EUROPEAN
— DATA SCIENCE —
ACADEMY

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

D4.2 Key Figures of Attained Audience

Deliverable Editor:	**Christopher Phethean (University of Southampton)**
Other contributors:	**Elena Simperl (University of Southampton)**
	Gareth Beeston (University of Southampton)
Deliverable Reviewers:	**Simon Bullmore (ODI)**
	Ali Syed (Persontyle)
Deliverable due date:	**31/07/2015**
Submission date:	**31/07/2015**
Distribution level:	**PUBLIC**
Version:	**1.0**

Change Log

Version	Date	Amended by	Changes
0.1	18/06/2015	C Phethean	Website, Twitter and Process Mining MOOC details
0.2	22/06/2015	C Phethean	LinkedIn groups spreadsheet, press coverage, real-world events
0.3	24/06/2015	C Phethean	Dashboard, Qualitative summary, Conclusion, Executive Summary
0.4	10/07/2015	C Phethean	Minor additions before submission for review
0.5	24/07/2015	C Phethean	Final revisions
1.0	31/07/2015	A Tumilowicz	Final QA

Table of Contents

List of Tables

List of Figures

EUROPEAN DATA SCIENCE ACADEMY

1. Executive Summary

This deliverable summarizes and presents the measures of communication and engagement with the public by EDSA in the first six months of the project. We report on a mix of real-world events (such as offline meetings) and online communication channels – following the community engagement plan outlined in D4.3 at M3 of the project. The aim of that plan was to target a variety of channels in order to solicit feedback from stakeholders including researchers, industry-specialists, policy-makers and decision makers from a range of sectors. This deliverable therefore reports on this work to date and provides some initial insights into our online community, offline event attendance and into the uptake of the first EDSA-branded course. By doing this we demonstrate the variety of events in which the EDSA consortium is participating in order to promote the project, and to source feedback and input on the data science skills gap. It also puts in to place the services and tools required in order to track the strength of the supporting community, and to monitor the dissemination rate of key EDSA learning materials, which will help us to continuously adapt to the changing needs and demands of the community as the project progresses. To conclude we also present a qualitative assessment of these results and what this means for our community-building work going forward with the remainder of the project.

2. Community Engagement Plan Recap

In D4.3 a real-world and online community engagement plan was presented. The aim of this was to outline a set of engagement tools and practices that the project could utilize in order to engage the community. These included a combination of both real-world events (such as offline, face-to-face meetings) and online communication and dissemination activities. Online activities were separated based on those that were broadcast-oriented and were primarily aimed to use to disseminate information, and those that were responsive and conversation-oriented and targeted at increasing communication and collaboration amongst the community. These activities will allow the EDSA consortium to gather feedback regarding the learning materials and training on offer, as well as to build, grow and establish the EDSA brand and the community that is aware of this. Ultimately this will allow us to maximize uptake of EDSA materials, and put in place a recognizable brand presence that will contribute to the sustainability and longevity of EDSA's legacy. As the curriculum is still in development, many of these channels have not yet fully been exploited. However with the curricula for topics now being developed with D2.1 (Data science curricula 1), which will also begin the process of learning resource production, it will now become possible to pursue many of these plans.

3. Channels and Activities

3.1 Website

The EDSA website provides a centralized location for all of the project's activities. It offers a comprehensive overview of the main activities taken by project partners, along with details about the project and any activities planned for the future (Figure 1). Its design and structure was discussed in more detail in D2.1 (Data science curricula 1), when the website was setup for the project. Since then, we have added Google Analytics to the site, which will allow us to measure in detail the journey of visitors to and across the site to gain an understanding of where the community is particularly interested in EDSA.
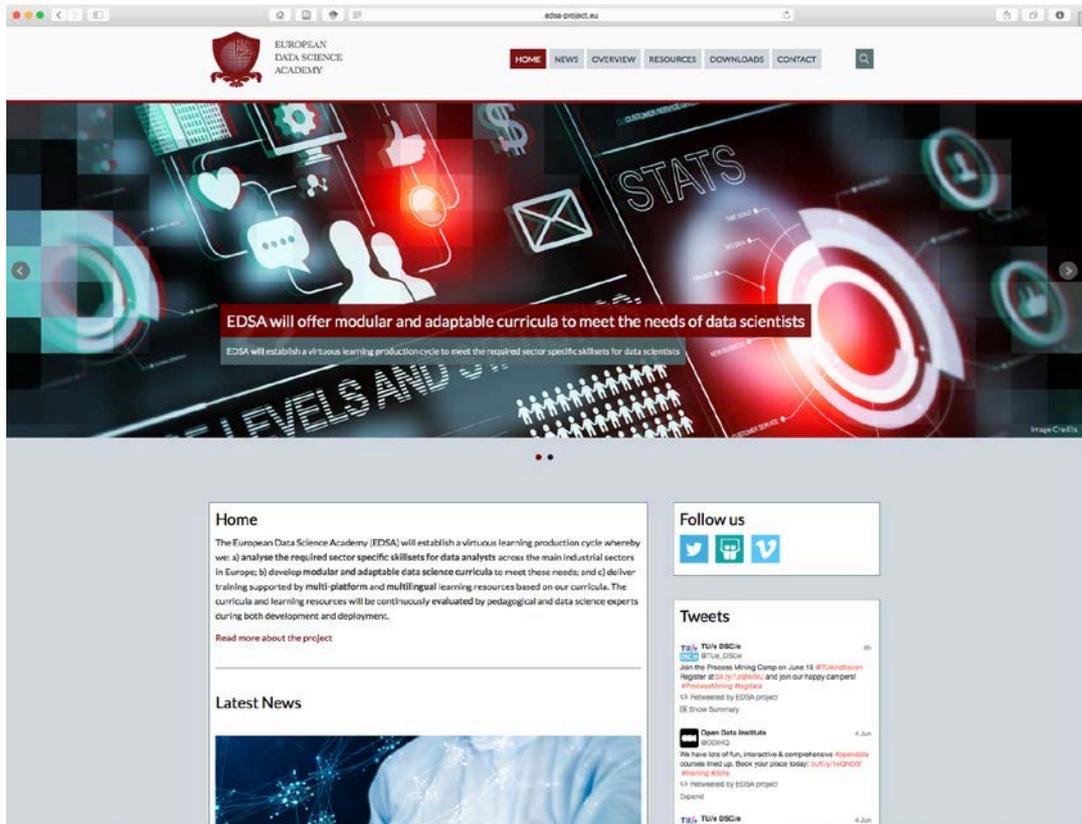
Figure 1 - EDSA project website homepage

For the activity preceding the installation of Google Analytics, the following statistics are available from Sawmill reports[1] relating to visitor statistics:

Table 1 - Website Visitor Statistics (1st Feb 2015 – 10th Jun 2015)

Metric	Total	Avg/day
Hits	267320	2056
Visitors	6628	50
Page Views	49950	384
Sessions	16930	130

Table 1 lists a number of general statistics for the website. The most revealing statistic here is the 'visitors' or 'unique hosts' that access the page, with on average 50 per day. We can also calculate that these visitors accessed around 7.5 pages on the site on average (49950/6628), and that per session a user tends to visit just under 3 pages (49950/16930). As more content is added to the site, and with the addition of Google Analytics, we aim to see which areas of the site (such as the curriculum, dashboards, publications etc.) attract the most visitors, and understand the journey that visitors take

---

[1] https://www.sawmill.net

EUROPEAN DATA SCIENCE ACADEMY

across the site. Figure 2 depicts the tracking of the number of sessions (visits) to the website over a one month period to the 9th July 2015, using data now available from Google Analytics, and the service also reveals that in that time period the site received 1499 sessions from 1229 users who each visited 2.29 pages on average per session (total of 3437 page views), which shows that these statistics show some similarity with the previous data we were collecting from Sawmill.
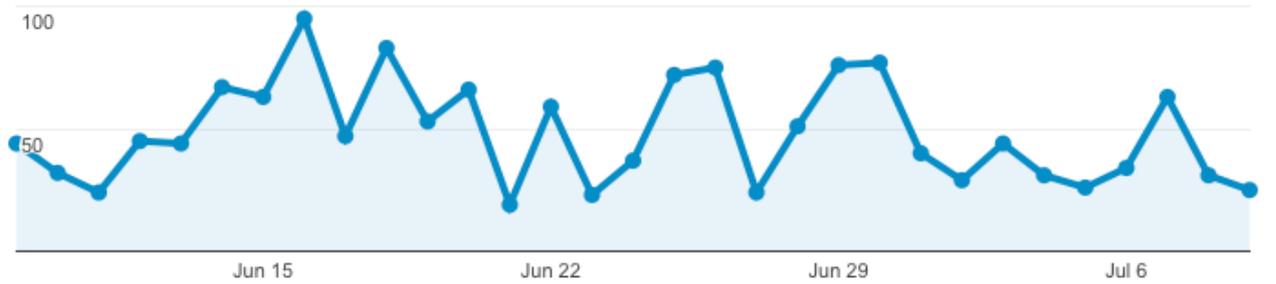
Figure 2: Initial Session Statistics from Google Analytics for 9th June - 9th July 2015

Table 2 - Website Visitor Country Statistics (1st Feb 2015 - 10th Jun 2015)

Country	Hits	Percentage of Total (Hits)
United Kingdom	59085	22.1%
Germany	41149	15.4%
United States	30794	11.5%
Netherlands	13261	5.0%
Spain	12334	4.6%
France	9565	3.6%
Belgium	8561	3.2%
Italy	6668	2.5%
India	5922	2.2%
Greece	4402	1.6%
Others	75579	28.3%

Table 2 helps to reveal the international audience that the website is receiving. While at the moment the UK provides the highest number of hits to the site, there are also large contributions from other EU member states including Germany, The Netherlands and Spain. A significant proportion of hits are also arriving from the US, where similar data science skills shortages are predicted and therefore similar challenges are faced around data science training.
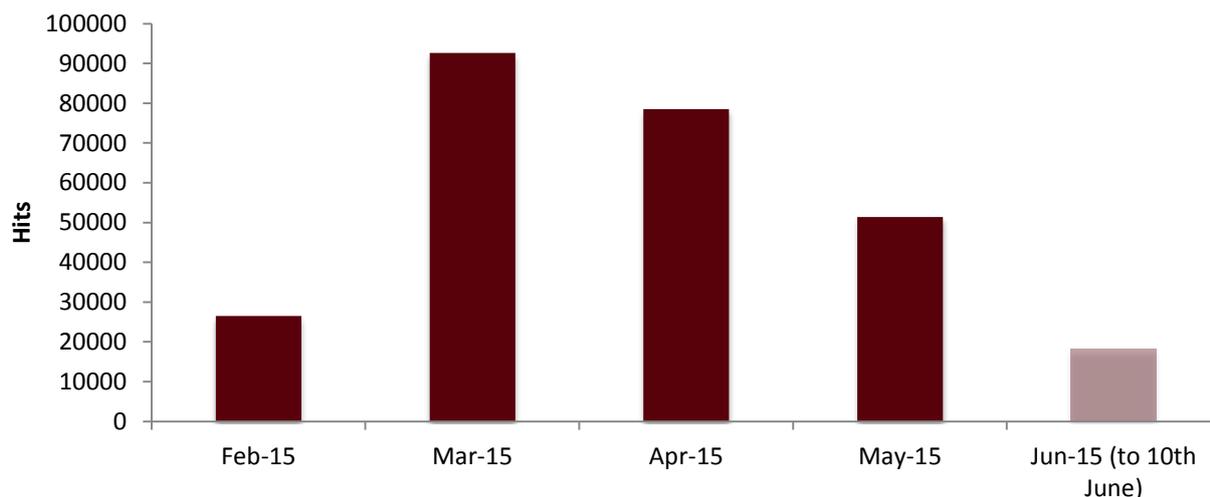
Figure 3 - Website hits over time

Figure 3 displays the number of website hits recorded since the creation of the site in February 2015. The figures for June are based on the data available as of 10th June and therefore do not reflect the entire audience for that month. As the number of deliverables released increases (especially following M6 of the project), and curricula items become available with associated learning resources, we expect the number of hits to rise and reflect the availability of these items.

3.2 Twitter

The EDSA Twitter account (@edsa_project) has been actively used throughout the first six months of the project. Using Twitter's built-in analytics service (http://analytics.twitter.com) it can be seen that the audience of the account has grown steadily to now stand at over 600 followers as of 5th June 2015 (See Table 3 and Figure 4). Using the same service, we can also see the resulting engagement with tweets from the EDSA account (Figure 5) to get an idea about the size of the audience they reach, and how many people go on to interact with the message further. Since the project kick-off meeting in March 2015, the statistics for the statistics for EDSA's tweet impressions are displayed in Table 4.

Table 3 - Twitter Follower Growth

Period	New Twitter Followers
March 2015	44
April 2015	70
May 2015	28
June 2015	35

EUROPEAN DATA SCIENCE ACADEMY
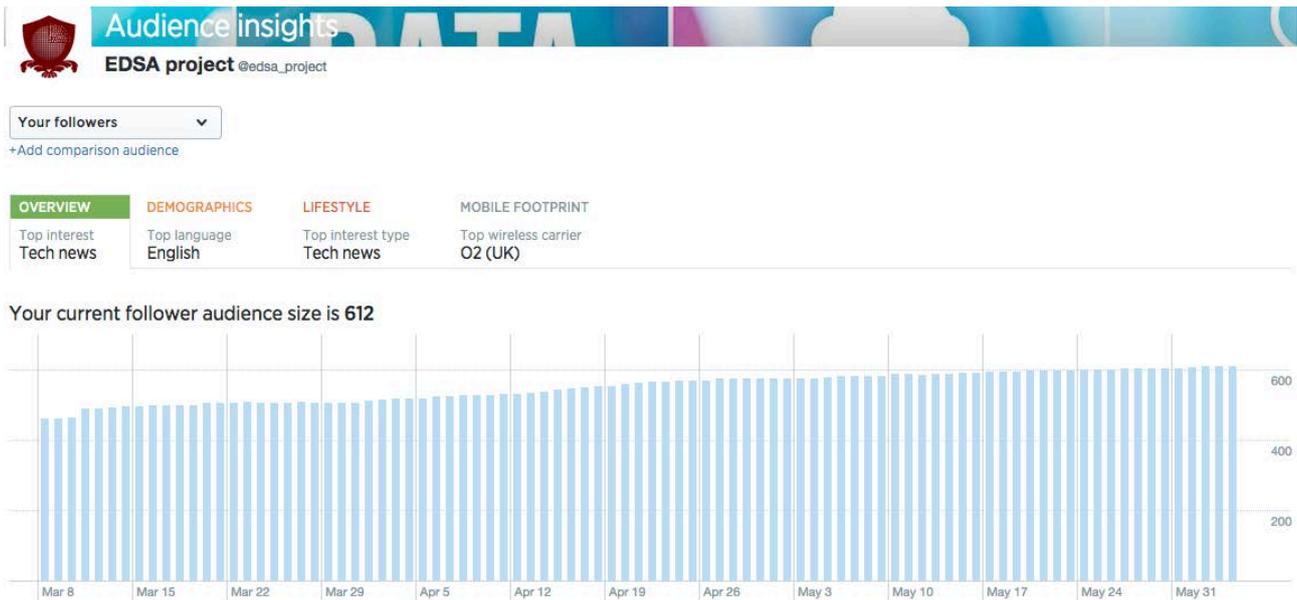
Figure 4 - Recent growth of EDSA Twitter audience

Table 4 - Tweet impression statistics

| Period	Tweet Impressions per day	Profile visits
March 2015	13	652
April 2015	96	823
May 2015	57	325
June 2015	161	605
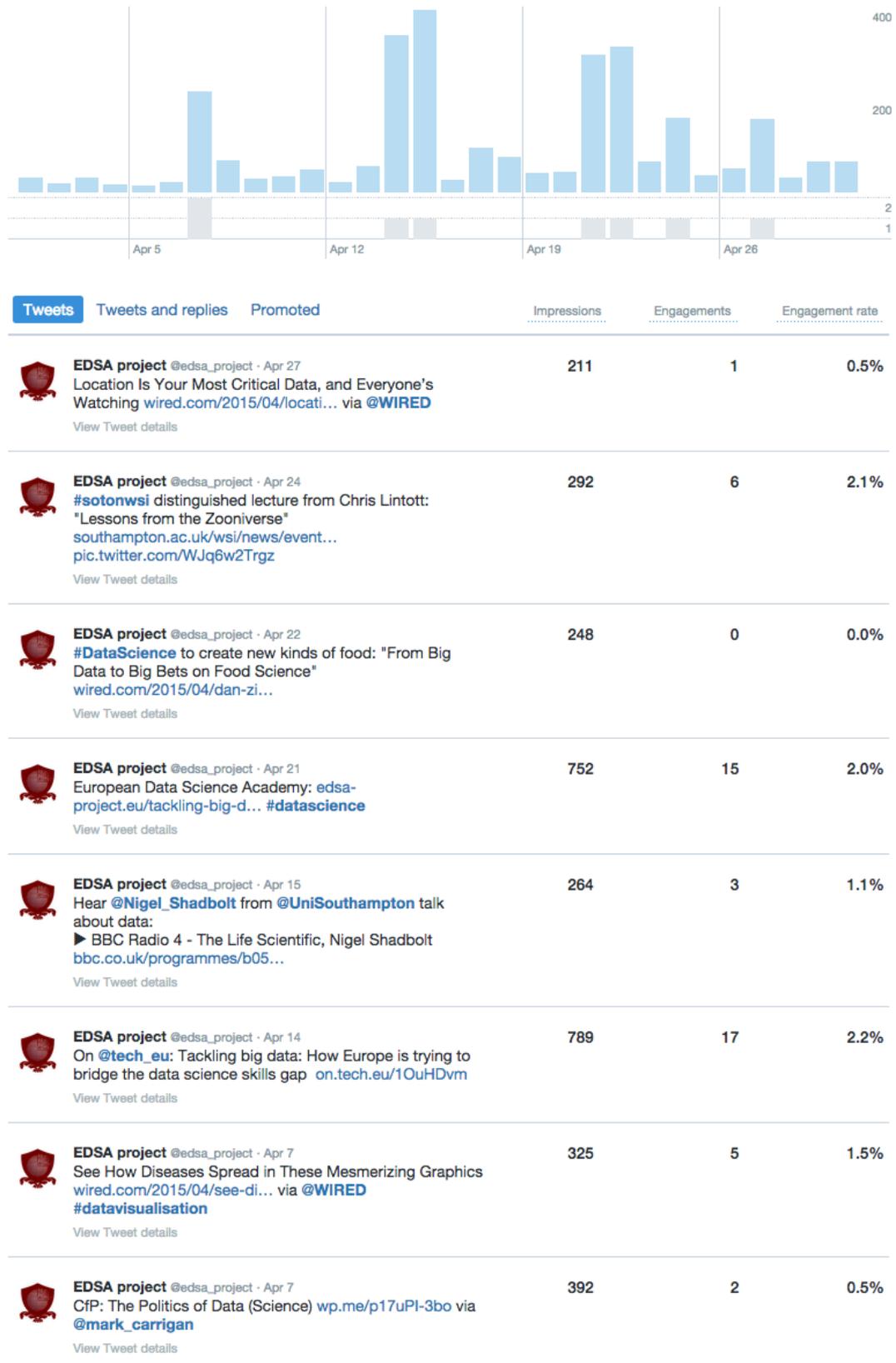
Figure 5 - Tweet Engagement Statistics for April 2015

3.3 LinkedIn

For LinkedIn we have carried out an assessment of some of the most active and populous community groups on the site. An entire copy of a table outlining our findings from this is available in Section 8.1 (Appendix A), whereas Table 5 below lists the top 5 groups based on the number of members that they contain. These groups offer a diverse spectrum of data science subjects and professionals, which is encouraging in that we can use them to target not only a sizeable audience, but a broad audience that will provide us with feedback based on a number of different backgrounds, sectors and job roles.

Table 5 - Top LinkedIn community groups related to EDSA based on number of members (as of 15/05/2015)

Group Name	Activity	Type	Member Count	Audience/Members
Advanced Business Analytics, Data Mining, and Predictive Modelling	Very Active	Professional Group	180,113	Data science, big data, visualization, business analytics, predictive modeling, data mining, web analytics, six sigma, econometrics, business intelligence, computational finance, quant, operations research, machine learning, data analysis, data warehousing, risk management, cloud computing, text mining, BI.
Business intelligence Professionals	Very Active	Professional Group	152,016	BI professionals
Big Data, Analytics, Strategy, Finance, Innovation	Very Active	Networking Group	131,392	CEO, CIO, CTO, CDO etc.
Big Data and Analytics	Very Active	Networking Group	126,761	People interested in big data
Business Intelligence, Big Data, Analytics, MIS Reporting and Database Group	Very Active	Professional Group	95,840	Professionals

3.1 Course Material

3.1.1 Process Mining MOOC (TU/e)

While the majority of course content is in development following the curricula design in D2.1 (Data science curricula 1), the first EDSA-marketed course has been run in the form of a MOOC. TU/e ran the second iteration of their Process Mining MOOC on Coursera from April – June 2015, which was purposely branded as an EDSA MOOC. Audience statistics from this course (see Table 6) showed that it received over 17000 learners visiting the course page, and over 3000 who submitted an exercise as part of their participation in the course.
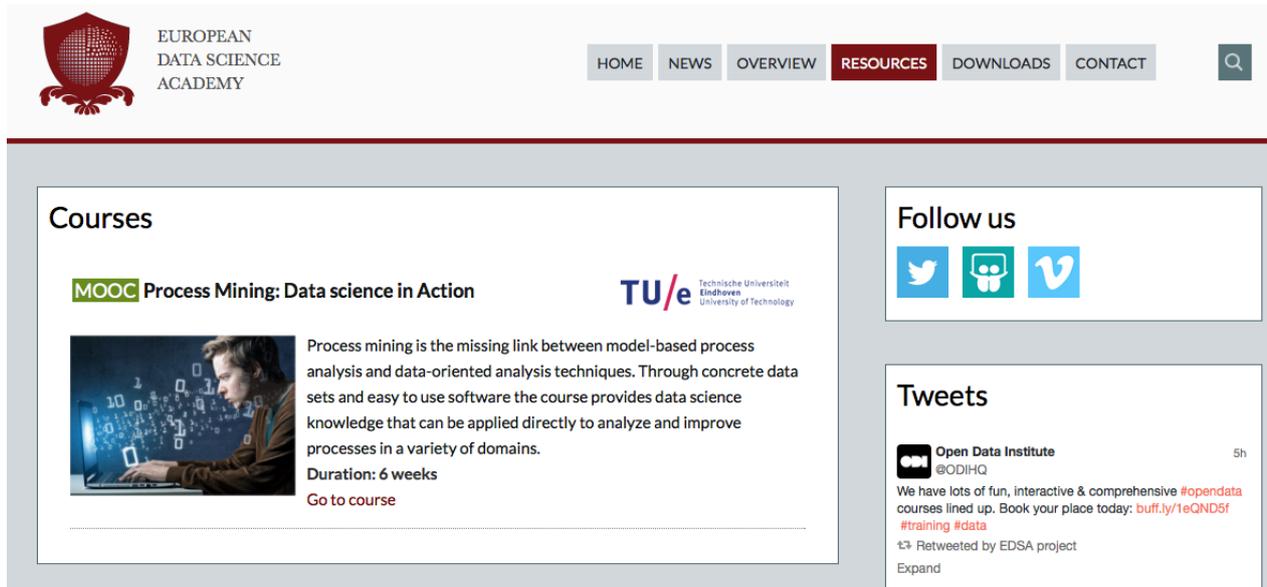
Figure 6 - EDSA website listing of Process Mining MOOC

Table 6 - Audience Statistics for Process Mining MOOC (as of 2nd June 2015)

Metric	Session 2
Start/end dates	April 1 '15 – June 3, '15
# Registered	24558
# Visited course page	17089
# Watched a lecture	11631
# Browsed forums	2899
# submitted an exercise	3212
# Certificates (normal/distinction)	916 (572 / 344)

3.2 Real-World Events

One of the main points of D4.3's real-world and online community engagement plan was an emphasis on promoting EDSA through a combination of online communication and real-world event attendance. Partners from across the consortium have been attending such events since the start of the project, speaking about and promoting the work that EDSA is carrying out. Details of these events are shown in Table 7. As discussed in the real-world and online community engagement plan (D4.3), these events cover two important areas related to the project: data-science themed events, and e-Learning themed events. Therefore this categorization is continued within Table 7 to demonstrate the variety of events that the consortium has been participating in.

Table 7 – EDSA Real-World Event Participation

Event	Date	Location	Event Type	Attended By	Link
LAK'15	March 2015	Poughkeepsie, NY, USA	Conference (e-Learning)	Fraunhofer IAIS – paper presentation	http://lak15.solaresearch.org
EIT ICT Labs Partner Event	April 2015	Trento	Meeting (data science)	KTH - presentation	https://www.eitdigital.eu/news-events/events/article/eit-ict-labs-partner-event-2015/
Smart Data - Deutschland und Europa auf dem Weg zu einer digitalen Datenökonomie	April 2015	Berlin	Conference (data science)	Fraunhofer IAIS – booth on data science training and promotional material sharing	http://www.bmwi.de/DE/Service/veranstaltungen,did=693428.html
PAPIS Connect	May 2015	Paris	Conference (data science)	Project details disseminated by Persontyle	http://www.papis.io/connect
VOR Divsion ICT and HE Symposium Research on MOOCs	May 2015	Wageningen, NL	Research symposium (e-Learning)	TU/e - presentation	http://www.vorsite.nl/nl/divisies-en-themagroepen/vor-divsion-ict-and-he-symposium-research-on-moocs-26-mei-2015.html
ODI Open Data Meetup	Feb 2015, May 2015	London	Meet-up (data science)	ODI	http://www.meetup.com/Open-Data-London/
Open University visit by UK HE Commission	June 2015	Open University, UK	Meeting	OU – presentation	n/a
Data Literacy Workshop at WebSci'15	June-July 2015	Oxford, UK	Conference (e-Learning, data science)	Southampton – talk and flyer dissemination	http://www.dataliteracy.eita.org.br/

3.3 Press Coverage

Since the start of the EDSA project, a number of articles have been released to the press detailing the project's scope, aims and ambitions (Table 8). These initially began with press releases from partners regarding the start of the project, along with news articles explaining what the funding was going towards. More recently, EDSA has been featured in an article on 'tech.eu' describing the initiative to fill the data science skills gap.

Table 8 - EDSA Press Coverage

Date	Coverage Type	Link
November 2014	Press Release (ODI)	http://theodi.org/news/odi-helps-unlock-11m-for-open-data-innovation-in-europe
November 2014	News Articles	For example: http://www.theguardian.com/technology/2014/nov/04/eu-commits-144m-to-support-open-data-across-europe
January 2015	Press Release (OU)	http://www3.open.ac.uk/media/fullstory.aspx?id=28449
February 2015	Press Release (Southampton)	http://www.southampton.ac.uk/mediacentre/news/2015/Feb/15_35.shtml#.VYPhL2BhhRo
March 2015	Course Advertisement – Process Mining MOOC (TU/e)	http://www.kdnuggets.com/2015/03/coursera-process-mining-data-science-action.html
April 2015	Online Article (Tech.eu)	http://tech.eu/features/4339/data-science-big-data-europe/
Tbc 2015	Press Release (Southampton alumni site)	Tbc.

4. Dashboards

As part of the project, we are creating a number of dashboards to present and disseminate information about EDSA. One of these will display information related to demand analysis in WP1, showing up-to-date results regarding this work and allowing anyone interested in data science to be able to see where the demand for particular skills resides. Another dashboard will present learning analytics from WP3 to demonstrate the engagement learners are showing with the teaching materials, and to help us revise and restructure the curricula and course content. Finally, we will develop a dashboard using data from online engagement channels (such as Twitter, the Website and LinkedIn) to showcase the levels of community engagement such as conversations happening about EDSA, and the number of downloads and views of particular dissemination materials. This will facilitate the ongoing monitoring of the effectiveness of our engagement strategies, allowing us to adjust in an agile manner to ensure that we continue to communicate productively with the EDSA audience.

EUROPEAN DATA SCIENCE ACADEMY

5. Qualitative Summary and Lessons Learned

The preceding discussion has revealed encouraging statistics for the engagement with the EDSA project. As expected, there was an apparent spike of interest around and following the project kick-off meeting in March, with Figure 3, Table 3 and Table 4 all reflecting this through website hits, Twitter follower growth and Tweet impression statistics respectively. This also marked the main initiation of the community building and dissemination campaign, which will have drawn in a large number of viewers. As the project progresses, we will be able to better understand the baseline level of engagement to expect on these two sites, and through knowing this we will implement strategies in order to increase it. After the spike in activity following the kick-off meeting, it is expected that a similar spike will occur when the first version of the curriculum is released and EDSA branded learning materials begin to be released. One area which will be targeted going forward is the diversification of audience to the website from around the EU – currently the UK, USA and Germany contribute higher proportions of website hits, and it is imperative that we attract other EU member states to increase their visits in order to address the skills gaps across the Union, and not just in the largest and most obvious markets.

The first of these courses – the Process Mining MOOC from TU/e – received nearly 25000 registered students, reflecting the potential for MOOCs to attract massive class sizes. Learning analytics being undertaken in WP3 are looking at the different student journeys through the materials, in order to better understand the differences between those students who go on to complete the end of course certificates and those who don't. We have seen from work such as that by Kizilcec et al. (2013) that learners on MOOCs do not always follow traditional patterns of viewing content and completing examination procedures and many students often labelled as 'noncompleting' are actually just learning through the course in different ways. Therefore the statistics for MOOC engagement presented above do not tell the whole story, and instead provide an overview of the audience size obtained for the course in general – the work in WP3 will help us to establish overall engagement and success rates for the course, and future EDSA courses.

Having established the top community groups on LinkedIn within which it would be suitable for EDSA to discuss course material, disseminate materials and solicit feedback, we are now in a position to take advantage of this network as the course material begins to be created as part of WP2. It is encouraging that within the top groups discovered that are relevant to EDSA there is a mix of topics including data mining, business intelligence and big data analytics that are of particular focus in each, and that each cater to different groups of user. This means that we can target a wide range of feedback for the work in EDSA, helping to ensure that we assess the demand for skills accurately and produce learning resources that are suitable to address them.

6. Conclusions

In this deliverable, we have provided the first assessment of the audience that we have reached through our communication, dissemination and general community-building efforts. We have provided summaries of the audiences obtained through our online channels: the website and Twitter, as well as an assessment of community groups on LinkedIn where there are existing audiences that the project will target. Next, an overview of the audience for the first EDSA-branded course was provided, showcasing early statistics around the engagement with an EDSA MOOC. Finally we offered an overview of real-world events that the EDSA consortium has attended since the start of the project, where each event has offered an opportunity for dissemination of the project, as well as a list of press releases covering the project. By carrying out this analysis now, and throughout the project, we place a high level of importance on the having informed measurements and as such we are developing analytic dashboards to reflect our commitment to monitoring these trends. This will help us to continue and improve this work for the remained of the project, during which we expect engagement rates to increase significantly.

7. References

René F. Kizilcec, Chris Piech, and Emily Schneider. 2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13), ACM, New York, NY, USA, 170-179. DOI=10.1145/2460296.2460330

8. Appendices

8.1 Appendix A.- LinkedIn Data Science Groups

The following pages contain a table showing an overview of data science groups on the LinkedIn social network. This helps to provide us with an idea of which communities are the most populous and active and therefore make suitable channels through which to disseminate EDSA updates.

EUROPEAN DATA SCIENCE ACADEMY

Group Name	Activity	Url	Type	Member Count	Audience/Members	Owner/Contact	Created
Advanced Business Analytics, Data Mining, and Predictive Modeling	Very Active	https://www.linkedin.com/groups/Advanced-Business-Analytics-Data-Mining-35222/about	Professional Group	180,113	Data science, big data, visualization, business analytics, predictive modeling, data mining, web analytics, six sigma, econometrics, business intelligence, computational finance, quant, operations research, machine learning, data analysis, data warehousing, risk management, cloud computing, text mining, BI.	Vincent Granville	September 28 2007
Business intelligence Professionals	Very Active	https://www.linkedin.com/groups?gid=40057&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431697147111%2CVSRPtargetId%3A40057%2CVSRPcmpt%3Aprimary	Professional Group	152,016	BI professionals	Rajasekar Nonburaj	October 1 2007
Big Data, Analytics, Strategy, Finance, Innovation	Very Active	https://www.linkedin.com/groups?gid=1814785&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431697207786%2CVSRPtargetId%3A1814785%2CVSRPcmpt%3Aprimary	Networking Group	131,392	CEO, CIO, CTO, CDO etc	Josie K	February 20 2009
Big Data and Analytics	Very Active	https://www.linkedin.com/groups?gid=4332669&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431697276071%2CVSRPtargetId%3A4332669%2CVSRPcmpt%3Aprimary	Networking Group	126,761	People interested in big data	Sarah Howes	March 1 2012

Business Intelligence, Big Data, Analytics, MIS Reporting and Database Group	Very Active	https://www.linkedin.com/groups/Business-Intelligence-Big-Data-Analytics-23006/about	Professional Group	95,840	Professionals	Sanjay Mehta	September 24 2007
Python Community	Very Active	https://www.linkedin.com/groups?gid=25827&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431698680204%2CVSRPtargetId%3A25827%2CVSRPcmpt%3Aprimary	Networking Group	85,491	Professional developers and users	Danny Adari	September 25 2007
Data Mining, Statistics, Big Data, Data Visualization, and Data Science	Very Active	https://www.linkedin.com/groups/Data-Mining-Statistics-Big-Data-152247/about	Professional Group	71,253	Data mining and stats professionals	Jon Francis	July 25 2008
Hadoop Users	Very Active	https://www.linkedin.com/groups/Hadoop-Users-988957/about	Professional Group	65,555	Hadoop experts and professionals	Jay Kreps	October 7 2008
Big Data, Analytics, Hadoop, NoSQL, & Cloud Computing	Active	https://www.linkedin.com/groups/Big-Data-Analytics-Hadoop-NoSQL-3990648/about	Professional Group	35,892	Interests in: Big Data, NoSql and Cloud Computing. Like Big Data Analytics, BI, SAS, Big Data Hadoop, Big Data MapReduce, YARN, PIG, Hive, HQL, Oozie, Sqoop, Flume, Kafka, Spark, Big Data NoSQL, HBase, Casandra, MongoDB, HUE, Pentaho, Teradata, BigTable, DataFu, Mahout, Big Data Hadoop Distributions MapR, Cloudera, HortonWork, Zookeeper,	VenkataHari Shankar	July 6 2011

					Cloud Computing, SaaS, PaaS, IaaS & Virtualization, Amazon Web Services Mapr M5, EC2 MapReduce, S3, DynamoDB, Amazon EMR (Elastic MapReduce) uses Big Data Hadoop		
Python Professionals	Active	https://www.linkedin.com/groups?gid=1846027&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A10359904814316998734220%2CVSRPtargetId%3A1846027%2CVSRPcmpt%3Aprimary	Networking Group	33,963	Professional developers and users	Frank Tang	March 13 2009
Data Warehouse, Big Data, Hadoop, Predictive Analytics	Very Active	https://www.linkedin.com/groups?gid=1824590&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A10359904814316997426953%2CVSRPtargetId%3A1824590%2CVSRPcmpt%3Aprimary	Professional Group	33,717	Professionals	Chris Rosser	February 4 2009
Machine Learning Connection	Very Active	https://www.linkedin.com/groups?gid=70219&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431696832774%2CVSRPtargetId%3A70219%2CVSRPcmpt%3Aprimary	Networking Group	24,661	Academic leaders	Shane Threatt	March 12 2008
Pattern Recognition, Data Mining, Machine Intelligence and	Active	https://www.linkedin.com/groups?gid=961087&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431697032405%2CVSRPtargetId%3A961087%2CVSRPcmpt%3Aprim	Professional Group	22,113	Those interested in pattern recognition, data mining, machine intelligence and leaning.	Belur V. Dasarthy	October 2 2008

Learning		ary					
Machine Learning and Data Science	Very Active	https://www.linkedin.com/groups?gid=4298680&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431695942156%2CVSRPtargetId%3A4298680%2CVSRPcmpt%3Aprimary	Professional Group	21,796	Big data practitioners	Richard Snee	Feburary 10 2012
Research Methods and Data Science	Very Active	https://www.linkedin.com/groups/Research-Methods-Data-Science-1895501/about	Professional Group	18,631	Researchers and analysts	Alex Liu	April 10 2009
IBM Big Data and Analytics	Very Active	https://www.linkedin.com/groups/IBM-Big-Data-Analytics-4014567/about	Professional Group	15,117	Professionals	Bruce Weed	July 22 2011
Big Data, Analytics and Data Science Training	Very Active	https://www.linkedin.com/groups?gid=4989164&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431696153151%2CVSRPtargetId%3A4989164%2CVSRPcmpt%3Aprimary	Professional Group	14,934	Schools, trainees and practitioners	Vincent Granville	May 2 2013
Big Data, Hadoop, Analytics and Cloud Technology Professionals	Active	https://www.linkedin.com/groups?gid=1346907&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431698341687%2CVSRPtargetId%3A1346907%2CVSRPcmpt%3Aprimary	Networking Group	14,181	Cloud computing professionals	Nareshkar Pakanati	November 24 2008
RDataMining: R and Data	Very Active	https://www.linkedin.com/groups/RDataMining-R-Data-	Professional Group	12,712	everyone using R	T=Yanchang Zhao	August 31 2011

Mining		Mining-4066593/about					
Big Data Community	Active	https://www.linkedin.com/groups?gid=4520336&trk=vsrp_groups_res_name&trkInfo=VSRPsearchId%3A1035990481431697634647%2CVSRPtargetId%3A4520336%2CVSRPcmpt%3Aprimary	Professional Group	12,489	business analysts, computer and data scientists, data architects, statisticians, and executives.	Vincent Granville	July 8 2012
Big Data Anlaytics and Hadoop	Very Active	https://www.linkedin.com/groups/Big-Data-Analytics-Hadoop-4488721/about	Professional Group	10,079	Learning and exchanging ideas	Vamsi Krishna Pannala	June 14 2012
KDNuggets Analytics, Data Mining and Data Science	Very Active	https://www.linkedin.com/groups/KDnuggets-Analytics-Data-Mining-Data-54257/about	Professional Group	9,129	Data science professionals and researchers	Gregory Piatetsky-Shapiro	February 4 2008
Data Visualization	Very Active	https://www.linkedin.com/groups/Data-Visualization-2244682/about	Professional Group	8,873	People in arts and sciences	Mayo Todorovic	August 28 2009
Data Science and Machine Learning	Moderately Active	https://www.linkedin.com/groups/Analytics-Big-Data-Data-Science-4377095/about	Professional Group	6,625	Business Analytics, Data Mining, Big Data, Data Science, Predictive Modeling, Quant, Operations Research, Text Mining, Web Analytics.	Vincent Granville	March 29 2012
The Big Data Forum	Active	https://www.linkedin.com/groups/Big-Data-forum-6534193/about	Professional Group	6,593	Big data professionals	Dez Blanchfield	October 23 2013
Python Data Science and Machine	Moderately Active	https://www.linkedin.com/groups?gid=4388870&trk=vsrp_groups_res_name&trkInfo=VSRPs	Networking Group	5,288	Data scientists	Lynn Bender	April 6 2012

Learning		earchId%3A103599048143169 6890054%2CVSRPtargetId%3A 4388870%2CVSRPcmpt%3Apri mary					
Data Science Community	Active	https://www.linkedin.com/gro ups/Data-Science-Community-3063585/about	Networking Group	5,248	Data scientists	Harsh Singhal	May 19 2010
Pivotal Data Science Group	Active	https://www.linkedin.com/gro ups?gid=4505921&trk=groups %2Fabout-r-subgr-subgrpname&goback=%2Eanb_ 4298680_*2_*1_*1_*1_*1_*1	Professional Group	4,717	Big data practitioners	Tim Madison	June 27 2012
Data Science Collective	Active	https://www.linkedin.com/gro ups/Data-Science-Collective-7483515/about	Networking Group	2,650	Job opportunities	Niall Wharton	March 21 2014
Big Data, Hadoop, Data Science, Statistics, Analytics, Machine Learning, Deep Learning	Active	https://www.linkedin.com/gro ups/Big-Data-Hadoop-Data-Science-7434866/about	Networking Group	2,208	Learning and exchanging ideas	Chris Pearson	January 6 2014
Data Mining and Machine Learning	Moderately Active	https://www.linkedin.com/gro ups/Data-Mining-Machine-Learning-4543555/about	Professional Group	2,044	Machine learning, data mining and NLP	Willson He	July 24 2012