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

D4.3 Real-World and Online Community Engagement Plan

Deliverable Editor:	**Christopher Phethean (University of Southampton)**
Other contributors:	**Elena Simperl (University of Southampton)**
Deliverable Reviewers:	**Alexander Mikroyannidis (The Open University)**
	Mandy Costello (ODI)

Deliverable due date:	**30/04/2015**
Submission date:	**30/04/2015**
Distribution level:	**Public**
Version:	**1.0**

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

Change Log

Version	Date	Amended by	Changes
0.1	30/03/2015	C Phethean	Main content
0.2	31/03/2015	C Phethean	Removed plan for wiki. Added section on MOOCs. Added section on dashboards.
0.3	01/04/2015	C Phethean	Added reviewers
0.4	15/04/2015	C Phethean	Added tables of existing online channels, plus details of the EDSA Twitter, Vimeo and SlideShare accounts in line with reviewer feedback. Added List of Tables.
0.5	16/04/2015	C Phethean	Implemented changes suggested in reviewer feedback.
0.6	22/04/2015	C Phethean	Wording changes, final modifications.
1.0	27/04/2015	A Tumilowicz	Final QA

Table of Contents

List of Tables

List of Figures

1. Executive Summary

This deliverable contains the initial plans for the EDSA project to engage the community through a combination of real-world events (consisting of offline, face-to-face meetings) and online communication and dissemination services. As the data science community is inherently broad, consisting of researchers, industry-specialists, policy-makers and decision makers all invested in a range of data-related topics, it is vital that EDSA engages with the community where and when possible in order to bring in rich and varied feedback from every angle. Attending offline events such as conferences and meet-ups where this will be realised forms one part of our engagement strategy; the other is based around the online community. Using a mixture of broadcast channels (for disseminating news and updates on the project) and responsive channels (for encouraging feedback and conversation) we aim to develop a strong level of support around the EDSA agenda, which can be leveraged in order to tailor EDSA activities to the community. This engagement activity is therefore an essential component of the EDSA project and will be vital in order to ensure that on-going feedback is received, and that the resources developed for the project are disseminated to a wide-ranging audience.

2. Introduction

The community engagement plan will present the methods through which it is feasible - within the project's constraints - to enable the consortium to:

- Develop and promote EDSA as a brand;
- Disseminate project outputs, such as deliverables, reports and publications, along with participatory materials, including those for WP1 to encourage participation in the demand analysis.
- Encourage use of the EDSA learning materials;
- Solicit feedback from the community on the content and structure of the curriculum, and the content and delivery of the learning materials.

By providing an environment whereby stakeholders can discuss and comment upon the developments within the project from outside, we will facilitate collaboration that will ensure the materials produced for EDSA remain in line with the demands and expectations elicited throughout the demand analysis in WP1. Because the data science environment is fast changing it is imperative that this input is maintained throughout the project so that outputs remain closely aligned with the demands of the community.

Since data science is a broad subject area, encompassing numerous communities from academia, industry and government, it is essential that EDSA both seeks out face-to-face meetings with the community, and performs a prolonged engagement strategy that encompasses a variety of social communication channels. As such, the community engagement plans covered in this document are broken down into those focusing around 'real-world' engagement, taking advantage of existing events which already bring the community together, and online engagement to provide up-to-date information and regular interaction with stakeholders.

3. Events for Real-World Community Engagement

With such a focus - and indeed a reliance - on community collaboration and engagement in this project, it is imperative that members of the EDSA consortium are actively attending real-world events to disseminate news of the project, and to offer training relating to the EDSA curriculum. These two categories of event are now focused upon to discuss the various activities through which EDSA will engage with the community in a real-world setting.

3.1 Conferences and Meet-ups

In order to create connections with key stakeholders in the data science agenda, we will attend a variety of events. As highlighted by reports into the data science skills gap, there is the need for the community to establish the key concepts that need to be covered in the EDSA curriculum, and interacting with these individuals at face-to-face events will be invaluable in carrying out this process. Two themes will be focused upon here, each vitally important due to the nature of EDSA: data science as a subject itself and eLearning as a means by which to deliver this training and education.

3.1.1 Data Science-themed Events

The European Data Forum[1] will be an ideal venue to achieve many of the community engagement plans described above. Bringing together a mix of industry, academia and policymakers interested in Big Data and the Data Economy, this event is exemplary of the type of meet-up and discussions that EDSA needs to be involved with. Likewise, the ICT 2015 meeting[2] will bring together the EU's leading ICT-based researchers, which presents a great opportunity for networking and to broaden the current EDSA community. The EIT ICT Labs Partner Event 2015[3] also offers a fantastic chance for networking with other labs researching innovative areas of ICT in which data science will be an essential component, and will help us to promote information exchanges with other related projects.

The Research Data Alliance (RDA) holds plenary meetings regularly every six months. The next event, in September 2015 will occur in Paris[4], bringing a global community of big data specialists and data scientists together to discuss and hear about a wide range of developments in the area of 'data'. Also in Paris, the Papis Connect[5] event offers the chance to meet with other data scientists - particularly those in the machine learning field - and with them talk to decision makers in order to help them understand the value and potential benefits of using data. The EDSA consortium already has registered representation and participation at this event through Persontyle. In addition, Persontyle will also host a number of community-focused events including their Open Data Science Fellows program, and a Deep Learning meet-up in London, and the ODI will organise and run an Open Data London meet-up to discuss a number of topics related to open data with policy makers, researchers and entrepreneurs among the intended audience

Fraunhofer IAIS will also represent the consortium at a number of events. There will be opportunities to engage with a range of community categories - such as managers, decision makers and researchers -

[1] http://2015.data-forum.eu
[2] http://ec.europa.eu/digital-agenda/ICT2015
[3] http://www.eitictlabs.eu/news-events/events/article/eit-ict-labs-partner-event-2015/#allView
[4] https://rd-alliance.org/plenary-meetings/rda-sixth-plenary-meeting.html
[5] http://www.papis.io/connect

interested in big data at the BITKOM Big Data Summit[6] and CeBIT 2016[7]. Likewise the Smart Data Conference (April, 2015)[8] will provide a similar audience discussing the potential innovation opportunities and research agendas for the digital data economy.

3.1.2 eLearning-themed Events

It will also be important to collaborate with researchers at the forefront of developments in e-Learning techniques and technologies. As an example, a primary focus in the first year will be at Online Educa, in Berlin in December 2015. Events such as this will allow us to promote the EDSA agenda, and gain new insights from a community that overlaps with EDSA in the learning and training delivery side of the project, rather than the data-oriented subject side. This will help us to ensure that as well as having a relevant curriculum, we also take advantage of the best possible approaches to delivering technology-facilitated education and training.

3.2 Focus Groups

Focus groups will play an integral role in the demand analysis for WP1 and – as well as helping to elicit the different requirements between industry sectors – they serve an additional function in allowing direct face-to-face interaction with communities of managers and data scientists from each of these sectors. These will therefore form an important part of our engagement strategy, as lasting and valuable relationships can be developed through these events. Being oriented around smaller groups, these events also allow for the possibility of richer and deeper interaction with each participant, and this will demonstrate the importance of having the face-to-face element in this strategy.

3.3 Training and Dissemination of EDSA curriculum

As well as providing electronic delivery of EDSA learning materials, each EDSA partner will also carry out data science training in the real world, offered through their existing channels of delivery and utilising their individual specialisations. This will be carried out through formalised, structured courses, such as the University of Southampton's MSc in Data Science, but also through more bespoke and tailored training offers to interested stakeholders, such as government, industry and public-sector organisations – we will aim to create relationships with these organisations through the engagement strategy and then develop these further by providing such training. Through this we aim to engage with organisations as much as possible on a partner-by-partner basis to showcase individual strengths but also the EDSA community and curriculum as a whole.

[6] http://www.bitkom-bigdata.de/programm/2015
[7] http://www.cebit.de/home
[8] http://www.bmwi.de/DE/Service/veranstaltungen,did=693428.html

4. Channels for Online Community Engagement

In parallel with the real-world community engagement events described above, we distinguish two groups of online community engagement strategies: those through which we will broadcast information around the EDSA project and disseminate updates, and those through which we will monitor discussions and engage in responsive conversation to draw people towards the EDSA learning materials. While there is naturally a lot of overlap between the activities carried out on each, the differing requirements around long-term engagement and feedback-solicitation mean that each channel requires consideration for the possibility of both outcomes.

4.1 Broadcast Channels

In order to develop the community around EDSA it will be necessary to disseminate relevant information about the project, including milestones and developments, to the public so that we can establish the EDSA brand identity and inform interested parties about the latest news. As such, an inherent part of the communication and engagement plan must take into account the process of generating awareness around the brand, so that engagement can be built upon this and ultimately lead into take-up of the rich resources we produce in the MOOCs, iBook, webinars and other materials.

By separating this strategy from that of engaging the community in discussion around feedback and collaboration, we can also break down the success metrics employed to assess the community performance. This will be covered more in Section 4, but allows us to track the sizes of these communities as a measure of the audience and the potential awareness that can be generated through their utilisation.

4.1.1 Social Media

Fundamentally, social media channels including Twitter and LinkedIn can be employed within this project to increase awareness of the activities being undertaken, and of the results and outputs produced. These services can provide the first point of call for the project, providing a presence and identity on the Web whereby people can quickly and easily find the latest news, and subsequently be drawn to work of EDSA. As such these sites - especially Twitter - are particularly suited to the dissemination and sharing of updates, but by carrying these out in a regular and informative manner means that we can create a unified social presence that creates prolonged and established engagement with the project. For LinkedIn we will investigate the relevant data science groups on the site in which data scientists currently discuss topics of this nature; by getting involved with these through our own EDSA account we can reach a specific subset of the community and lead them towards either consuming the learning materials, or providing feedback to inform our demand analysis or to improve the curriculum. On both Twitter and LinkedIn, we will also use the EDSA account as an aggregator for what each of the partners is posting about – rather than duplicating their work and ignoring their existing, established communities, we will monitor their activity and use this to promote EDSA and encourage further awareness of the project where possible. While both popular channels, we do not initially plan to focus on either of Facebook or Google Plus, although the strategy will remain agile and flexible to incorporate these networks should it become apparent that an EDSA presence is required.

Table 1 - Example LinkedIn groups to follow

LinkedIn Group	URL	Current Members
Advanced Business Analytics, Data Mining and Predictive Modeling	https://www.linkedin.com/groups/Advanced-Business-Analytics-Data-Mining-35222	> 175000
Big Data, Analytics and Data Science Training	https://www.linkedin.com/groups?home=&gid=4989164&trk=anet_ug_hm	> 14000
Machine Learning and Data Science	https://www.linkedin.com/groups?home=&gid=4298680&trk=anet_ug_hm	> 21000
KDnuggets Analytics, Data Mining, and Data Science	https://www.linkedin.com/groups/KDnuggets-Analytics-Data-Mining-Data-54257	> 8000

4.1.2 Mailing Lists

As with social media channels such as Twitter, discussed above, mailing lists can be utilised to disseminate details of updates to the project. We cover this channel in more detail as a responsive channel due to the opportunity for extended, longer-length responses and discussions to be carried out, and provide some examples of lists to follow. However as a broadcast mechanism, setting up a mailing-list will allow us to share more information about new milestones that the project reaches, and this can be used as part of the unified social strategy to either direct people to the project website, encourage communication on the list, or point people to the social media channels where further and more rapid communication may have occurred. We will create a project-based mailing list that will be publicly available so that anyone interested in EDSA can sign up and hear the latest news about the project.

4.1.3 SlideShare

For the presentations produced by the EDSA partners and used in the webinars and other learning materials, we will have a dedicated EDSA account through which these can be made available to the public. This will ensure that the learning materials are available to members of the community who are already aware of the project, but can also allow further potential users to discover the project and the courses that we offer. As such, while this site will primarily focus on disseminating our presentation material, it will also act as a broadcast channel to inform the wider community about EDSA.

4.1.4 Video Channel

With an account on Vimeo, an online video distribution channel, EDSA has the opportunity to disseminate rich media content as well as the textual and static updates provided through social media channels such as Twitter and LinkedIn, discussed above. Videos that we will release may include screencasts or screen captures with voice-over explanations, as tutorials, or recordings of presentations around the content of the core curriculum. We will push to make all videos released on

Vimeo available under a Creative Commons CC-BY license so that the community is free to share and reuse the content if attribution is given. This will help to spread the reach of the EDSA content, and widen the community around the project. In addition, using the VideoLectures.net platform in WP3 for delivering content and analytics, we have another channel through which to provide this type of content and increase awareness of the learning materials produced by the EDSA consortium. Webinars will be provided of the core curriculum as a means of disseminating the essential elements of the curriculum, forming the online equivalent of the real-world training described earlier. Feedback from these events, as with all channels, will be fed into the overall process of curriculum development and refinement focused upon in WP2. This will ensure that the curriculum and learning materials remain up-to-date and in line with the demands and desires of the wider community.

Table 2 - Social Media Account Details

Social Media Platform	Link
Twitter	https://twitter.com/edsa_project
Vimeo	https://vimeo.com/edsa
SlideShare	http://www.slideshare.net/edsa-project

4.2 Responsive Channels

While utilising many of the same platforms as the broadcast channels above, the responsive strategies through online channels will aim to increase collaboration and conversation around EDSA-related topics. We seek to engage with experts and interested practitioners from outside of the project so a discussion environment that supports this is essential. These channels build on the awareness generated from the broadcast work and seek to encourage productive discussion that elicits feedback and suggests which can be used to further refine and improve EDSA material throughout the project.

4.2.1 Mailing List

As discussed above, the EDSA mailing-list that will be set up will provide considerable opportunities for broadcasting outputs of the project. In addition to this, this channel affords further contributions from the community, enabling conversations to form in response to the mails, and also to allow community-contributed messages to spark new and further conversations.

Mailing lists that will be targeted include:

- rda-edu-ig@rda-groups.org: The Research Data Alliance mailing list for the Education and Training on handling of research data special interest group.
- discuss@ml.jhu.edu: The John Hopkins University machine learning group mailing list.
- euopendata@lists.okfn.org: Open data in Europe for the Open Knowledge Foundation.
- open-education@lists.okfn.org: Open education – including educational data and educational resources, from the Open Knowledge Foundation.
  public-lod@w3.org: The Linking Open Data project mailing list from the W3C

4.2.2 Community Blog

The community blog will facilitate the publication and promotion of community-contributed articles around the data science topic. By providing this, and encouraging submissions from within the wider public, we can increase the overall community and strengthen the EDSA brand image. A blog allows the community itself to broadcast relevant information and share their ideas, but through the

comments feature allows for a discussion to occur around the post where other members of the community, and members of the EDSA consortium can reply and engage. The blog will be hosted on either WordPress or Blogger, both of which allow robust, Web-based editing systems for publishing content and are simple and fast to setup and maintain.

4.2.3 Social Media

As well as offering an incredibly valuable channel for disseminating regular updates and information about the EDSA project, social media enable us to remain in contact with the community. This means that we can encourage engagement on these sites around the content that we share about the project, and by sustaining this can produce a community which is driven around collaboration and contributions from both the EDSA consortium itself, and the data science community interested in the outputs.

4.2.4 MOOCs

Primarily utilising the FutureLearn[9] platform, we will provide particular sections of the EDSA curriculum as Massive Open Online Courses (MOOC). These will be available online and contain all the learning resources – including articles, exercises and quizzes – necessary for a learner to gain an understanding in a particular area. The first MOOC related to EDSA has already been published by TU/e on the Coursera site[10] and exemplifies this by offering a 6-week online course in process mining. By using outlets such as these, each EDSA partner will be able to offer a dedicated course in areas that they specialise in related to data science. On platforms such as FutureLearn, education is largely community-driven and there is a strong emphasis on soliciting discussions and contributions from the learners themselves in order to share knowledge and improve understanding between classmates. Additionally from the comments and discussions received on these sites we can incorporate the feedback of students who are actually experiencing the courses that EDSA offers, and can subsequently shape them to fit the needs of the community.

4.2.5 Videolectures

As discussed above in the video channels for broadcasting online, videolectures.net will be used by the project to deliver content, targeting a range of user groups. The videolectures platform allows us to collect feedback on the content so that an analysis of viewing patterns can be carried out. We will tie this in with our broader analysis plans so that we can develop evidence-based best practices in WP3 that can be used to improve the curricula, content and delivery of training material. Commenting facilities allow viewers to provide qualitative feedback, as well as to begin discussions around the content with other viewers.

[9] https://www.futurelearn.com

[10] https://www.coursera.org/course/procmin

5. Measuring Engagement Effectiveness

At 18 months, D4.4 will provide an interim report on participation in events and feedback regarding EDSA's participation in both real-world and online activities. Depending on these results, the engagement plan presented in this document may be adapted in order to ensure that the potential for engagement in as many events and channels as possible is realised. Finally at 36 months, a final community engagement and networking report will be delivered in D4.5 that will discuss the overall result of EDSA's engagement and community building efforts, including the feedback received from the community. We will use this to outline a plan for potential adoption of successful events and processes in other projects by producing a longevity plan as part of D4.5.

5.1 Real-World Engagement Assessment

Reporting on the impact of real-world engagement activities will assess the number of events attended, and what category (e.g. data science-related, e-Learning or training) they fall into. Furthermore we will also assess the number of participants and the number of attendees of the events where applicable to gauge the potential networking that could have resulted. We will encourage feedback from the community, and when hosting real-world training events, we will gather this specifically through surveys in order to obtain data on:

- The current role or position of each attendee;
- Their experience with data science topics;
- Their level of understanding of the particular training topic before the event;
- Their subsequent understanding of the topic;
- The quality of the teaching materials;
- Their awareness of the wider EDSA agenda;
- Their level of interest in the EDSA iBook, MOOCs, webinars and other learning materials;
- What they believe the EDSA curriculum needs to cover.

5.2 Online Engagement Assessment

For the broadcast channels discussed above, it is notoriously difficult to measure the resulting engagement of this type of communication[11] and therefore a number of generally accepted measures of awareness will be adopted in order to gain an impression of the 'reach' of this work. This will consist of:

- The size of the community - number of followers, members, viewers or registered 'signups' of the channel;
- The spread of EDSA content - number of shares (e.g. retweets) of announcements and of EDSA content.

[11] http://searchenginewatch.com/sew/how-to/2358553/want-to-measure-social-roi-start-with-these-5-cross-channel-metrics

- Conversations - the number of replies or comments received in response to particular pieces of content. Some analysis can be carried out on this data to assess how regularly these come from 'repeat engagers', or whether most users just engage once.

For the responsive channels, many of these same metrics will be employed to judge the effectiveness. However a greater emphasis will be placed on measuring the community contributions to each, such as the frequency of blog posts, and number of messages sent by each communicating member of one of the communities. We will also monitor web server logs and analytics regarding the main project website's access, to determine the level of traffic generated by all of these activities.

5.3 EDSA Dashboards

In order to disseminate certain elements of the EDSA project's work, we will produce a number of interactive dashboards that present data in a way that is engaging and intuitive for the public. A sample of this is available at the EDSA website[12] (Fig.1) where data is taken from LinkedIn job postings and processed to visually show the demand in various EU countries for a range of related skills.

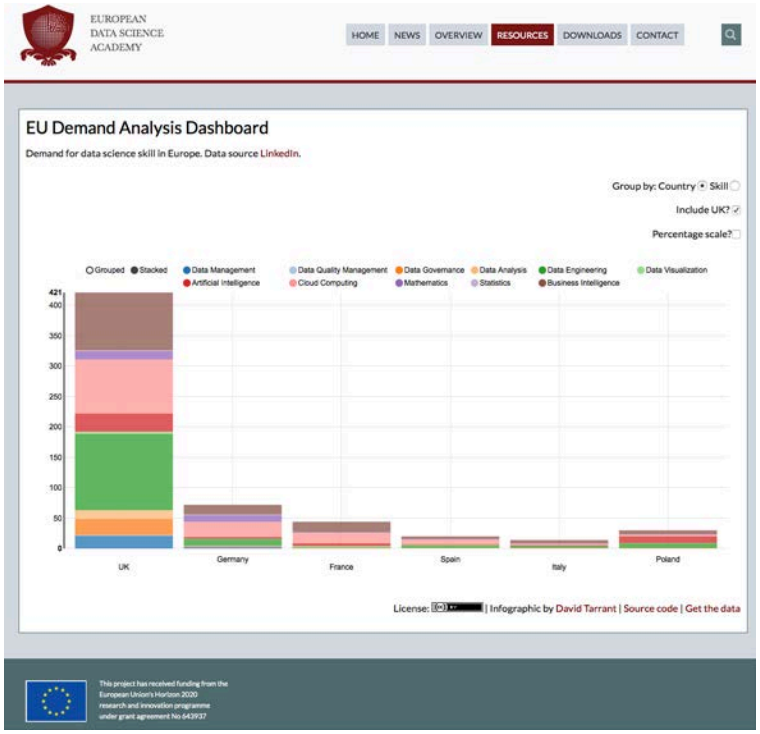

Figure 1 - EDSA Dashboard for Demand Analysis

This dashboard will showcase up-to-date information regarding the demand for skills that EDSA will develop through the training provided. Providing this resource will allow the community to quickly assess the demand for particular skills and notice the relevance and importance of the EDSA curriculum (Examples in Fig.2 and Fig.3). Therefore the visualisation of the demand analysis carried out in WP1 will create a fantastic opportunity to encourage more engagement from communities

[12] http://edsa-project.eu/resources/dashboard/

already invested and interested in the data science agenda. The dashboard will feature a 'click to contribute' button to allow visitors to the site to contribute data for their own country or sector, allowing the information base used to create the visualisations to grow and develop based on the contributions of the community. To further enhance interaction around this, social sharing buttons will allow visitors to the dashboard to share links to the visualisations of the data, helping to increase its publicity. The demand analysis dashboard will also provide its own set of engagement analytics in terms of the number of visits, proportions of new and returning visitors, and location (country) of visitor, which will allow us to assess the reach of this production.

We will provide additional dashboards for different purposes. Using the data described above and obtained from the online engagement channels, we can provide an overview of community engagement in order to showcase the discussions and volume of conversation happening around EDSA. This will provide insights into the level of community involvement with the overall project. Furthermore we will offer a dashboard to present learning analytics sourced from WP3 that will demonstrate the engagement that learners are showing with the teaching materials. This will help in the assessment of where students are engaging with the course, and which parts require adaptation to improve retention rates.

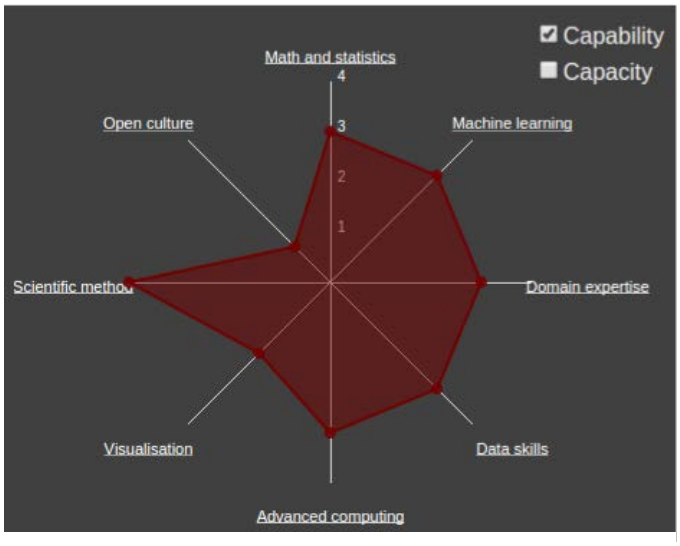

Figure 2 - EDSA Dashboard: data science capability radar diagram

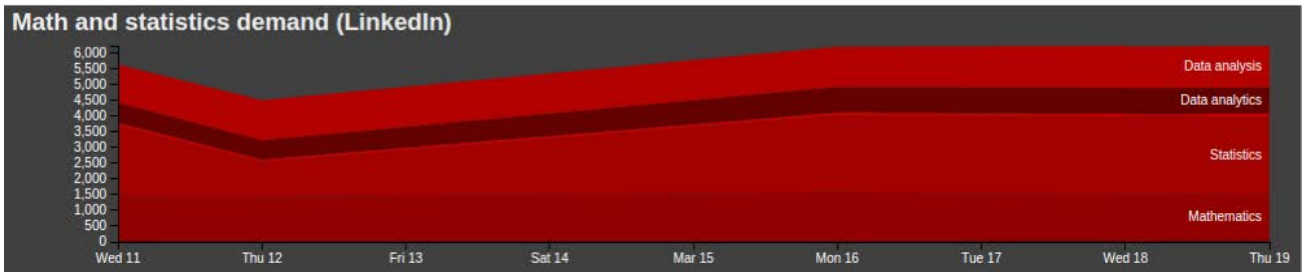

Figure 3 - EDSA Dashboard: jobs demand by skill in a particular country

6. Conclusion

In this deliverable, we have outlined the strategy for engaging with communities related to the EDSA project through a combination of online and real-world activities. The aim of this is to promote the EDSA brand and in doing so increase the community involvement in the demand analysis for data science (WP1), the take up of the EDSA curriculum, and the use of the training materials we will provide. By successfully engaging the community we hope to receive feedback on our approach that will enable us to respond in an agile manner to the changing demands of the project. In distinguishing different forms of engagement, through offline and online channels to start with, we have provided a repertoire of means for communication and considered for each the potential to judge the level of engagement generated.