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

D3.1 Report on the delivery of video-lectures, webinars and face-to-face trainings

Deliverable Editor:	**Angi Voss (IAIS)**
Other contributors:	**OU, SOTON, JSI, ODI, TU/e**
Deliverable Reviewers:	**Inna Novalija (JSI)**
	Patrick Mukala (TUe)
Deliverable due date:	**31/07/2015**
Submission date:	**31/07/2015**
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	03/06/2015	A.Voss, IAIS	Initial version
0.2	16/06/2015	A.Voss, IAIS	Updates after project meeting
0.3	22/06/2015	Alexander Mikroyannidis, OU	Updates regarding the Summer School
0.4	23/06/2015	A.Voss, IAIS	Evaluation added to 3.3, introduction and recommendations rewritten
0.5	08/07/2015	A.Voss, IAIS	Consolidated version after section on video lectures was filled by JSI
0.6	20/07/2015	A.Voss, IAIS	Changes due to the reviews
0.7	27/07/2015	A.Voss, IAIS	Changes due to final review
0.8	30/07/2015	A.Voss, IAIS	Update of the video table
1.0	31/07/2015	Aneta Tumilowicz	Final QA

Table of Contents

List of Tables

1. Executive Summary

This deliverable reports on the courses delivered in WP3. 32 video lectures, one MOOC and 27 face-to-face trainings related to the project have been provided by the partners in Tasks 3.1-3.3 since the start of the project.

These courses constitute candidate content for WP2. In the next 12 months the contribution of this material to the curriculum will become clearer. Right now, the comparison in this report has started a discussion whether open data should be added as another subject area.

Also in the next 12 months, feedback from the delivered trainings will be collected and analysed. Then the follow-up deliverable D3.2 will be able to give recommendations to the courses delivered and the content from WP2. Right now, the views of the video lectures indicate a high interest in big data analytics and applications. The latter might be candidates for sector-specific learning materials.

The learning analytics will also tell us how to distinguish learners who dropped out from those who completed a training. Right now, the number of participants is on a good track compared with the target numbers planned.

2. 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, video lectures and face-to-face training.
2. Gather feedback data about the effectiveness of learning from the courses.
3. Analyse this data and obtain indications of how to improve the curricula, content, form of deployment and

WP3 produces two series of deliverables. While the first series addresses objectives 1 and 3 above, the second one addresses objective 2:

1. D3.1 (month 6), D3.2 (month 18) and D3.4 (month 36) shall give an overview of the courses delivered and give recommendations for the curricula, learning resources, and future delivery.
2. D3.3 (month 18) and D3.5 (month 36) contain the results of the analysis of feedback data obtained from the delivered courses.

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

The delivered courses may contribute content to WP2. For that purpose, they are associated with target groups and subject areas from the curriculum of WP2. The delivery of courses will also produce feedback data for T3.4 and D3.3 in month 18. Since now, in month 6, there are not yet any learning analytics, this deliverable cannot contain any conclusions drawn on their basis.

Consistently, this report will focus on objective 1 and report on the trainings delivered by the partners in the project so far.

3. 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 video lectures of T3.1, eLearning (active) corresponds to the MOOCs and webinars of T3.2, while face-to-face trainings are covered by Task 3.3.

The next table compares the target figures with the number of persons reached so far. The number of views of the video lectures was obtained within the last 2,5 months.

Table 1. Target audience figures

Activity	Overall target	Audience reached in the project
eLearning (passive)	200,000	373
eLearning (active)	50,000	24,558 registered students
Face-to-face-training	2,500	356 participants

The next three subsections report on the courses delivered in the three formats. They contain tables with the following structure: Title of course, provider or author, date & length, number of participants (or views in the case of videos), subject area and target group.

The subject areas proposed for the curriculum in WP2 are:

- foundations of data science
- foundations of big data
- statistical/mathematical foundations
- programming & computational thinking
- data management & curation
- big data architecture
- distributed computing
- stream processing
- data analytics (machine learning / data mining / basic analytics)
- big data analytics
- process mining
- data visualization
- finding stories in open data

Comparing these subject areas with the topics of the courses delivered and reported below, the following can observed:

- Two courses are about open data, which could be included as an additional subject area in WP2.
- Several video lectures are about applications in industry. They might be candidates for sector-specific courses.
- One face-to-face course about the business potentials of big data analytics might be related to foundations of big data and big data science.

- No courses have been delivered on programming & computational thinking, distributed computing and stream processing. Here courses from other providers should be identified for WP2. However, distributed computing and stream processing could be covered by a partner in the project who is not involved in WP3.

The target groups identified in WP1 are:

- managers, product developers and business experts
- data-skilled persons: data managers, curators and data engineers
- data analysts
- system architects and application developers

Most courses delivered are suitable for all sectors. Only one course explicitly addresses media and advertising. However, the video lectures about applications in industry might be candidates for sector-specific learning materials. Candidates sectors are:

- aerospace and defence
- agriculture
- automotive
- consumer services
- construction and engineering
- consultancy
- data and information systems
- energy
- finance, insurance, real estate
- government and public sector
- health
- manufacturing
- media and advertising
- mining
- professional services
- telecoms
- transport
- any

In the next deliverable D3.2 the relation between the courses delivered and the content to be produced in WP2 can be made more specific.

3.1 Video lectures at JSI

VideoLectures.NET is an award-winning free and open access educational video lectures repository. The lectures are given by distinguished scholars and scientists at the most important and prominent events like conferences, summer schools, workshops and science promotional events from many fields of Science. The portal is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high quality didactic contents not only to the scientific community but also to the general public. All lectures, accompanying documents, information and links are systematically selected and classified through the editorial process taking into account also users' comments.

The portal presently offers over 20k of educational videos distributed through 579 categories. Several categories fall also into data science context, such as big data, computational linguistics, data mining, data modelling, data visualization, information extraction, knowledge extraction, machine learning etc.

covered by thousands of videos from most prominent EU and Worldwide conferences. The computer science category covers more than 55% of all videos presently hosted on the portal.

Since the start of the project videos from the 12th ESWC Conference in Portoroz 2015 have been provided on the platform (http://videolectures.net/eswc2015_portoroz/). 32 videos relevant to data science are listed in the next table. Except for the videos on applications in industry, they are neither sector-specific nor do they address a particular target group, The total number of views between May 15 and July 30 is 373.

Table 2. EDSA related video lectures from ESWC 2015

Title	Authors	Date; length	Views	Subject area (categories)
Why big data matters – a lot	Viktor Mayer-Schönberger	15/7/15; 1:12:13	35	Big data
Combining Statistics and Semantics to Turn Data into Knowledge	Lise Getoor	15/7/15; 1:15:55	13	Data management (semantics)
What Crowdsourcing Tells Us about Cognition: the Case of Anaphora	Massimo Poesio	15/7/15; 0:54:44	6	Data analytics (machine learning)
Large Scale Rule-Based Reasoning Using a Laptop	Martin Peters	15/7/15; 0:20:16	10	data analytics
A Context-Based Semantics for SPARQL Property Paths over the Web	Olaf Hartig	15/7/15; 0:14:21	4	data analytics
Distributed and Scalable OWL EL Reasoning	Raghava Mutharaju	15/7/15; 0:16:48	7	data analytics
Linked Data-as-a-Service: The Semantic Web Redeployed	Wouter Beek,	15/7/15; 0:24:59	28	Application (Industry)
Standardized and Efficient RDF Encoding for Constrained Embedded Networks	Sebastian Käbisch	15/7/15; 0:31:01	1	Application (Industry)
Processing Aggregate Queries in a Federation of SPARQL Endpoints	Dilshod Ibragimov,	15/7/15; 0:17:21	4	data management, big data (semantics)
Cooperative Techniques for SPARQL Query Relaxation in RDF Databases	Géraud Fokou,	15/7/15; 0:19:36	6	data management, big data (semantics)

Query Execution Optimization for Clients of Triple Pattern Fragments	Joachim Van Herwegen,	15/7/15; 0:19:27	15	data management, big data (semantics)
From Symptoms to Diseases - Creating the Missing Link	Heiner Oberkampf,	15/7/15; 0:19:20	4	Application (Industry)
Crowdmapping Digital Social Innovation with Linked Data	Harry Reeves Halpin,	15/7/15; 0:21:45	6	Application (Industry)
Supporting Open Collaboration in Science through Explicit and Linked Semantic Description of Processes	Felix Michel,	15/7/15; 0:20:31	6	Application (Industry)
Automating RDF Dataset Transformation and Enrichment	Mohamed Ahmed Sherif	15/7/15; 0:16:22	16	data analytics (machine learning)
Assigning Semantic Labels to Data Sources	Ramnandan Krishnamurthy,	15/7/15; 0:18:20	4	data analytics (machine learning)
Inductive Classification Through Evidence-Based Models and Their Ensembles	Giuseppe Rizzo	15/7/15; 0:18:36	8	data analytics (machine learning)
HDT-MR: A Scalable Solution for RDF Compression with HDT and MapReduce	José M. Giménez-García	15/7/15; 0:23:30	2	data management, big data (semantics)
A Survey of HTTP Caching Implementations on the Open Semantic Web	Kjetil Kjernsmo	15/7/15; 0:30:31	22	data management, big data (semantics)
A Compact In-Memory Dictionary for RDF data	Hamid R. Bazoobandi	15/7/15; 0:29:23	1	data management, big data (semantics)

Semi-supervised Instance Matching Using Boosted Classifiers	Mayank Kejriwal	15/7/15; 0:21:40	1	data management (ontology matching & summarization)
Requirements for and Evaluation of User Support for Large-Scale Ontology Alignment	Valentina Ivanova	15/7/15; 0:22:25	4	data management (ontology matching & summarization)
RDF Digest: Efficient Summarization of RDF/S KBs	Georgia Troullinou	15/7/15; 0:20:18	15	data management (ontology matching & summarization)
Troubleshooting and Optimizing Named Entity Resolution Systems in the Industry	Panos Alexopoulos	15/7/15; 0:19:25	5	Application (Industry)
Using Ontologies For Modeling Virtual Reality Scenarios	Mauro Dragoni	15/7/15; 0:20:05	18	Application (Industry)
Using Semantic Web Technologies for Enterprise Architecture Analysis	Maximilian Osenberg	15/7/15; 0:18:26	7	Application (Industry)
PADTUN - Using Semantic Technologies in Tunnel Diagnosis and Maintenance Domain	Vanja Dimitrova	15/7/15; 0:26:24	4	Application (Industry)
Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing	Eugene Cherny	15/7/15; 0:22:30	3	Application (Industry)
Desperately Searching for Travel Offers? Formulate Better Queries with Some Help from Linked Data	Chun Lu	15/7/15; 0:33:21	76	Application (Industry)
Heuristics for Fixing Common Errors in Deployed schema.org Microdata	Robert Meusel,	15/7/15; 0:19:14	6	data management (linked data)
A Comparison of Data Structures to Manage URIs on the Web of Data	Ruslan Mavlyutov	15/7/15; 19:03	26	data management (linked data)

Quality Assessment of Linked Datasets using Probabilistic Approximations	Jeremy Debattista	15/7/15; 20:37	10	data management (linked data
		In total 32	**In total 374**	

Many more lectures and tutorials related to data science can be found at Videolectures.NET that were recorded earlier. In particular, for the big data topic there are 6 events, 141 lectures, 32 keynotes. For the machine learning category 106 events, 2218 lectures, 158 tutorials, 283 keynotes, 2 interviews are listed and for the data mining category 26 events, 576 lectures, 38 tutorials, 71 keynotes, 1 interview are present.

3.2 Webinars and MOOCs

Rather than webinars, the partners have MOOCs for active online learning. With 24,558 registered persons the reported MOOC for process mining already constitutes a good step towards the target audience of 50,000 at the end of the project. Future learning analytics will show how many of the registered users can be counted as serious participants

Table 3. Webinars and MOOCs

Title	Partner	Date, length	Number of participants	Subject area	Target group	Sector
Process Mining: Data Science in Action	TU/e	3/4/15 - 6/6/15. 8 weeks	24,558 registered students	Process Mining	non-specific	non-specific

The course “Process Mining: Data science in Action” 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.

3.3 Face-to-face trainings

The 27 face-to-face trainings reported in the next table have reached 356 persons from the planned target of 2,500 at the end of the project. The courses do not target specific sectors, except for the one in the last row on multimedia analytics, which addresses the sector “media and advertising”.

Table 4. Face-to-face trainings

Title	Partner	Length, dates	Number of partici-pants	Subject area	Target group
Finding stories in open data	ODI	1 day: 27/1/15, 10/3/15	12	Data visualisation	non-specific
Open data in practice	ODI	3 days: 3-5/2/15, 12-14/3/15	20	Open data	non-specific
Data anonymisation workshop	ODI	0.5 days: 12/2/15, 26/2/15, 9/3/15, 25/3/15	72	Open data	non-specific
5th ESWC Summer School	OU	31/08 - 05/09/2015	Up to 50	Data management and curation	Data-skilled persons
Big Data analytics: QMiner AILab training	JSI	04/02/2015	20	Big data analytics	computer scientists, data analysts
Basic Analytics	IAIS	2 days: 23-24/2/15, 18-19/2/15, 26-27/5/15, 1-2/6/16, 10-11/6/15	54	Data analytics	data analysts
Big data architecture	IAIS	2 days: 25-26/2/15, 4-5/3/15, 9-10/3/15 20-21/5/15, 8-9/6/15,	61	Big data architecture	System architects and application developers

		29-30/6/15,			
Big data analytics	IAIS	1 day: 27/2715, 22/5/15, 12/6/15, 6/3/15, 1/7/15,	48	Big data analytics	data analysts and application developers
Linked enterprise data integration	IAIS	1 day: 16/4/15,	4	Data management	data engineers
Business potentials of big data analytics	IAIS	2 days: 29-30/4/14, 1-2/6/15, 22-23/6/15	30	Foundations of big data and data science?	business experts, managers
Multimedia analytics	IAIS	2 days: 15-16/6/15,	5		data analysts
In total: 27			**In total: 356**		

ESWC Summer School

The OU will organise this year's ESWC Summer School in Crete from August 31st to September 5th 2015. The overall goal for this event is to provide intensive training and networking opportunities to data-skilled researchers and professionals. Specifically for EDSA, the main outcome of the event will be testing the project's learning materials and exercises and gathering feedback from the participants of the event. The ESWC Summer School will be therefore used by EDSA primarily as a testing channel for the project's curricula, instead of a delivery channel.

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.

Courses at Fraunhofer

Fraunhofer IAIS offers different face-to-face seminars on data science and big data targeting managers, data analysts and engineers of big data systems. Each seminar is limited to 10 participants and has two trainers who are practicing experts in the subject. This allows us to respond to individual needs in the group. Seminars are also given inhouse. The courses listed above have been held since two years. In 2015 - 2016 they will be extended stepwise to a comprehensive data scientist training programme with contributions from several other institutes of Fraunhofer's Big Data Alliance. New sector-

independent seminars target managers and data managers and will include aspects of data security and privacy, which is very important in Germany.

Funded by EIT Digital (www.eitdigital.eu/) several sector-specific blended learning modules are being developed in the fields of smart energy systems, smart cities, smart spaces and smart production. Further seminars are being developed for life sciences and health care.
Fraunhofer will offer a personal certification programme for data scientists at three levels – basic, specialist and expert – targeting persons who seek formal qualifications for their career building. The Fraunhofer training programme is being structured accordingly into a basic seminar, specialist seminars for the different target groups, and sector-specific seminars as part of the expert level.

4. Conclusions

At this stage of the project it can be seen that the curriculum in WP2 does not cover all courses delivered in this WP - and vice versa. Notably, a subject area “open data” might to be missing in the curriculum. Most probably it would address the target group of data skilled persons.

Most courses delivered do not focus on specific sectors, whose specific needs still have to be identified in WP1. However, several video lectures about applications in industry are candidates for sector-specific contents.

On the basis of learning feedback and its analysis the follow-up deliverable D3.2 will be able to give recommendations for the courses delivered and beyond them for the content of WP2.