

Project acronym: EDSA

Project full name: **European Data Science Academy**

Grant agreement no: 643937

D5.4 Project Exploitation Report

Deliverable Editor: Philip Horgan (ODI)

Other contributors: Ryan Goodman (ODI)

All partners

Deliverable Reviewers: Aba-Sah Dadzie (OU), Simon Scerri (Fraunhofer IAIS)

Deliverable due date: **31/01/2018**

Submission date: **31/01/2018**

Distribution level: P

Version: 1.0



Change Log

Version	Date	Amended by	Changes
0.1	28/11/2017 Philip	Philip Horgan	Initial structure
0.2	07/12/2017	Philip Horgan	Incorporating partner contributions
0.3	18/12/2017	Philip Horgan	Drafting content on the impact of the project
0.4	07/01/2018	Ryan Goodman	Drafting sustainability of the initiative
0.5	29/01/2018	Philip Horgan	Editing following comments, feedback and further input from partners
0.6	31/01/2018	Philip Horgan	Final review and edit
1.0	31/01/2018	Alexander Mikroyannidis	Final QA

Table of Contents

Cł	nange Log	2
Τa	able of Contents	3
Li	st of Tables	4
Li	st of Figures	4
1.	Executive Summary	5
	Exploitation by project partners and end users	5
	Progress towards project targets	5
	Future Opportunities and Sustainability	5
2.	Exploitation by project partners	6
	2.1 Fraunhofer	6
	2.2 ideXlab	6
	2.3 JSI	7
	2.4 KTH	8
	2.5 ODI	8
	2.6 The Open University (OU)	9
	2.7 University of Southampton	9
	2.8 T/Ue	. 10
	2.9 Persontyle	. 10
	2.10 Training provision	. 11
3.	Exploitation of training courses and EDSA website	. 13
	3.1 Introduction	. 13
	3.2 Exploitation of training courses	. 13
	3.3 Website	
	Exploitation of the dashboard, learning pathways and courses	14
	Website sustainability	15
4.	Demand for skills vs. Supply of training courses	. 15
5.	Project impacts	. 16
6.	Sustainability of the initiative through the online Academy	. 19
7.	Future opportunities	. 19
	References	. 20
8.	Appendices	. 21
	Appendix A Output and Exploitation classifications	. 21
	Project output classifications	21
	Appendix B EDSA Event Participation	. 22

List of Tables

Table 1: Summary of training exploitation	12
Table 2: Summary participants	13
Table 3: Use of website functionality	14
Table 4: Website content views	15
Table 5: Skills in demand, and courses offered	16
List of Figures	_
Figure 1: Joh demand and supply of videolectures	7

1. Executive Summary

This report details three areas of interest, covering: the exploitation of EDSA project outputs by project partners and end-users such as individuals and business, the achievements of the project against impact targets set out in the project grant agreement, and reflections on future opportunities for the project and sustainability of the achievements of EDSA.

Exploitation by project partners and end users

The EDSA project outputs have been exploited by project partners in a variety of ways. Insights from the demand and supply analysis have been used in the development of new curricula and resources by project partners (or have been used to enhance existing courses), such as in the development of a curriculum for a new Master's program "Autonomous System" for students of KTH and EIT Digital Master School. Curricula and Resources developed in the programme have been incorporated into training project partners' home institutions, such as the PhD and MSc programs at Jožef Stefan International (JSI) Postgraduate School. Overall, project partners have developed 30 new or modified courses based on the outputs from the EDSA program, all offered through the EDSA course website.

A large number of individuals have taken part, or viewed, the courses and training materials developed by the project partners, as well as identified and curated content from other institutions. Similarly, the website dashboard and learning pathways has been utilised to investigate the demand for skills within the data science job market and to link to course details relevant to those skills.

Progress towards project targets

Six impact targets were agreed in the project grant agreement, covering the dashboard, curricula, cost and ease of access to data science training, quantity of training delivered and trainees participating, training outcomes and the development of an ecosystem of educational resources and services.

New curricula, courses and resources have been developed including a wealth of multi-modal videolectures and paid for, and free online and offline courses, all accessible through facilitated learning pathways to guide the development of skills across courses.

An effective dashboard has been developed linking automated skills analysis (from job descriptions) and courses offered, curated from online and offline sources.

Training participant numbers have significantly exceeded targets (see above) across the three categories of active and passive e-learners, and face to face participants:

- a target of 200,000 passive e-learners achieved with 270,000 viewers of video learning content;
- a target of 50,000 active e-learners achieved with 116,000 students registered on MOOCs
- a target of 2,500 face to face training participants achieved with 4,780 participants.

Comparison of the skill demand analysis and the curated courses (including those developed specifically for or during the project by EDSA partners) reveals that there are a large number of courses on offer from a range of training providers for the vast majority of priority skills demanded. A small number of topics (for example in the top 20 demanded skills) are covered by less than 10 courses each, but the majority of topics are covered by between 20 and 60 courses each.

Future Opportunities and Sustainability

Some targets, such as the development of strong and sustainable ecosystems can only be achieved in the long term, through for example, participants working through multiple courses on a learning pathway, or the future collaboration with others within the data science training space. The EDSA project has demonstrated its contribution to these impacts, and partners have committed to sustaining the project's online services for one year after the end of the project (see the Memorandum of Understanding).

During this period, it is anticipated that new collaborations and partnerships will be developed to utilise the outputs of the project in building a long term sustainable ecosystem. Partners have begun experimenting with new innovations and developments such as micro accreditation, badging and certification in the data science space and have agreed to co-design a set of open badges for data science

which will sit on a blockchain run from the Big Data Value Ecosystem (BDVE). Moreover, a collection of European and UK initiatives such as the <u>European Association for Data Science</u>, <u>Data Pitch's</u> Data Innovation Academy and the recent UK government announcement on the 'Institute of Coding' see's new opportunities for partners to further support Data Science across Europe.

2. Exploitation by project partners

The following sections details the exploitation of project outputs by each project partner, followed by summary statistics across the project. Partner details reflect plans articulated in D5.8 Individual Exploitation Plans, July 2017. Details cover the project period to January 2018.

2.1 Fraunhofer

Exploitation plans detailed the desire to enhance face to face courses with online material for self-study and to use the demand data collected through EDA for new analysis and business scenarios.

Fraunhofer has enhanced two of their face-to-face courses with EDSA material. The EDSA material is available online for self-study at the EDSA portal. The online learning material developed for EDSA was designed to be introductory and therefore will not require regular (expensive) updates and was produced from the face-to-face courses to ensure consistency. The additionally supplied material serves several purposes: as a motivation to subscribe to a course, as a preparation to read before the course, and for further learning after the course. The exploitation of the online courses for self-study started immediately after they were published in EDSA.

Fraunhofer contributed to the information extraction pipeline for both demand (job postings) and supply (online courses) data relevant to EDSA. The data has been routinely analysed to identify skill demand in the current job market related to the data science domain. In addition to fulfilling the project's needs by providing data that can be explored through the use of the EDSA dashboard, Fraunhofer has exploited the availability of this data to investigate additional analysis, specifically time series analysis, by considering sequential snapshots of demand data over a time period. These efforts are still in progress.

In addition to providing the latest picture of the demand landscape, longitudinal (time series) analysis can expose trends in the sector, thus enabling both educators and trainees to identify emerging skills and groups of skills that are increasingly sought together in candidates. We intend to showcase the time series analysis (in itself not a planned/required project output), in the dashboard following integration (planned for the first quarter of 2018). The same analysis can also be commercialised to provide services that can help employers, educators, and prospective data scientists to identify skill sets that are emerging or increasing in popularity, based on an automated background/long-term study of the demand landscape. We intend to investigate commercialisation options further over the coming years.

2.2 ideXlab

Exploitation plans detailed the desire to couple the existing open innovation with new services based on online learning resources, and, through its contributions to the EDSA project, to engage new clients with specific data science training needs or existing clients with new needs.

ideXlab explored the potential of online learning resources as a natural extension of open innovation. This was done by offering educational resources to innovating companies, to upskill their employees, and to intermediation expert services. While open innovation provides easy access to external expertise, online learning resources provide an efficient means to increase internal expertise. In the project, ideXlab was particularly interested in the relationship between its current open innovation core business (as an intermediary between companies and external experts) and training material. ideXlab tested the markets interest in this coupling between open innovation and online training with a qualitative study of coupling from open innovation / ideXlab clients to learning resources platforms, and from learning resources platforms to the ideXlab platform. The response to the first was quite positive, while the response to the second was less positive, presumably as the user profile and motivation to



visit a learning platform is too far from a typical open innovation user. This investigation is now complete.

Through its inclusion in the EDSA project, ideXlab engaged with new clients or separate organisations within existing clients, with specific needs in data science training. ideXlab reused and promoted whatever dissemination material was made available by the project. The model used in this project is also applicable to other domains in high demand (e.g., biomimetics, additive manufacturing, smart materials, Internet of Things) where course material would lead to developing new business opportunities. This activity contributed to the promotion and dissemination of the EDSA project outputs.

These actions are completed.

2.3 JSI

Exploitation plans detailed the desire to enhance the operability of the VideoLectures.NET portal, extending the PhD and MSc programs at Jožef Stefan International Postgraduate School and to internationalise through multilingual services.

Within the VideoLectures.NET portal, JSI has semi-automatically categorized thousands of videolectures, assigning the relevant videolectures to the newly created Data Science category. At the VideoLectures.NET portal we have created an EDSA project page with a description of the project and recommended videolectures and where viewers interested in Data Science can get information about EDSA. Within the demand analysis task, we have developed a connection between demand analysis and supply of videolectures at the VideoLectures.NET portal.



Figure 1: Job demand and supply of videolectures

Unless otherwise noted, Data Science videolectures presented on VideoLectures.NET portal use Creative Commons Attribution Non-commercial No Derivative Works licenses. JSI applications developed within the EDSA project are based on BSD-licensed tools.

The audience of the VideoLectures.NET portal continues to increase (in particular, we observe around 450 viewers interested in data science coming from the EDSA website following the recommendations of videolectures).

We shall provide maintenance of the achieved results according to the agreement between project partners.

JSI has performed internal training in the area of Data Science (described in Deliverable D3.4), targeted at the Artificial Intelligence Laboratory (Jožef Stefan Institute) staff members and PhD and MSc students

at Jožef Stefan International Postgraduate School. Internal training and EDSA data science materials fill the content gap for existing courses at the Jožef Stefan International Postgraduate School. For instance, the specialized internal trainings in the area of Big data (Big data in finance, Big data in official statistics) provided the attendees the overview of Big data tools and technologies tailored to specific domains. These actions are completed.

The demand and supply analysis applications developed by JSI along with the VideoLectures.NET portal provides multilingual services in the demand side analysis of job advertisements in different European languages, and dissemination in non-English speaking parts of the world. The multilingual demand and supply analysis data obtained through the EDSA demand and supply analysis and connected to VideoLectures.NET portal fills the existing internationalization and dissemination gap. The dashboard itself is provided in English, while job descriptions, search terms, and associated course suggestions are multilingual. The VideoLectures.NET portal provides a set of lectures with subtitles available in different languages.

2.4 KTH

Exploitation plans detail KTHs desire to improve the quality and distribution of KTH's learning material.

KTH provides courses that are part of the master programs "Software Engineering of Distributed Systems" at KTH and "Cloud Computing and Services "and "Data Science" at EIT Digital. The courses (Distributed Systems, Part 1&2, Distributed Artificial Intelligence and Intelligent Agents, Data Intensive Computing Foundations, Scalable Machine Learning and Deep learning, Advanced Topics in Distributed Computing, Programming Web Services) are parts of the "Distributed Computing" and "Data Intensive Computing" modules in the EDSA curricula. All courses (except "Scalable Machine Learning and Deep learning" course) are video recorded and put on EDSA portal and available via YouTube. Three of the existing courses were enhanced by taking into account results of demand and supply analysis.

Learning materials from the enhanced courses are available on the EDSA portal for self-study. A curriculum for a new Master program "Autonomous Systems" was developed using insights from the EDSA demand and supply analysis. In particular, the program specialization "Intelligent Autonomous Systems" accommodates a number of data science courses reported on the EDSA portal. The new Master's program will be offered both for students of KTH and EIT Digital Master School and we expect to continue performing adjustments for the courses under development.

The self-study learning materials available on the EDSA portal serve as advertisements for our courses and programs. They provide an in-depth overview of the educational opportunities available in our courses and increase the visibility of our programs for external students. As a result, we expect an increased number of applicants to our study programs.

2.5 ODI

Exploitation plans detail ODI's desire to improve the content and delivery of training.

ODI enhanced their 'Finding Stories with Data' course with EDSA material to offer a blended opportunity for learners. The course is available online for self-study at the EDSA portal, as well as offering the option to undertake the course in a face-to-face environment. The second course created via EDSA, will look to be available in the same blended fashion. This content will be continually updated to maintain relevance in line with demand.

The ODI's expected impact from this activity is to increase the ROI on data science projects by improving the way that data is understood across all levels of organisations. The ODI generates revenue from training courses in multiple formats, including public and bespoke face-to-face training and via development of customised online material for clients. The ODI also develops free and openly available online content, which leads to participants signing up for paid-for training. New courses developed as a result of the EDSA project will follow the same commercial model. ODI intends to use the new course content, and offer this as a face to face course, in addition to repackaging the online content into customised commercial products. It is anticipated that this will take place on a bespoke basis, based on the demand for this content from public and commercial organisations.



2.6 The Open University (OU)

Exploitation plans detail The Open University's desire to increase visibility of EDSA learning materials.

The OU has made the EDSA learning pathways available via the EDSA dashboard and the EDSA courses portal. The EDSA eBook has also been made available via the EDSA courses portal. There have been regular updates of the EDSA eBook throughout the duration of the project. The eBook and other learning materials, including the learning pathways, will remain available online via both the EDSA dashboard and the EDSA courses portal, as a sustainable result beyond the end of the project. The EDSA learning materials are distributed using Creative Commons licenses, thus allowing the reuse, repurposing and republishing of the materials. Commercial use of the materials may also be allowed depending on the licences used.

Beyond the end of the project, the OU will integrate the developed curricula, educational resources and learning pathways into the courses taught at the OU. The project's educational resources will be reused, repurposed and further developed by existing and emerging OU courses. Following the established tradition of open learning, the resources will be offered via open educational channels, including OpenLearn and iTunes U. Interactive eBooks and courses will be published based on the project's learning resources, delivered through a wide range of platforms, including iOS and Android devices, as well as desktop computers. Suitable materials will be added to our FutureLearn MOOC offerings and project results will also be taken up by periodic summer schools and various seminars organised by OU researchers across the world.

More specifically, the EDSA learning materials will be reused and repurposed in order to teach computer science skills as part of the UK's new national initiative, the Institute of Coding. The Institute of Coding brings together a consortium of leading UK universities and will strengthen the UK's position globally in computing and IT, address the UK digital skills gap and create opportunities for more computer science graduates. The OU will be leading the Institute's theme on university learning, aiming at influencing computer science teaching in universities nationally.

2.7 University of Southampton

Exploitation plans detail the University of Southampton's desire to deliver data science graduate courses and to conduct professional data science training.

The University of Southampton has expanding the graduate programmes it offers, launching a MSc in Data Science in September 2015, making use of the curricula developed for Foundations of Data Science and Data Visualisation in WP2. This includes two new face-to-face modules related to the modules released within EDSA: Foundations of Data Science and Data Visualisation.

Having been awarded a HEFCE grant, the University of Southampton is now offering Data Science 'conversion' courses, allowing graduates from disciplines such as Physics and Chemistry to take modules that can equip them with the technical skills required for studying data science. The courses teach graduates the discipline-specific techniques about data analysis along with the general programming and computational knowledge that will allow them to take advantage of big data technologies and cloud computing.

Data science courses are varied and can focus on numerous different elements of the wider discipline. By basing the MSc programme on the EDSA curriculum, particularly drawing on experience developing the curriculum for Foundations of Data Science, we ensure that a broad coverage of the data science pipeline is provided to students, who can then choose to specialise their knowledge in a number of optional modules covering topics such as data mining and machine learning. On the conversion courses, students can choose from a mix of modules within their original discipline and from within the department of Electronics and Computer Science, allowing them to combine their new technical knowledge with continued development of their own subject expertise and domain knowledge.

Providing training that is accessible to the public regardless of their location or circumstances is essential in order to allow anyone to acquire the skills necessary to become a data scientist and D1.4 indicates that data science professionals need to continue to develop skills. EDSA significantly helped

SOTON to build an international profile in data science, and as a result SOTON were able to secure private investment to develop a curriculum for Continuing Professional Development (CPD) courses. We have created the <u>Southampton Data Science Academy</u>, which provides courses informed by the demand analysis carried out in WP1, and are aligned to the curriculum (WP2). These courses will be a mixture of online, and face to face, professional, paid-for courses that will target those wishing to upskill into data science areas. By offering high-quality, paid-for training targeted at professionals, as well as graduates, we ensure that we cover a broad range of audiences.

The first course, Data Science Fundamentals (Technical) was released in November 2016, and Data science fundamentals (non-technical) were released in November 2017.

The series of online courses cover the following topics:

- Data science fundamentals (technical)
- Data science fundamentals (non-technical)
- Data science for digital marketing
- Cybersecurity for data science (technical)
- Cybersecurity for data science (general)
- Artificial intelligence (in Data Science?)

In addition, a one week, face to face course on *Data analytics in the built environment*, was delivered in London in October 2017 with an engineering company as part of a corporate masters. It is intended to continue working with this company, and it is intended to deliver a similar course next year on machine learning and AI.

To date, 203 people have been trained online, with 136 obtaining a pass, and 25 people offline. As professional courses, these are charged for and revenue is split between the University of Southampton and the private investors.

There are still Continuing Professional Development (CPD) courses which are being developed. The current development schedule for the remainder of the courses is as follows:

•	Cybersecurity for data science (general):	Q1 2018
•	Cybersecurity for data science (technical)	Q2 2018
•	Artificial intelligence	04 2018

2.8 T/Ue

Exploitation plans detail TU/e's desire to improve visibility through providing EDSA courses, and to enhance their online courses by applying learning analytics.

Through the EDSA project TU/e has been able to produce three Massive Online Open Courses (MOOCs), attracting over 60,000 students between July 2016 and January 2018. This provides great exposure to both EDSA and TU/e.

TU/e has used the learning analytics developed within the EDSA project to improve their online courses. Applying the learning analytics tools, we identified popular and less popular videos and quiz (questions). We analysed the results and were able to make slight improvements to the courses to improve student satisfaction. We expect the sales of certificates to increase (slightly) by improving the quality of the courses. For example, the first session of the 'Introduction to process mining with ProM' attracted 4875 people and sold 13 certificates. The course we launched after this, 'Process mining in healthcare' attracted 1232 people in the first session, selling 29 certificates. We believe this is partly because of the lessons learned from the earlier courses.

2.9 Persontyle

Exploitation plans detail Persontyle's desire to create Machine Learning talent for Europe by designing, delivering and scaling professional training programmes, both face to face and online, in the form of MOOCs, to design new training programmes for data scientists.



Persontyle have designed new machine learning and deep learning training programs based on the feedback received from industry and EDSA demand analysis for industry professionals. Through the programmes, students and researchers can start practicing and experimenting with machine learning and deep learning. Persontyle's primary focus is to enable the existing workforce, and develop new talent, so that they can benefit from machine learning and apply it to create new value for industries, societies and governments. In the last eighteen months we have conducted 6 face-to-face training programs and a total of 179 participants from across Europe have attended these training programs.

Persontyle's aim is to make the delivery of training programs sustainable, and to scale their reach and target markets by generating revenue from multiple channels, including public and corporate training programs, and via development of customised training for clients. Persontyle also conducts free community learning programs and Meetups. Persontyle will continue to deliver machine learning and deep learning training programs across Europe on a demand led basis.

Face to face, instructor-led workshops were delivered during the EDSA project. Feedback from the attendees and output of the demand analysis have helped Persontyle in the design of two new EDSA MOOCs - Advanced Machine Learning and Deep Learning for Data Scientists in partnership with Future Learn. Both of the MOOCs will cover advanced Machine Learning and Deep Learning techniques used in data science, and the roles they play within large-scale data projects, enabling talent and skilled digital workforce to address market demand of today and in the future. We anticipate scheduling and conducting 2 MOOCs in 2018, the first MOOC by March 2018 and second by September 2018.

The courses examine a large number of advanced supervised, unsupervised, semi-supervised and reinforcement learning algorithms, examining how these are related to the simpler techniques they are built from and how models generated by them can be tuned and evaluated. They also examine feature engineering, dealing with missing data and analysis of sufficiency of data. MOOCs are balanced between theory and practice, and participants will gain experience applying all algorithms covered. We expect to generate revenue from sales of certificates.

Persontyle has designed a 2-day EDSA bootcamp titled "Machine Learning for Data Scientists" to get industry professionals and researchers started on their machine learning journey. Participants will learn, through a practical approach, machine learning models, methods and algorithms. Participants attend the bootcamp from all over Europe, bringing with them a range of experience in Machine Learning and Data Science – from absolute beginners to experienced practitioners. Learners reported that the blend of practical tasks and theory was well balanced, and that they appreciated the chance to attempt a variety of methods in R. When asked, all said they intended to put the new knowledge gained from the bootcamp to practical use, and would recommended EDSA's training to others.

We will continue to deliver these bootcamps across major cities in the EU with the support of industry and university partners. The primary aim in conducting these bootcamps is to address a skills gap, providing the training and practical understanding of machine learning necessary to meet the demand for specialist roles in Europe. Persontyle will continue conducting Machine Learning for Data Scientists bootcamp across Europe in 2018 on a demand basis.

2.10 Training provision

The table below summarises the extent of the exploitation of the curricula and resources developed through the EDSA project by the project partners as training providers.

Table 1: Summary of training exploitation

			Planned
Purpose of exploitation	Metric	Actual (to Nov 2017)	(to end Jan 2018/19)
Develop new course curricula, course resources and delivery methods	Number of new courses (curricula, resource sets) developed based on demand and supply analysis	9 courses	2
Identify and curate courses	Number of courses (curricula, resource sets) collected and curated based on demand and supply analysis	Approximately 2,000 collected online resource, 11,700 videolectures	Not available
Enhance existing curricula, resources or delivery methods	Number of partner's existing courses adapted based on insights from demand and supply analysis	7	3
Deliver training courses based on the curricula, course resources or delivery methods generated through this project	Number of courses run using EDSA project output (curricula, resources, delivery methods)	24	11
Customise the materials generated in the project and deliver training	Number of EDSA course materials (curricula, resources) customised	3	
Customise the materials generated in the project and deliver training	Number of courses run using customised project outputs (curricula, resources or delivery methods)	10	6
Enhance partners existing training course materials based on curricula, resources or delivery methods generated in this project	Number of partner's existing course materials (curricula, resources) enhanced	6	
Enhance partners existing training course materials based on curricula, resources or delivery methods generated in this project	Number of courses run using existing curricula, resources or delivery methods, enhanced by the project outputs	14	9
Promote courses to appropriate trainees through learning pathways	Number of courses offered through the EDSA platform	30	

Full details of the training courses produced and usage figures can be found in the deliverable report 'D3.4 Report on the delivery of videolectures, webinars and face-to-face trainings'



3. Exploitation of training courses and EDSA website

3.1 Introduction

There are two broad categories of exploitation of project outputs by end users such as individuals and business - the participation in courses, either face to face, via e-learning or utilising the wealth of videos gathered on VideoLectures.NET, and use of the website to identify relevant jobs, skill sets, learning pathways and course details. Details of these exploitations are outlined below.

3.2 Exploitation of training courses

The technical annex of the project agreement details target audience size for the delivery of different training formats. These targets and actual reach are detailed in below. E-Learning (passive) corresponds to the videolectures of project agreement task T3.1, e-Learning (active) corresponds to the MOOCs and e-learning courses of T3.2, while face-to- face training is covered by T3.3.

Data for the table below can be found in deliverable D3.4 Report on the delivery of videolectures, webinars and face-to-face trainings.

Table 2: Summary participants

Activity	e-Learning (passive) / Videolectures (views counted)	e-Learning (active) / MOOCs (registered students counted)	Face-to-face- training (participants counted)
Audience reported in D3.1, month 6	1,005 for 2015 (selected Data Science videolectures)	24,558	356
Total audience reported in D3.2, month 18	31,733 for 2015 (selected Data Science videolectures)	50,089	2,018
Audience reported since month 18	59,596 (number of accesses of the data science videolectures, from M18 to M36)	65,865	2,764
Total audience reached	271,333 (number of accesses of the data science videolectures from 2015-2017) 2,632,318 (number of views in Data Science category)	115,954	4,782
Overall target	200,000	50,000	2,500

For videolectures the target was 200,000; videos in the data science category were viewed 271,333 times from 2015-2017 (and 2,632,318 times since the start of the platform). In the category of MOOCs and e-learning we attracted in total 115,954 participants, which is more than twice the promised number of 50,000. In the category of face-to- face courses we reached 4,782 participants, which is almost twice the promised 2,500.

3.3 Website

The EDSA website offers opportunities for users to engage with the outputs produced from the project. Directly linked to the website, is the EDSA courses portal (courses.edsa-project.eu), which provides a centralized location for all of the courses offered by the EDSA project. This portal offers users courses that have been developed by the EDSA project, including courses owned and managed by affiliated institutions. The courses offered will be continually updated to reflect new and relevant courses that are aligned to EDSA's aims.

The dashboard component of the website (edsa-project.eu/resources/dashboard/) allows independent users such as data science candidates the opportunity to explore the demand landscape, for example, to identify gaps in their skill sets, and allows industry users to review trends in their domain. Skill sets in demand are identified through a job search of 4.8 million job descriptions (Jan 2018), and links made either directly to related training courses and video lectures (through the VideoLectures.NET portal), or via the four Learning pathways (Data Analytics, Data Science Engineering, Data Management and Business Process Management).

Exploitation of the dashboard, learning pathways and courses

The following statistics illustrate the extent of the use of different areas of the website focussing on the resource dashboard, the learning pathways and overall course pages

Table 3: Use of website functionality

Dashboard, Learning pathways and courses section (figures, Jan 2018)					
Number of page views (home page /resources/dashboard/)	4,917				
Number of unique page views (home page /resources/dashboard/)	3,671				
Number views of the learning pathways section of the website (/courses.edsa-project.eu/mod/checklist/*, since November 2015)	2,205				
Number of unique page views of all course pages (http://courses.edsa-project.eu/)	67,579				

The following table illustrates the preferred content areas for users within the dashboard and course sections of the website.

Table 4: Website content views

Pre	erred content areas	Unique page views						
Nun	ber of unique page views of the page 'Download the EDSA eBook'	1,495						
Тор	Top 5 courses by unique page view figures of course information page							
	. Big Data Architecture	1,069						
2.	Essentials of Data Analytics and Machine Learning	977						
3.	Foundations of Big Data	919						
4.	Big Data Analytics	813						
5.	Foundations of Data Science	772						

Website sustainability

To ensure longevity and maintenance of the website, the Open University (OU) is committed to hosting the project website until March 2019, with all partners contributing to its existence and promotion. Based on the engagement metrics as highlighted in D4.5 Final community engagement and networking report, the immediate focus will be on the promotion of EDSA's services to allow for the exploitation of project results by the partners, as well as other interested parties.

In order to increase engagement in the Academy and its online services, each partner has agreed to promote the EDSA brand via in-kind contributions. The website will be used to secure acquisitions of new customers, through sign up commitments, and will offer users the opportunity to exploit the openly available results of the project.

4. Demand for skills vs. Supply of training courses

An analysis of the top 20 demanded skills and subjects, and the quantity of courses on offer for each (December 2017), provides a overview of the match between the demand for skills and the supply of training courses. The analysis (see below), shows that, training is provided that cover the top 20 demanded skills and subject, however there is some inconsistency between the ranking of the skills and subjects and the quantity of courses on offer, suggesting priorities for future course development.

Within the top 20 demanded skills and subjects, many of skills and subjects are covered through a range of 20-50 courses. A few topics - 'analysis', 'business intelligence', are covered by less than 10 courses, particularly noteworthy is the subject of 'analysis', which is the 7th most in-demand skill or subject. 'Databases', 'data mining', 'php' and 'analytics' are covered by less than 20 courses of which the subject 'database' is noteworthy, being the most demanded subject.

It is also noteworthy that several of the subject with the highest number of courses are not the most demanded - "business" (63 courses), is the twentieth most demanded subject. "Machine learning" and "javascript" (56 courses each) are higher in the list of skills and subjects demanded - at fifth and eighth respectively, and "c++", the sixteenth most demanded skills and subject, is served by 49 courses.

The analysis also highlights that much of the demand is comprised of skills and subjects under broad categories, such as "database" and "computer science", followed by specific programming languages

such as java and C++. It is worth noting the opportunities that this analysis highlights, particularly in the development of further courses for 'analysis', 'business intelligence', 'databases' and 'data mining'.

Table 5: Skills in demand, and courses offered

Skill / subject name	Number of occurrences (job posting)	Courses on offer through EDSA dashboard
database	330823	13
statistics	228673	28
sql	221319	22
computer science	160416	31
machine learning	146075	56
design	137422	28
analysis	130858	3
javascript	107788	56
business intelligence	99844	8
data mining	97509	11
data analysis	93450	29
data science	92876	36
artificial intelligence	85264	35
linux	84638	24
java	84413	28
C++	83503	49
php	82715	14
analytics	80858	14
big data	70642	25
project management	68059	63

Across a total of 6916 courses related to data science, 2134 are online courses. Further details of the analysis of skills demanded and training courses supplied, are outlined in report <u>D1.5 Demand and supply analysis report</u>.

5. Project impacts

The EDSA business case outlined the impacts that the project would work towards. In the following section we detail progress made against the planned impact. It is worth noting that some changes in the project design and activities took place in response to the project review in October 2016. The review made a range of recommendations including 'to link the skills side not only to self-developed training materials, but to link it to existing (external) high-quality learning resources including MOOCs'.



However, these changes are not necessarily reflected in the intended impact statements which come from the original business case.

Impact 1.1

Planned: A dashboard, as a free service for EU education and professional training institutions, will offer near-real-time demand monitoring and an analysis of important topic trends.

Actual: The EDSA dashboard provides real time analysis of European job descriptions, as of January 2018, there were 53,401 data science jobs, out of 4,887,507 jobs on the site. Job descriptions can be further searched for different terms and users are able to view the locations of the data science jobs and the respective data science skills required. Further detail on the functionality of the dashboard can be found in <u>D1.5 Demand and supply analysis report</u>.

Impact 1.2

Planned: The EDSA curricula (core curriculum plus sector-specific) are well aligned with the skill demand and remain relevant over time. Dissemination and networking with major players increases visibility of the project.

Actual: The comparison of top 20 demanded skills and the provision of courses through the EDSA dashboard (see section 4 above) demonstrate that there is overall a good alignment between provision and skill demand.

Dissemination and networking activities have included a combination of events, online communication and dissemination activities, with support from a network of ambassadors. EDSA has engaged with 42 workshops, conferences and other events, since January 2015, the list of events can be found in Annex B. Communication and dissemination has taken place through the project website, a quarterly newsletter, Twitter, Linkedin, Slideshare, VideoLectures.NET, Vimeo and Youtube networks. A European Network of Ambassadors has been established, with 12 individuals across 10 countries, promoting the work of the project. See report D4.5 for further details of the community engagement and networking.

Impact 2.1

Planned: Cost of big data education is significantly lowered. Individuals, SMEs and large organizations have easy access to learning material and knowledge. New players can enter the big data economy more easily. Support for anytime, anywhere, ubiquitous learning.

Actual: As detailed above EDSA has produced a range of courses delivered by the partners in the project supporting anytime, anywhere ubiquitous learning. As detailed in section 3.2, MOOCs/e-learning and F2F courses have reached far more participants than promised. 11,500 lectures and tutorials were added to the VideoLectures.NET portal during EDSA with six different MOOCs and one e-learning course offered in the second half of the project alone. By providing industry with a large volume of learning resources in several formats, EDSA is proving its value in reducing the cost of big data education through offering users freely accessible courses and learning materials. Self-guided learning i.e. e-learning and videolectures, are openly sourced resources free for use. By making resources available online, partners remove the need to recoup facilitation costs through course attendance fees. This in turn increases motivation for more users to undertake the course, anytime and anywhere, as access to training is increased.

Impact 2.2

Planned: Data science courseware will be openly available in relevant EU languages, supported and maintained by sustainable communities.

Actual: The demand and supply analysis applications developed by JSI along with the VideoLectures.NET portal provide multilingual services and dissemination in non-English speaking parts of the world. The demand analysis is applied to jobs advertisements in different European languages.

Results of analysis within the project suggested a desire from trainees for learning resources in English, with a secondary need for learning resources in German. For example; 54 percent of all interview and survey participants considered training that has been translated from English as not required, only 18 percent think it is essential. Based on these results, the consortium have looked to produce content in English and to a lesser extent, in German. Additionally, bespoke F2F courses can be tailored by partners to accommodate the immediate and local community if needed.

In order to ensure the maintenance and sustainability of the resources, the project partners have committed to run the online services via in-kind contribution, for one year after the end of the project (see the Memorandum of Understanding for further details).

By using trends sourced from the dashboard, the partners will be able to sustain and develop these learning resources to ensure EDSA's offer remains relevant to its market and the Data Science community. EDSA has committed to publishing its learning resources openly under Creative Commons 4.0 (CC BY 4.0) license, to facilitate the maximum use of the resources.

Impact 3.1

Planned: More than 200 days of training delivered directly to 2,500 professionals in 6 countries. More than 3000 students attend EDSA courses at the consortium universities.

Actual: Please see section 3.2 for statistics of those trained

Impact 3.2

Planned: More than 50,000 learners participating directly in EDSA online training activities, the indirect reach (downloads, visits) of the comprehensive online portfolio of educational material will be at least an order of magnitude higher.

Actual: As detailed in section 3.2. approximately 120,000 participants were registered on e-Learning or MOOC courses or participated in face-to-face training. There were 271,333 accesses of the data science videolectures (2015-2017) and 2,632,318 views in the Data Science category.

Impact 4.1

Planned: EDSA alumni will have a rounded portfolio of transferable skills and a comprehensive understanding of big data from a technical and business perspective.

Actual: At this stage it is not possible to detail the long-term results of the use of the EDSA dashboard and the courses and resources developed by project partners. However, interim results show that there are a high number of individuals engaging in training (see training statistics detailed in section 3.2), and a strong synergy between the most highly demanded technical and business skills, and courses offered (see section 3.5). These results, alongside the focus on providing a large number of curated courses, dashboard functionality that links jobs, skill set and courses, dashboard functionality of learning pathways, and demand analysis for course and curricula designers, suggests that the programme is on track to deliver this impact. A follow-on assessment of EDSA alumni would provide further details of the end user impact.

Impact 4.2

Planned: An ecosystem of high-quality cross-lingual, cross-sector, cross-national Open Educational Resources and services will emerge as a direct outcome of our efforts.

Actual: The project has generated a huge collection of resources and services. As report D3.4 Report on the delivery of videolectures, webinars and face-to-face trainings outlines, both, MOOCs/e-learning and F2F courses have reached far more participants than promised. In the second half of the project, FutureLearn became the platform for EDSA MOOCs. Use of the courses through the platform, has been significantly higher than anticipated, with 116,000 e-learning participants and 4,780 face to face training participants, compared to targets of 50,000 and 25,000 respectively.



6. Sustainability of the initiative through the online Academy

The overall aim of EDSA has been to reduce the data literacy skills gap by improving data science training. Recent years have shown that the supply of data science training across Europe has grown exponentially. As a direct result from this, the market has become crowded, with a number of different offerings, which can become confusing for learners. The aim of EDSA was to bring clarity to the market by connecting learners to the training they need. Learners and managers will be able to understand what skills are in demand via market trends. Learners will be able to make informed decisions about the best way to develop these skills and will be able to access the high-quality, multimodal learning solutions designed by EDSA, which support industry's demand for skills. Connecting users to high-quality learning has the potential to make a long-term impact on EU economies by building skills and capability in data science. There is therefore value in making the EDSA approach viable beyond the term of the project.

In order to find a suitable approach to making EDSA a viable offering long term, the project has undertaken a number of market consultations with industry representatives to determine what characteristics will service the market best. The feedback received has provided valuable input into determining the most suitable option for an EDSA offer that could be sustained, along with ways that EDSA's offer may evolve.

Feedback from the testing phase shows that a revenue-generating model for EDSA is not the most suitable. Although there is indeed an appetite for an online service, industry representatives feel the space is quite crowded with competitors and with this choice, comes the ability to seek alternative training providers who offer multiple course topics, rather than specialising in one theme.

That being said, the most viable option for sustainability of the online Academy is for partner continuation. This is the most suitable option, as EDSA is perfectly aligned with what the training providers in the consortium are already doing. The outputs can look to be preserved through the further funding opportunities that the partners have identified in D5.2.

Following this decision, each partner has committed to an agreement to run the online services via inkind contributions. These contributions have been outlined and agreed within a Memorandum of Understanding (MoU) and will be signed by all partners following the end of the project. This agreement states the Academy will be maintained and promoted for at least one year beyond the end of the project by the relevant and voluntary partners. Following year 1 of operation (after the project), partners will evaluate success, before agreeing the strategic next steps for the online Academy. For further details about the agreement, please refer to the MoU.

7. Future opportunities

At the start of the EDSA journey, two key needs were identified, that of a lack of comprehensive analysis of the skills in demand in the data science sector, and lack of curricula and courses to meet those needs. The project set out to meet these two needs.

The initial analysis of job descriptions provided a wealth of information on the needs of industries across Europe, and based on the findings, EDSA partners developed new curricula and courses or adapted and improved existing courses based on new findings or materials. Demand analysis became increasingly automated and comprehensive as the project progressed, allowing development of curricula and courses to be increasingly matched to the identified needs. Partners also expanded the number of courses on offer, bringing in new investment and partnerships such as in the establishment of the Southampton Data Science Academy.

The 2016 project review shifted the focus away from the development and delivery of courses by EDSA partners, shaped by the demand analysis, towards a brokering and matching service between jobs, skill demand and courses offered by a range of suppliers, including those developed through EDSA partners. Following the development of these features, the EDSA dashboard now provides the critical link between automated demand analysis from live jobs, required skill sets and the supply of training courses. The automated demand analysis will drive future curricula development education providers, while the dashboard provides the interface for citizens to identify the skills in demand and find

appropriate training. Aggregating courses together along with our process model and use of tools that respond to actual job requirements, have improved the connection to industry and ensures that courses will deliver where there is need.

Looking forward, we anticipate continued growth for data science technical and business skills with that continued demand for education and training, face to face, online, blended and self-directed learning. The European Commission stated that the period of 2013 -2020 will see industry needing a further 346,000 Data Scientists, highlighting the need for related training opportunities in a multimodal and multilingual format (Curtis & Hopper, 2017). From this, we view EDSA's role as acting as a broker for the sector, offering a service for people to understand how well their courses match industry needs. The project partners are committed to sustaining the EDSA online Academy, and in order to fulfil that role, new collaboration opportunities and new innovations and developments have been investigated. Partners have begun experimenting with micro accreditation, badging and certification in the data science space and have agreed to co-design a set of open badges for data science which will sit on a blockchain run from the Big Data Value Ecosystem (BDVE). The badges will take an input from our courses/curricula and the dashboard. Moreover, the recent UK government announcement on the 'Institute of Coding' see's partners involved with improving graduate employability in Computer and Data Science.

Furthermore, Partner involvement with European Initiatives such as the <u>European Association for Data Science</u> and <u>Data Pitch's</u> Data Innovation Academy, support the partners' decision on sustainability as EDSA is aligned to partner activities. For example, the Data Innovation Academy will focus on Webbased training built upon a rich collection of learning materials available at Southampton University around Data Science topics, complemented by ODI's programme around data-driven innovation.

The partnerships outlined above, represent future partner funded activities. These activities are aligned with EDSA, and due to this alignment, partner continuation is indeed the most suitable option for sustainability of the Academy and partner exploitation of project outputs.

References

J Curtis & C Hopper, 2017, *Data scientist jobs: Where does the big data talent gap lie?* http://www.itpro.co.uk/careers/28929/data-scientist-jobs-where-does-the-big-data-talent-gap-lie-1 [Accessed: 24/01/2018]



8. Appendices

Appendix A.- Output and Exploitation classifications

Project output classifications

- 1) Training
 - a) Curricula
 - b) Course resources
 - c) Course delivery methods
 - d) Best practice knowledge on training delivery based on training analytics
- 2) EDSA website incorporating
 - a) Demand and Supply side analysis
 - i) Demand & insights in skills and skills-training from individuals
 - ii) Demand for skills articulated by business in Job descriptions
 - iii) Website dashboard comparing demand for skills and supply of training
 - b) Learning pathways
 - c) Access to appropriate training courses (training supply)

Exploitation classifications

Exploitation by data science training providers

- 1) Exploitation of demand and supply side analysis
 - a) To develop new course curricula, course resources and delivery methods developed
 - b) To enhance existing curricula, resources or delivery methods
- 2) Exploitation of new curricula, course resources or delivery methods
 - a) Deliver training courses based on the curricula, course resources or delivery methods generated through this project
 - b) Customise the materials generated in the project and delivering training to meet a particular need
 - c) Enhance existing training course materials based on curricula, resources or delivery methods generated in this project
- 3) Exploitation of EDSA website
 - a) To promote courses to appropriate trainees through learning pathways

Exploitation by citizens/website users:

- 4) Exploitation of website
 - a) To identify demands for data science skills
 - b) To identify learning pathways
 - c) To identify appropriate courses to build required skill sets

Appendix B.- EDSA Event Participation

Event	Date	Location	Event Type	Attended By	Interaction with Event	Number of Attendees	Link
Data science focus group, UK government organisatio n	January , 2015	London, UK	Focus Group	ODI	Ran focus group		N/A
LAK'15	March 2015	Poughke epsie, USA	Confer ence (e- Learni ng)	Fraunhof er IAIS	Paper presentatio n		http://lak15.s olaresearch.or g
EIT ICT Labs Partner Event	April 2015	Trento, Italy	Meeti ng (data scienc e)	КТН	Presentatio n		https://www.e itdigital.eu/ne ws- events/events /article/eit-ict- labs-partner- event-2015/
Smart Data - Deutschland und Europa auf dem Weg zu einer digitalen Datenökono mie	April 2015	Berlin, Germany	Confer ence (data scienc e)	Fraunhof er IAIS	Ran booth on data science training and promotional material sharing		http://www.b mwi.de/DE/Se rvice/veransta ltungen,did=69 3428.html
PAPIS Connect	May 2015	Paris, France	Confer ence (data scienc e)	Pesontyl e	Project details disseminate d		http://www.p apis.io/connec t
VOR Divsion ICT and HE Symposium Research on MOOCs	May 2015	Wagenin gen, Netherla nds	Resea rch sympo sium (e- Learni ng)	TU/e	Presentatio n		http://www.v orsite.nl/nl/di visies-en- themagroepen /vor-divsion- ict-and-he- symposium- research-on- moocs-26-mei- 2015.html
ODI Open Data Meetup	Feb 2015,	London, UK	Meet- up (data	ODI	Ran meet up		http://www.m eetup.com/Op

	May 2015		scienc e)				en-Data- London/
Open University visit by UK HE Commission	June 2015	London, UK	Meeti ng	OU	Presentatio n		N/A
Data Literacy Workshop at WebSci'15	June- July 2015	Oxford, UK	Confer ence (e- Learni ng, data scienc e)	UoS	Presentatio n and flyer disseminati on		http://www.d ataliteracy.eita. org.br/
Data science focus group, UK health organisatio n	July, 2015	London, UK	Focus Group	ODI	Ran focus group		N/A
ESWC Summer School	Septem ber, 2015	Kalamaki , Greece	Summ er School	UoS, OU	Ran summer school		http://summe rschool2015.es wc- conferences.or g/
ODI Network Event	Septem ber, 2015	London, UK	Netwo rking event	ODI	Organised event		
ICT 2015	October , 2015	Lisbon, Portugal	Confer ence (Data, IT, Innov ation)	Fraunhof er, UoS, OU	Conducted interviews, promoted dashboard and ran workshop. Showing insights of the European data market that we have acquired.	6,000	https://ec.eur opa.eu/digital- single- market/en/ict 2015
Interdiscipli nary Mathematic s	October , 2015	Uppsala, Sweden	Meetu p	Persontyl e	Attended		N/A
VOILA, 2015	October , 2015	Bethlehe m, USA	Works hop	OU	Presented paper		http://voila20 15.visualdataw eb.org/

Deep Learning London	Novem ber, 2015	London, UK	Meetu p	Persontyl e	Organised meetup		http://www.m eetup.com/De ep-Learning- London/
European Data Forum	Novem ber, 2015	Luxembo urg, Luxembo urg	Confer ence (data scienc e)	UoS, OU, Fraunhof er	Moderated session, ran session and presented dashboard. Showing insights of the European data market that we have acquired.		http://2015.da ta-forum.eu/
ODI Pre- Summit Training Day	Novem ber, 2015	London, UK	Traini ng	ODI	Ran training		N/A
ODI Summit	Novem ber, 2015	London, UK	Summ it – open data	ODI	Organised event		N/A
EDSA PMB	Decemb er, 2015	London, UK	Meeti ng	All	Organised event	30	N/A
E-Skills 2015	Decemb er, 2015	Luxembo urg, Luxembo urg	Confer ence	OU	Panel member		http://eskills4j obs.lu/progra mme
SIMPDA 2015	Decemb er, 2015	Vienna, Austria	Symp osium	TU/e			http://simpda 2015.di.unimi.i t/
e- Leadership Skills Expert Workshop	January , 2016	Brussels, Belgium	Works hop (data scienc e)	UoS	Attended workshop		N/A
EDISON Workshop	Februar y, 2016	Brussels, Belgium	Works hop (data scienc e)	UoS, OU	Attended workshop		http://edison- project.eu/dat a- infrastructure- competences- and-skills- framework- european-and-

							global- challenge
Big Data Summit	Februar y, 2016	Hanau, Germany	Confer ence (big data)	Fraunhof er	Attended	500	http://www.bi tkom- bigdata.de/
Fraunhofer' s Big Data Alliance	Februar y, 2016	Sankt Augustin, Germnay	Annua l meeti ng (data scienc e)	Fraunhof er	Attended	40	N/A
PAPI's Connect	March, 2016	Valencia, Spain	Confer ence (IT)	Persontyl e	Attended		http://www.p apis.io/connec t/
ODI Network Event	March, 2016	London, UK	Netwo rking event	ODI	Ran event		N/A
Open Data Science	March, 2016	Lancaste r, UK	Works hop	ODI	Attended		N/A
Data science focus group	March, 2016	London, UK	Focus group	ODI	Ran focus group		N/A
BigDataValu e Association	March, 2016	The Hague, Netherla nds	Summ it	Fraunhof er	Attended		http://www.b dva.eu/
WWW Conference	April, 2016	Montreal , Canada	Confer ence (Web)	OU, UoS	Ran workshop		http://microp osts2016.seas. upenn.edu/
Web Science 2016, Data- Driven Innovation Workshop	May, 2016	Hannove r, Germany	Works hop (innov ation, data scienc e)	UoS	Ran workshop and distributed flyers	20	http://websci1 6.org/
ODI Node Gathering	May, 2016	London, UK	Netwo rking event	ODI	Organised event		N/A
EDSA PMB	June, 2016	Dubrovni k, Croatia	Meeti ng	All	Organised event	30	N/A
European Data Forum	June, 2016	Eindhove n, Netherla nds	Confer ence	UoS, OU, Persontyl e, TU/e	Organised event, presented project		http://2016.da ta-forum.eu

EDISON Data Science Champions Conference	July, 2016	Brocken hurst, UK	Confer ence (data scienc e)	UoS	Keynote	50	N/A
Essex Big Data and Analytics Summer School	Septem ber, 2016	Essex, UK	Summ er School	ODI	Organised event		N/A
ODI Pre- Summit Training	October , 2016	London, UK	Traini ng	ODI	Ran event		N/A
ODI Summit	Novem ber, 2016	London, UK	Summ it – open data	ODI	Organised event		N/A
ODI Pre- Summit Training	October , 2017	London, UK	Traini ng	ODI	Ran event		N/A
ODI Summit	Novem ber, 2017	London, UK	Summ it – open data	ODI	Organised event		N/A