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

D3.3 Report on the Evaluation of Course Content and Delivery 1

Deliverable Editor:	**Aba-Sah Dadzie (OU)**
Other contributors:	**Inna Novalija (JSI), Erik Novak (JSI), Rémi Brochenin (TU/e), Joos Buijs (TU/e), Alexander Mikroyannidis (OU)**
Deliverable Reviewers:	**Angi Voss (Fraunhofer) / Simon Bullmore (ODI)**
Deliverable due date:	**31/07/2016**
Submission date:	**29/07/2016**
Distribution level:	**Public**
Version:	**1.0**

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

Change Log

Version	Date	Amended by	Changes
0.1	27/05/2016	Aba-Sah Dadzie	ToC & analysis of EDSA LRM
0.2	03/06/2016	Inna Novalija	JSI VideoLectures Data, Data Analysis for VideoLectures, Statistical Analysis, Visual Exploration & Analysis, Prototype Interface, Prototype Implementation
0.3	06/07/2016	Aba-Sah Dadzie Rémi Brochenin	ToC edits, content
0.4	01/07/2016	Aba-Sah Dadzie	restructuring & consolidation
0.5	15/07/2016	Aba-Sah Dadzie	addressing review comments
0.6	22/07/2016	ALL	addressing review comments and finalising deliverable
0.7	25/07/2016	Aba-Sah Dadzie	submission version for scientific review
0.8	28/07/2016	Elena Simperl	Scientific review
1.0	29/07/2016	Aneta Tumilowicz	Final QA

Table of Contents

List of Tables

List of Figures

1. Executive Summary

This deliverable reports the start of work in EDSA Task 3.4: the initial evaluation of course content and delivery, by studying student interaction with online course material. This is to obtain insight into student engagement and behaviour, and the impact of student behaviour, demographics and other (external) factors on course completion and success.

Deliverable 3.3 reports exploratory analysis of event data from (i) the EDSA project's portal providing access to open, online, self-study courses in Data Science; (ii) a Coursera MOOC on a selected topic in Data Science; (iii) a portal containing learning resources captured on video on a broad variety of topics including Data Science. Each dataset is analysed independently using one or more of three different analysis techniques - statistical analysis, visual analysis and process mining. All three techniques highlighted common interest in or popularity of specific resources and courses overall; to validate both the approach taken and the appropriateness of each technique employed for the datasets and the analysis required we apply the three techniques to the most popular course in the EDSA Online Courses portal (by event and enrolment).

Triangulating our results we see both overlapping findings and information revealed only using a single approach, each of which takes a particular perspective on the data. Overall, student activity patterns show intermittent access to courses over time, with high attrition shortly after enrolment, and gradually, but not uniformly, declining toward the latter stages of a course. Access to course material varies, with some resources seeing significantly more access than others. Only some courses include assessment material; further analysis as Task 3.4 progresses will examine in more detail the impact of student behaviour and other factors on performance in both individual courses and in general for online self-study. We aim also to identify potential criteria for assessing what learning has occurred where explicit, objective assessment and/or course grades are not available.

The outcomes of Learning Analytics in Work Package 3 are to lead ultimately to evidence-based best practice that will guide (re)design of course content and delivery and programme curricula in the field of Data Science, and to aid where necessary tailoring of learning material to fit particular contexts. We envisage also that our findings will contribute to guiding students in selecting courses that meet their requirements for self-improvement and as part of the processes of skills training and job-seeking.

Because the analysis reported here is preliminary we indicate mainly further directions for research and analysis based on patterns and trends recognised as we carried out the exploratory analysis of the three datasets. Where strong evidence exists for making early recommendations we suggest also potential paths to explore further to reinforce these.

1. Introduction

The objective of WP3 in EDSA is to:

1. Deploy the course material developed in WP2 for different target groups and in different environments: webinars, video lectures and face-to-face training.
2. Gather feedback about the effectiveness of learning from these courses.
3. Analyse feedback and other data generated during course delivery, to feed into improving content and/or form of deployment, as well as into the design of new courses.

Work Package 3 (WP3) produces two series of deliverables. The first series addresses objectives 1 and 3 and the second objectives 2 and 3. Deliverable 3.3 (D3.3) is the first in the second set of deliverables, within Task 3.4 (T3.4), and focuses on initial analysis of the content of and interaction with online learning material produced within the EDSA project and by project partners on topics related to EDSA's remit.

The deliverable describes first the three event datasets on courses being delivered and relevant learning material - both in progress and complete - and the analysis approaches employed to obtain an overview of the data, followed by more in-depth analysis of selected courses and student paths. The ultimate aim is to provide a picture of student behaviour across all and in different courses and how this relates to completion, time to completion and student grades.

We triangulate the results obtained from the different analysis approaches to obtain broader reaching and more reliable conclusions on student behaviour. We aim to measure the impact of type of learner (e.g., professional or practitioner vs. college or university student), course and content delivery and topic, among others, on student success. Based on the results of our analysis and project discussions feeding into the deliverable we propose a foundation for a structured process to guide Learning Analytics (LA) within EDSA, to ensure:

1. event (user interaction) data collection captures all aspects of student activity required to obtain a sufficiently comprehensive map of behaviour that allows effective provision of feedback to both students and course instructors/owners,
2. learning analytics within the project contributes to the state of the art in the field, with, where possible, use cases and benchmarks that can be released for further reuse, as Linked Open Data,
3. the outcomes of research on learning analytics for online, self-study courses in WP3 cross-fertilise other related work in EDSA, key being (i) the analysis of face-face and blended courses, and (ii) Data Science job demand and skill analysis, to identify where and how skill-driven training can aid the closing of the skills gap recognised in the project.

1.1 Outline

Section 2 provides descriptions of the three event datasets analysed in this deliverable, licenses for reuse and the data publication plan for data produced within the EDSA project. Section 3 details the data analysis carried out for the three datasets, using one or more of the three approaches: statistical analysis, visual exploratory analysis and process mining. All three approaches are then applied to the course in the EDSA Learning Management System that saw the highest enrolment and recorded the largest number of events overall - "Foundations of Big Data". Triangulating the results obtained here (section 4.1.2) allows us to validate both the analysis approaches employed and begin to point to the

formalisation of the learning analytics process to be followed within EDSA. The deliverable concludes with a discussion of the findings of this initial exploratory analysis (section 4) and plans (section 5.1) for learning analytics for the second half of the project.

2. Online Learning Event Datasets

The analysis carried out for the first phase of Task 3.4 looks at three main datasets:

1. the EDSA project hosts a Learning Management System (LMS) that collects event data generated through interaction with the courses offered from the EDSA online courses portal[1],
2. the VideoLectures.NET[2] portal hosted by JSI that provides open access to learning material as online video,
3. the Coursera MOOC "Process Mining: Data Science in Action" delivered by TU/e.

This section provides a description of each dataset, detailing event data collected, addressing of privacy concerns and licenses for reuse within the project and by third parties.

2.1 EDSA Project Data: Learning Locker

As described in D2.4, the learning material produced by the project is delivered via the EDSA online courses portal[1]. The portal is an LMS based on the open-source Moodle[3] software. It hosts the full set of learning material (presentations, webinars, text, quizzes, etc.) for the EDSA self-study courses, to which learners can enrol and study at their own pace. Learners may register to access the portal using an existing social media account, such as Google, Facebook and LinkedIn, or request an EDSA account. The portal also lists other types of EDSA courses, both online and not - MOOCs, blended and face-to-face (f-f) courses. For each course there is a brief overview and a link to the dedicated course page on the website of the EDSA partner or associate EDSA partner that offers the course; this data is also included with corresponding event logs .

In order to collect data for Learning Analytics from the portal, we have deployed an open-source Moodle plugin[4] that conforms with the xAPI (or Tin Can API) specification. xAPI provides a framework for capturing statements such as "someone does an action to/with something", e.g., "John has initiated an experiment". It offers an extensible vocabulary of:

- *verbs*[5], which describe the action performed during the learning experience, e.g. "answered", "asked", "interacted".
- *activities*[6], which describe something with which an Actor interacts, e.g. a "course", "question", "simulation".

[1] http://courses.edsa-project.eu
[2] http://videolectures.net
[3] https://moodle.org
[4] https://moodle.org/plugins/logstore_xapi
[5] http://adlnet.gov/expapi/verbs
[6] http://adlnet.gov/expapi/activities

The following code snippet shows an example xAPI statement in JSON declaring that a user has *enrolled onto* a course (key event action in bold):


```
"actor": {
  "name": "Test user_name",
  "account": {
    "homePage": "http:\/\/www.example.com\/user_url",
    "name": "1"
  }
},
"context": {
  "platform": "Moodle",
  "language": "en",
  "extensions": {
    "http:\/\/www.example.com\/context_ext_key": {
      "test_context_ext_key": "test_context_ext_value"
    },
    "http:\/\/lrs.learninglocker.net\/define\/extensions\/info": {
      "https:\/\/moodle.org\/": "1.0.0",
      "https:\/\/github.com\/LearningLocker\/Moodle-Log-Expander": "1.0.0",
      "https:\/\/github.com\/LearningLocker\/Moodle-xAPI-Translator": "1.0.0",
      "https:\/\/github.com\/LearningLocker\/xAPI-Recipe-Emitter": "1.0.0"
    }
  },
  "contextActivities": {
    "grouping": [
      {
        "id": "http:\/\/www.example.com\/app_url",
        "definition": {
          "type": "http:\/\/id.tincanapi.com\/activitytype\/site",
          "name": {
            "en": "Test app_name"
          },
          "description": {
            "en": "Test app_description"
          },
          "extensions": {
            "http:\/\/www.example.com\/app_ext_key": {
              "test_app_ext_key": "test_app_ext_value"
            }
          }
        }a
      }
    ],
    "category": [
      {
        "id": "http:\/\/www.example.com\/source_url",
        "definition": {
          "type": "http:\/\/id.tincanapi.com\/activitytype\/source",
          "name": {
            "en": "Test source_name"
          },
          "description": {
```

```
                "en": "Test source_description"
              }
            }
          }
        }
      ]
    },
    "instructor": {
      "name": "Test instructor_name",
      "account": {
        "homePage": "http:\/\/www.example.com\/instructor_url",
        "name": "1"
      }
    }
  },
  "timestamp": "2015-01-01T01:00Z",
  "verb": {
    "id": "http:\/\/www.tincanapi.co.uk\/verbs\/enrolled_onto_learning_plan",
    "display": {
      "en": "enrolled onto"
    }
  },
  "object": {
    "id": "http:\/\/www.example.com\/course_url",
    "definition": {
      "type": "http:\/\/lrs.learninglocker.net\/define\/type\/moodle\/course",
      "name": {
        "en": "Test course_name"
      },
      "description": {
        "en": "Test course_description"
      },
      "extensions": {
        "http:\/\/www.example.com\/course_ext_key": {
          "test_course_ext_key": "test_course_ext_value"
        }
      }
    }
  }
}
```


The xAPI statements collected from the EDSA online courses portal are stored in an instance of the Learning Locker[7] software, the reference open source Learning Record Store (LRS). It offers a data repository designed to store learning activity statements generated by xAPI compliant learning activities. Among its features is the ability to visualise data via dashboards, scalability, functionality for generating custom reports, its RESTful API, as well as the functionality for exporting data as CSV and JSON. Figure 1 shows a screenshot of the dashboard for the EDSA Learning Locker[8].

[7] http://learninglocker.net
[8] http://analytics.edsa-project.eu

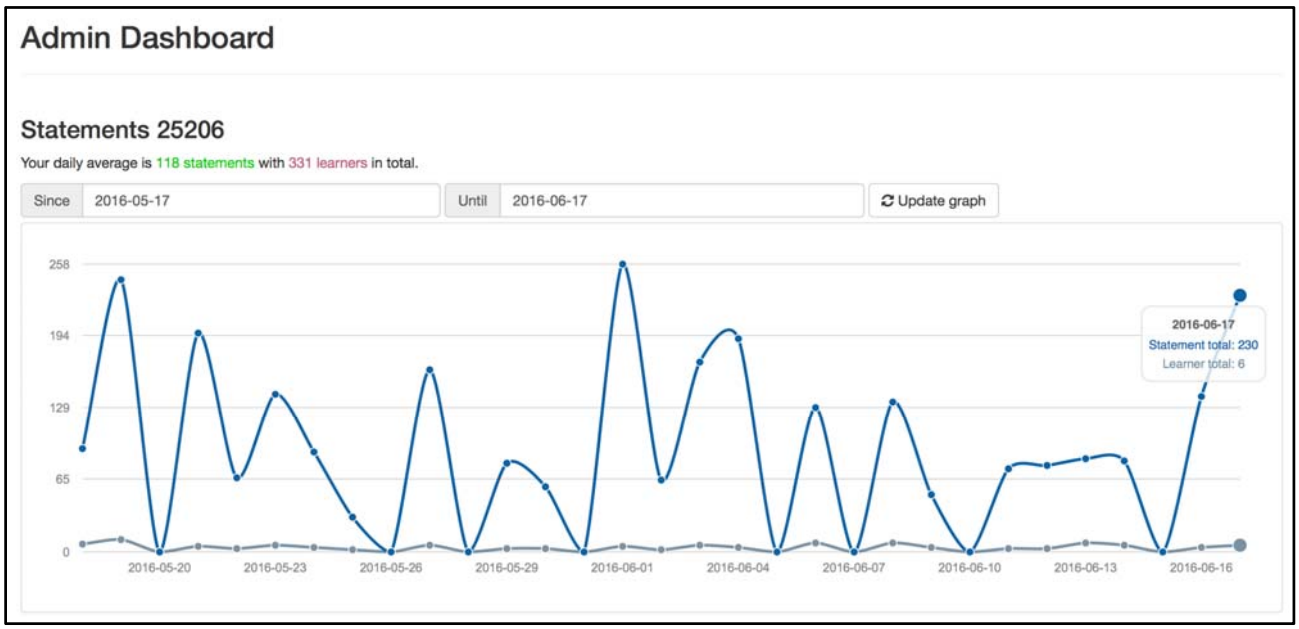


Figure 3.2: The EDSA Learning Locker dashboard.

2.1.1 Data Publication Plan

The Learning Analytics dataset from the EDSA online courses portal has been published[9] under a Creative Commons Attribution 4.0 license and will be updated quarterly. The dataset consists of xAPI statements describing the interaction of users with the EDSA online courses portal. It is anonymised to prevent identification of users, but with distinct IDs that allow the tracking of individual user paths.

2.2 Other Data in use by the EDSA Consortium

2.2.1 JSI VideoLectures Data

VideoLectures.NET[2] is a free and open access educational video lectures repository. It contains over 20,000 lectures by scholars and scientists at research events including conferences, summer schools, workshops and science promotional events.

Lectures on the portal fall into various categories, including Data Science, as established within the EDSA project.

[9] https://alexmikro.github.io/learning-analytics-dataset-from-the-edsa-online-courses-portal

For the first round of Learning Analytics tasks on VideoLectures, we use a set of *raw log* files extracted from the VideoLectures.NET portal containing events between Sep 2012 and Dec 2015. The log files processed include the following information:

- ID of entry
- timestamp
- Session ID
- log entry type
- lecture
- other information, including event type, IP address, location (if captured).

Figure 1 presents a snapshot of a VideoLectures raw log file.

```
# id,time,session,log,lecture,ref_type,ref_id,ref_name,"data"

101220275,2015-01-01
00:00:01.622361,101220276,,seehealth2010_golob_bpas,vl.lecture,11547,seehealth2010_golob_bpas,

101220276,2015-01-01 00:00:01.637593,101220276,session,,vl.log,101220276,,"ip=188.165.15.126
key=8bxdr82qbswggputjnocc52034t4isop"

101220277,2015-01-01 00:00:05.724054,101220278,-
,mit7012f04_introduction_biology,vl.lecture,6305,mit7012f04_introduction_biology,

101220278,2015-01-01 00:00:05.743161,101220278,session,,vl.log,101220278,,"lc=AU
ip=123.125.71.90 key=em0n1ml4mbydy8v55y40z4ieh9ge5od6"

101220279,2015-01-01 00:00:06.423510,101138550,-
,wsdm08_yang_aksbch,vl.lecture,4309,wsdm08_yang_aksbch,

101220280,2015-01-01 00:00:09.773501,101220281,-
,icml08_larochelle_cud,vl.lecture,5281,icml08_larochelle_cud,

101220281,2015-01-01 00:00:09.798654,101220281,session,,vl.log,101220281,,"ip=95.82.45.167
key=fgf5m0dccxbnhma26aayjo92u6pt7tvc"
```

Figure 3.2: Snapshot of a VideoLectures raw log file.

Processing the raw logs extracts key event types such as:

- *view* - a user accessing a lecture web page,
- *download* - the user downloading a presentation, e.g., a video, from the lecture web page,
- *search* - the user performing a search using the portal.

We also collected log files capturing the behaviour of selected users watching selected lectures - referred to as *ranges* log files.

Ranges log files (see Figure 3) present the actions of the user while watching videos, such as moving forward or backwards on the player, skipping some video section, etc.

{"id": 101220316, "time": "2015-01-01T00:01:07Z", "session": "fxkr0jtofwbsm57wo8nibnta1mvjwahm", "username": null, "ranges": [0, 25654, 859046, 862966, 1180104, 1189517, 1102009, 2307009], "video": {"id": 11018, "part": 1, "lecture": {"id": 10825, "type": "vl", "slug": "yalephys200f06_shankar_lec23"}}}

{"id": 101220399, "time": "2015-01-01T00:03:45Z", "session": "zvcyr4hteql2ha777wirs2ikjwsre2uz", "username": null, "ranges": [0, 33697], "video": {"id": 4842, "part": 7, "lecture": {"id": 4336, "type": "vtt", "slug": "mlss08au_lucey_linv"}}}

{"id": 101220411, "time": "2015-01-01T00:04:08Z", "session": "nmg1ycikq82ennvw10vce9iv2kn7u1zf", "username": null, "ranges": [351977, 360658, 3605976, 3657013, 3911988, 3917335, 3920258, 3922223], "video": {"id": 5335, "part": 1, "lecture": {"id": 5262, "type": "vit", "slug": "icml08_collins_spp"}}}

{"id": 101220423, "time": "2015-01-01T00:04:26Z", "session": "liwah46wrs9jzxhwoc84d17jd89a27a5", "username": null, "ranges": [1637170, 2290000], "video": {"id": 4730, "part": 1, "lecture": {"id": 4645, "type": "vl", "slug": "mit801f99_lewin_lec01"}}}

{"id": 101220498, "time": "2015-01-01T00:06:02Z", "session": "tgszt77hwshb2zbs4gaohioohp363vxg", "username": null, "ranges": [1215575, 1391116, 1058963, 1281640], "video": {"id": 4749, "part": 1, "lecture": {"id": 4664, "type": "vl", "slug": "mit801f99_lewin_lec20"}}}

{"id": 101220625. "time": "2015-01-01T00:09:35Z". "session":

Figure 1.3: Snapshot of a VideoLectures ranges log file.

2.2.2 Data Publication Plan

Summaries of the VideoLectures data are available for download[10], with a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 license, in line also with requirements for reuse in EDSA.

2.2.3 *TU/e MOO*

The datasets we analyse here were obtained from Coursera for the MOOC ''Process Mining: Data Science in Action''[11] for 43,218 registered students. These datasets are centred around the students participating in the MOOC and the stream of click events they generated on the course web pages. The dataset structure comprises:

- *clickstream* data,
- *student* data,
- *course structure* data.

[10] https://github.com/innanoval/edsa-videolectures-statistics-dataset-1/tree/gh-pages/data
[11] https://www.coursera.org/learn/process-mining

Clickstream data: during a course, students visit the course website to, among others, watch lecture videos and answer quizzes. As students click through the website to look up these videos and quizzes, they leave a trail of click events, collectively called a clickstream. Each event in a clickstream can be associated with, e.g., a particular lecture or a particular quiz submission. Pages visited by a student are recorded as a *page view* action, in addition to interaction with other course material, e.g. accessing a lecture video is recorded as a *video* action.

Student data: for each student we have information on the exact time registered for the course and their final course grade. Registration records also include whether a student registered for the special (paid) signature track, in order to obtain a verified certificate. The course grade consists of two parts: the normal grade and a distinction. In addition, each student is assigned an achievement level based on grades obtained. Where a student does not complete the course exams, the achievement level is absent. Where a student completes the exams but with a *normal grade* below the pass mark this is recorded as *failing* the course. Students with normal grades above both the normal pass mark and the distinction grade achieve the level *distinction*.

Course structure data: in Coursera MOOCs, lectures and quizzes are grouped into sections, each typically spanning a week. Each section is visible to the students at a predetermined time (the *open time*), in order to give structure to the course. Within a section, lectures and quizzes may have their own open time, to further guide students to follow a particular study rhythm. Quizzes have deadlines (the *close time*) and may be attempted multiple times up to a given submission maximum before this deadline.

2.2.4 Data Publication Plan

Due to restrictions set by Coursera the data used for the analysis of this course cannot be made available for reuse.

3. Data Analysis

As T3.4 progresses we will continue to examine online learning resources using different approaches, including those reported in this deliverable, to generate multiple views on the data and therefore increase coverage in our analysis, to obtain a more complete picture of the patterns and trends within the data. To validate both our approach and our results we:

1. employ multiple analytical techniques on a single dataset – the EDSA Learning Locker – to:
    a. identify differences in patterns due to differences in perspective taken,
    b. validate, through triangulation, the results obtained for each approach and the approaches themselves,
    c. identify additional requirements for data collection and analysis necessary to provide confident recommendations on course and curriculum design, and to match user (skill) profiles and interests to learning material.
2. employ the same analytical technique on independently collected datasets – to identify differences in patterns and trends due to:
    a. learner type,
    b. course type,
    c. content delivery mechanism,
    d. commitment or effort required of students.

Sections 3.1, 3.2 and 3.3 report exploratory analysis of the three datasets described in sections 2.1, 2.2.1 and 2.2.3 respectively, using one or more of statistical analysis, visual analysis and process mining. The section concludes (in 3.4) by applying all three analytical approaches to the data for one course in the EDSA LMS, "Foundations of Big Data"[20]. The aim of this exercise is to generate overviews of data content, to provide an initial picture of student interaction with the resources available in each data sets, and identify recurring and potentially anomalous patterns of behaviour. This is to help us discover how student behaviour maps to identification of resources relevant to their needs and interests, subsequent engagement with the material and other students and instructors, course completion and performance. We summarise here how we employ the three techniques, with examples of relevant work in the field. We report in each sub-section additional information on where data structure and/or content limit the depth of analysis or the technique in use.

Statistical analysis: this is to provide summaries of the data structure and content, and help to identify regions of interest (ROIs) and optimal techniques to employ for more detailed analysis. Of the three datasets analysed in this deliverable only the VideoLectures dataset contains sufficient data to carry out statistically significant analysis; section 3.1 therefore provides relatively more detailed statistical analysis, while sections 3.2 and 3.3 report only basic summaries, focusing more on other analytical approaches.

Visual analysis: this is to employ visualisation as a means to reveal patterns and trends within the data, to remove cognitive load to more intuitive perception and aid especially in-depth analysis [Thomas & Cook 2005]. We make use of different visualisation techniques to support or drive our analysis; this approach is not unknown - Conde et al., (2015) demonstrate the construction of a visual analytics dashboard to support the application of learning analytics to informed decision-making in a real use case. The study, which analysed data collected over five years, demonstrates the power in utilising multiple visualisation-driven analytics techniques to support analysis as well as reporting and the reuse

of findings to the different participants, as in our case, to students and instructors. Bakharia et al., (2016a), as part of research toward developing a conceptual framework for LA, report findings from a visual analytics dashboard developed to aid enquiry-based evaluation of learning data. They aim, as we do, to feed the results of LA into the design of learning material.

Process mining: this is to support the construction of a model describing the learning process as defined by course designers and/or instructors in the formal course structures. This is to, among others, aid process discovery, conformance checking and enhancement of the models in use, using a set of algorithms, tools and techniques to compare event data during actual interaction with a course against this defined structure [van der Aalst 2011]. Mukala et al., (2015), for instance, demonstrate the application of process mining to a MOOC, to investigate student behaviour as recorded by events captured during interaction with course material. The study provides insight into differences and similarity across categories of students throughout the course and how this correlates to completion and performance.

Appendix A contains snapshots of the data extracted from the EDSA Learning Locker (see section 2.1) used in our analysis.

3.1 Learning Analytics Task employing VideoLectures

In order to analyse user behaviour at the VideoLectures.NET portal, we employ a set of data analysis techniques, using web-based, interactive prototypes hosted on the VideoLectures Learning Analytics Dashboard[12] to provide statistical analysis and visual exploration features.

3.1.1 Statistical Analysis

This analysis is performed from four perspectives:

- aggregated perspective for all lectures
- perspective of a single lecture
- aggregated perspective of all viewers
- perspective of single viewer.

In addition, we have developed a set of metrics for viewers and lectures that provide insight into user behaviour on the VideoLectures.NET portal.

Lecture metrics (see Figure 4.8):

- number of viewers
- number of views

[12] http://learninganalytics.videolectures.net

- avg (average) moves forward
- avg moves backward
- avg time spent
- avg time in % (considering full video time as 100%)
- std dev (standard deviation), time spent
- std dev, time in %
- std dev, moves forward
- std dev, moves backward.

Viewer metrics (see Figure 4.9):

- number of views
- avg time spent
- avg time in %
- std dev, time spent
- std dev, time in %
- avg moves forward
- std dev, moves forward
- avg moves backward
- std dev, moves backward.

The analysis presented here considers a set of 1,000 most popular videos (by views) relevant to EDSA topics for the period Sep 2012 to Dec2015. User interaction resulted in:

- **7,055,427** views
- **3,020,090** downloads
- **1,045,860** visitors
- **866,912** searches within the portal.

The VideoLectures portal is independent of the EDSA project, with data covering a much wider range of topics and fields. However, the broad range of topics covered means that it also captures learning material of relevance to EDSA; Figure 4.1 shows a search cloud generated from popular terms used to retrieve resources during the collection period. We see that interest extends to skills core to Data Science as identified by research in EDSA, some of the more frequently used including "machine learning", which corresponds also to the acronym *MLSS* - the "Machine Learning Summer School", statistics, data mining, deep learning, the programming language python which saw high frequency of occurrence in selected regions in the EU in the demand analysis carried out in WP! (see D1.4). We also see additional related terms, including the product "hadoop", good knowledge of which contributes to the analysis of "big data", another key EDSA skill/term.

Figure 2.1: Search cloud showing interest in learning material available from the VideoLectures.NET portal for data collected Sep 2012 - Dec 2015 - including a number of key skills in Data Science.

Conde et al., (2015) in their learning analytics dashboard illustrate the use of a semantic tag cloud to map evolution of topics in course discussion fora. As we progress through T3.4 we will carry out similar analysis, using periodic snapshots of interest as expressed through search for learning material to follow how this changes specifically for Data Science terms, as the field grows and demand for relevant skills becomes more clearly defined.

3.1.2 Visual Exploration & Analysis

This section details the prototype interface developed for interactive analytics tasks on the VideoLectures portal - the interactive, web-based *VideoLectures Learning Analytics Dashboard*[10] and the *VideoLectures Explorer Tool*.

***VideoLectures Explorer*:** is a tool for exploring the lectures published on VideoLectures.NET. The dashboard enables the user to search through lectures and find similarities between them, e.g., to find lectures of a specific category or presenter of interest.

Database: the database used for the visualisation and basic statistical analysis contains data from all lectures on VideoLectures.NET. For each lecture the database contains the lecture's title and description, the name of the presenter and the organisation where the lecture was held, the language in which it was presented, its publication date, video duration, number of views and the scientific categories the lecture belongs to. The database was constructed using the VideoLectures API[13].

[13] http://videolectures.net/site/api/docs

***Search Options*:** these are in two parts: the basic and the advanced search (see Figure 5). Basic search options consist of a single input, where the user can enter the names of presenters, organisations or categories.

The advanced search options may be accessed by clicking on the button next to the search button. It expands the advanced options which can be used for a more refined search. This allows the searcher to specify whether input search terms are to be matched to categories, presenters or organisation. Further, it enables search by number of views and the language the lecture was presented in.

The search fields provide autocomplete functionality to aid identification of matching keywords and terms.

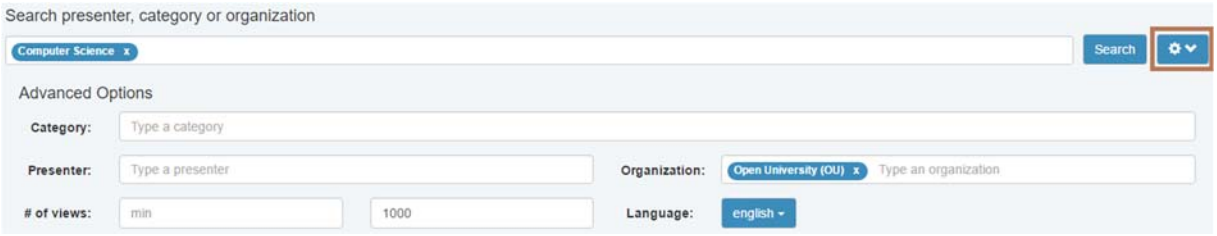


Figure 3: Search options for VideoLectures Explorer. The red square marks the advanced options button.

3.1.3 The Landscape

The main visualisation is the landscape. The snapshots in Figures 4.3, 4.4 and 4.5 show similarity between lectures in a category, by querying using three skills - "Data Mining", "Databases" and "Python", respectively - identified as core to data science as part of research in EDSA, and which demand analysis in EDSA found to be in high demand for job roles in Data Science (see D1.1 and D1.4). Each lecture is represented by a point with size mapping to number of views. Similarity between lectures is mapped to distance between points; more similar lectures are brought closer together. Hovering over a point brings up a tooltip containing information about the lecture: its title and description, the name of the presenter and organisation where the lecture was held, the language in which the lecture was presented, which scientific categories it belongs to, its duration, when it was published and the number of views since the last database update. Any data attribute for which values are not available is omitted from the tooltip. Landmarks show areas populated with lectures that are of the same category. The user can zoom in and out of the landscape to enable more detailed search. While zooming the size of the points get larger so it's easier for the user to hover over the points and get the lecture information.

In Figures 4.3, 4.4 and 4.5 we see that querying on the three skills highlights other relevant lectures categorised under those used in the queries - "Data Mining", "Databases" and "Python". We also see additional skills co-occurring with these, including "Big Data" and "Machine Learning" other skills with high demand also listed in the core set of Data Science skills in D1.4.

Because the term "Data Science" is quite new it has only recently been included as a category in the VideoLectures.NET database. The data is currently being updated to tag relevant lectures with this new category, after which querying on the category will return relevant results. We therefore show three queries for existing categories in Computer Science that fall also under Data Science. Figure 4.3 shows the landscape generated for "data mining", one of the most frequently occurring skills in job postings

across the EU (see also Figure 4, which shows "data mining" to be among the more popular search terms). We see a number of hits, with closely related terms including "databases" and "knowledge extraction", a task which "machine learning" supports. The resource selected is a tutorial on "Big Data".

Figure 4.3: The landscape of lectures retrieved using the keyword "Data Mining", showing other closely related terms. The overlay maps to the large point shown in the lower, right-hand corner - a resource with a relatively large amount of views compares to all other hits.

Figures 4.4 and 4.5 show the landscapes for "Databases" and "Python" respectively. The lecture highlighted in Figure 4.4 is on "Data Mining" in the health industry; other related topics in the landscape include "machine learning". Figure 4.5, which shows one of the most frequently occurring terms under programming and scripting languages, "python" in the demand analysis (D1.4) retrieves terms directly related to programming and also applications and other skills that employ programming, such as "machine learning".

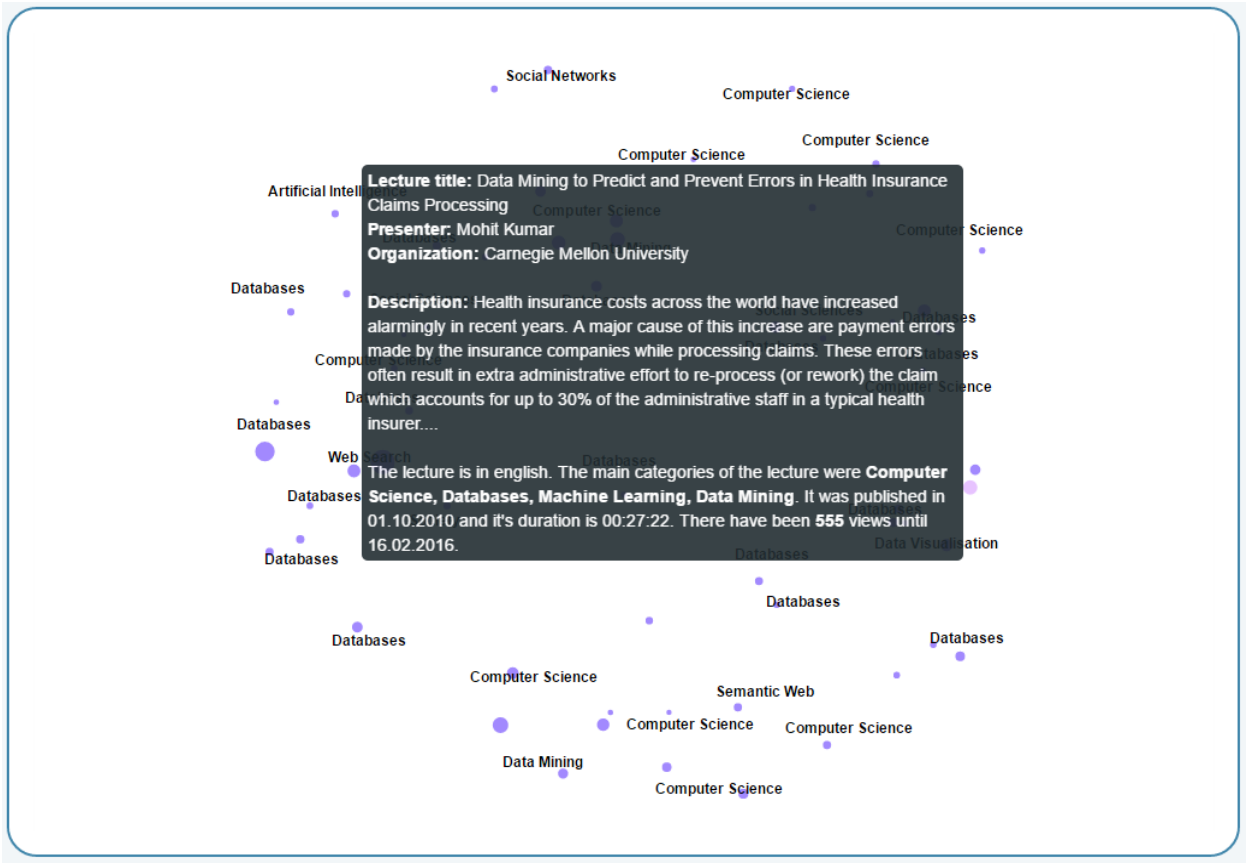


Figure 4.4: The landscape of database lectures, created using the keyword 'Databases'.

Figure 4.5: The landscape of python lectures, created using the keyword 'Python'.

The *Landscape* view has as its aim to map the queried lecture dataset into a two-dimensional vector space to allow plotting on the computer screen. For this we use different machine learning algorithms. First, by using the bag-of-words model[14], we represent the queried dataset as vectors. We then construct a term-document matrix using these vectors. Using Latent Semantic Indexing[15] we reduce the amount of information noise that is contained in the dataset. After that we use multidimensional scaling to map the the term-document matrix into a two-dimensional vector space. The coordinates are then scaled such that the coordinate values are between zero and one, which makes the final visualisation process easier. After that we add the lecture detail to corresponding coordinates, which are then used to visualise the landscape. This procedure is written using the QMiner data analytics platform[27].

Landmarks are constructed by first randomly distributing landmark locations around the landscape. A radius is then set for each landmark location and lecture points within this radius are collected. One of the categories that the bounded lecture points all fall within is randomly selected and the landmark name set to the category.

When the landscape is generated the dashboard also shows an additional information window (see Figure 4.6). This is in two parts: the first shows the query information, containing the names of presenters, organisations and categories, the minimum and maximum number of views and the language the user used to query the data. The second part contains basic statistics about the queried data: the number of lectures in the queried data, the total number of lecture views and the scientific categories with the number of their occurrences in the queried data. Clicking on a category in the dashboard automatically queries the data using the category name, and the information window is updated along with the landscape view.

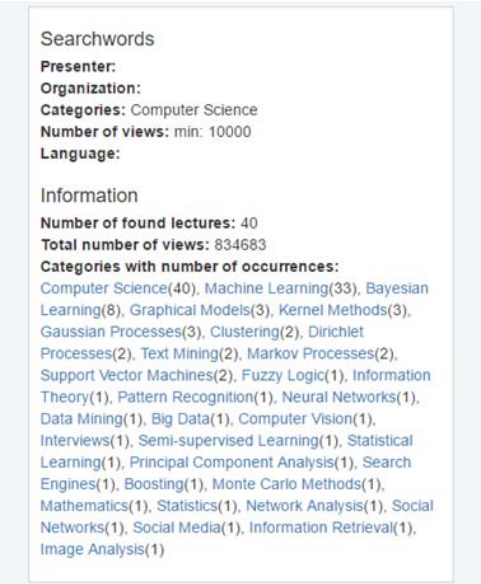


Figure 4.6: The additional information window. Created using the 'Computer Science' category with minimum number of views set to 10000.

[14] https://en.wikipedia.org/wiki/Bag-of-words_model
[15] https://en.wikipedia.org/wiki/Latent_semantic_analysis

3.1.4 The VideoLectures Learning Analytics Dashboard

We describe functionality for the prototype interface built for the VideoLectures Learning Analytics Dashboard:

- all lecture / per lecture / all viewers / per viewer perspectives,
- basic statistics on downloads, views and searches,
- trends for downloads, views and searches,
- metrics for visitors and lectures,
- the *search cloud*;
- interactive search for particular lecture/visitor related information, statistics and trends.

Figure 4.7 shows the trend for views on the VideoLectures.NET portal from late 2012 to Jan 2016. Figure 4.8 shows the trend for downloads at the portal, while Figure 4.9 shows the graph for user search. All graphs are interactive – the user can dynamically restrict the time period (by clicking and dragging a selection on the graph) – the view will automatically adjust to the new query.

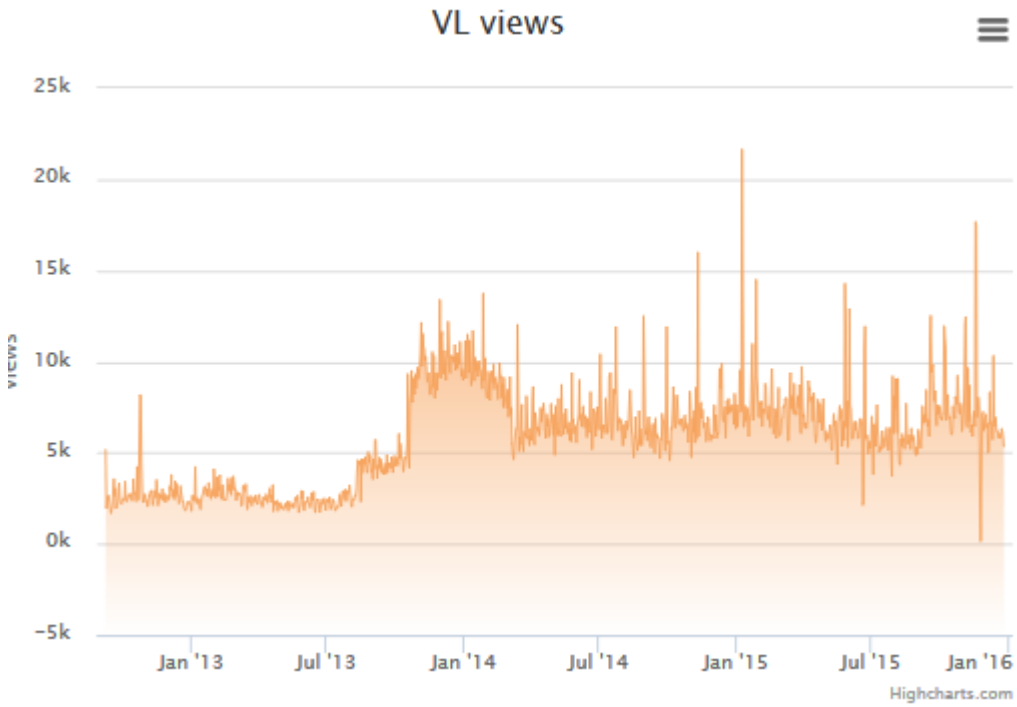


Figure 4.7: VideoLectures View Trends.

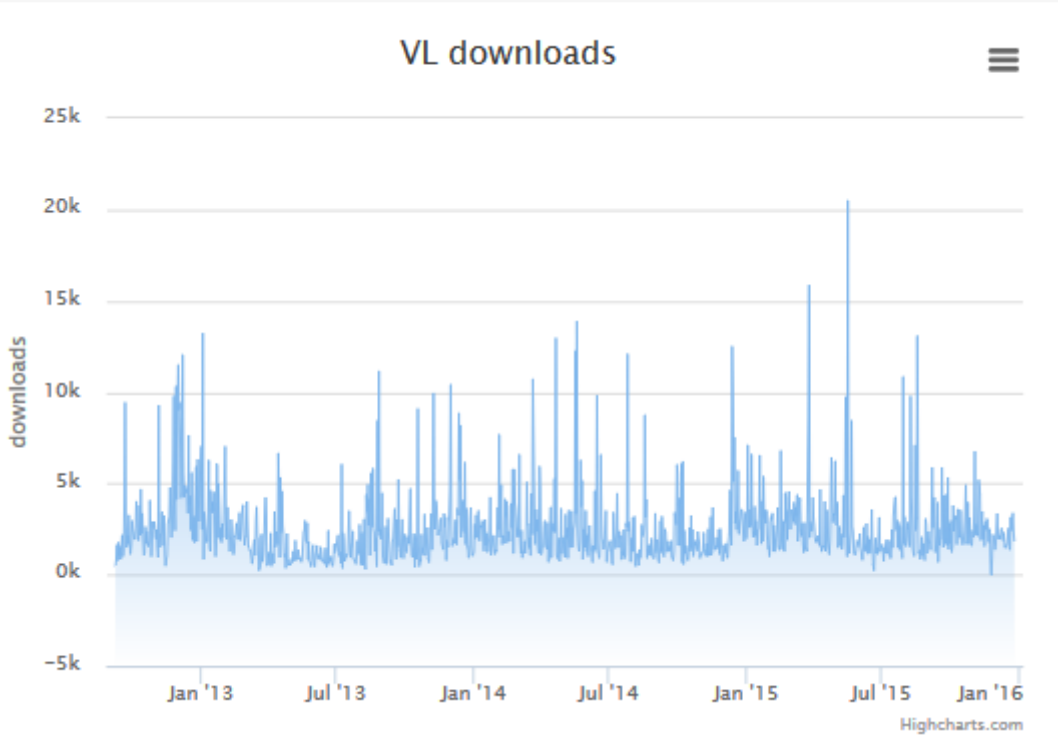


Figure 4.8: VideoLectures Download Trends.

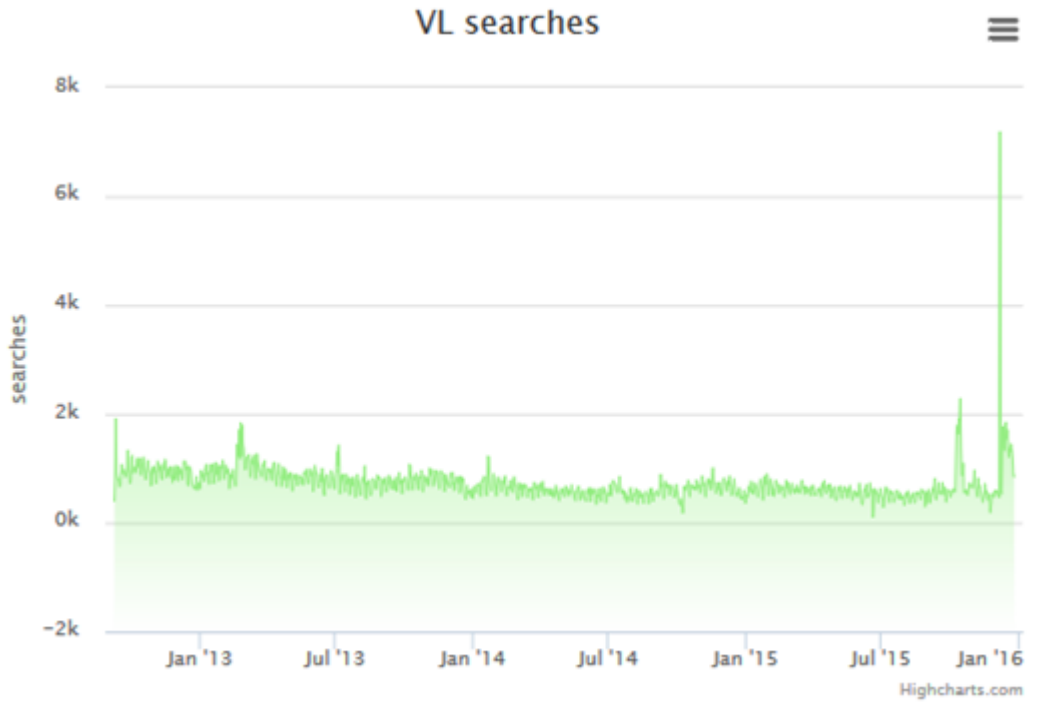
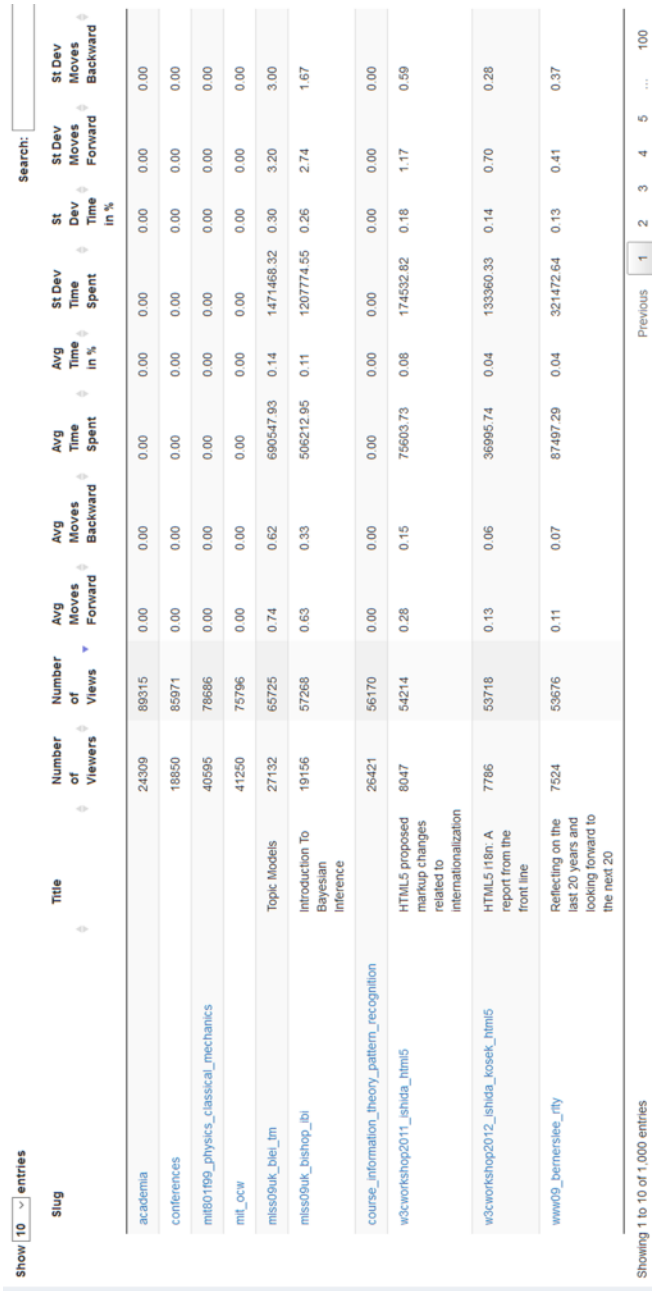


Figure 4.9: VideoLectures Search Trends.

Figures 4.7, 4.8 and 4.9 show trends generated based on VideoLectures.NET logs before the Data Science category was created on VideoLectures.NET portal, and therefore show overall trends for interaction

with resources available from the portal. As corresponding resources are tagged we will be able to extract trends specific to the Data Science category.

Figure 4.10 and Figure 4.11 presents metrics tables for lectures and viewers respectively at the portal, as detailed in the discussion of metrics in section 4.1.1. To allow comparison against a specific metric, table columns are sortable.


Show 10 entries

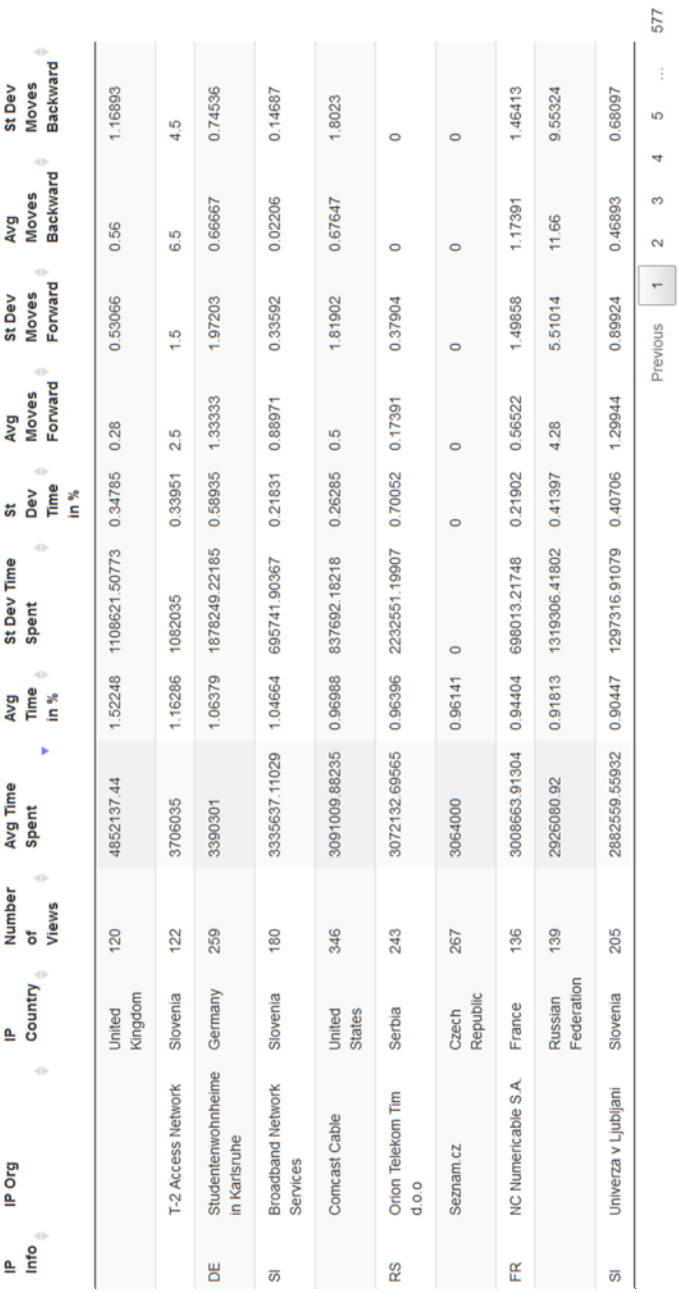
Search:

Slug	Title	Number of Viewers	Number of Views	Avg Moves Forward	Avg Moves Backward	Avg Time Spent	Avg Time in %	St Dev Time Spent	St Dev Time in %	St Dev Moves Forward	St Dev Moves Backward
academia		24309	89315	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
conferences		18850	85971	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mit801f99_physics_classical_mechanics		40595	78686	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mit_ocw		41250	75796	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mlss09uk_blei_tm	Topic Models	27132	65725	0.74	0.62	690547.93	0.14	1471468.32	0.30	3.20	3.00
mlss09uk_bishop_ibi	Introduction To Bayesian Inference	19156	57268	0.63	0.33	506212.95	0.11	1207774.55	0.26	2.74	1.67
course_information_theory_pattern_recognition		26421	56170	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
w3cworkshop2011_ishida_html5	HTML5 proposed markup changes related to internationalization	8047	54214	0.28	0.15	75603.73	0.08	174532.82	0.18	1.17	0.59
w3cworkshop2012_ishida_kosek_html5	HTML5 i18n: A report from the front line	7786	53718	0.13	0.06	36995.74	0.04	133360.33	0.14	0.70	0.28
www09_bernersiee_rthy	Reflecting on the last 20 years and looking forward to the next 20	7524	53676	0.11	0.07	87497.29	0.04	321472.64	0.13	0.41	0.37

Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 ... 100

Figure 4.10: VideoLectures Lecture Metrics.



IP Info	IP Org	IP Country	Number of Views	Avg Time Spent	Avg Time in %	St Dev Time Spent	St Dev Time in %	Avg Moves Forward	St Dev Moves Forward	Avg Moves Backward	St Dev Moves Backward
		United Kingdom	120	4852137.44	1.52248	1108621.50773	0.34785	0.28	0.53066	0.56	1.16893
	T-2 Access Network	Slovenia	122	3706035	1.16286	1082035	0.33951	2.5	1.5	6.5	4.5
DE	Studentenwohnheime in Karlsruhe	Germany	259	3390301	1.06379	1878249.22185	0.58935	1.33333	1.97203	0.66667	0.74536
SI	Broadband Network Services	Slovenia	180	3335637.11029	1.04664	695741.90367	0.21831	0.88971	0.33592	0.02206	0.14687
	Comcast Cable	United States	346	3091009.88235	0.96988	837692.18218	0.26285	0.5	1.81902	0.67647	1.8023
RS	Orion Telekom Tim d.o.o	Serbia	243	3072132.69565	0.96396	2232551.19907	0.70052	0.17391	0.37904	0	0
	Seznam.cz	Czech Republic	267	3064000	0.96141	0	0	0	0	0	0
FR	NC Numericable S.A.	France	136	3008663.91304	0.94404	698013.21748	0.21902	0.56522	1.49858	1.17391	1.46413
		Russian Federation	139	2926080.92	0.91813	1319306.41802	0.41397	4.28	5.51014	11.66	9.55324
SI	Univerza v Ljubljani	Slovenia	205	2882559.55932	0.90447	1297316.91079	0.40706	1.29944	0.89924	0.46893	0.68097

Figure 4.11: VideoLectures Viewers Metrics.

The interface enables search and focus on a lecture/viewer of interest; Figure 4.12 presents the information, statistics and trend for a selected lecture, on “Deep learning in Natural Language Processing”.

Figure 4.12: VideoLectures Information Trend: focus on lecture on 'Deep Learning in Natural Language Processing'.

Figure 4.13, similarly, shows the information, statistics and trend for a selected viewer.

Figure 4.13: VideoLectures Viewer Information/Trend.

Figure 4.14 shows a map of visits. A high number of visits coming to VideoLectures.NET portal originate in the USA, followed by Canada, China and Australia.

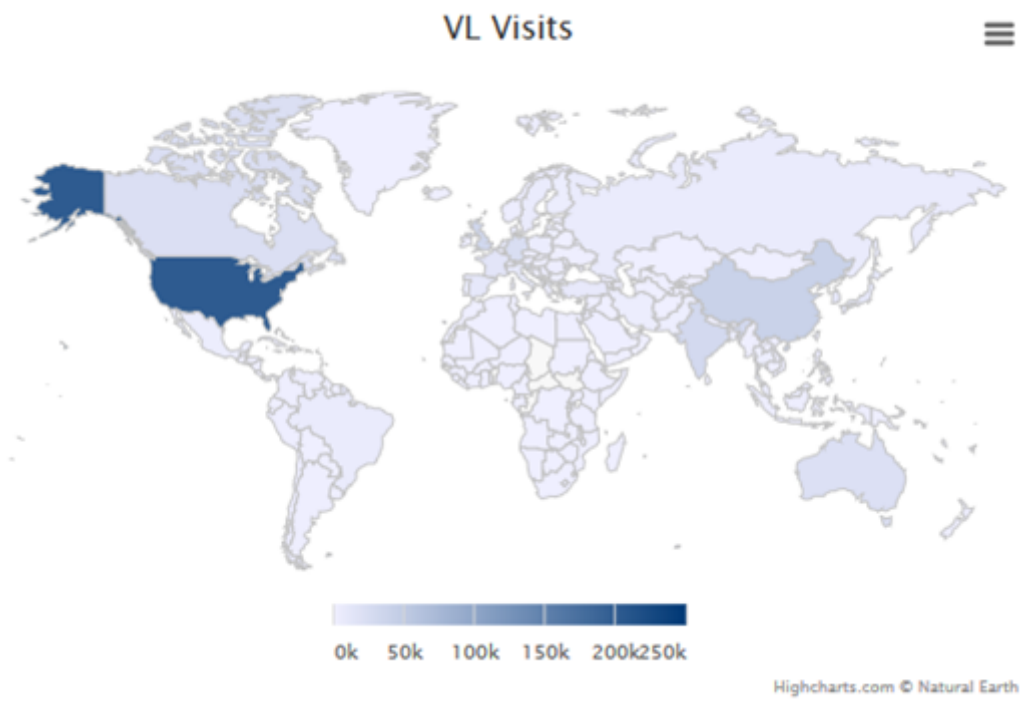


Figure 4.14: VideoLectures Visit Statistics.

3.1.5 Summary

By adopting statistical and visual analysis techniques on the VideoLectures logs, we are able to explore how viewers in general behave on the VideoLectures.NET portal. By mapping IP addresses to locations we obtain some demographic information - which countries and organisations viewers come from, how they view, search and download material stored on the site. We have identified that most viewers come to the portal from the USA, China, UK, India, Germany, Canada, Australia, Slovenia and France. The top five groups of users by view come from, in descending order:

- United States/Google - 41,822
- Germany - 35,519
- Iran/Rasaneh Avabarid Private Joint Stock Company - 23,033
- United States/Yahoo! - 21,928 and 20,320 from two distinct IPs
- Russian Federation/Yandex enterprise network - 18,931.

Diver et al., (2015) perform a similar mapping of IP address to location, to retrieve also further demographical information, on language, employment, internet access and typical bandwidth and speeds, to compare how the distribution of engagement and achievement to this background information. Kizilcec & Halawa (2015) also examine correlation between geographical location and dropout and performance. Our exploratory analysis has highlighted the need to collect additional user demographic information; as T3.4 progresses we will carry out a deeper investigation of users based on the current and more detailed information we are able to obtain, to identify how user backgrounds impact engagement, completion and performance.

VideoLectures.NET sees more access from outside the EU; as part of the contribution to the EDSA project, we aim to to attract more visitors from the EU. We therefore now publish a regular

VideoLectures newsletter on the EDSA project website with a list of recommended videos based on popularity over the previous two to three months. We are also building direct links between the EDSA demand analysis dashboard (see D1.4) and the VideoLectures.NET portal. This is to feed information on demand and skills analysis into the identification of further categories in Data Science, and therefore improve the recommendation of relevant learning resources to match queries in the demand analysis dashboard for learning material on a topic (skill or skill set of interest) or a user's skill profile.

Across all viewers we find that the videos seeing the most interest - the most popular, are:

- by the number of viewers:
  - MIT OpenCourseWare (OCW)
  - tutorials on Deep Learning
  - MLSS (the Machine Learning Summer School)
- by time spent watching:
  - videos on Deep Learning
- by time spent as a percentage of video length:
  - KDD 2014 (the ACM conference on Knowledge Discovery and Data Mining)
  - the 2013 W3C workshop on "Making the Multilingual Web Work"
  - ISWC 2014 (the International Semantic Web Conference).

We are able also to map activity over time for each viewer and each resource (video).

VideoLectures.NET, having been built prior to and independent of EDSA, provides an independent view on interest in a variety of topics including Data Science. Building these dynamic tools for the portal allows us to monitor access to learning material relevant to EDSA. With the definition of this new category in the database and as we obtain more data on this relatively new field we can start to restrict our analysis of the data to resources categorised as Data Science and also those in other categories tagged with skills and terms we have identified in EDSA as relevant to Data Science. This will allow us to explore in more depth what contributes to popularity of specific resources and categories in the field. We will feed any insight thus obtained into the overall learning analytics task as T3.4 progresses, and specifically into further categorisation of online learning material, both in the VideoLectures.NET database and by linking also to related courses and selected course material delivered by EDSA.

3.2 Learning Analytics Task employing the EDSA Learning Locker

This task employs the event data captured in the LMS described in section 2.1, covering course activity captured from 17 Nov 2015 and up to 13 Jun 2016. We explore here student behaviour for all courses, and then for selected courses and students in the EDSA LMS.

Similar to the analysis carried out for the VideoLectures.NET data, we examine the data from four perspectives, corresponding to those in section 3.1:

- aggregated perspective for all courses
- perspective from a single course
- aggregated perspective of all students
- perspective of a single student.

We start with a summary of data content in section 3.2.1, reporting counts for key data attributes - events records, number of users overall and per selected courses. The information on events, students and courses that record relatively high or low activity (peaks and outliers, respectively), as well as what at least initially appears to be the general picture is used to guide initial exploration. Beyond this we do not carry out further statistical analysis as there is not sufficient data to date on interaction with courses hosted by EDSA for this approach to provide further insight into student behaviour. Importantly, also, more detailed statistical analysis here would not paint a representative picture of activity, as high attrition is seen in all courses, occurring, with rates highest soon after enrolment onto a course. This is not unusual in online self-study - Diver et al., (2015) and Kizilcec & Halawa (2015), among others, report similar observations; both studies delve deeper into student behaviour to uncover potential reasons for this and the impact of student behaviour overall in such courses on achievement and completion.

The analysis process here is therefore driven by visual exploratory discovery, feeding the output of the statistical analysis into the construction of simple interactive visual analytics modules to explore content in more depth and thus reveal ROIs that in turn point to optimal methods to use for further, more detailed analysis.

3.2.1 Basic Statistical Analysis

Processing the data to feed into the visual analysis process, the following summary information was obtained:

- total event data records for the time period: **7891**
- no. of courses accessed: **27**
- unique user IDs: **322** - note that a single user may log into the system using different IDs, from social media or using the ID obtained by registering with the EDSA LMS.
- IP addresses in use: **477** - as for user IDs, a single user may log in from different IPs, or multiple users may record the same IP address, for example, if they use machines in the same institution. Further work in T3.4 will cross-reference IP addresses with user IDs, to:
    1. merge user IDs
    2. group users by institution/affiliation and track any trends that might pertain to this attribute, such as higher completion rates by, say, professionals/practitioners in the field for a specific course or course category.
- total activities captured: **7196**, out of the eight activity types. Of these eight, one, *started*, recorded no occurrences. Note that *enrolled onto* is recorded only for some courses as the action is not implemented by all systems hosting online courses, including Moodle. Nor is the action captured for MOOCS as enrolment or the equivalent is captured on the host site. The breakdown by activity type, illustrated also in Figure 4.17, listed in reverse order of frequency follows:
    - "viewed" - **5562**
    - "logged in to" - **731**
    - "enrolled onto"- **455**
    - "registered to" - **336**
    - "logged out of" - **80**
    - "answered" - **17**
    - "completed" - **15**
    - "started" - **0**

Table 4.1 shows counts for event types for the top 5 courses by activity count and enrolments. These are, from the bottom, "Process Mining"[16], followed by "Big Data Analytics"[17], which saw relatively intense activity over the second half of the period. Third, "Big Data Architecture"[18], was the only other course, with Big Data Analytics", that captured students' quiz results. "Essentials of Data Analytics and Machine Learning"[19] saw consistent activity over the collection period, recording the second highest number of views for 105 unique user IDs. Finally, "Foundations of Big Data"[20] saw the highest activity overall, with 1634 events generated by 117 unique user IDs from Feb 2016 to the end of the collection period in mid Jun 2016. We include also the number of unique user IDs recording events for each course with the enrolment count; the latter exceeds the former as some students enrolled more than once onto the same course. Further investigation is required to determine why this occurred.

Figure 4.17 provides an overview of all activity over time for all courses, including those listed in Table 4.1. (Note that the entry at the top in Figure 4.17 - "EDSA Online Courses" - is the default login/logout point for the EDSA Online Courses portal and not a course in itself.)

Table 4.1: Top five courses by activity and enrolment from launch of portal to 13 Jun 2016.

Activity	Foundations of Big Data	Essentials of Data Analytics and ML	Big Data Architecture	Big Data Analytics	Process Mining
viewed	1,507	1,243	739	563	575
enrolled onto (unique user IDs)	127 (117)	113 (105)	69 (64)	76 (59)	36 (35)
answered (quiz)	-	-	9	8	-
completed (quiz)	-	-	9	6	-

Browsing the statistical summaries we find that student activity counts vary significantly, between students on a single course and for the same student (ID) across multiple courses. The summaries however still help us identify users of potential interest - one of the most active IDs, *social_user_163*, recorded 454 events[21] over two and a half months, the majority of which were views. The remainder were 10 logins and enrolment onto courses including the two of the most popular: 'Big Data Analytics" and "Big Data Architecture".

[16] http://courses.edsa-project.eu/course/view.php?id=15
[17] http://courses.edsa-project.eu/course/view.php?id=33
[18] http://courses.edsa-project.eu/course/view.php?id=27
[19] http://courses.edsa-project.eu/course/view.php?id=25
[20] http://courses.edsa-project.eu/course/view.php?id=26
[21] Note that nodes in the visualisations may overlap where there is intense activity over a very short time period. These snapshots do not distinguish overlapping using additional visual encoding. For social_user_163, for instance, intense viewing activity occurred over short bursts and is therefore not obvious in the overview snapshots; however zooming into the ROIs in the view (section 4.2.2.2) reveals this detail.

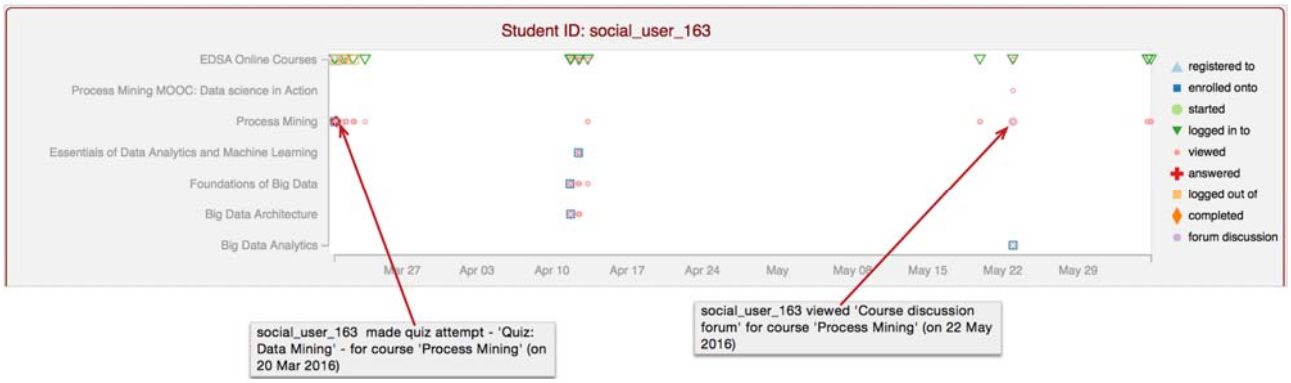


Figure 4.15: Event activity for a user ID recording mainly viewing activity after enrolment in a sub-set of the six courses interacted with.

Another user ID, *social_user_304*, recorded 65 events in about an hour on a single day, across eight courses. After enrolling onto six of these courses the user returned to the second - "Big Data Analytics". Almost half of the events for this user were on this course alone: they paged through the course content and made two attempts at answering quizzes. The user ended their session by viewing content in the course "Big Data Architecture".

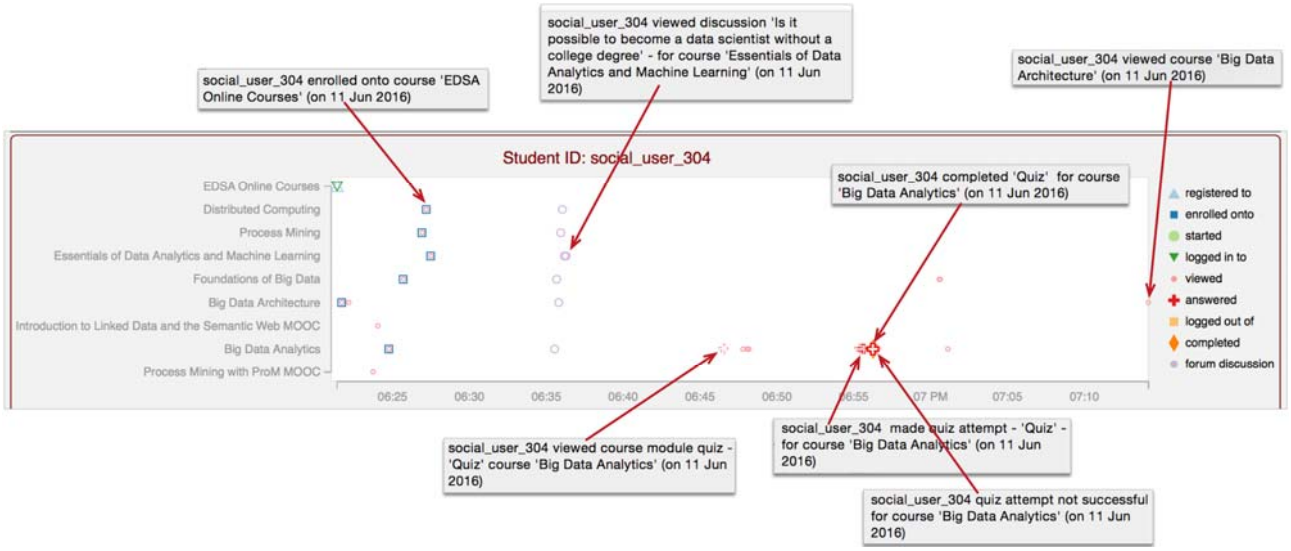


Figure 4.16: Event activity for an ID recording intense activity over roughly one hour on a single day across 8 courses. This included attempting a quiz in the course that saw the highest activity.

3.2.2 Visual Exploratory Analysis

Each activity record is timestamped and with a record of the IP address and device/browser type used to access the resource[22], allowing the mapping of a user's path through the system overall and for a course or activity of interest. Figure 4.17 shows the activity pattern for all courses during the data collection period. It should be noted that a single point as viewed on this layout may represent more than one instance of the same activity occurring at the same time; we currently do not provide additional visual cues to distinguish these. Zooming in to an ROI however provides the extra space to reveal detail for clusters.

We look next at the interaction of multiple users across all and selected courses, to identify ROIs to examine in more detail, from the perspective of a course and/or a user.

3.2.3 Course-centric Perspective

Figure 4.17 highlights which courses recorded intense activity over the whole or part of the period. Registration and logging in and out are not shown here for individual courses as these are captured for the user entering into the EDSA LMS (entry "EDSA Online Courses" at the top). Further analysis is required to link these to specific courses, by following a selected user's path.

Figure 4.17: Overview - Patterns seen for event data for all courses in the EDSA LRM - each point represents at least one occurrence of the corresponding event for a course per date. Selected events for the course 'Big Data Architecture' are highlighted, with detail including the user triggering it in the popup.

Some courses see fairly regular bursts of activity throughout the collection period. Regardless of density of activity this bursty pattern recurs, with periods of up to several days with no interaction. More detailed analysis beyond M18 will identify potential causes of these patterns and the potential impact of course structure (including staggered release of modules), day of the week or holiday periods such as Christmas and New Year that see downtime across wide regions of the world.

[22] Further investigation of IP addresses and access device will be carried out in the second half of the project to obtain more information on student demographics (see also section 6.1).

View events may be further broken down into

- course module views
  - url, e.g., pointers to other online material such as webinars and course feedback pages
  - (course module) page, e.g., pointers to syllabi, extra information on the course, webinars
  - (course module) resource - the components of a course, such as pages addressing course topics
- discussion fora
  - discussions on a specified topic
  - discussion forum views
- feedback pages
- quizzes
  - views
  - attempts.

We focus in this analysis on *viewed* activity events in discussion fora and to do with quizzes. The latter differ from *answered* and *completed* events explicitly recorded for quizzes and which provide detail on quiz questions and results. Where a *viewed* event concerns quizzes, however, this does not include any detail beyond the course name and in some cases a more informative quiz label. We distinguish between these and quiz *viewed* events by using a cross with a broken border in pink and red for quiz views and attempts respectively (a small pink circle is used for *viewed* and a red cross for *answered* events respectively). (Appendix A contains a snapshot of the results obtained for a query following a named user, with examples of the level of detail on quiz information for the three event types.)

Figure 4.15 illustrates this additional encoding for the user with ID *social_user_163*.

3.2.4 Focus on a Single User

To allow a more detailed inspection of the activity of *social_user_163* (overview in Figure 4.15) follow the path of this user, one of the most active, for the collection period. We look in more detail at other selected users in section 3.4.1, to examine further patterns of behaviour.

We zoom first into the start of the period of activity (20 Mar) to reveal detail on the activity here. Figure 4.18 narrows down to the first cluster seen, over the first four days. This reveals five clusters each with a distinct log in/out action, and several quiz views and attempts for the course "Process Mining" in the first burst. Activity for other courses falls after this period.

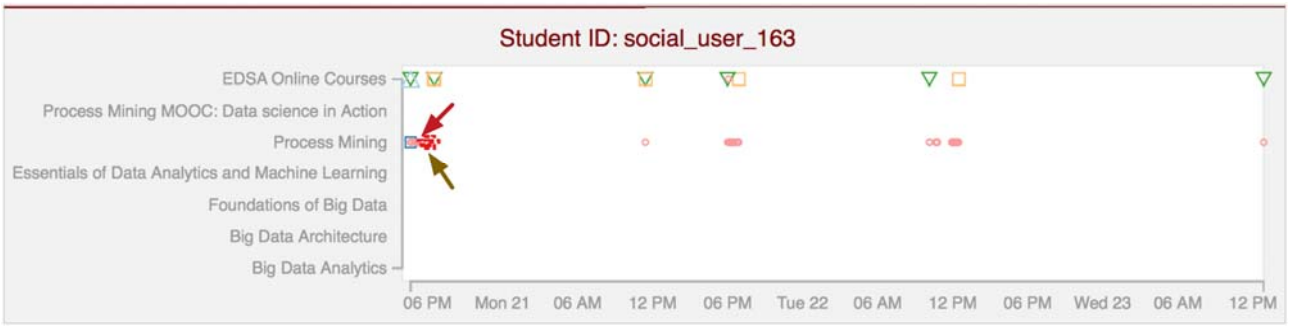


Figure 4.18: zooming into the first four days of the collection period for one of the most active users (see Figure 4.15) where quite intensive activity was recorded. The red arrow points to a quiz view and the olive to attempts at answering quiz questions.

There is activity on only one course - “Process Mining”; we therefore filter out all other courses accessed by *social_user_163* and zoom in further to the first two days, and then the first four hours (top and bottom respectively, Figure 4.19). This shows, after enrolment, quick views of a number of course modules, with two short and two relatively longer pauses, which may be detailed reading before moving on to another module. Shortly after the fourth break, this user appears to browse through and attempt a series of quiz questions (the broken borders on the cross icons indicate these are captured as quiz *viewed* events rather than *answered* events). A relatively long pause follows views of questions in the quiz 'Data Mining', followed by several further attempts at questions in this quiz. Another pause follows after viewing of a quiz question before the user logs out of the system.

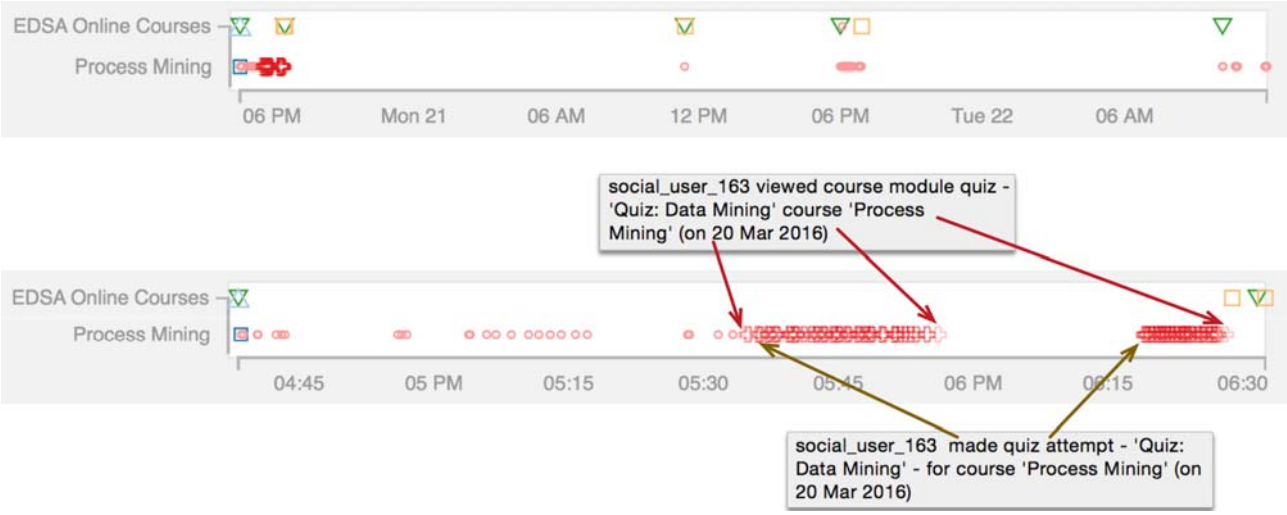


Figure 4.19: Zooming in further into the first two days, then the first two hours of activity for social_user_163 to reveal detail on activity during the latter period.

After this burst of activity there are four further short views of course material over the following four days. There is no further activity for the next three weeks, after which *social_user_163* views further material in 'Process Mining' and enrols onto three of the four most popular courses (see overview in Figure 4.15).

No further activity is recorded for another five weeks, after which this user views additional material in "Process Mining". After entering its discussion forum *social_user_163* briefly views material in the "Process Mining MOOC: Data science in Action" but does not engage any further with it. After a break of approximately a fortnight the user returns to view additional material in "Process Mining", with the last record on the 4th of June (for the collection period ending 13 Jun).

3.2.5 Summary

By carrying out simple statistical analysis during data processing we were able to, with the aid of the visual overview of the complete dataset (in Figure 4.17), identify initial ROIs to explore in more detail.

The exploratory visual analysis shows one recurring trend: even for the most active and persistent students: periods of intense activity followed by long breaks. Different factors may account for this, including defined course structure and the staggered release of material typically used to structure self-study in online courses. Additional potential factors include difficulty with elements in the course that may require additional study, as well as other (external) factors more difficult to identify and/or control,

such as demographics and commitments outside study. Some studies, e.g., Diver et al., (2015), link delays in accessing course material (once available) to dropout. Kizilcec & Halawa (2015) examine the impact of greater difficulty accessing help from other students and instructors in online learning on, among others, poor results in or incomplete assessment (quizzes, assignments and exams). We take a closer look (in section 3.4.1) at access to discussion fora to discover what correlation may exist between interaction with these fora and continued engagement with a course. Further work in T3.4, as we obtain also information on course completion, will involve further investigation to confirm whether these and other factors contribute to the patterns observed and what impact they have on student performance.

A second frequently occurring pattern is short periods viewing course material after enrolment, followed by abandonment of the course (Figure 4.26 highlights this pattern for one course, and the analysis carried out in section 3.3 using process mining reports similar findings). This may be attributed to the need to view course material to determine suitability or confirm interest - some courses do not allow browsing before enrolment. Further investigation of the material viewed in such instances is necessary to confirm that this is the case and/or what other factors influence early dropout. The findings from this more detailed investigation should feed into guiding students in the first instance in selecting courses aligned to their needs and interests.

3.3 Learning Analytics Task employing Process Mining

Process mining provides a set of algorithms, tools and techniques to analyse event data [van der Aalst 2011]. Three main perspectives offered by process mining include process discovery, conformance checking and process enhancement. Discovery techniques allow the enactment of process models from log data. Conformance checking attempts to verify conformity to a predefined model and identify deviations, if any. Enhancement provides for models to be improved based on results of process discovery and conformance checking.

Given the online-based format and nature of MOOCs, it is possible to track student activities by following the individual clicks they make on course web pages. The data generated can give us insight into how and when students follow lectures and how they prepare for exams. We analyse a MOOC hosted on the Coursera platform - "Process Mining: Data Science in Action'" delivered by TU Eindhoven.

3.3.1 Preprocessing: Building an Event Log

We are interested in analysing student behaviour based on the trails of click events they generate. Before we could use process mining to analyse this behaviour, we first needed to map the MOOC data to an event log. There are two things we needed to specify for this mapping: what constitutes an *event*, and what makes a *case* (i.e., a sequence of events).

As we are focusing on the behaviour of students, we consider each student as an individual *case*. The clickstream a student generates is the basis for the *events* in this *case*. This highlights the separation between *case* and *event*. For the analysis reported here we will primarily focus on action events of type *page view*.

We filtered the MOOC data to obtain a view of the lecture watching behaviour of students. For each case, we stored the data available about the student, including their course grade data. For each clickstream

event, we created an event belonging to the corresponding case (based on the student user id) and stored this with the reference lecture as the event name.

Based on different data attributes, we can determine several students groups. First, we group students that failed the course or successfully obtained a certificate, with the latter split into a normal certificate or one with distinction. The second grouping attribute is whether or not a student enrolled in the signature track.

3.3.2 Visualisation of Learning Behaviour

In this section, we use the dotted chart and a process mining discovery algorithm (Fuzzy Miner) to visualise the event data and discover the actual learning process. The aim is to visually provide insights on the overall MOOC and profile student behaviour throughout the duration of the MOOC. We consider three important dimensions in this analysis: the general lecture videos viewing habit, the quiz submission behaviour. These insights can help to understand how students study and what impact such behaviours have on their involvement in the MOOC.

Visualising Viewing Behaviour: we make use of the dotted chart (see Figure 4.20) to visualise the path followed by students while viewing videos. This provides a broad representation of student watching behaviour throughout the course.

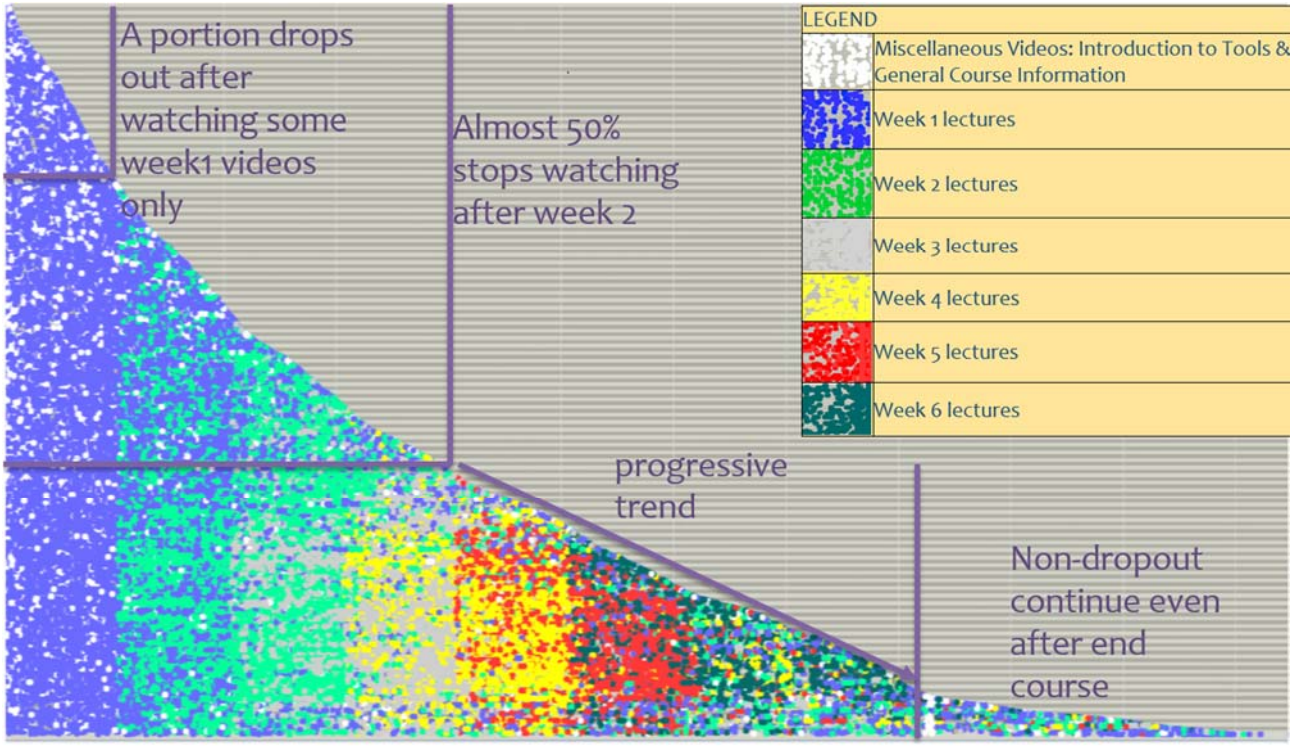


Figure 4.20: Dotted Chart depicting general viewing behaviour throughout the duration of the MOOC.

In Figure 4.20 the dotted chart depicts the viewing behaviour for all students who registered for the MOOC, focusing on when and how they watch videos. The *x*-axis depicts the time expressed in weeks, while the *y*-axis represents students. Seven different colours represent different events at a given point in time as carried by students. The white dots show the timing when students viewed miscellaneous videos (two videos on course background and introduction to tools). All videos for Week 1 are depicted

with blue dots, green dots represent videos for Week 2, gray dots show the distribution for videos in Week 3, yellow dots show lecture views for Week 4, Week 5 videos are seen in red and Week 6 lecture videos are depicted by dark green dots. One can note that the videos from week 2 are never watched in week 1; this results from the organisation of the course website, where students cannot see these videos in week 1 (and similarly for later weeks).

Looking at Figure 4.20 we observe that a significant number of students drop out throughout the duration of the course. Also, many stop watching after the first week and about 50% of students drop out by the end of the second week of the course. Further, not all students watch the videos in sequence: although all watch Week 1 before watching Week 2, even toward the end of the course we still see many watching Week 1 videos. This indicates that some videos are watched repeatedly and that a number of students join the course late. This represents an interesting pattern - where students joining late may be looking at selected course material, and/or students who joined at the start are reviewing some material. This pattern could be investigated by delving more deeply into analysis of student behaviour.

In order to get detailed insight into this trend, we also group the students into sub-groups based on their profiles. We make this classification based on their final performance (distinction, normal and fail) and the type of the certificate they sign up for (signature or non-signature track). We consider only distinction students on the signature track as an illustration in Figure 4.21.

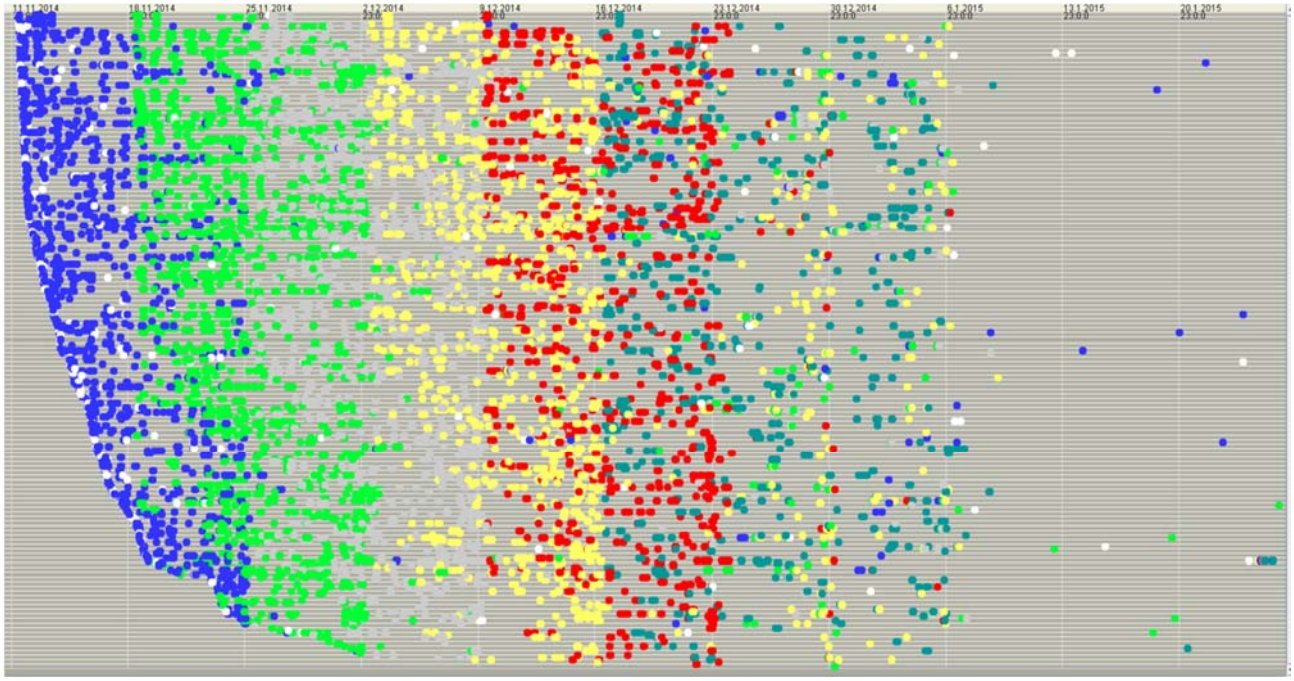


Figure 4.21: Dotted Chart for Distinction students on Signature Track.

In Figure 4.21, these students follow a sequential pattern as they watch the videos. Some join a little late at Week 2 or Week 3, but the general trend remains that most of them watch videos sequentially as they are made available. This can be seen by looking at the demarcation imposed by respective lecture video colour. This is also captured by the process models depicted in Figure 4.22. Successful students typically follow videos sequentially or with orderly loops (the repetition of a set of videos in their intended order) while unsuccessful students appear to be volatile and unpredictable in their watching pattern (the visualisation of this process is exemplified by the model shown on the right side of Figure 4.22). We look next at the process models depicting both successful and unsuccessful student learning paths.

Process Discovery: entails learning a process model from the event log. One can make use of an event log as input to a number of process mining algorithms in order to visualise and enact the real behaviour (sequential steps) of students. We use the fuzzy miner algorithm [Gunther & van der Aalst 2007] to mine our dataset. Out of all student process models we consider for illustrative purposes two extremes: the distinction students on the signature track and failing students not on the signature track. The resulting models are displayed in Figure 4.22.

The models in Figure 4.22 indicate that distinction students tend to have a more structured learning process, with a single path where possible loops are highlighted. On the contrary, failing students follow an unstructured learning process that exemplifies the volatility and unpredictability of their learning patterns. Although the fuzzy miner only shows the most dominant behaviour, Figure 4.22 still shows that there are many alternative paths through a course.

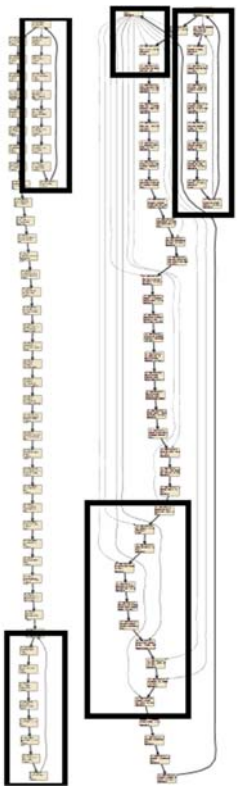


Figure 4.22: Process Models for Signature Track Distinction Students with few "loopbacks" (left) vs. Non-Signature Track Failing students with "loopbacks and deviations" (right).

3.3.3 Quantification of Learning Behaviour

We perform conformance checking using the normative model in Figure 4.23 to analyse viewing behaviour for all sub-groups of students. By exploring the alignment details we are able to extract details about the overall watch status (moment at which a video is seen, such as "early") and viewing habits. We make use of the alignment on the model to understand whether a video was watched early (earlier than in the intended order), or late (later than in the intended order) – or regularly watched as intended. We also use temporal information to study viewing habits (daily, weekly, in batches, etc.)

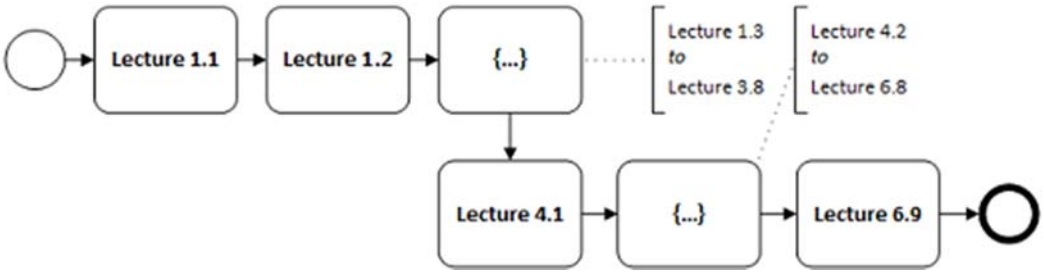
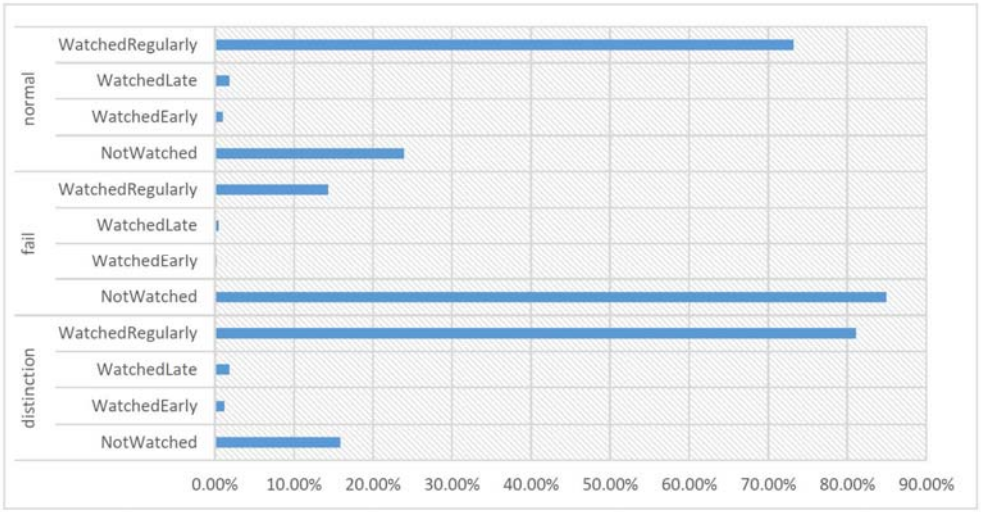


Figure 4.23: Normative model.

Watch Status: Figure 4.24 gives a view on the overall video status for both signature and non-signature track students. This indicates that successful students appear to be more committed to watching videos than unsuccessful students. We note however that unsuccessful students also fail to watch some videos because they would have dropped out before some resources were made available. Further, demarcation is observed between signature-track and non-signature track students - the former show more signs of commitment to watching videos across the different sub-groups in comparison to their non-signature track counterparts. Successful students with the signature track certificate watched regularly more than 80% of the videos, while those who failed watched 45% of the videos. The non-signature track students who were successful watched on average 80% of videos regularly, while the students in this group who failed watched only 15% of the videos. These trends are shown in Figure 4.24 for the signature track (top) and non-signature track (bottom) students.



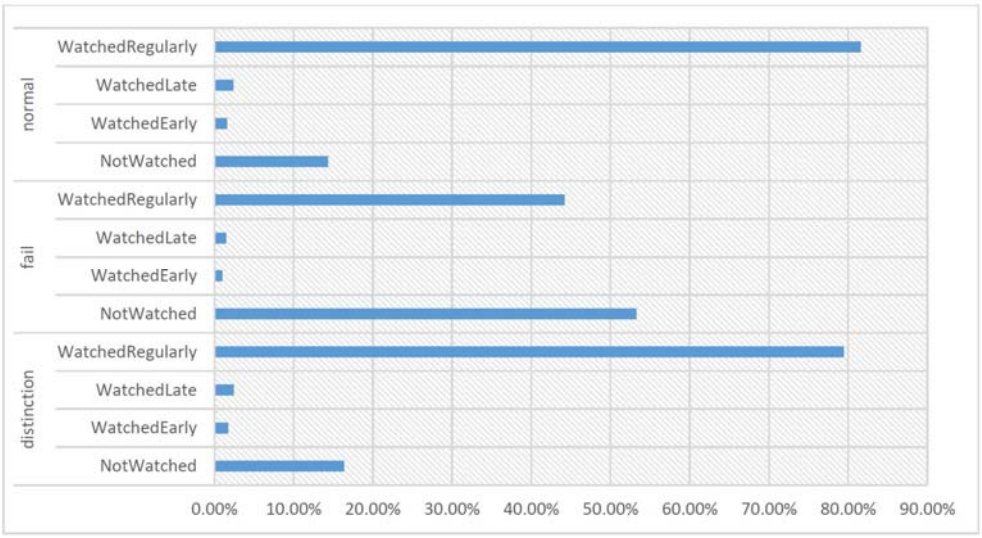
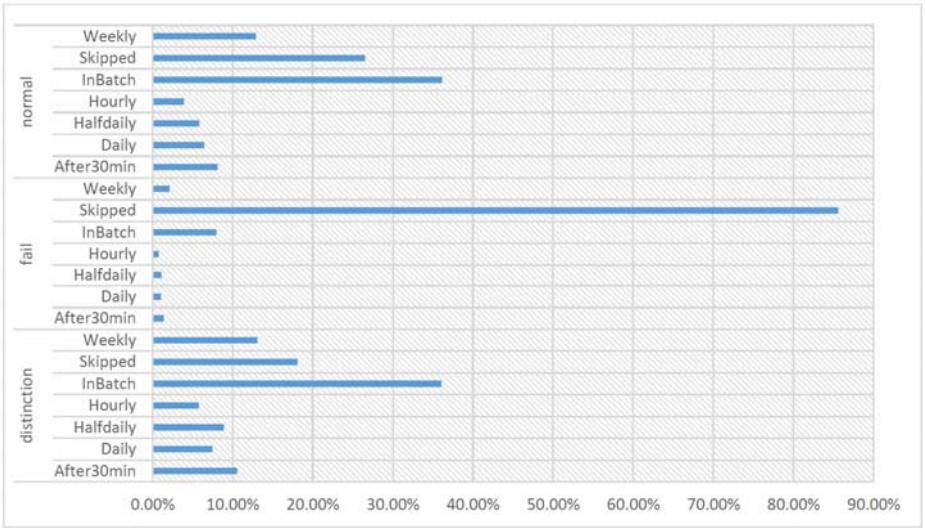


Figure 4.24: Overall watch status for the entire student population, Signature Track (bottom) and non-Signature Track (top).

Viewing Habit: The viewing habit measure describes the time commitment in the student learning behaviour. Figure 4.25 indicates that for the most part successful students watch videos more in a batch and do not waste a lot of time between videos. Unsuccessful students skip more videos than they watch.



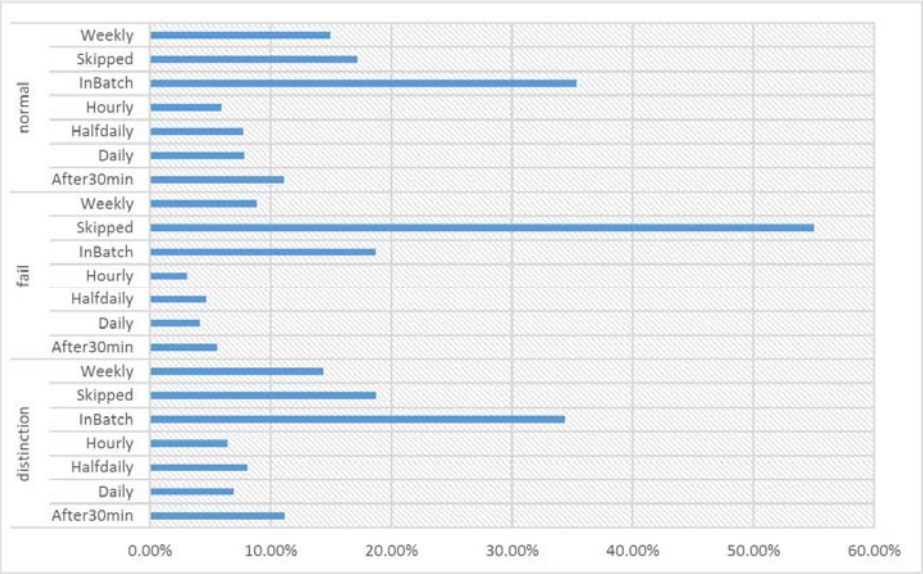


Figure 4.25: Overall representation of viewing habits, Signature Track (bottom) and Non-Signature Track (top).

3.3.4 Summary

Taking our Coursera MOOC as a case study, we show the added value of process mining on MOOC data. Our results demonstrate that the way students watch videos as well as the interval between successively watched videos are related to their performance. Results indicate that successful students follow a sequentially structured watching pattern while unsuccessful students are less predictable and watch videos in a less structured way. Moreover, student learning behaviour can be described from two dimensions: watch status and viewing habit. In general, student viewing habits are determined by the time between successive videos while the watch status is determined by alignment to the normative model.

A more detailed description of the results can be found in [Mukala et al., 2015]. The implications of these results will be studied in further detail as T3.4 progresses. It appears that many students prefer not to follow the intended order of a course. This presents a possible path for improvement of course delivery: trying to cater to the needs of students for whom the path prescribed by the instructor does not seem relevant.

3.4 Focus on the EDSA Online Course "Foundations of Big Data"

Analysis of the event data captured to date in the EDSA LRM showed activity that is at times sparse and at others extremely dense, looking at all data, individual courses and individual learner IDs. One course generated enough events to allow analysis using all three approaches reported in this section: "Foundations of Big Data". We therefore triangulate the independent analysis carried out on the event data for this course, to provide a single foundation on which to begin to build a framework for learning analytics in EDSA and provide pointers, and ultimately, benchmarks that will contribute to the state of the art in the field.

Sample data from the event logs for this course can be found in Appendix B.

3.4.1 Statistical & Visual Analysis

The right-hand side of Figure 4.26 shows the trend seen for the two activities recorded for the course - enrolment onto a course and views (course module pages, discussion fora and feedback forms). From top to bottom, users are ordered by date enrolled, followed by total activity count. The two snapshots on the left show, respectively, event data the 20 most and least active users (by total activity count). An interesting observation is that users who persist beyond their initial session very often access discussion fora soon after enrolment (on average the highest occurrence of this action across all users), and again further into the course.

The first event recorded for the course was a view (by *social_user_5*) on 16 Dec 2015, after which no more events were recorded till 02 Feb 2016 (vertical delineator). Incidentally, this point falls shortly after formal publicising of the EDSA online courses portal at the end of Jan 2016. Further investigation of the interaction path of *social_user_5* and others who accessed the portal prior to this is necessary to identify any differences in behaviour before and after this point for other courses.

Beyond 02 Feb 2016 enrolment and views gradually increased through to the end of the data capture period on 13 Jun 2016.

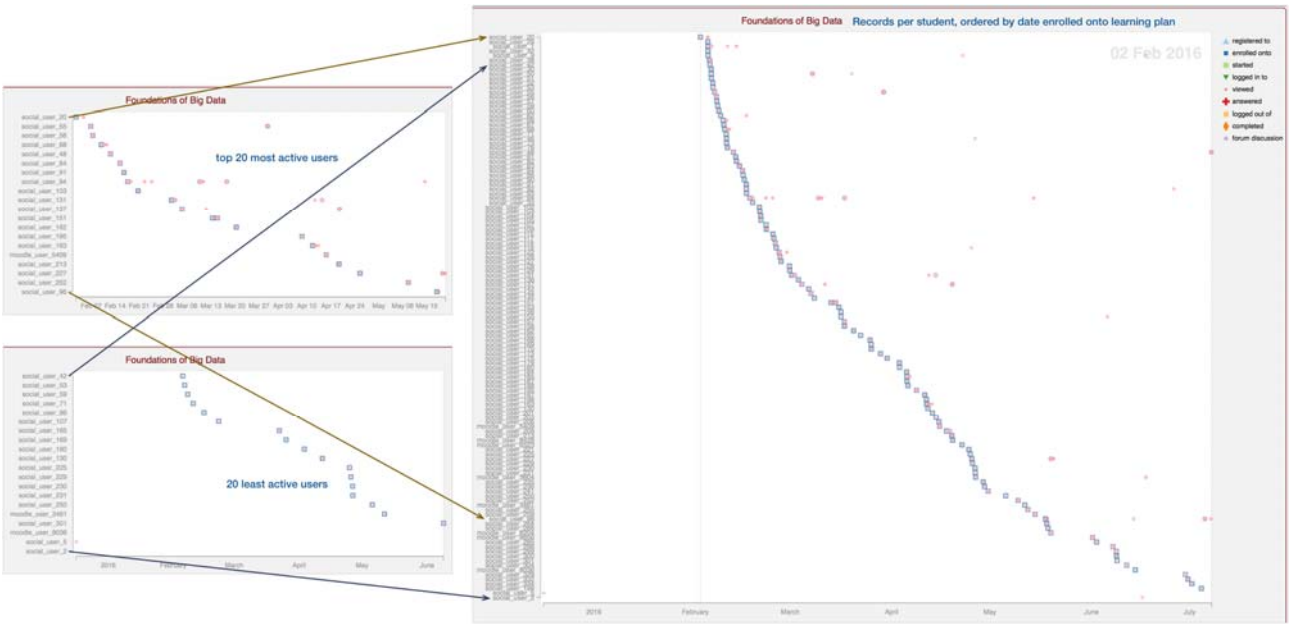


Figure 4.26: Overview of the event activity for the 117 unique user IDs for the course "Foundations of Big Data", focusing also on the 20 most and least active learners respectively.

There is no data on quizzes for this course; we do however see some interaction with the course's discussion fora (the arrows in Figure 4.27 point to a sub-set of the nodes for view events associated with discussion fora), very often coinciding with enrolment onto the course. The only labelled discussion for this course was "Greetings", with discussion content described as "A Moodle discussion". Further forum activity fell under the topic "Course discussion forum" - described simply as "A module" discussion. A "General discussion forum" for the EDSA portal (rather than a specific course) asks students to join the discussion forum to obtain more information on Data Science and EDSA courses. A second discussion labelled "EDSA courses" and described as "A Moodle discussion" also falls under the general course portal.

The LMS does not provide further information on discussion content. We aim to investigate further the kinds of discussions in the "Course discussion forum", as well as other specific discussions started, to determine what additional impact both the social nature of the activity and the actual discussions have on engagement and performance.

Figure 4.27: Flattening the layout on the right in Figure 4.26 to illustrate the bursty interaction seen across all courses. The arrows point to records of interaction with discussion fora.

Figure 4.15 (section 3.2.1) shows the activity recorded for all courses which saw activity by *social_user_163*, the most active user on this course - see Figure 4.28, which zooms into the snapshot in the top, left, Figure 4.26. Figure 4.29 focuses on activity by *social_user_163* and *social_user_20*, the first to enrol onto the course ("Foundations of Big Data").

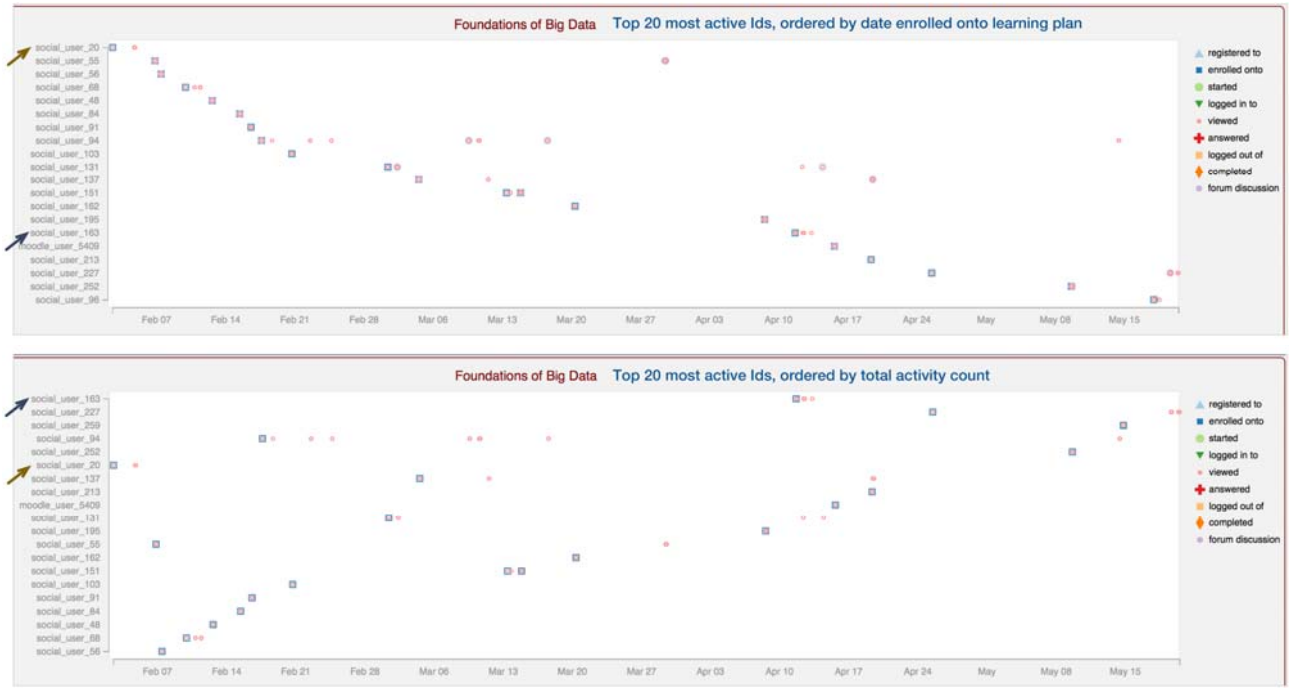


Figure 4.28: Activity trends for the top 20 most active learners, ordered first by enrolment date (top) then by total activity count (bottom), highlighting the top user in each case.

We see again activity by user in short bursts; some users return to the course after a break, but the majority appear to drop out after the initial interaction period. (See also Table 4.2, which shows the total activity count for the 30 most active users for this course, with very quick dropoff in interaction.)

Looking again at *social_user_163* we see (in Figure 4.29) intense activity immediately after enrolment, followed by two shorter spells of activity halfway through the 36 hours this user spent on the course. *Social_user_20*, the first to enrol onto the course, records activity over 2 days, but for only 22 (more) views, most toward the end of that time.

Figure 4.29: Focus on *social_user_163* and *social_user_20* - see also Figure 4.28.

Most user activity occurs in short bursts over a short total time span - minutes to hours to a few days, as seen for even the active users in Figure 4.29. The three users in Figure 4.30 break this trend, spanning, for the longest, three months, and with one and a half months each for the other two.

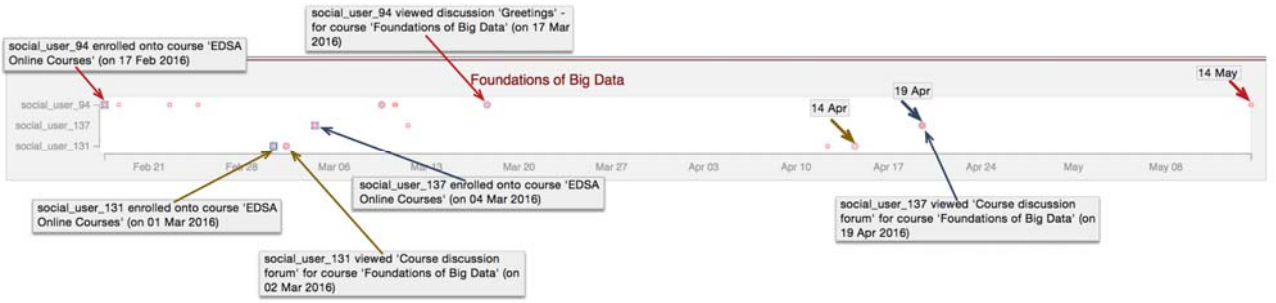


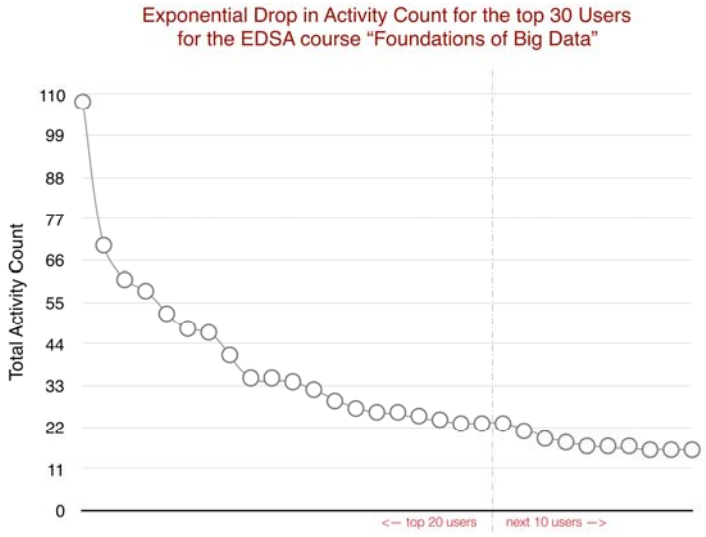
Figure 4.30: Focus on *social_user_94*, *social_user_131* and *social_user_137*, whose activity spanned longer periods than average.

We modify sort order to largest number of events for this course only - the results are shown in Table 4.2 and Figure 4.31 for the top 30 IDs. Activity drops away quickly from from 108 events for *social_user_151* to under 20 for the last eight. *Social_user_163* drops to third position, after *social_user_94*. These users, especially *social_user_151* engage with the course discussion fora, in addition to viewing course material (see Figures 4.32 and 4.33). As is the norm engagement with the course is clustered into a series of short, relatively intense activity events for users such as these, who persist beyond the initial enrolment and module resource views.

Table 4.2: From the most active, the top 30 most active users on the course "Foundations of Big Data - activity count drops from 108 to under 20 for the last eight.

User ID	Total events for "Foundations of Big Data"		User ID	Total events for "Foundations of Big Data"
social_user_151	108		social_user_91	26
social_user_94	70		social_user_56	25

social_user_163	61		social_user_84	24
social_user_55	58		social_user_50	23
social_user_137	52		social_user_20	23
social_user_131	48		social_user_109	23
social_user_227	47		social_user_122	21
social_user_68	41		social_user_144	19
social_user_252	35		social_user_38	18
social_user_103	35		social_user_170	17
social_user_48	34		social_user_157	17
oodle_user_5409	32		social_user_1	17
social_user_213	29		social_user_197	16
social_user_162	27		social_user_168	16
social_user_195	26		social_user_60	16



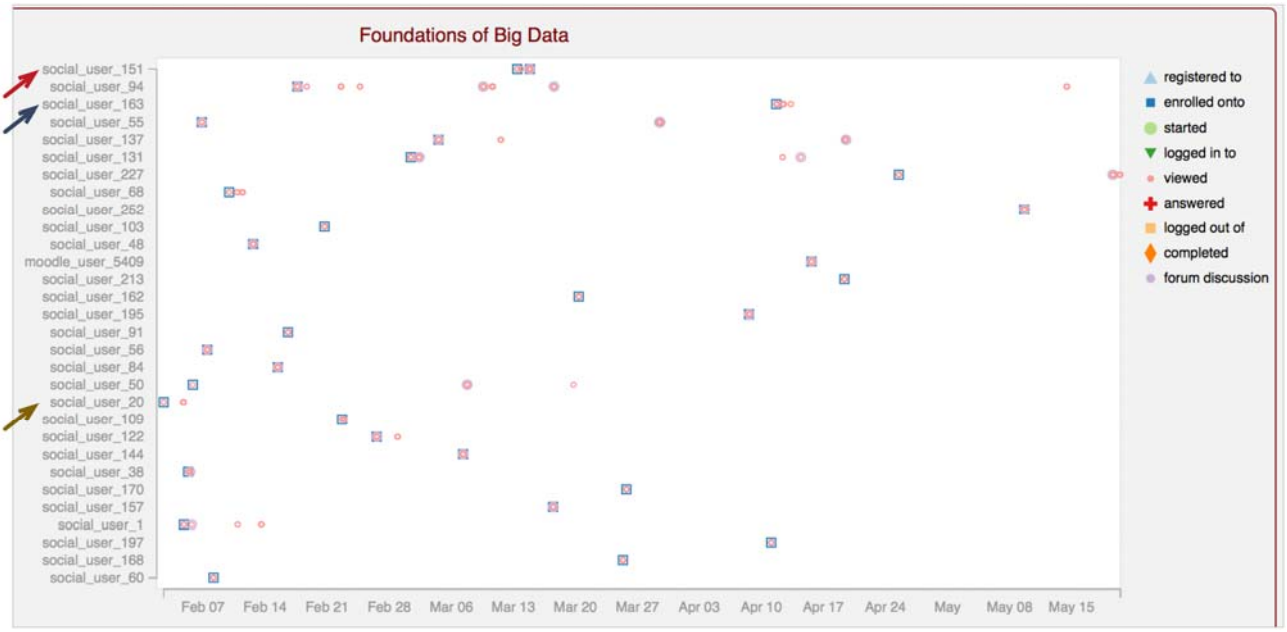


Figure 4.31: Activity trends for the top 30 most active learners in this course only, ordered by total number of events per user ID for this course. The chart at the top shows the exponential drop in activity for the data in Table 4.2, while the bottom details activity type for these users.

We see one cluster in the centre of Figure 4.31 for the activity of *social_user_151* - zooming into this region at the top, Figure 4.32, shows activity in three bursts over almost two days. The first two clusters (middle) show initial enrolment followed by a view event. After a break of almost nine hours the user triggers another view event. After another pause of about a day we see most of the activity of this user (bottom), including interaction with the course's discussion fora and enrolment (for a second time) onto the same course, more than halfway through this session.

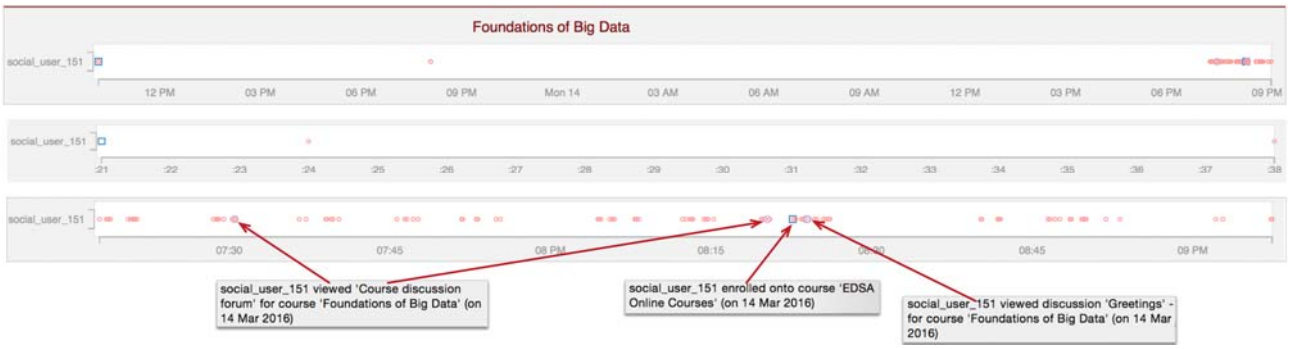


Figure 4.32: Zooming in to the first (middle) and last (bottom) of three clusters for social_user_151, the most active user for the course.

While with a third fewer events recorded than the most active user for this course, *social_user_94* breaks the trend by engaging with the course over two months. Zooming in from the overview for the top 30 users (in Figure 4.31) to activity for only *social_user_94*, Figure 4.33 shows the norm, bursty interaction. We examine the first few days of interaction (top two, with increasing zoom into the start), the fourth

and the final cluster of events. In addition to course module resources we see fairly regular interaction with the discussion fora.

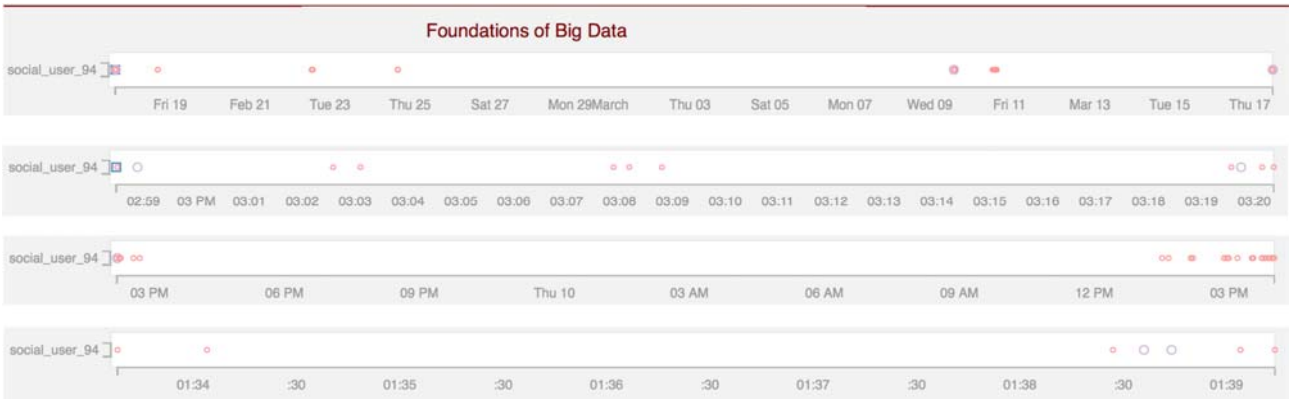


Figure 4.33: Zooming in to the clusters to reveal detail for the second most active user, social_user_94, for the course "Foundations of Big Data".

In addition to the methods used above for visual analytics, by analogy to the analysis carried out using the VideoLectures Learning Analytics dashboard (section 3.1) and to aid linking of the outcomes of analysis we created a tool for analysing the EDSA Learning Locker logs.

Figure 4.34 presents a visualisation of activity over time for the course of interest, the EDSA online course “Foundations of Big Data”.

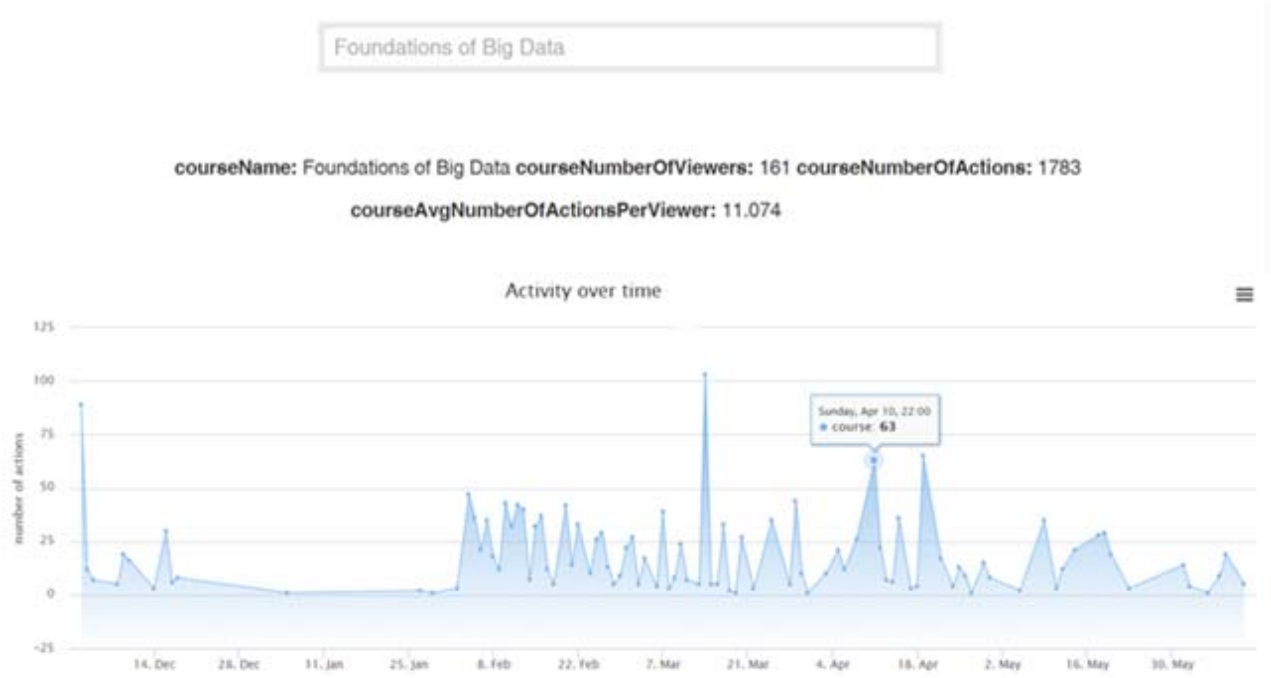
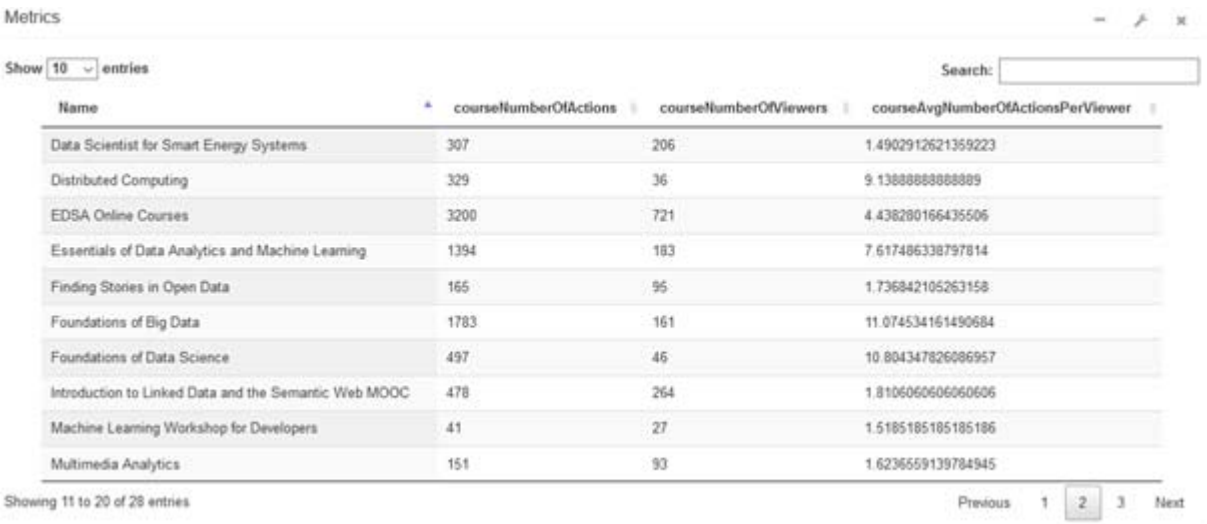


Figure 4.34: Activity over time for the course “Foundations of Big Data”.

Figure 4.35 shows metrics for a number of EDSA online courses, including "Foundations of Big Data". The graph in Figure 4.34 and the metrics here confirm that this is the most popular course (discounting that for the entry to the portal labelled as "EDSA Online Courses"), with over 1700 actions and 161 viewers to date, and on average 11 actions per viewer.


Metrics

Show 10 entries Search:

Name	courseNumberOfActions	courseNumberOfViewers	courseAvgNumberOfActionsPerViewer
Data Scientist for Smart Energy Systems	307	206	1.4902912621359223
Distributed Computing	329	36	9.13888888888889
EDSA Online Courses	3200	721	4.438280166435506
Essentials of Data Analytics and Machine Learning	1394	183	7.617486338797814
Finding Stories in Open Data	165	95	1.736842105263158
Foundations of Big Data	1783	161	11.074534161490684
Foundations of Data Science	497	46	10.804347826086957
Introduction to Linked Data and the Semantic Web MOOC	478	264	1.8106060606060606
Machine Learning Workshop for Developers	41	27	1.5185185185185186
Multimedia Analytics	151	93	1.6236559139784945

Showing 11 to 20 of 28 entries Previous 1 2 3 Next

Figure 5.35: Courses Metrics for a number of courses available from the EDSA Courses portal, including our focus - "Foundations of Big Data".

3.4.2 Process Mining

Finally, we apply process mining to the same course. The data studied is extracted from a set of page visits from the courses portal[1] on the EDSA project website, each visit comprising the URL of the page visited, accompanied by the ID of the user and the timestamp of the visit.

An *event* here is defined as a page visit. A *resource* is one of a pdf file, a video or a quiz. Out of the events, we discard visits to the homepage or a course list page, keeping exclusively the visits to resources. We also discard events from the users "admin", "demo" or "guest". Finally, eliminating the 36 users who visited only one resource, 121 users remain. These spent a median time of 25 minutes between their first and last page visited on the whole website. This low value reflects the fact that many users just browse through a few of the resources before ending their visit to the course section of the website; the upper limit for the time span on this course to this point is three months.

"Foundations of Big Data" was found to be the most popular course - by number of events, enrolment and access of resources across the course. All resources in this course were accessed at least 10 times. We see much lower coverage and overall access in all other courses in the EDSA portal, where the majority of resources were accessed less than six times per course. "Foundations of Big Data" is the only course that currently provides sufficient data to obtain meaningful results using process mining. To aid human interpretation of our results we match each event to its corresponding (human-readable) name, e.g., a URL to the title of a course module.

3.4.3 Normative model

For users who visited at least two resources in the course, 27% are repeat visits to the previous resource accessed. 82% of event changes are to the next expected event, as prescribed by course order/structure. We also observe that users drop out of the course at any time, except in the middle of late chapters (while they could if so they wanted).

Figure 4.36 shows the model built manually to match the patterns observed, which follows the path prescribed by the course designer or instructor. The model is linear, with the events ordered as expected for the course on the second line. The leftmost circle starts the process, and the lone circle at the bottom (fourth line) terminates the process. The rectangles on the first line each represent one resource in the course, any of whose events may be repeated. The grey rectangles on the third line allow to drop out the course at any time, leading to the end of the process on the fourth line.

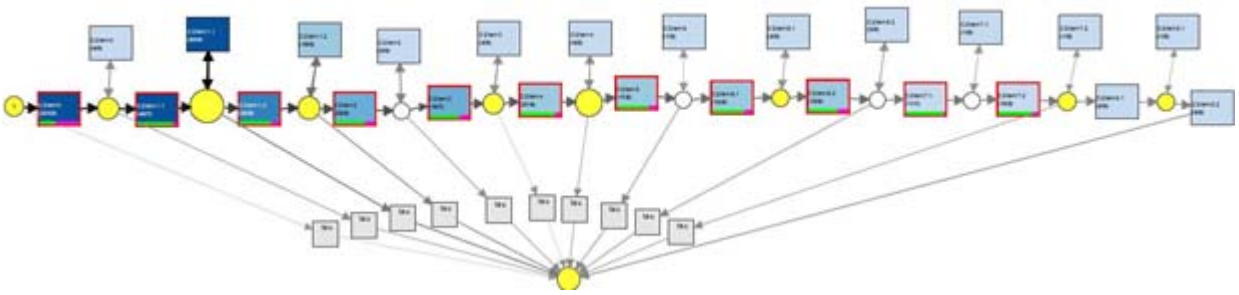


Figure 4.36: Model illustrating patterns observed in the course "Foundations of Big Data".

3.4.4 Analysis

User actions match well the expected model. In the detailed view (Figure 4.37), the green and pink lines at the bottom of each state node represent the proportion of events taking place as expected (green) vs. events not taking place as expected (pink). Where there is no green/pink line all events took place as expected (see the events on the first line that represent repetitions). Another indication of difference between the data and the model is the size of the yellow circles: when they are larger, more irrelevant events take place. The pink line represents events that are in the model but not in the data, while the yellow circles represent events that are in the data but not in the model.

The first event, the course syllabus (first rectangle on the second line) shows an exception to our prediction: the majority of users skip the syllabus. We see therefore a predominantly pink line at the bottom of this event. A second exception is the large yellow (third) circle on the second line: many events following viewing of the first element of the curriculum are followed by an arbitrary further item, which may represent the wish of users to have an indication of where this course is going before going on or giving up.

The rest of the normative model is followed as expected. Standard metrics for the quality of the model report precision as 0.7, and the fitness as 0.9.

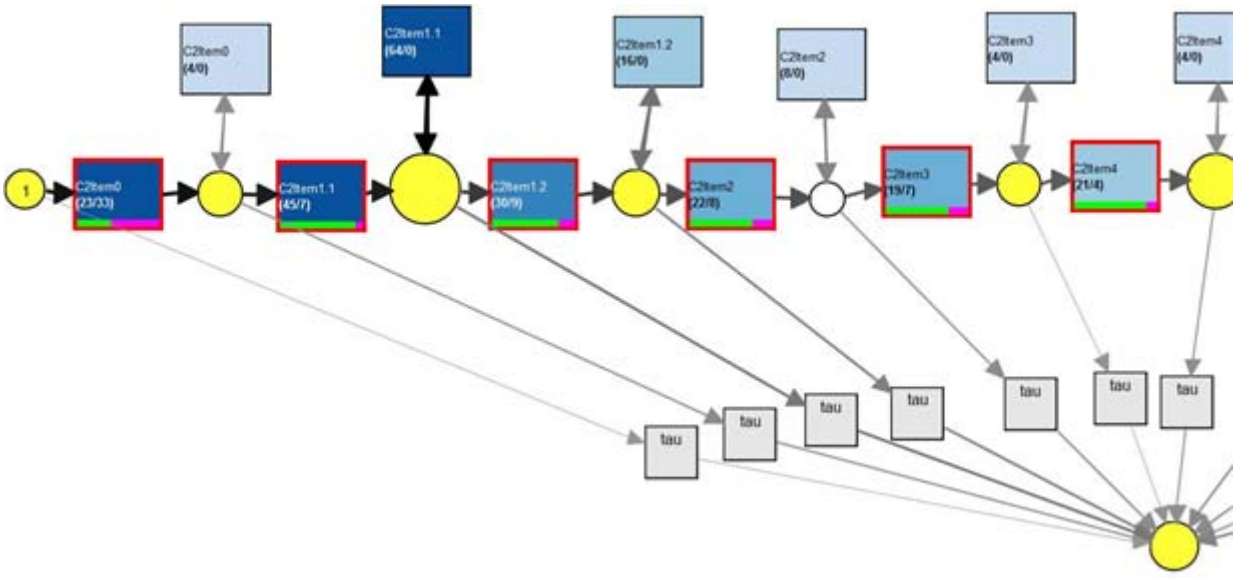


Figure 4.37: Detail of the model for the course "Foundations of Big Data", showing, largely, conformance to course structure as prescribed, with two exceptions to the rule at the start of the process.

We discuss in section 4 the results obtained for this course and for the other data on online courses analysed in this section, and point to further requirements for the detailed analysis required to obtain deeper understanding of student behaviour, potential causes of these and how we can feed our findings into improving the student experience.

.

4. Discussion

4.1 Initial Overviews of Student Online Learning Behaviour

Investigating student behaviour for the data collected for three independent sets of online course material before honing in to a single focus - the most popular course (by event count and enrolment) available from the EDSA Online Courses portal, using the three different approaches and perspectives, provides broader insight into user actions. The first two analysis approaches provide overviews of student behaviour from the perspective of the student accessing the material. The third approach investigates behaviour from the perspective of the course as prescribed in the curriculum, before overlaying actual behaviour on top of this.

The different perspectives thus obtained in this preliminary analysis reveal different ROIs, each of which trigger more detailed investigation that should help to understand better user behaviour and potential reasons for actions taken - why a course is initially accessed, why a user is likely to drop out and when.

The Coursera MOOC and the EDSA Online Courses portal require students to commit to a number of study sessions to complete a course. These courses all require viewing and/or reading of lecture and supplementary learning material, some additionally with requirements for completing assignments, quizzes and in some cases, exams to both complete and successfully pass the course. The VideoLectures portal on the other hand allows users to select and view single, independent lectures, tutorials and other learning material across a broad range or within a specified category(ies).

Our analysis starts with the generation of broad overviews of interaction with the material, looking at topics that attract interest, what within the broader topic or individual course is accessed, and what other actions users carry out during their study sessions. Information on user demographics is limited, due mainly to privacy issues; further work will seek ways to capture, without violating ethics or privacy, anonymous or collective information on user backgrounds, in order to feed this into identifying potential (external) triggers for student actions that negatively impact their ability to successfully complete a course.

For the self-study courses we find, overall, enrolment occurring through to the final stages of courses, but with high attrition soon after individual students' initial interaction with a course - a trend seen in online self-study [Diver et al., 2015; Mukala et al., 2015]. Dropout continues as courses progress both for early and later entrants; however, where students persist through to the later stages of a course they are more likely to complete it. Interaction with course material generally occurs in distinct batches, with periods of no or very little activity seen throughout all courses, even the most popular. This may be due to staggered release of course information; we will investigate this further to identify other contributing factors and any impact this may have on completion and success. Common to all three sets of learning resources and online self-study, however, is the very low bar to and cost of enrolment, especially when compared to f-f courses; relevant studies show that this may also contribute to the the high dropout rate, especially where other competing factors such as time and additional commitments of students override even interest and a desire for self-improvement [Diver et al., 2015; Kevan et al., 2016]. This is

reflected also in the comparison of interaction, performance and completion of the (paying) signature track vs. (free) non-signature track students in section 3.3.

We discuss in more detail in section 4.2 our findings for a single course - that attracting the most interest in the EDSA Online Courses portal. Triangulating the results obtained from the different analysis approaches both reveals insight that a single perspective alone may not, and helps us to identify further techniques that will augment our analysis.

4.2 Triangulation of Results for EDSA Course "Foundations of Big Data"

Our findings indicate that users first enrol onto a course to access its content, and then initially browse course material, sometimes as prescribed by the course structure, or by dipping in and out of a number of course modules, seemingly to get a feel for the course. Further analysis of individual user paths confirmed the initial statistical analysis that indicated variation in user activity. After the initial exploration of course material we see behaviour split into main two cohorts - *drop out* or continued interaction with course material (albeit with further drop out down the line) - see Figures 4.20 and 4.26. On average, fairly short total interaction time is seen per user for a course, even for the most popular. This however increases when considering only users who persist beyond the initial interaction period, from time spans of a few minutes to two hours or more in a single day or session. Overall, users typically access course material using batch mode and sequentially - as per the prescribed order for the course. Again, where students continue to access course material beyond the initial interaction session, the bursty or batch access mode persists over several days, and as an exception, to up to a few months as seen for the users with IDs *social_user_94, social_user_131* and *social_user_137* (see Figure 4.31).

Being online, self-study courses, interaction with other students is naturally restricted. We see some interaction with discussion fora at different points in time for all courses; students who accessed the discussion fora fell tended to be those recording relatively higher total activity count and interaction time on the course. This reinforces potential for inter-student and student-instructor discussion to motivate retention. Kizilcec & Halawa (2015) found similar patterns of activity. Diver et al., (2015) further found in their large scale study MOOCs a link between dropout rate and lower frequency of access to discussion fora, noting also the potential for additional benefit in the social activity. Further analysis in T3.4 will examine what impact of this interaction, with its added social element, has on retention, final course grades. We will also, where we are able to access the content of discussions, investigate how this may be used to determine where students encounter difficulty with course material and why (see also [Bakharia et al., 2016a; Conde et al., 2015]).

While the most popular course, no quiz information is reported for the course "Foundations of Big Data". We therefore do not have information on student grades, and rely only on access to course material as an indication of progress and completion. As T3.4 progresses and more student interaction is captured we will revisit other courses with and without student grading information to map how behaviour as seen in individual and across all courses impacts performance and retention. ROIs identified in each perspective will be further investigated using, where possible, all three analysis approaches, to increase

coverage in our analysis and provide more detail to answer also additional questions raised during our initial analysis.

4.3 Contribution to Face-Face And Blended Courses in EDSA

Learning analytics in WP3 looks at

1. face-face (f-f) and blended courses, where attendance is restricted by student location,
2. online courses, which remove restrictions due to physical access/location, but with the attendant disadvantage of restricted, remote access to instructors and fellow students.

This deliverable focuses on analysis of event data collected from the latter, while D3.2 focuses on the former. While in-person courses benefit from direct contact with teaching staff and other students online, self-study courses have at best remote, often asynchronous access to other participants. We are therefore reliant on event interaction logs and formal, online feedback forms to collect information on students that may be used to aid the interpretation of their behaviour. Conversely, the event logs provide objective records of student interaction with course material, and consequently, maps of relatively detailed student behaviour that is available only through experience and intuition in f-f instruction, and even then only to a limited extent, outside formal assignments and exams. F-F and blended courses on the other hand tend to have fairly detailed information on student demographics, information that is more difficult to collect in online courses, especially where students are not registered for a (paid) certificate of attainment or other qualification. By combining information collected about student backgrounds, interaction and behaviour in these two contexts we envisage more comprehensive, in-depth analysis of student behaviour and the factors that influence this.

We acknowledge that differences in design and delivery of online and f-f courses prevent straightforward mapping of interaction and behaviour. This is also the case within online and f-f or blended courses on different topics and with designed for different contexts, due also to teaching style and relevant learning facilities available. Diver et al., (2015) acknowledge this in their comparison of two MOOCS. Nguyen (2015) compares online learning to f-f courses, to identify what contributes to effective online learning, in order to continue to improve on this form of study. The remainder of the project will see closer synergy between analysis in the two parts of WP3 as outlined in the introduction (section 1), to ensure we achieve the aims of Learning Analytics and those of the project as a whole.

4.4 Synergy with EDSA Demand Analysis Task

A key goal of the analysis of data science job (and skill) demand (WP1) is to determine where training of new Data Scientists and retraining of practitioners in the field or similar fields is needed to close skill gaps identified. The demand analysis task provides information on demand along three key criteria:

1. over time
2. across location (and corresponding spoken or working language)
3. by skill or skill set.

We also consider the impact of a fourth criterion - domain or industry sector, on categorisation and ranking or relevance of skills seen as core to Data Science.

Information gained in the demand analysis task on the categorisation and ranking of skills is to feed into (individual) course and programme design, to allow students to create learning plans combining

courses on skills core to their target role type(s) and domain or sector with those that teach skills typically required along with the former. The demand analysis task aims also to build in support to guide end users in constructing "skill profiles". This will be based also on information collected from practitioners in the field about their perspectives on skills core to their roles and domain, as well as capability within the existing workforce and capacity of the job market to absorb these skills.

The learning analytics task, on the other hand, aims to derive information on effectiveness of course content and delivery and the impact of student demographics (individual and group) and additional factors such as motivation on course completion and grades. By demonstrating (relative) importance of selected skills to a role or job type and optimal paths to follow to these, analysis results from the demand analysis task should complement findings in learning analytics, to allow more informed decisions to be made in course and curricula design.

Information obtained on interest and engagement with specific courses and categories of courses will also be fed into recommendation of learning material, both complete courses that require a good degree of commitment, with the opportunity to obtain verifiable qualifications, and standalone and bite-size learning resources - lectures, tutorials and presentations, among others, that require lower effort but are able also to supplement more structured learning in related courses.

4.5 Initial Proposal for a 'Learning Analytics Framework' for EDSA

The overall aim of Learning Analytics in the EDSA project is to follow evidence-based best practice:

- for course design (content and delivery),
- to structure and tailor student feedback to guide selection of courses, and following this, encourage retention and provide the resources required to ensure student completion and success (both the individual and the cohort),
- to provide formal measures and benchmarks describing student behaviour, performance and retention for reuse within and outwith the project, based on:
    - (in addition to) course material:
    - data on student consumption of material (passive engagement),
    - student co-construction of data - through interaction in discussion fora, completion of assignments, open-ended quizzes and exam responses (active engagement).

We aim to use this information and our analysis to obtain reliable, data-driven insight into student behaviour and identify critical and more general factors that impact success [see, among others, Bakharia et al., 2016a; Diver et al., 2015; Kevan et al., 2016; Nguyen 2015; Wells et al., 2016]. Based on the preliminary analysis detailed in this deliverable we have identified core information required to achieve this aim, and differences in granularity as well as gaps in the current data collection or reporting functionality in the core datastores we feed from for our analysis. To succeed here we rely on detailed, comprehensive information on

- student demographics (anonymised and collated where necessary to protect privacy) to support definition of and clustering based on student types. Demographical data must include:
    - age and gender,
    - geographical location (home, school and/or work),
    - highest level of educational attainment,
    - for in-work students - occupation along with domain of specialisation and current industry sector.

- detailed objective and subjective information on student behaviour collected across different course types and delivery mechanisms (self-study or guided/structured online, f-f and blended courses), including but not limited to:
    - access to course material with indication of type, time spent on each, sub-levels of interaction where applicable,
    - relationships between course material, indicating also where required (core/) or supplementary.

We therefore propose the construction of a formal framework for Learning Analytics in EDSA, with potential for reuse beyond the project remit, as follows:

1. the definition of course types; our analysis revealed clear distinction between the three datasets we investigated - EDSA-specific courses, MOOCs and VideoLectures. Different criteria may be used for defining these; further discussion within the project and further study of the literature should add to those seen here - standalone learning material as lectures, tutorials, workshop, papers; complete courses broken down into defined sub-topics with or without explicit assessment and options for certification. This is to allow us to more effectively manage student expectation of course content, commitment and certification type, among others.
2. the definition of a structure for each core or prescribed event, action or case captured during student interaction with a course,
    a. to support the derivation of metadata based on this structure for the data collected,
    b. and guide effective (re)use of analysis results.
3. clear distinction between resources in a course based on:
    a. format and delivery mechanism, e.g., video vs. text page vs. downloadable material (e.g., images or printable PDFs of a complete module),
    b. content type, e.g., course module vs. course description (e.g., syllabi) vs. discussion forum vs. supplementary information (e.g., external webinars, related datasets and experiments),
    c. availability of material and requirements for completing assignments, quizzes and exams.

As T3.4 progresses we will refine the EDSA LA framework, with an aim to provide a structured workflow for EDSA and outwith the project. We envisage the framework will also support the identification of optimal analysis and results presentation techniques for different target user types - students, instructors and other interested parties such as course design teams and policy-makers in higher education, on-the-job skills training and the recruitment market.

5. Conclusions

This deliverable reports initial work in T3.4, on the "design and deployment of learning analytics" within EDSA, to aggregate, harmonise and analyse event data generated through student interaction with online courses created for the EDSA project and by project partners and third parties, to train individuals to develop capability required to work in assorted jobs in Data Science across the EU. The aim is to monitor student behaviour in order to obtain insight into the student experience and the impact of student behaviour, demographics and interests on course completion and level of success or attainment. Findings from T3.4 are to feed into the design of new courses and curricula for Data Science programmes, to enable tailoring of content and delivery mechanisms to the student context. Further, information obtained about demographic and other external factors that lead to non-completion or failure or otherwise negatively impact the student experience will feed into the development of both student and instructor-specific feedback and support mechanisms.

Initial, exploratory analysis in T3.4 served three main purposes:

1. to obtain an understanding of the structure and content of the key data stores we must rely on for learning analytics in WP3, and therefore collect additional requirements for data capture and reuse, to ensure we obtain data that will enable us to answer the questions posed in T3.4.
2. to build data overviews that helped us to obtain a picture of student behaviour, and that will continue to feed into identifying additional tools and techniques to employ for more detailed analysis of interesting patterns and trends revealed in the bigger picture.
3. to provide early indicators of the links from this task to other work in EDSA, first, analysis also of student behaviour in the f-f and blended courses also delivered in EDSA. To meet overall project goals we must also cross-fertilise findings in T3.4 with those in demand and skill analysis tasks in WP1, to guide selection of courses to meet students' interests and skill development requirements for target job roles, and to improve course completion and success, and ultimately, success in filling job roles in Data Science.

5.1 Next Steps

Based on our findings to date and lessons learnt in our initial analysis we conclude the deliverable by presenting our plans for completing the work in T3.4 during the second half of the project.

The visualisation-driven dashboards and other visual representations employed in process mining (described in the analysis in section 3) paint a picture of interest and engagement across the three course types we explored. The VideoLectures dashboards specifically also indicate where analysis in T3.4 aligns with EDSA work on (data science skill and job) demand analysis.

To ensure we build on and contribute to the state of the art in LA as T3.4 progresses we must expand on the exploratory analysis that looks at user interaction at the course and system level, to investigate in more detail sub-levels in the more general viewing behaviour reported in this deliverable. We aim to feed the findings obtained to date into more in-depth analysis employing other analytical and data mining techniques, and considering also relevant research on distance learning, learning at scale, technology-enhanced learning and educational data mining (EDM). This more detailed analysis will examine the dynamics in dedicated courses and other learning resources in Data Science. We will also

cluster and merge users based on additional demographic and contextual data, e.g., source IP, information describing students collected at registration and from course feedback forms, to identify patterns influenced by student background and context, and how behaviour differs between course types and learning resources as defined in section 4.5.

The Learning Analytics community recognises the importance of access to open data and resources in advancing research in the field, as well as for practical, successful delivery of online courses. Practical research and application of LA rely on real use cases and real or synthetic data (typically real data anonymised to allow reuse without violating privacy or ethics as in the EDSA LMS). VideoLectures.NET provides an independent, open resource containing learning resources that we have already started to investigate. To ensure our analysis approach and outcomes are reusable and replicable we must examine also other relevant use cases and resources. We will examine pilots of Learning Analytics tools embedded into or feeding directly from LMS such as OU Analyse, which is carrying out live analysis of data from the Virtual Learning Environment (VLE) of the Open University in the UK to provide guidance on the structuring of feedback to distance learners, to aid retention and feed into the university's aim to improve the student experience through its "Students First: Strategy for Growth". Another example is Edinburgh University's LARC project[23] (the Learning Analytics Report Card), which directly involves students in the collection and analysis of data on their learning activity and progress.

WP3 in EDSA also considers the analysis of student interaction and behaviour in face-face and blended courses. This preliminary report considers only online data; as Task 3.4 progresses we will look also at the outcomes of analysis and lessons learnt in courses that see physical interaction between students and their instructors, to identify where similar behaviour is observed, and where success using one method of delivery may fed into overcoming limitations in the other.

As part of the contribution of the EDSA project to research and practical application of Learning Analytics we are exploring, in conjunction with the definition of the framework discussed in section 4.5, the construction of semi-automated workflows for learning analytics. To work toward this goal we will employ the techniques presented in this deliverable - statistics, visual analytics and process mining - with the visualisation of input data and analysis results, and also other techniques we find to augment further detailed analysis. We will build on the insight gained through the exploratory analysis carried out to this point, to continue to explore options toward both (automated) data mining and analysis to handle data on a scale much larger than that employed here, and for human-driven exploration, analysis and management of "own" data.

We will also explore the design and construction of intuitive user interfaces to guide the presentation and reuse of findings for different audiences. For example, an instructor would download data on a course they are delivering along with student interaction and demographic data from their platform of choice (e.g., the EDSA portal, FutureLearn or Coursera), and feed this into the tools we provide them with to trigger, transparently, semi-automated scripts and analysis modules that would generate

[23] http://www.de.ed.ac.uk/project/learning-analytics-report-card

reports on student progress. They would then be provided with intuitive, low-effort tools for creating custom reports to answer queries typical in online education delivery, and to guide them in obtaining answers to questions outside those typically posed within this scope. Corresponding support would be provided to students and policy and decision-makers to feed the findings from learning analytics into study, their regular work and into occasional, but related tasks.

Finally, In addition to expanding our data collection process to include additional metadata as detailed in section 4.5, we aim to feed into our analysis findings from research on learning analytics in other related projects (as discussed above), and link to open datasets (such as the LAK Dataset & Challenge[24]; see also Dietze et al., (2015)) collected as part of such research. The ultimate aim is to provide, based on the more in-depth analysis and wider coverage this broader scope will enable, evidence-based guidelines for further research and for practical use of the outcomes of Learning Analytics in EDSA, to augment course delivery and learning, and improve the student experience. Linking to existing datasets will also support an important goal in EDSA and EU-funded research, to provide as Linked Open Data data generated and the outcomes of research.

[24] http://meco.l3s.uni-hannover.de:9080/wp2

6. References

[van der Aalst 2011] van der Aalst, Wil. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer

[Bakharia et al., 2016a] Bakharia, Aneesha, Corrin, Linda, de Barba, Paula, Kennedy, Gregor, Gašević, Dragan, Mulder, Raoul, Williams, David, Dawson, Shane and Lockyer, Lori. (2016). A conceptual framework linking learning design with learning analytics. In *Proc., Sixth International Conference on Learning Analytics & Knowledge* (LAK '16), pp. 329-338.

[Bakharia et al., 2016b] Bakharia, Aneesha, Kitto, Kirsty, Pardo, Abelardo, Gašević, Dragan and Dawson, Shane. (2016). Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In *Proc., Sixth International Conference on Learning Analytics & Knowledge* (LAK '16), pp., 378-382.

[Conde et al., 2015] M. Á Conde, F. J. García-Peñalvo, D. A. Gómez-Aguilar and R. Therón. (2015), Exploring Software Engineering Subjects by Using Visual Learning Analytics Techniques, *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 10(4), pp. 242-252.

[Dietze et al., 2015] Dietze, S., Taibi, D., d'Aquin, M. (preprint - 2015), Facilitating Scientometrics in Learning Analytics and Educational Data Mining – the LAK Dataset, *Semantic Web Journal*.

[Diver et al., 2015] Diver, P., Martinez, I. (2015). MOOCs as a massive research laboratory: opportunities and challenges, *Journal of Distance Education*, 36(1), pp. 5-25.

[Gunther & van der Aalst 2007] Gunther, Christian and van der Aalst, Wil. (2007). *Fuzzy Mining - Adaptive Process Simplification Based on Multi-perspective Metrics*, Springer.

[Kevan et al., 2016] Kevan, Jonathan M., Menchaca, Michael P. and Hoffman, Ellen S. (2016). Designing MOOCs for success: a student motivation-oriented framework. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (LAK '16), pp. 274-278.

[Kizilcec & Halawa 2015] Kizilcec, R. F. and Halawa, S. (2015). Attrition and Achievement Gaps in Online Learning. *Proc., 2nd ACM Conference on Learning @ Scale* (L@S '15), pp.57-66.

[Kuzilek et al., 2015] Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z. and Wolff, A., (2015). OU Analyse: Analysing At-Risk Students at The Open University, *Learning Analytics Review*, LAK15-1.

[Mukala et al., 2015] Mukala, Patrick, Buijs, Joos C. A. M., Leemans, Maikel, van der Aalst, Wil M. P. (2015). Learning Analytics on Coursera Event Data: A Process Mining Approach. *Proc., 5th International Symposium on Data-driven Process Discovery and Analysis* (SIMPDA 2015), pp. 18-32.

[Nguyen 2015] Nguyen, T. (2015). The Effectiveness of Online Learning:Beyond No Significant Difference and Future Horizons, *Journal of Online Learning and Teaching*, 1(2), pp. 309-319.

[Thomas & Cook 2005] Thomas, J.J. and Cook, K.A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE CS Press

[Wells et al., 2016] Wells, Marc, Wollenschlaeger, Alex, Lefevre, David, Magoulas, George D. and Poulovassilis, Alexandra. (2016). Analysing engagement in an online management programme and implications for course design. In *Proc., Sixth International Conference on Learning Analytics & Knowledge* (LAK '16), pp. 236-240.

7. Appendices

7.1 Appendix A.- List of Queries

Events for online courses delivered by EDSA or through its website are captured in the EDSA Learning Locker (see section 2.1). For the analysis reported in this deliverable on EDSA courses, the following (key) queries were run, with sample results as below. In addition to queries providing complete dumps for a specified user(s), a course, or all users/courses we use aggregate queries to flatten the data structure and ease the extraction of data specific to targeted analysis, based on the findings from the start of the analysis process used to identify ROIs to examine in more detail. We list below key queries, with snapshots of the results obtained.

7.1.1 Follow a Named User

```
db.statements.aggregate([
                { "$match": { "statement.verb.id" : { $exists: true },
                 "statement.actor.name" : { $in: [ "social_user_163", "social_user_20", "social_user_94", "social_user_131", "social_user_137",
"moodle_user_3270", "social_user_202", "social_user_189", "social_user_304", "social_user_95" ] }
                 }
                }, {
                "$group": { "_id": {
studentId: "$statement.actor.name",
 activity: "$statement.verb.id",

courseId: "$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.courseid",
courseName:
"$statement.context.contextActivities.grouping.definition.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_cours
e.fullname",
courseUrl:
"$statement.context.contextActivities.grouping.definition.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_cours
e.url",

courseUrlAccessed: "$statement.context.contextActivities.grouping.id",

 activityLocationId: "$statement.context.contextActivities.parent.id",
 activityDefinitionType: "$statement.context.contextActivities.parent.definition.type",
 activityDefinitionName: "$statement.context.contextActivities.parent.definition.name.en",

 quizResult: "$statement.result.score.raw",
quizSuccess: "$statement.result.success",
quizCompletion: "$statement.result.completion",

 activityTriggerId: "$statement.object.id",
 activityTriggerDefinitionType: "$statement.object.definition.type",
 activityTriggerDefinitionName: "$statement.object.definition.name.en",
 activityTriggerUrl: "$statement.object.extensions.url",
activityTriggerExternalUrl:
"$statement.object.definition.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_module.externalurl",

activityContextEventname:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.eventname",
activityContextComponent:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.component",
activityContextAction:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.action",
activityContextObjectTable:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.objecttable",
```

```
activityContextObjectid:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.objectid",
activityContextOtherId:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.other",
activityContextTimecreated:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.timecreated",
activityContextOriginatingDevice:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.origin",
activityContextObject:
"$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.object",

ip: "$statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.ip",
timestamp: "$timestamp",
updatedAt: "$updated_at"
                        }
                    }
        }
]])
```

7.1.2 Results Snapshot

```
{
        "_id" : {
                "studentId" : "social_user_304",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "15",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/course/view.php?id=15",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/course",
                "activityTriggerDefinitionName" : "Process Mining",
                "activityContextEventname" : "\\core\\event\\course_viewed",
                "activityContextComponent" : "core",
                "activityContextAction" : "viewed",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1467709645),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course",
                "ip" : "131.155.69.119",
                "timestamp" : ISODate("2016-07-05T09:07:25Z"),
                "updatedAt" : ISODate("2016-07-05T09:07:26.504Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_304",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "33",
                "courseName" : [
                        "EDSA Online Courses",
                        "Big Data Analytics"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=33"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=33"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/quiz/view.php?id=340",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/quiz",
```

```
                "activityTriggerDefinitionName" : "Quiz",
                "activityContextEventname" : "\\mod_quiz\\event\\course_module_viewed",
                "activityContextComponent" : "mod_quiz",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "quiz",
                "activityContextObjectid" : "7",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1465667786),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course_module",
                "ip" : "145.120.15.201",
                "timestamp" : ISODate("2016-06-11T17:56:26Z"),
                "updatedAt" : ISODate("2016-06-11T17:56:27.310Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_304",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "27",
                "courseName" : [
                        "EDSA Online Courses",
                        "Big Data Architecture"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/forum/view.php?id=288",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/forum",
                "activityTriggerDefinitionName" : "Course discussion forum",
                "activityContextEventname" : "\\mod_forum\\event\\course_module_viewed",
                "activityContextComponent" : "mod_forum",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "forum",
                "activityContextObjectid" : "19",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1465666548),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course_module",
                "ip" : "145.120.15.201",
                "timestamp" : ISODate("2016-06-11T17:35:48Z"),
                "updatedAt" : ISODate("2016-06-11T17:35:48.935Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_304",
                "activity" : "http://adlnet.gov/expapi/verbs/answered",
                "courseId" : "33",
                "courseName" : [
                        "EDSA Online Courses",
                        "Big Data Analytics"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=33"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=33",
                        "http://courses.edsa-project.eu/mod/quiz/attempt.php?attempt=30"
                ],
                "activityLocationId" : [
                        "http://courses.edsa-project.eu/mod/quiz/view.php?id=340"
                ],
                "activityDefinitionType" : [
                        "http://lrs.learninglocker.net/define/type/moodle/quiz"
                ],
                "activityDefinitionName" : [
```

```
                        "Quiz"
                ],
                "quizSuccess" : false,
                "quizCompletion" : false,
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/question/question.php?id=96",
                "activityTriggerDefinitionType" : "http://adlnet.gov/expapi/activities/cmi.interaction",
                "activityTriggerDefinitionName" : "10",
                "activityContextEventname" : "\\mod_quiz\\event\\attempt_reviewed",
                "activityContextComponent" : "mod_quiz",
                "activityContextAction" : "reviewed",
                "activityContextObjectTable" : "quiz_attempts",
                "activityContextObjectid" : "30",
                "activityContextOtherId" : "a:1:{s:6:\"quizid\";s:1:\"7\";}",
                "activityContextTimecreated" : NumberLong(1465667775),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "attempt",
                "ip" : "145.120.15.201",
                "timestamp" : ISODate("2016-06-11T17:56:14Z"),
                "updatedAt" : ISODate("2016-06-11T17:56:15.862Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_202",
                "activity" : "http://adlnet.gov/expapi/verbs/answered",
                "courseId" : "27",
                "courseName" : [
                        "EDSA Online Courses",
                        "Big Data Architecture"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27",
                        "http://courses.edsa-project.eu/mod/quiz/attempt.php?attempt=18"
                ],
                "activityLocationId" : [
                        "http://courses.edsa-project.eu/mod/quiz/view.php?id=325"
                ],
                "activityDefinitionType" : [
                        "http://lrs.learninglocker.net/define/type/moodle/quiz"
                ],
                "activityDefinitionName" : [
                        "Quiz"
                ],
                "quizResult" : NumberLong(1),
                "quizSuccess" : true,
                "quizCompletion" : true,
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/question/question.php?id=82",
                "activityTriggerDefinitionType" : "http://adlnet.gov/expapi/activities/cmi.interaction",
                "activityTriggerDefinitionName" : "6",
                "activityContextEventname" : "\\mod_quiz\\event\\attempt_reviewed",
                "activityContextComponent" : "mod_quiz",
                "activityContextAction" : "reviewed",
                "activityContextObjectTable" : "quiz_attempts",
                "activityContextObjectid" : "18",
                "activityContextOtherId" : "a:1:{s:6:\"quizid\";s:1:\"6\";}",
                "activityContextTimecreated" : NumberLong(1461099312),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "attempt",
                "ip" : "77.11.16.208",
                "timestamp" : ISODate("2016-04-19T20:55:11Z"),
                "updatedAt" : ISODate("2016-04-19T20:55:12.841Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_202",
                "activity" : "http://adlnet.gov/expapi/verbs/completed",
                "courseId" : "27",
                "courseName" : [
                        "EDSA Online Courses",
```

```
                        "Big Data Architecture"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=27",
                        "http://courses.edsa-project.eu/mod/quiz/attempt.php?attempt=18"
                ],
                "quizResult" : 9.5,
                "quizCompletion" : true,
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/quiz/view.php?id=325",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/quiz",
                "activityTriggerDefinitionName" : "Quiz",
                "activityContextEventname" : "\\mod_quiz\\event\\attempt_reviewed",
                "activityContextComponent" : "mod_quiz",
                "activityContextAction" : "reviewed",
                "activityContextObjectTable" : "quiz_attempts",
                "activityContextObjectid" : "18",
                "activityContextOtherId" : "a:1:{s:6:\"quizid\";s:1:\"6\";}",
                "activityContextTimecreated" : NumberLong(1461099312),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "attempt",
                "ip" : "77.11.16.208",
                "timestamp" : ISODate("2016-04-19T20:55:12Z"),
                "updatedAt" : ISODate("2016-04-19T20:55:12.841Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_163",
                "activity" : "http://adlnet.gov/expapi/verbs/registered",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu",
                "activityTriggerDefinitionType" : "http://id.tincanapi.com/activitytype/site",
                "activityTriggerDefinitionName" : "EDSA Online Courses",
                "activityContextEventname" : "\\core\\event\\user_created",
                "activityContextComponent" : "core",
                "activityContextAction" : "created",
                "activityContextObjectTable" : "user",
                "activityContextObjectid" : NumberLong(166),
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1458491893),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "user",
                "ip" : "185.18.231.50",
                "timestamp" : ISODate("2016-03-20T16:38:13Z"),
                "updatedAt" : ISODate("2016-03-20T16:38:13.865Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_163",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
                ],
```

```
                "activityTriggerId" : "http://courses.edsa-project.eu/course/view.php?id=26",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/course",
                "activityTriggerDefinitionName" : "Foundations of Big Data",
                "activityContextEventname" : "\\core\\event\\course_viewed",
                "activityContextComponent" : "core",
                "activityContextAction" : "viewed",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1460454440),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course",
                "ip" : "192.168.101.23",
                "timestamp" : ISODate("2016-04-12T09:47:20Z"),
                "updatedAt" : ISODate("2016-04-12T09:47:20.606Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_94",
                "activity" : "https://brindlewaye.com/xAPITerms/verbs/loggedout/",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu",
                "activityTriggerDefinitionType" : "http://id.tincanapi.com/activitytype/site",
                "activityTriggerDefinitionName" : "EDSA Online Courses",
                "activityContextEventname" : "\\core\\event\\user_loggedout",
                "activityContextComponent" : "core",
                "activityContextAction" : "loggedout",
                "activityContextObjectTable" : "user",
                "activityContextObjectid" : "97",
                "activityContextOtherId" : "a:1:{s:9:\"sessionid\";s:26:\"rfh9sd79au2crus7bv6ovt4us0\";}",
                "activityContextTimecreated" : NumberLong(1457530739),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "user",
                "ip" : "195.96.242.52",
                "timestamp" : ISODate("2016-03-09T13:38:59Z"),
                "updatedAt" : ISODate("2016-03-09T13:38:59.337Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_137",
                "activity" : "https://brindlewaye.com/xAPITerms/verbs/loggedin/",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu",
                "activityTriggerDefinitionType" : "http://id.tincanapi.com/activitytype/site",
                "activityTriggerDefinitionName" : "EDSA Online Courses",
                "activityContextEventname" : "\\core\\event\\user_loggedin",
                "activityContextComponent" : "core",
                "activityContextAction" : "loggedin",
                "activityContextObjectTable" : "user",
                "activityContextObjectid" : "140",
                "activityContextOtherId" : "a:1:{s:8:\"username\";s:15:\"social_user_137\";}",
                "activityContextTimecreated" : NumberLong(1457096110),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "user",
                "ip" : "88.98.46.53",
                "timestamp" : ISODate("2016-03-04T12:55:10Z"),
                "updatedAt" : ISODate("2016-03-04T12:55:10.415Z")
        }
}
```

7.1.3 Capture User Activities

The same user may log in using multiple identities or multiple users may log in from a single location. These queries, by sorting on IP address, capture both cases, and we use further analysis to distinguish between the two.

Queries as in 9.1 may be used to extract the data required.

7.1.4 Group by Event within a Course

This set of queries follows activity within courses, to map the path of the initial group registered for a course through to completion, in order to capture interaction with lessons and coursework, and fall-off as the course progresses.

Queries as in 9.1 but with the following match criteria for, e.g., the course with ID 26 - "Foundations of Big Data" as well as events to the login prompt for the EDSA Online courses portal - which captures login, logout and registration events for all other course. This query also filters out events logged by the users *admin*, *guest* and *demo*.

```
{ "$match": { "statement.verb.id" : { $exists: true },
"statement.context.extensions.http://lrs&46;learninglocker&46;net/define/extensions/moodle_logstore_standard_log.courseid" : { $in: [ "1", "26" ] },
"statement.actor.name" : { $nin: [ "admin", "guest", "demo" ] }
}
}
```

7.2 Appendix B.- Sample Snapshots of Event Data for Course "Foundations of Big Data"


```
{
	"_id" : {
		"studentId" : "social_user_48",
		"activity" : "http://id.tincanapi.com/verb/viewed",
		"courseId" : "26",
		"courseName" : [
			"EDSA Online Courses",
			"Foundations of Big Data"
		],
		"courseUrl" : [
			"http://courses.edsa-project.eu",
			"http://courses.edsa-project.eu/course/view.php?id=26"
		],
		"courseUrlAccessed" : [
			"http://courses.edsa-project.eu",
			"http://courses.edsa-project.eu/course/view.php?id=26"
		],
		"activityTriggerId" : "http://courses.edsa-project.eu/mod/forum/view.php?id=287",
		"activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/forum",
		"activityTriggerDefinitionName" : "Course discussion forum",
		"activityContextEventname" : "\\mod_forum\\event\\course_module_viewed",
		"activityContextComponent" : "mod_forum",
		"activityContextAction" : "viewed",
		"activityContextObjectTable" : "forum",
		"activityContextObjectid" : "18",
		"activityContextOtherId" : "N;",
		"activityContextTimecreated" : NumberLong(1467900935),
		"activityContextOriginatingDevice" : "web",
		"activityContextObject" : "course_module",
		"ip" : "88.0.206.95",
		"timestamp" : ISODate("2016-07-07T14:15:35Z"),
		"updatedAt" : ISODate("2016-07-07T14:15:36.082Z")
	}
},
{
	"_id" : {
		"studentId" : "moodle_user_3804",
		"activity" : "http://id.tincanapi.com/verb/viewed",
		"courseId" : "26",
		"courseName" : [
			"EDSA Online Courses",
			"Foundations of Big Data"
		],
		"courseUrl" : [
			"http://courses.edsa-project.eu",
			"http://courses.edsa-project.eu/course/view.php?id=26"
		],
		"courseUrlAccessed" : [
			"http://courses.edsa-project.eu",
			"http://courses.edsa-project.eu/course/view.php?id=26"
		],
		"activityTriggerId" : "http://courses.edsa-project.eu/mod/page/view.php?id=253",
		"activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/page",
		"activityTriggerDefinitionName" : "Learning objectives & syllabus",
		"activityContextEventname" : "\\mod_page\\event\\course_module_viewed",
		"activityContextComponent" : "mod_page",
		"activityContextAction" : "viewed",
		"activityContextObjectTable" : "page",
		"activityContextObjectid" : "22",
		"activityContextOtherId" : "N;",
		"activityContextTimecreated" : NumberLong(1466171578),
		"activityContextOriginatingDevice" : "web",
		"activityContextObject" : "course_module",
		"ip" : "131.211.81.80",
		"timestamp" : ISODate("2016-06-17T13:52:58Z"),
		"updatedAt" : ISODate("2016-06-17T13:52:58.356Z")
	}
},
{
	"_id" : {
		"studentId" : "social_user_304",
```

```
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses",
                        "Foundations of Big Data"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/forum/view.php?id=287",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/forum",
                "activityTriggerDefinitionName" : "Course discussion forum",
                "activityContextEventname" : "\\mod_forum\\event\\course_module_viewed",
                "activityContextComponent" : "mod_forum",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "forum",
                "activityContextObjectid" : "18",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1465666540),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course_module",
                "ip" : "145.120.15.201",
                "timestamp" : ISODate("2016-06-11T17:35:40Z"),
                "updatedAt" : ISODate("2016-06-11T17:35:41.125Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_296",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses",
                        "Foundations of Big Data"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26",
                        "http://courses.edsa-project.eu/mod/forum/view.php?id=287"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/forum/discuss.php?d=5",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/forum_discussions",
                "activityTriggerDefinitionName" : "Greetings",
                "activityContextEventname" : "\\mod_forum\\event\\discussion_viewed",
                "activityContextComponent" : "mod_forum",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "forum_discussions",
                "activityContextObjectid" : "5",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1465302695),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "discussion",
                "ip" : "82.150.248.28",
                "timestamp" : ISODate("2016-06-07T12:31:35Z"),
                "updatedAt" : ISODate("2016-06-07T12:31:36.264Z")
        }
},
{
        "_id" : {
                "studentId" : "moodle_user_9856",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses",
                        "Foundations of Big Data"
```

```
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/url/view.php?id=315",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/url",
                "activityTriggerDefinitionName" : "Your feedback",
                "activityTriggerExternalUrl" : "http://courses.edsa-project.eu/mod/feedback/view.php?id=285&courseid=26",
                "activityContextEventname" : "\\mod_url\\event\\course_module_viewed",
                "activityContextComponent" : "mod_url",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "url",
                "activityContextObjectid" : "76",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1464765251),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course_module",
                "ip" : "94.224.94.171",
                "timestamp" : ISODate("2016-06-01T07:14:11Z"),
                "updatedAt" : ISODate("2016-06-01T07:14:12.091Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_55",
                "activity" : "http://id.tincanapi.com/verb/viewed",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses",
                        "Foundations of Big Data"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu",
                        "http://courses.edsa-project.eu/course/view.php?id=26"
                ],
                "activityTriggerId" : "http://courses.edsa-project.eu/mod/url/view.php?id=315",
                "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/url",
                "activityTriggerDefinitionName" : "Your feedback",
                "activityTriggerExternalUrl" : "http://courses.edsa-project.eu/mod/feedback/view.php?id=285&courseid=26",
                "activityContextEventname" : "\\mod_url\\event\\course_module_viewed",
                "activityContextComponent" : "mod_url",
                "activityContextAction" : "viewed",
                "activityContextObjectTable" : "url",
                "activityContextObjectid" : "76",
                "activityContextOtherId" : "N;",
                "activityContextTimecreated" : NumberLong(1459247700),
                "activityContextOriginatingDevice" : "web",
                "activityContextObject" : "course_module",
                "ip" : "86.174.237.52",
                "timestamp" : ISODate("2016-03-29T10:35:00Z"),
                "updatedAt" : ISODate("2016-03-29T10:35:00.388Z")
        }
},
{
        "_id" : {
                "studentId" : "social_user_30",
                "activity" : "http://www.tincanapi.co.uk/verbs/enrolled_onto_learning_plan",
                "courseId" : "26",
                "courseName" : [
                        "EDSA Online Courses"
                ],
                "courseUrl" : [
                        "http://courses.edsa-project.eu"
                ],
                "courseUrlAccessed" : [
                        "http://courses.edsa-project.eu"
```

```
                    ],
                    "activityTriggerId" : "http://courses.edsa-project.eu/course/view.php?id=26",
                    "activityTriggerDefinitionType" : "http://lrs.learninglocker.net/define/type/moodle/course",
                    "activityTriggerDefinitionName" : "Foundations of Big Data",
                    "activityContextEventname" : "\\core\\event\\user_enrolment_created",
                    "activityContextComponent" : "core",
                    "activityContextAction" : "created",
                    "activityContextObjectTable" : "user_enrolments",
                    "activityContextObjectid" : NumberLong(54),
                    "activityContextOtherId" : "a:1:{s:5:\"enrol\";s:4:\"self\";}",
                    "activityContextTimecreated" : NumberLong(1454701946),
                    "activityContextOriginatingDevice" : "web",
                    "activityContextObject" : "user_enrolment",
                    "ip" : "196.3.88.52",
                    "timestamp" : ISODate("2016-02-05T19:52:26Z"),
                    "updatedAt" : ISODate("2016-02-05T19:52:27.107Z")
            }
}
```


7.3 Appendix C.- Third Party APIs and Analysis Tools Used

D3.js[25] is employed in a number of visualisations described in this deliverable, including the ***Landscape*** for the VideoLectures.NET portal. The snapshots shown in section 3.2 are taken from a visual analytics prototype built using D3.js. The tool ProM[26] was used for the process mining tasks.

7.3.1 VideoLectures Learning Analytics Dashboard

Qminer[27] - a data analytics platform for processing large-scale, real-time streams containing structured and unstructured data. The platform for data storing, processing and dashboard server side support.

Node.js[28] - a JavaScript runtime built on Chrome's V8 JavaScript engine. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient. Node.js is used for server creation for learning analytics tools for Videolectures.NET.

Highcharts[29] **-** graphs for graphical implementation and support for dynamic interaction.

DataTables[30] - a plug-in for the jQuery Javascript library. It is a highly flexible tool, based upon the foundations of progressive enhancement, and will add advanced inte

[25] https://d3js.org
[26] http://promtools.org
[27] http://qminer.ijs.si
[28] https://nodejs.org/en
[29] http://www.highcharts.com/about
[30] https://datatables.net