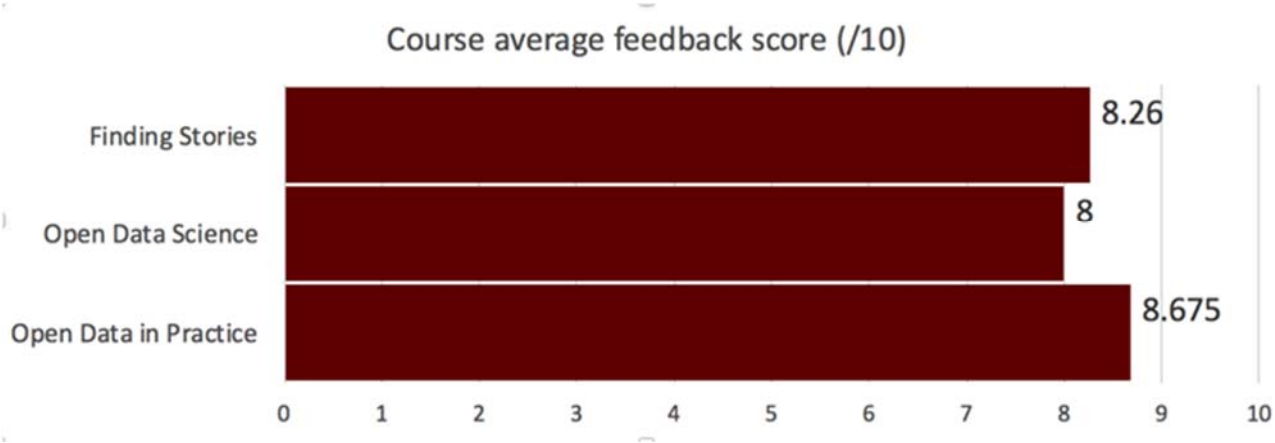

Figure 5: Average feedback score

The three courses listed above and delivered throughout 2015 and 2016 have received an average participant rating of 8/10. Open Data in Practice, the ODI's longest running course, has the highest participant rating. 92% of participants would recommend the ODI's courses to a friend or colleague.

92%

of people would recommend our courses to others

Open Data in Practice recommendations:

1. Maintain diversity of course content
2. Extend activity length
3. Describe as 'cultural' rather than technical.

100% of participants stated that the course materials and slides were presented clearly. Participants gave varied answers about their biggest takeaway, and one participant claimed that the variety and 'diversity' of the content covered on the course is what makes the course valuable for participants. Two

participants suggested there could be clearer instruction about activities and more time for activity completion in order to improve the pace of the course. A further recommendation for Open Data in Practice has been to frame the course as 'more cultural than technical'. Having received a cultural introduction to open data, participants should then be guided to the EDSA online and classroom courses for more technical guidance.[2]

Open Data Science recommendations:

1. Reframe course as beginner level course
2. Reduce course content
3. Extend activity length.

43.75% of participants agreed that they were unclear of how the course content could be applied to their jobs. Of these participants, 80% have a medium-high level of technical knowledge. The recommendation would be to change the course level to 'beginners' and emphasise how skills taught on the course could be applied to the jobs of the audience. 50% of participants would prefer more time to work on exercises and to reflect on the content. This may be counterbalanced by reducing down some of the course's content and extend the length of activities. Our next step in the project is to focus on providing a pathway for more technical people to follow. As the 'introductory' online courses have been most popular on the EDSA course portal at the time of writing, the goal is to create a technical introductory online course.

Finding Stories in Open Data recommendations:

1. Clarify required level of pre-course knowledge on course description
2. Include more activities with data manipulation tools
3. Change intended audience to NGOs

73.33% of participants agree that the course had a good pace and mix of activities. Participants who had problems with the pace of the course were at different levels of ability, and as a consequence needed more or less time. This can be counterbalanced by specifying more clearly what level of knowledge participants should have before attending. Data cleaning and visualisation are the most valuable topics taught on the course, where the two topics make up 52.63% of the topics that participants will most take away with them. In course feedback comments, 47.37% of participants stated that having access to and learning how to use tools was the most valuable learning. Another popular topic in the course is finding data, which comes in at 21.05% in the feedback. Despite the fact that the course is marketed as a data journalism course, only 6.1% of attendees came from newspapers or other journalism organisations. By contrast, 71.4% of attendees belonged to charities and NGOs who wanted to gain more technical skills. Based on the above evidence, the next steps for the project would be to develop two different data journalism materials; one will be more technical and skills based, one will be more journalism focussed.

2.3.4 Courses at Soton

In addition to the MOOC "Introduction to Linked Data and the Semantic Web" mentioned above, the University of Southampton began delivering a face-to-face Data Science MSc course in September 2015, with two new modules forming a key component of this which relate to their assigned modules in the

[2] Online and face to face courses can be found here: http://courses.edsa-project.eu/

EDSA curriculum: Foundations of Data Science and Data Visualisation. Each module was implemented based on the curricula designed for EDSA. 15 students were registered on the MSc programme itself, although the two new modules were also open to students on similar courses including computer science and artificial intelligence.

Southampton are also involved in the preparation of the ESWC Summer School and as with the OU will use this as an opportunity to test elements of the data science curriculum.

Table 9: Courses at Southampton

Title	**Foundations of Data Science**	**Data Visualisation**
Course characteristics		
Stage	Foundations	Interpretation & use
Experience	Students/Graduates	Students/Graduates
Level	Basic	Basic
Length	1 semester (3.5 months)	1 semester (3.5 months)
Delivery since M7		
Start date of delivery	01/10/2015	25/01/2016
Number of participants	31	42

Based on our experiences of delivering both modules, we have revised the syllabus for "Foundations of Data Science" as part of D2.2. This is based on the experiences of the module teachers after the first instance of delivery, with a particular need to reduce the number of topics and highlight more particular aspects of the data science pipeline.

Professional Courses

As a further variation of their postgraduate courses, Southampton are currently preparing a number of Continuing Professional Development (CPD) courses in data science, with the first 'Foundations of Data Science' online CPD course to go live in Autumn 2016, as outlined in their exploitation plan in D5.3. A number of specialised courses will then follow, including data science for marketing, finance and healthcare. Each course will run for four weeks, will be offered online through the Canvas platform, and will be targeted towards professionals seeking to upskill in order to meet the skills requirements of data science jobs.

2.3.5 Courses at OU

After the 5th ESWC Summer School in 2015 the OU will also organise this year's ESWC Summer School in Dubrovnik from September 5th to 10th 2016. The overall goal for this event is to provide intensive training and networking opportunities to data-skilled researchers and professionals. This year's motto

will be 'The rise of the data scientist'. In particular, the school will feature a mixture of invited talks and tutorials on several aspects of data science, including statistics, exploratory data analysis, data publishing and interlinking, machine learning, and data visualization as well as group projects, in which the participants will have the chance to apply what they have learned by developing their own data science research ideas and demos as part of a team.

The summer school is open to anyone studying in a Semantic Web or Data Science related postgraduate course or is at an early stage of a Semantic Web or Data Science related career. Places will be limited to 50 in order to ensure that all participants receive quality time with their tutors. Accepted participants will be obliged to attend the whole week. As in previous editions of this event, all lectures will be recorded and will be made available online via videolectures.net.

Table 10: Courses at OU

Title	6th ESWC Summer School
Course characteristics	
Stage	Storage & processing
Target group	Data-skilled persons
Experience	Practitioners: Professional and research experience required on data science
Level	Advanced
Length	6 days
Delivery since M7	
Start date of delivery	05/09 - 10/09/2016
Number of participants	Up to 50

Specifically for EDSA, the main outcome of the ESWC summer school is testing the project's learning materials and exercises and gathering feedback from the participants of the event. The ESWC Summer School is therefore used by EDSA primarily as a testing channel for the project's curricula, instead of a delivery channel. Last year's ESWC summer school offered us the opportunity to present the project's curriculum to a data science professionals and gather feedback about its potential and its limitations. This year's summer school will also offer the opportunity to participants to try some of the learning materials produced by EDSA and offer their feedback on them.

2.3.6 Courses at JSI

JSI continued to provide periodical internal trainings. They can be seen as a source of topics and material for future EDSA modules.

Table 11: Courses at JSI

Title	Rapid Development of Data Mining Applications in Node.js	Nowcasting	Personalized mobility patterns	Data mining for social good
Course characteristics				
Stage	Analysis	Analysis	Analysis	Analysis
Sector	Information Technologies	Information Technologies/Finance/Economics	Information Technologies/Transport	Information Technologies
Target group	IT+DA: computer scientists, data analysts	IT+DA: computer scientists, data analysts	IT+DA: computer scientists, data analysts	IT+DA: computer scientists, data analysts
Experience	Student	Student	Student	Student
Level	Basic	Basic	Basic	Basic
Length	1 day	1 day	1 day	1 day
Delivery				
Start date of delivery	02/12/2015	15/04/2015	13/05/2015	09/09/2015
Number of participants	20	20	20	20

Our experience shows that having internal trainings is a useful way to detect and cover the gap of formal education. In particular, the new and emergent technologies and methods presented at the internal trainings can be exploited as sources for extension of EDSA training materials, PhD and Master educational programs (in particular, at Jožef Stefan International Postgraduate School).

2.3.7 Courses at KTH

KTH provides courses that are part of the master programs "Software Engineering of Distributed Systems" at KTH and "Cloud Computing and Services "at EIT Digital. The reported courses are parts of the "Distributed Computing" module in the EDSA curricula.

Table 12: Courses at KTH

Title	**Distributed Systems, Part 1**	**Distributed Artificial Intelligence and Intelligent Agents**	**Programming Web Services**	**Distributed Systems, Part 2**
Course characteristics				
Stage	Storage & processing	Storage & processing	Storage & processing	Storage & processing
Target group	IT: Students of KTH and EIT Digital Master programs, IT professionals	IT: Students of KTH and EIT Digital Master programs, IT professionals	IT: Students of KTH and EIT Digital Master programs, IT professionals	IT: Students of KTH and EIT Digital Master programs, IT professionals
Experience	Students	Students	Students	Students
Level	Master/advanced	Master/advanced	Master/advanced	Master/advanced
Length	1.5 months	1.5 months	1.5 months	1.5 months
Delivery since M7				
Start date of delivery	20/01/2016	01/11/2015	20/01/2016	25/03/2016
Number of participants	35	42	32	30

Our experience shows that clustering related courses into an educational module is positive for the learners. Therefore we updated the EDSA course curricula so that all the above mentioned courses are included into a module on "distributed computing".

2.3.8 Courses at TU/e

As a University, TU/e provides courses to students, and has a broad and increasing variety of courses related to data science. TU/e held the following data science courses at the Master level. The table is split into two parts for readability purposes.

Table 13: Courses at TU/e

Title	Advanced process mining	Advanced Data Analysis	Web information retrieval and data mining	Introduction to process mining	Visualization
Course characteristics					
Stage	Analysis	Analysis	Foundations	Analysis	Interpretation & use
Experience	Students	Students	Students	Students	Students
Level	Advanced	Advanced	Basic	Basic	Basic
Length	17.5 days in 2.5 months	17.5 days in 2.5 months	17.5 days in 2.5 months	17.5 days in 2.5 months	17.5 days in 2.5 months
Delivery since M7					
Start date of delivery	01/04/15	01/09/15	01/09/15	01/09/15	01/11/15
Number of participants	93	31	97	25	96

Title	Statistics for big data	Foundations of data mining	Principles of data protection	Big data and experiments for urban analysis	Data engineering
Course characteristics					
Stage	Foundations	Foundations	Storage & processing	Analysis	Storage & processing
Sector					

Target group					
Experience	Students	Students	Students	Students	Students
Level	Basic	Basic	Basic	Basic	Basic
Length	17.5 days in 2.5 months	17.5 days in 2.5 months	17.5 days in 2.5 months	35 days in 5 months	17.5 days in 2.5 months
Delivery since M7					
Start date of delivery	01/11/15	01/02/16	01/09/15	01/02/16	01/02/16
Number of participants	20	70	79	10	49

Among the 570 participants, 241 were not Dutch, 107 were female (about 19%), and 463 male. As usual in this field, there was a majority of male participants. The proportion varied among topics, and for instance "advanced data analysis" had a majority of female participants. No clear pattern emerges from the graph below, showing the proportion of female participants in the total of participants per course.

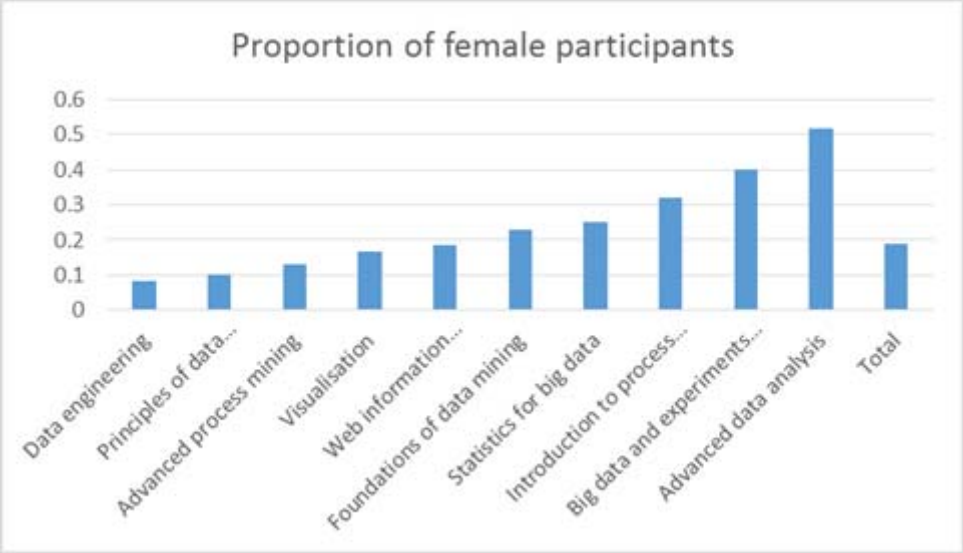

Figure 6: Proportion of female participants

Among the participants, the average age was about 25 year old for all of the courses. The age of the participants ranged from 20 to 37 years old. The graph below shows the proportion of participants from other countries than the Netherlands. We found that the courses with a higher proportion of foreign students tend to have a better satisfaction rate on this very small data sample. We can also note that "introduction to process mining" had a large majority of foreign students (as well as the best reported satisfaction rate).

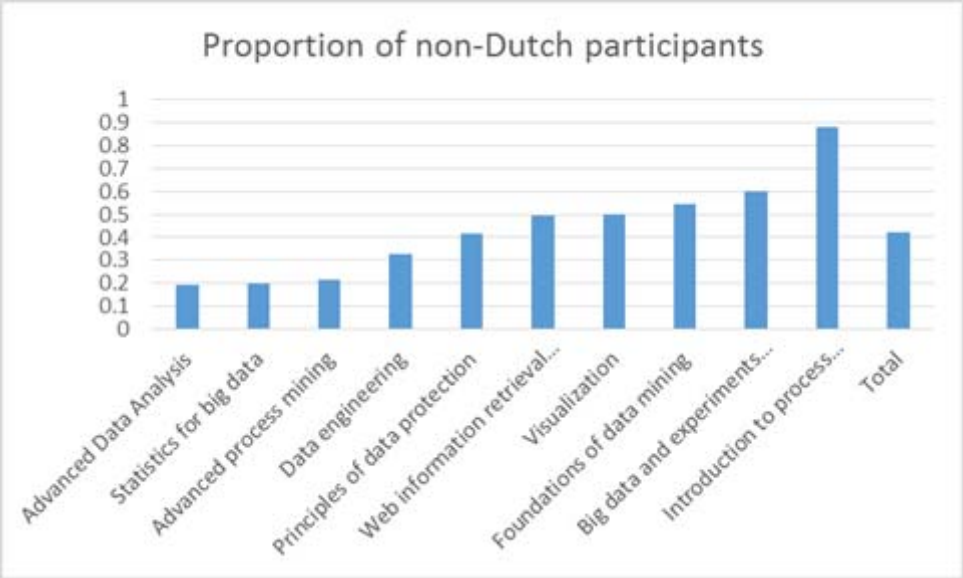

Figure 7: Proportion of non-Dutch participants

3. Summary and Conclusions

The preceding section has presented a range of courses delivered by the partners or offered as EDSA online courses. Apart from ca. 700 videolectures and 2 MOOCs we have 32 F2F courses. Among the latter we have a good balance between basic and advanced level courses (17:14). As the consortium is dominated by universities, we clearly have more F2F courses for students than for practitioners (21:12). Most F2F courses do not focus on a particular sector: there are only four courses especially for the sector "Data and Information Systems, one for "Media", one for "Financial & Insurance Services" and one for "Transport".

Figure 8 shows the number of F2F courses per stage of the data science core curriculum. Notably, only four courses are foundational and five on "interpretation & use". An explanation for few foundational academic courses could be that students have obtained the foundational in earlier, more general courses, such as statistics or mathematics, while professional courses often require basic knowledge as given in their audience of practitioners. The category "interpretation & use" may be rather sector-specific. Here it may be difficult to obtain disclosable data for examples, or demand is comparatively low. As a recommendation, production of sector-specific learning material with a focus on "interpretation & use" in EDSA could deal with the first factor. Judging by the popularity of videolectures, bioinformatics, data visualisation, image analysis might be worth considering as these fields for possible extension of our curriculum.

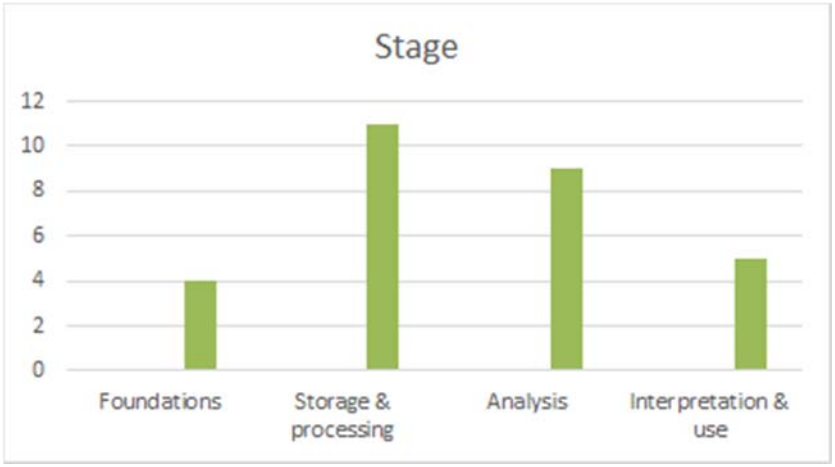

Figure 8: Number of delivered F2F courses per stage

Most courses address data scientists in any roles, as the next figure shows. Many courses address IT-educated persons, which may be due to the number of computer science universities in the consortium. The fact that only 1 course is for business experts/product developers/managers might again be related to the low number of sector-specific courses. Finally, the fact that only one course focuses on data skilled persons supports the extension of the core curriculum to the topic of semantic web and linked data.

Figure 9: Number of delivered F2F courses per role of the target group

The number of female participants was reported by TU/e in Section 2.3.8 for its academic courses and by by Fraunhofer in Section 2.3.1 for its professional courses. With an average of 19% at TU/e and 16% in most courses of Fraunhofer with peaks of 30% in "Visual Analytics" and 25% in the new certificate these figures seem quite comparable.

Table 14 traces the online courses produced for the EDSA core curriculum to courses delivered by the partners. Where EDSA courses have emerged from preceding courses of the partners, lessons learned from ongoing deliveries may be applicable to the EDSA courses and lead to revisions, as is the case for Soton's course "Foundations of Data Science and KTH's course on "Distributed Computing".

As far as F2F courses are concerned, these courses may now be combined with the online material available through EDSA. EDSA courses can be offered to the participants optionally for preparation, as done by Fraunhofer IAIS, or as introductory or follow-up courses, as planned by ODI and Soton. Online material could be used for self-assessment, given to interested persons before they decide to register for a course, or for rehearsal and preparation for an exam. And of course, EDSA courses can be used for promotion.

The table also shows that some EDSA courses have been created from scratch. These materials can also be exploited for new F2F courses, blended courses or learning paths.

Table 14: Relation between the EDSA core curriculum, as far as release in June 2016, and the courses delivered

Stage	Topic	Preceding course of a partner	EDSA release (June 2016)	Exploitation, reuse	Partner
Foundations	Foundations of Data Science		Self-study	For continuous professional development	Soton
			revised		
	Foundations of Big Data		Self-study		JSI
	Big Data Architecture	F2F	Self-study	Optional preparation, rehearsal, self-assessment	IAIS
	Distributed Computing	F2F	Self-study		KTH
			revised		
	Linked Data and the Semantic Web	F2F	MOOC		Soton
Analysis	Machine Learning, Data Mining and Basic Analytics		Self-study		Personyte
			revised		
	Big Data Analytics	F2F	Self-study	Optional preparation, rehearsal, self-assessment	IAIS
	Process Mining	MOOC	Self-study		TU/e
			MOOC		
Interpretation and Use	Data Visualisation and Storytelling	F2F		Forming F2F/online learning paths	ODI Soton

Not all courses delivered by the partners are precursors or successors of EDSA courses. They constitute a source for extending the curriculum. Topics not yet covered I the curriculum are:

- Nowcasting
- Open data in practice
- Open Data Science
- Personalized mobility patterns
- Data mining for social good
- Web information retrieval and data mining
- Big data and experiments for urban analysis /project
- Analysing experimental data

Appendix 1: Videolectures

The number of views of all videos in the data science category is 2 497 880, with 31 733 views in the years 2015-2016. Table 15 presents a list of top data science videos (by views).

Table 15: Top Data Science Videolectures Viewing Statistics (Data Science Videolectures published in 2015-2016)

Publishing Date	Lecture Title	Lecture URL	Views
28.07.2015	Deep Reinforcement Learning	http:// videoLectures.net/rldm2015_silver_reinforcement_learning	3072
5.12.2015	Two high stakes challenges in machine learning	http:// videoLectures.net/icml2015_bottou_machine_learning	848
28.07.2015	Basics of Computational Reinforcement Learning	http:// videoLectures.net/rldm2015_littman_computational_reinforcement	740
5.12.2015	Natural Language Understanding: Foundations and State-of-the-Art	http:// videoLectures.net/icml2015_liang_language_understanding	393
24.02.2016	It's Learning All the Way Down	http:// videoLectures.net/iccv2015_lecun_learning	329
10.02.2016	Multi-Task Recurrent Neural Network for Immediacy Prediction	http:// videoLectures.net/iccv2015_chu_neural_network	319
28.07.2015	Quickly Learning to Make Good Decisions	http:// videoLectures.net/rldm2015_brunskill_good_decisions	258
5.12.2015	Bayesian Time Series Modeling: Structured	http:// videoLectures.net/icml2015_fox_structured_representations	237

	Representations for Scalability		
23.02.2016	Convex Optimization with Abstract Linear Operators	http:// videoLectures.net/iccv2015_boyd_convex_optimization	232
5.12.2015	Advances in Structured Prediction	http:// videoLectures.net/icml2015_daume_structured_prediction	215
28.07.2015	Natural RLDM: Optimal and Subptimal Control in Brain and Behavior	http:// videoLectures.net/rldm2015_daw_brain_and_behavior	205
23.02.2016	Borrowing New Ideas from Human Vision	http:// videoLectures.net/iccv2015_lowe_human_vision	188
10.02.2016	Fast R-CNN	http:// videoLectures.net/iccv2015_girshick_fast_r_cnn	169
10.02.2016	Deep Neural Decision Forests	http:// videoLectures.net/iccv2015_kontschieder_decision_forests	165
28.07.2015	Generalization and Exploration via Value Function Randomization	http:// videoLectures.net/rldm2015_van_roy_function_randomization	160
5.12.2015	Modern Convex Optimization Methods for Large-scale Empirical Risk Minimization	http:// videoLectures.net/icml2015_schmidt_risk_minimization	149
6.03.2015	NLPGo: Better research through better tools	http:// videoLectures.net/cmuseminars_hodson_nlpgo	144
15.07.2015	Desperately Searching for Travel Offers? Formulate Better	http:// videoLectures.net/eswc2015_lu_linked_data	139

	Queries with Some Help from Linked Data		
27.09.2015	Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP)	http:// videoLectures.net/icml2015_wilson_ kernel_interpolation	134
2.07.2015	Dense 3D Face Alignment from 2D Videos in Real-Time	http:// videoLectures.net/fgconference2015 _jeni_2d_videos	125
10.02.2016	Ask Your Neurons: A Neural-based Approach to Answering Questions about Images	http:// videoLectures.net/iccv2015_malino wski_your_neurons	124
10.02.2016	Human Parsing With Contextualized Convolutional Neural Network	http:// videoLectures.net/iccv2015_liang_hu man_parsing	121
10.02.2016	Unsupervised Visual Representation Learning by Context Prediction	http:// videoLectures.net/iccv2015_doersch _visual_representation	121
28.07.2015	Proximal Reinforcement Learning: Learning to Act in Primal Dual Spaces	http:// videoLectures.net/rldm2015_mahad evan_dual_spaces	118
5.12.2015	Learning to Rank Using Gradient Descent	http:// videoLectures.net/icml2015_burges_ learning_to_rank	113
28.07.2015	Robots learning from human teachers	http:// videoLectures.net/rldm2015_thomaz _human_teachers	113
28.07.2015	A Neuroeconomics Approach to Pathological Behavior	http:// videoLectures.net/rldm2015_levy_pa thological_behavior	112

5.12.2015	Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift	http:// videoLectures.net/icml2015_ioffe_batch_normalization	106
5.12.2015	Welcome address	http:// videoLectures.net/icml2015_welcome_address	104
10.02.2016	Opening Ceremony	http:// videoLectures.net/iccv2015_opening_ceremony	103
10.03.2015	NLPGo	http:// videoLectures.net/cmuseminars_kambadur_nlpgo	103
29.10.2015	Data analytics involving text	http:// videoLectures.net/single_mladenic_data_analytics	103
27.09.2015	Approval Voting and Incentives in Crowdsourcing	http:// videoLectures.net/icml2015_shah_crowdsourcing	101
27.09.2015	A Relative Exponential Weighing Algorithm for Adversarial Utility-based Dueling Bandits	http:// videoLectures.net/icml2015_gajane_dueling_bandits	101
5.12.2015	Is Feature Selection Secure against Training Data Poisoning?	http:// videoLectures.net/icml2015_biggio_feature_selection	101
5.12.2015	Computational Social Science	http:// videoLectures.net/icml2015_wallach_social_science	99
15.07.2015	Assigning Semantic Labels to Data Sources	http:// videoLectures.net/eswc2015_krishnamurthy_data_sources	96

28.07.2015	What limits performance in decision making?	http:// videoLectures.net/rldm2015_pouget_decision_making	94
10.02.2016	Semantic Image Segmentation via Deep Parsing Network	http:// videoLectures.net/iccv2015_liu_li_image_segmentation	93
27.09.2015	Support Matrix Machines	http:// videoLectures.net/icml2015_luo_support_matrix_machines	91
27.09.2015	Large-Scale Markov Decision Problems with KL Control Cost and its Application to Crowdsourcing	http:// videoLectures.net/icml2015_malek_crowdsourcing	91
5.12.2015	Modeling Order in Neural Word Embeddings at Scale	http:// videoLectures.net/icml2015_gilmore_trask_modeling_order	89
6.03.2015	Cross-lingual Global Media Monitoring	http:// videoLectures.net/cmuseminars_mladenic_media_monitoring	88
9.12.2015	Deep Edge-Aware Filters	http:// videoLectures.net/icml2015_ren_deep_filters	88
27.09.2015	Training Deep Convolutional Neural Networks to Play Go	http:// videoLectures.net/icml2015_clark_deep_convolutional_neural_networks	87
27.09.2015	Accelerated Online Low Rank Tensor Learning for Multivariate Spatiotemporal Streams	http:// videoLectures.net/icml2015_yu_multivariate_spatiotemporal_streams	86
10.11.2015	Build it, and they will come: Applications of semantic technology	http:// videoLectures.net/iswc2015_horrocks_semantic_technology	86

2.07.2015	Perinasal Indicators of Deceptive Behavior	http:// videoLectures.net/fgconference2015 _dcosta_deceptive_behavior	86
10.02.2016	Discovering the Spatial Extent of Relative Attributes	http:// videoLectures.net/iccv2015_xiao_rel ative_attributes	80
10.02.2016	Learning Discriminative Reconstructions for Unsupervised Outlier Removal	http:// videoLectures.net/iccv2015_xia_outli er_removal	80
27.09.2015	Deep Unsupervised Learning using Nonequilibrium Thermodynamics	http:// videoLectures.net/icml2015_sohl_dic kstein_deep_unsupervised_learning	79
27.09.2015	An embarrassingly simple approach to zero-shot learning	http:// videoLectures.net/icml2015_romera _paredes_zero_shot_learning	79
27.09.2015	Multi-instance multi-label learning in the presence of novel class instances	http:// videoLectures.net/icml2015_raich_n ovel_class_instances	78
5.12.2015	Policy Search: Methods and Applications	http:// videoLectures.net/icml2015_neuman n_peters_policy_search	78
15.07.2015	Why Big Data Matters - a lot	http:// videoLectures.net/eswc2015_mayer_ schoenberger_big_data	78
8.05.2015	Ali bo računalnik po sposobnosti prehitel človeške možgane?	http:// videoLectures.net/sinapsa_repovs_m ladenic_racunalnik_ali_clovek	76
28.07.2015	Bootstrapping Skills	http:// videoLectures.net/rldm2015_manko witz_bootstrapping_skills	75

28.07.2015	The Online Discovery Problem and Its Application to Lifelong Reinforcement Learning	http:// videoLectures.net/rldm2015_li_discovery_problem	72
15.07.2015	Combining Statistics and Semantics to Turn Data into Knowledge	http:// videoLectures.net/eswc2015_getoor_turn_data	72
24.04.2015	Računalniška analiza velikih socialnih omrežij	http:// videoLectures.net/kolokviji_leskovec_analiza_socialnih_omrezij	71
22.10.2015	Introduction to the summer school	http:// videoLectures.net/eswc2015_simperl_introduction	70
10.02.2016	Holistically-Nested Edge Detection	http:// videoLectures.net/iccv2015_xie_edge_detection	70
9.12.2015	Jezikovna opremljenost slovenščine	http:// videoLectures.net/single_krek_jezikovna_opremljenost_slovenscine	67
10.02.2016	Deep Fried Convnets	http:// videoLectures.net/iccv2015_yang_fried_convnets	64
2.07.2015	Realistic Inverse Lighting from a Single 2D Image of a Face, Taken Under Unknown and Complex Lighting	http:// videoLectures.net/fgconference2015_shahlaei_complex_lighting	64
5.12.2015	An Empirical Exploration of Recurrent Network Architectures	http:// videoLectures.net/icml2015_jozefowicz_network_architectures	63

10.02.2016	Leave-One-Out Kernel Optimization for Shadow Detection	http:// videoLectures.net/iccv2015_yago_vicente_kernel_optimization	63
27.09.2015	Complete Dictionary Recovery Using Nonconvex Optimization	http:// videoLectures.net/icml2015_wright_nonconvex_optimization	63
28.07.2015	Reinforcement learning objectives constrain the cognitive map	http:// videoLectures.net/rldm2015_stachenfeld_cognitive_map	60
27.09.2015	Learning Deep Structured Models	http:// videoLectures.net/icml2015_chen_deep_structured_models	60
2.10.2015	Predstavitev projekta "Pravo v dobi velikih podatkov"	http:// videoLectures.net/okroglamizapravo2015_zavrsnik_pravo_veliki_podatki	58
27.09.2015	Functional Subspace Clustering with Application to Time Series	http:// videoLectures.net/icml2015_liu_functional_subspace_clustering	58
2.10.2015	Zakaj se sodniki motijo	http:// videoLectures.net/okroglamizapravo2015_leskovec_sodniki	58
27.09.2015	Improving the Gaussian Process Sparse Spectrum Approximation by Representing Uncertainty in Frequency Inputs	http:// videoLectures.net/icml2015_gal_frequency_inputs	58
2.10.2015	Okrogla miza "Ali je lahko računalnik boljši od sodnika"	http:// videoLectures.net/okroglamizapravo2015_racunalnik_sodnik	57

5.12.2015	From Word Embeddings To Document Distances	http:// videoLectures.net/icml2015_kusner_ document_distances	57
27.09.2015	Deterministic Independent Component Analysis	http:// videoLectures.net/icml2015_szepesv ari_component_analysis	54
10.02.2016	Where to Buy It: Matching Street Clothing Photos in Online Shops	http:// videoLectures.net/iccv2015_berg_str eet_clothing	53
15.07.2015	Linked Data-as-a-Service: The Semantic Web Redeployed	http:// videoLectures.net/eswc2015_beek_s emantic_web	53
28.07.2015	Practical RL: Representation, interaction, synthesis, and morality (PRISM)	http:// videoLectures.net/rldm2015_stone_ practical_rl	53
10.02.2016	On the Visibility of Point Clouds	http:// videoLectures.net/iccv2015_tal_poin t_clouds	53
27.09.2015	Consistent estimation of dynamic and multi-layer block models	http:// videoLectures.net/icml2015_xu_bloc k_models	52
5.12.2015	Social Interaction in Global Networks	http:// videoLectures.net/icml2015_kleinbe rg_social_interaction	52
10.02.2016	Mutual-Structure for Joint Filtering	http:// videoLectures.net/iccv2015_shen_joi nt_filtering	52

10.02.2016	Webly Supervised Learning of Convolutional Networks	http:// videoLectures.net/iccv2015_chen_supervised_learning	52
1.12.2015	securePART project with Alexandre Almeida	http:// videoLectures.net/esr2015_almeida_securepart_project	51
11.01.2016	Representation and Reasoning with Universal Schema Embeddings	http:// videoLectures.net/iswc2015_mccallum_universal_schema	51
10.11.2015	Networks of Linked Data Eddies: An Adaptive Web Query Processing Engine for RDF Data	http:// videoLectures.net/iswc2015_acosta_linked_data	51
15.07.2015	Low-cost Open Data As-a-Service in the Cloud	http:// videoLectures.net/eswc2015_dimitrov_open_data	51
15.07.2015	A Comparison of Data Structures to Manage URIs on the Web of Data	http:// videoLectures.net/eswc2015_mavlyutov_web_data	50
27.09.2015	Fixed-point algorithms for learning determinantal point processes	http:// videoLectures.net/icml2015_mariet_determinantal_point_processes	50
10.02.2016	Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books	http:// videoLectures.net/iccv2015_zhu_aligning_books	49

27.09.2015	Spectral MLE: Top-K Rank Aggregation from Pairwise Comparisons	http:// videoLectures.net/icml2015_suh_spectral_mle	49
27.09.2015	A Multitask Point Process Predictive Model	http:// videoLectures.net/icml2015_lian_predictive_model	49
10.02.2016	Structured Indoor Modeling	http:// videoLectures.net/iccv2015_ikehata_indoor_modeling	49
27.09.2015	Asymmetric Transfer Learning with Deep Gaussian Processes	http:// videoLectures.net/icml2015_kandemir_asymmetric_transfer_learning	48
27.09.2015	Multi-Task Learning for Subspace Segmentation	http:// videoLectures.net/icml2015_wang_subspace_segmentation	48
27.09.2015	The Ladder: A Reliable Leaderboard for Machine Learning Competitions	http:// videoLectures.net/icml2015_hardt_reliable_leaderboard	47
27.09.2015	Hidden Markov Anomaly Detection	http:// videoLectures.net/icml2015_goernitz_anomaly_detection	47
27.09.2015	Algorithms for the Hard Pre-Image Problem of String Kernels and the General Problem of String Prediction	http:// videoLectures.net/icml2015_rolland_string_prediction	47

Most popular video lecture in the data science category since the first publications in 2007.

Table 16: Top Data Science Videolectures Viewing Statistics (Data Science VideoLectures published in 2007-2016)

Publishing Date	Lecture Title	Lecture URL	Views
2.07.2007	Basics of probability and statistics	http:// videoLectures.net/bootcamp07_keller_bss	69891
25.02.2007	Machine Learning, Probability and Graphical Models	http:// videoLectures.net/mlss06tw_roweis_mlpgm	42814
2.11.2009	Markov Chain Monte Carlo	http:// videoLectures.net/mlss09uk_murray_mcmc	41402
25.02.2007	Gaussian Process Basics	http:// videoLectures.net/gpip06_mackay_gpb	39970
2.11.2009	Topic Models	http:// videoLectures.net/mlss09uk_blei_tm	34234
15.09.2009	A tutorial on Deep Learning	http:// videoLectures.net/jul09_hinton_deeplearn	32694
13.03.2008	Monte Carlo Simulation for Statistical Inference, Model Selection and Decision Making	http:// videoLectures.net/mlss08au_freitas_asm	28047
5.02.2008	Introduction to Support Vector Machines	http:// videoLectures.net/epsrcws08_campbell_isvm	23330
25.02.2007	Semisupervised Learning Approaches	http:// videoLectures.net/mlas06_mitchell_sla	16655
25.02.2007	Dirichlet Processes, Chinese Restaurant Processes, and all that	http:// videoLectures.net/icml05_jordan_dpcrp	16185
5.08.2010	Introduction to Machine Learning	http:// videoLectures.net/bootcamp2010_murray_iml	14814

12.03.2009	Challenges in Building Large-Scale Information Retrieval Systems	http:// videoLectures.net/wsdm09_dean_cblirs	14723
2.11.2009	Information Theory	http:// videoLectures.net/mlss09uk_mackay_it	14328
2.07.2007	Introduction to Machine Learning	http:// videoLectures.net/bootcamp07_guyon_itml	14197
4.07.2012	Big-Data Tutorial	http:// videoLectures.net/eswc2012_grobelnik_big_data	13507
2.11.2009	Deep Belief Networks	http:// videoLectures.net/mlss09uk_hinton_dbn	12872
25.02.2007	Generative Models for Visual Objects and Object Recognition via Bayesian Inference	http:// videoLectures.net/mlas06_li_gmvoo	12718
30.07.2009	An Overview of Compressed Sensing and Sparse Signal Recovery via L1 Minimization	http:// videoLectures.net/mlss09us_candes_ocsssrl1m	12702
25.02.2007	Text Classification	http:// videoLectures.net/mlas06_cohen_tc	11664
2.11.2009	Particle Filters	http:// videoLectures.net/mlss09uk_godsill_pf	11454
9.09.2011	Social Media Analytics	http:// videoLectures.net/single_leskovec_social	11013
20.08.2007	Introduction to bioinformatics	http:// videoLectures.net/mlss07_gunnar_intbio	9937
25.02.2007	A short Tutorial on Semantic Web	http:// videoLectures.net/training06_sure_stsw	9684

24.11.2008	How to Publish Linked Data on the Web	http:// videoLectures.net/iswc08_heath_hpldw	9580
12.01.2011	Optimization Algorithms in Machine Learning	http:// videoLectures.net/nips2010_wright_oaml	8864
12.01.2011	How to Grow a Mind: Statistics, Structure and Abstraction	http:// videoLectures.net/nips2010_tenenbaum_hgm	8681
25.02.2007	Učenje povzemanja besedil s pretvorbo v semantiÄŤno mreĹľo	http:// videoLectures.net/single_leskovec_diploma	8482
20.12.2008	Matplotlib	http:// videoLectures.net/mloss08_hunter_mat	8440
3.12.2007	Statistical techniques for fraud detection, prevention, and evaluation	http:// videoLectures.net/mmdss07_hand_stf	8295
3.09.2010	Introduction to Statistics	http:// videoLectures.net/cernstudentsummerschool09_cowan_is	8251
24.11.2008	Introduction to the Semantic Web	http:// videoLectures.net/iswc08_hendler_ittsw	8181
2.11.2009	Approximate Inference	http:// videoLectures.net/mlss09uk_minka_ai	7918
25.02.2007	Some Mathematical Tools for Machine Learning	http:// videoLectures.net/mlss03_burges_smtml	7889
7.08.2009	Multiple regression analysis	http:// videoLectures.net/ssmt09_kittel_mra	7887
3.03.2008	Group Theory and Machine Learning	http:// videoLectures.net/mlcued08_kondor_gtm	7785
5.05.2008	Learning in Computer Vision	http:// videoLectures.net/mlss08au_lucey_linv	7563

5.08.2010	Probability and Mathematical Needs	http:// videoLectures.net/bootcamp2010_anthoine_pmn	7365
25.02.2007	Text Information Extraction	http:// videoLectures.net/mlas06_nigam_tie	7138
28.01.2008	Mining Large Graphs: Laws and Tools	http:// videoLectures.net/ecml07_leskovec_mlg	6968
27.08.2007	Topics in image and video processing	http:// videoLectures.net/mlss07_blake_tiivp	6899
11.10.2010	The Importance of Reproducible Research in High-Throughput Biology: Case Studies in Forensic Bioinformatics	http:// videoLectures.net/cancerbioinformatics2010_baggerly_irrh	6895
25.01.2012	Alternating Direction Method of Multipliers	http:// videoLectures.net/nipsworkshops2011_boyd_multipliers	6665
25.02.2007	Computer Vision	http:// videoLectures.net/mlss04_blake_cv	6495
30.07.2009	Geometric Methods and Manifold Learning	http:// videoLectures.net/mlss09us_niyogi_belkin_gmml	6277
25.02.2007	Information Geometry	http:// videoLectures.net/mlss05us_dasgupta_ig	6097
19.01.2010	Sparse Methods for Machine Learning: Theory and Algorithms	http:// videoLectures.net/nips09_bach_smm	6083
9.10.2014	Deep Learning	http:// videoLectures.net/kdd2014_salakhutdinov_deep_learning	5923
26.08.2009	Tutorial on Learning Deep Architectures	http:// videoLectures.net/icml09_bengio_lecun_tldar	5855

1.09.2010	Cancer: A Computational Disease that AI Can Cure	http:// videoLectures.net/aaai2010_tenenbaum_cac	5837
30.08.2007	Sequential Monte Carlo methods	http:// videoLectures.net/mlss07_doucet_smcm	5692
1.04.2009	Computer vision	http:// videoLectures.net/ssll09_hartley_covi	5522
17.08.2012	How to Grow a Mind: Statistics, Structure and Abstraction	http:// videoLectures.net/aaai2012_tenenbaum_grow_mind	5495
26.08.2009	Online Dictionary Learning for Sparse Coding	http:// videoLectures.net/icml09_mairal_odlsc	5315
2.06.2008	15. CB2: Child Robot with Biomimetic Body	http:// videoLectures.net/aaai08_noda_crbb	5130
3.12.2007	Open Source Intelligence	http:// videoLectures.net/mmdss07_best_osi	5125
26.09.2008	Mining Massive RFID, Trajectory, and Traffic Data Sets	http:// videoLectures.net/kdd08_han_mmrfid	5114
25.02.2007	Information Retrieval and Text Mining	http:// videoLectures.net/mlss04_hofmann_irtm	5079
18.03.2011	NLP at Google	http:// videoLectures.net/russir2010_filippova_nlp	4933
25.02.2007	Where the Social Web Meets the Semantic Web	http:// videoLectures.net/iswc06_gruber_wswms	4883
14.08.2007	From Trees to Forests and Rule Sets - A Unified Overview of Ensemble Methods	http:// videoLectures.net/kdd07_elder_seni_fttf	4861

8.10.2007	Learning to align: a statistical approach	http:// videoLectures.net/ida07_ricci_lta	4806
31.03.2011	Data mining and Machine learning algorithms	http:// videoLectures.net/aibootcamp2011_balcazar_dmml	4778
16.01.2013	Quantum information and the Brain	http:// videoLectures.net/nips2012_aaronson_quantum_information	4753
19.01.2010	Understanding Visual Scenes	http:// videoLectures.net/nips09_torralba_uvs	4725
29.08.2007	Machine learning and finance	http:// videoLectures.net/mlss07_gyorfi_mlaf	4679
31.03.2011	Introduction to Machine Learning	http:// videoLectures.net/aibootcamp2011_quinn_iml	4662
25.02.2007	Link analysis with pajek	http:// videoLectures.net/acai05_kejzar_lasn	4568
24.09.2007	EEG Coupling, Granger Causality and Multivariate Autoregressive Models	http:// videoLectures.net/mda07_schloegl_eegc	4566
18.05.2009	Lecture 1 - The Motivation & Applications of Machine Learning	http:// videoLectures.net/stanfordcs229f08_ng_lec01	4553
25.02.2007	Statistical Learning Theory	http:// videoLectures.net/mlss03_bousquet_slt	4482
7.10.2010	Boilerplate Detection Using Shallow Text Features	http:// videoLectures.net/wsdm2010_kohlschutter_bdu	4391
26.08.2009	Convolutional Deep Belief Networks for Scalable Unsupervised	http:// videoLectures.net/icml09_lee_cdb	4379

	Learning of Hierarchical Representations		
12.08.2007	Text Mining and Link Analysis for Web and Semantic Web	http:// videoLectures.net/kdd07_grobelnik_tmala	4369
22.08.2012	Bayesian dynamic modelling	http:// videoLectures.net/isba2012_west_dynamic_modelling	4314
29.11.2007	Structure and Dynamics in Complex Networks	http:// videoLectures.net/eccs07_newman_sdc	4247
10.10.2008	Data Mining for Anomaly Detection	http:// videoLectures.net/ecmlpkdd08_lazarevic_dmfa	4200
22.06.2007	Bayesian models of human inductive learning	http:// videoLectures.net/icml07_tenenbaum_bmhi	4197
14.08.2007	Truth Discovery with Multiple Conflicting Information Providers on the Web	http:// videoLectures.net/kdd07_yin_tdwmci	4196
19.01.2010	Deep Learning in Natural Language Processing	http:// videoLectures.net/nips09_collobert_weston_dlnl	4192
14.08.2007	A Data Minerâ€™s Story â€“ Getting to Know the Grand Challenges	http:// videoLectures.net/kdd07_fayyad_dms	4170
25.02.2007	Latent Semantic Variable Models	http:// videoLectures.net/slsfs05_hofmann_lsvm	4103
1.10.2010	Large-scale Data Mining: MapReduce and Beyond	http:// videoLectures.net/kdd2010_papadimitriou_sun_yan_lsdm	4102

12.07.2010	Scikitlearn	http:// videoLectures.net/icml2010_varaquaux_scik	4049
30.07.2009	Matrix Completion via Convex Optimization: Theory and Algorithms	http:// videoLectures.net/mlss09us_candes_mccota	3865
19.07.2010	Detecting Text in Natural Scenes with Stroke Width Transform	http:// videoLectures.net/cvpr2010_epshtein_dtns	3856
22.11.2007	Dynamics of Real-world Networks	http:// videoLectures.net/thesis_leskovec_drn	3853
26.09.2008	The Future of Image Search	http:// videoLectures.net/kdd08_malik_fis	3845
17.10.2007	Text and web data mining	http:// videoLectures.net/ess07_grobelnik_twdmI	3794
1.04.2009	Computability And Incompleteness	http:// videoLectures.net/ssll09_martin_cai	3701
15.11.2007	POWERSET - Natural Language and the Semantic Web	http:// videoLectures.net/iswc07_pell_nlpsw	3678
5.12.2008	Sparse Geometric Super-Resolution	http:// videoLectures.net/etvc08_mallat_sgsr	3646
4.11.2008	Content Based Image Retrieval (CBIR)	http:// videoLectures.net/russir08_vassilieva_cbir	3623
13.05.2013	What is Machine Learning?	http:// videoLectures.net/mlss2012_lawrence_machine_learning	3575
20.03.2007	Data Mining Vs. Semantic Web	http:// videoLectures.net/solomon_milutinovic_dmv	3551

4.02.2013	Lecture 1: Object-Oriented Programming	http:// videoLectures.net/mit601s201_freeman_lec01	3523
14.05.2007	Python - jezik za nove čase	http:// videoLectures.net/rtk07_tori_p	3487
5.07.2007	Other ML/DM software (R, Weka, Yale)	http:// videoLectures.net/bootcamp07_belanche_mldm	3483
25.02.2007	Which Supervised Learning Method Works Best for What? An Empirical Comparison of Learning Methods and Metrics	http:// videoLectures.net/solomon_caruana_wslmw	3480
9.08.2013	Learning Representations: A Challenge for Learning Theory	http:// videoLectures.net/colt2013_lecun_theory	3414

Appendix 2: Reference Sectors

EUROSTAT reference sectors are mapped to a smaller set of sectors in the project.

Table 8. Mapping of Sectors

Eurostat reference	Breakdown of reference	Adjusted survey sectors
A AGRICULTURE, FORESTRY AND FISHING Detail	Agriculture	Agriculture
B MINING AND QUARRYING Detail	Mining	Mining
C MANUFACTURING Detail	Manufacturing	Manufacturing
D ELECTRICITY, GAS, STEAM AND AIR CONDITIONING SUPPLY Detail	Energy (including electricity, gas and steam)	Energy
E WATER SUPPLY; SEWERAGE, WASTE MANAGEMENT AND REMEDIATION ACTIVITIES Detail	Water and waste management	Water and waste management
F CONSTRUCTION Detail	Construction	Construction
G WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES Detail	Wholesale and retail trade	Wholesale and retail
	Automotive repair	
H TRANSPORTATION AND STORAGE Detail	Transportation and support services including postal service	Transport
I ACCOMMODATION AND FOOD SERVICE ACTIVITIES Detail	Accommodation and food services	Accommodation and food services
J INFORMATION AND COMMUNICATION Detail	Publishing	Media
	Film, video and television production	
	Broadcast	
	News agency	
	Telecommunications	Data and information systems
	Computer programming	
	Information services including web services and data processing	
K FINANCIAL AND INSURANCE ACTIVITIES Detail	Financial and Insurance services	Finance and insurance services
L REAL ESTATE ACTIVITIES Detail	Real Estate	Real Estate

M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES Detail	Legal and accounting	Professional services
	Management consultancy	
	Architecture and engineering	
	Veterinary	
	Other including specialist design, translation and photography	
	Scientific research	Scientific and market research
	Advertising and market research	
N ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES Detail	Rental and leasing	Business administration services
	Employment	
	Security and investigation	
	Building and landscape services	
	Office administration	
	Travel	Tourism
O PUBLIC ADMINISTRATION AND DEFENCE; COMPULSORY SOCIAL SECURITY Detail	Public administration and defence	Public administration and defence
		Government and public sector
P EDUCATION Detail	Education	Education
Q HUMAN HEALTH AND SOCIAL WORK ACTIVITIES Detail	Human health and social work	Human health and social work
R ARTS, ENTERTAINMENT AND RECREATION Detail	Creative arts and entertainment	Arts, recreation and entertainment
	Libraries, archives and museums	
	Gambling	
	Sports activities, amusement and recreational activities	
S OTHER SERVICE ACTIVITIES Detail	Membership organisations	Consumer services
	Computer, personal and household goods repair	
	Dry cleaning	
	Hair and beauty	

	Funeral and related services	
	Physical and well-being activities	
	Other personal service activities including pet grooming, spiritualist services	
T ACTIVITIES OF HOUSEHOLDS AS EMPLOYERS; UNDIFFERENTIATED GOODS- AND SERVICES-PRODUCING ACTIVITIES OF HOUSEHOLDS FOR OWN USE Detail	Household activities	
U ACTIVITIES OF EXTRATERRITORIAL ORGANISATIONS AND BODIES Detail	Extraterritorial organisations and bodies	