ENVISAGE

ENhanced VIrtual learning Spaces using Applied Gaming in Education

H2020-ICT-24-2016

# D2.3 Visualization Strategies for Course Progress Reports

| | |
|---|---|
| **Dissemination level:** | Public (PU) |
| **Contractual data of delivery:** | Month 5, May 30th, 2017 |
| **Actual date of delivery:** | Month 6, June 13th, 2017 |
| **Work package:** | WP2 User data aggregation, shallow game analytics, and visualizing learning |
| **Task:** | T2.3 Data visualization for course progress reports |
| **Type:** | Demonstrator |
| **Approval status:** | Final |
| **Version:** | 1.0 |
| **Number of pages:** | 32 |
| **Filename:** | D2.3 Visualization strategies for course progress reports.pdf |

**Abstract**

This deliverable describes the initial visualization strategies developed for the ENVISAGE project and their implementation. Building from an overview of the state of the art in general Analytics and Game Analytics, it moves on to identify requirements and goals for learning analytics visualization, building from previous deliverables in the project. A number of visualization strategies are presented and their technical implementation is described.



co-funded by the European Union

## History

| Version | Date | Reason | Revised by |
|---|---|---|---|
| 0.1 | 2017-05-13 | First draft | Christoffer Holmgård |
| 0.2 | 2017-06-01 | Review first half | F. Hadiji (review) / C. Holmgård (revisions) |
| 0.3 | 2017-06-07 | Review first half | F. Hadiji (review) / C. Holmgård (revisions) |
| 0.4 | 2017-06-08 | Added details, deltaDNA | Fabian Hadiji |
| 1.0 | 2017-06-12 | Final version | Christoffer Holmgård |
| | | | |

## Author list


| Organization | Name | Contact Information |
|---|---|---|
| University of Malta | Christoffer Holmgård | holmgard@gmail.com |
| Goedle.io | Fabian Hadiji | fabian@goedle.io |
| | | |
| | | |
| | | |
| | | |

# Executive Summary

This deliverable describes the initial visualization strategies developed for the ENVISAGE project and their implementation. Building from an overview of the state of the art in general Analytics and Game Analytics, it moves on to identify requirements and goals for learning analytics visualization, building from previous deliverables in the project. A number of visualization strategies are presented and their technical implementation is described.

# Abbreviations and Acronyms

| | |
|---|---|
| **CSS** | Cascading Style Sheets |
| **DLA** | Deep Learning Analytics |
| **HTML** | HyperText Markup Language |
| **JS** | JavaScript |
| **LA** | Learning Analytics |
| **SLA** | Shallow Learning Analytics |

# Contents

# List of Figures

# 1 Introduction

The purpose of this deliverable is to outline strategies for visualizing learner performance in digital learning environments. In the case of ENVISAGE this primarily means virtual labs. The deliverable provides an initial implementation of visualization and reporting techniques for visual exploration of the collected data that aim to provide useful insights into the learners' profiles and actions.

While the actual deliverable is a Demonstrator; i.e. a software solution that aims to solve the problem of visualizing learning analytics. This companion document aims to provide the context for the delivered software, explaining the implementation and the design choices made during development. Additionally, this companion document details the software libraries used and technical implementation employed.

In the following Section, State of the Art in Game Analytics Visualization, we first provide a brief review of existing solutions from Game Analytics (GA). ENVISAGE is premised on the fact that methods and technology from GA can be transferred into the new domain of Learning Analytics (LA). Hence it may be helpful to identify the state of the art in GA before designing solutions for LA.

Then, in Section 3, Requirements for Visualizing Course Progress Reports, we enumerate and interpret requirements identified in the other deliverables of the ENVISAGE project. We do this to leverage the insight that digital learning environments may be different from digital games in their goals, mechanics, dynamic, and aesthetics. A translation of methods, practices, technology from GA into LA should take this into account, rather than import these wholesale with no modification. In Section 4, Goals of Learning Analytics Visualization, we draw together the information from Sections 2 and 3 and define the goals of LA in ENVISAGE and the derived visualization strategies. Section 5 describes the specific visualization strategies that we developed to achieve these goals, and Section 6 describes the technical architecture and implementation used to bring these visualizations to live, in concert with the other software components of the ENVISAGE project.

# 2 State of the Art in Game Analytics Visualization

A core idea for the ENVISAGE project is to leverage existing practices and ideas from GA for commercial digital games. To enable this, this section briefly reviews visualization strategies in existing commercial solutions for analytics. It covers both some general analytics solutions and some game specific analytics solutions.

To identify the most relevant (Game) Analytics services and solutions to include, we conducted a survey of offerings on-line using the Google search engine along with a review of the game development professionals' site Gamasutra.com. We also did an informal e-mail and in-person survey within our personal network of professional independent game developers, mobile game developers, AAA game developers, and game researchers. Our exploratory research was focused on identifying solutions targeted at (but not limited to) small to medium-size game developers. Through personal communication we learned that large game development studios are likely to either use the same tools as the small to medium sized game developers or to implement their own custom data collection and analytics solutions. In the latter case, the data from is typically handled by in-house analytics and design teams.

Using this approach, we identified 3 general analytics solutions commonly used by game developers as well as 2 predominant game analytics solutions. We also identified 1 research project with significant potential for both Game Analytics and Learning Analytics.

In the section below we describe and draw some general conclusions from these visualization solutions. We do not describe these solutions exhaustively, as that would be beyond the scope of this deliverable, but present the most important take-away from each service, focusing on visualization strategies.

## 2.1 General Analytics solutions

### 2.1.1 App Annie

App Annie [1] is a general analytics platform that provides a number services that may be of interest to commercial game developers. The service focuses on external metrics of App Store games' performances, such as rankings in the various app stores, ratings, and features – all analytics that are valuable to game developers tracking their games' performance. The service also provides analytics on in-app (or in this case in-game) advertising performance. The service does not provide insights into in-game events and as such cannot be used to inform game design or tweaks, except for at the abstract and general level. Given its focus on current data, historical data over time, and comparisons with other apps, App Annie mostly uses bar charts and line graphs to display the collected values. The services provides a number of features for filtering the collected data in terms of geographical regions, user segments/categories, and date ranges.

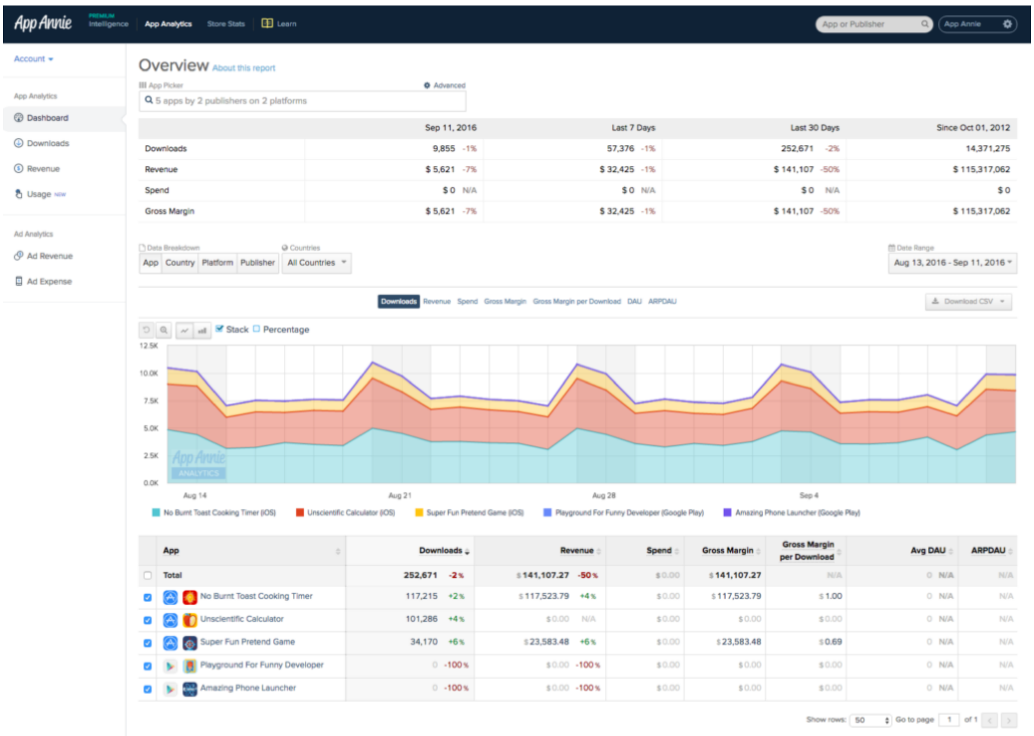



Figure 1: Longitudinal view in App Annie.

### 2.1.2 Facebook Analytics

Facebook Analytics addresses many of the same issues as App Annie with a similar focus, albeit with a focus not just on apps, but also on Facebook Apps – content developed specifically for the Facebook platform. As with App Annie, Facebook Analytics is focused on tracking the flow of users toward and away from content over time, and provides numerous time-line visualizations and aggregated statistics about users that break them down into categories.


Figure 2: Cross-sectional view with segments in Facebook Analytics.

In terms of visualization, Facebook Analytics mostly uses line graphs, bar charts, pie charts, and tables. The service also supports a Funnel editor and viewing component where an analyst can define a desired outcome (e.g. a purchase, also typically known as a conversion) and a number of steps preceding this outcome defining a process. The funnel view then allows the analyst to see how many users make to the desired end step of the process and how many users are lost along the way.

This may be of interest to learning analytics where educators may have a particular process in mind and would like to understand how many of their students make it the end of this flow, or which steps along the way are hurdles to the students.

### 2.1.3 Google Analytics

Google Analytics is also a general analytics platform. The service is generally directed at digital content and not limited to e.g. apps or games. Consequently, the service is advanced with many different visualizations and options for customization. In this brief overview, we try to draw together the most predominant visualization strategies offered on the service, rather than describe the full service in detail. Google Analytics features a virtual dashboard that allows for sub-sampling collected data by defining segments of users and filtering data in terms of time. Once a (sub)sample of users is selected, the dashboard allows for the display of metrics over time, alone or multiple metrics simultaneously. The metrics displayed are objectively defined measure of user behavior, such as e.g. how many users experience a certain piece of content, or how many users leave immediately after seeing a piece

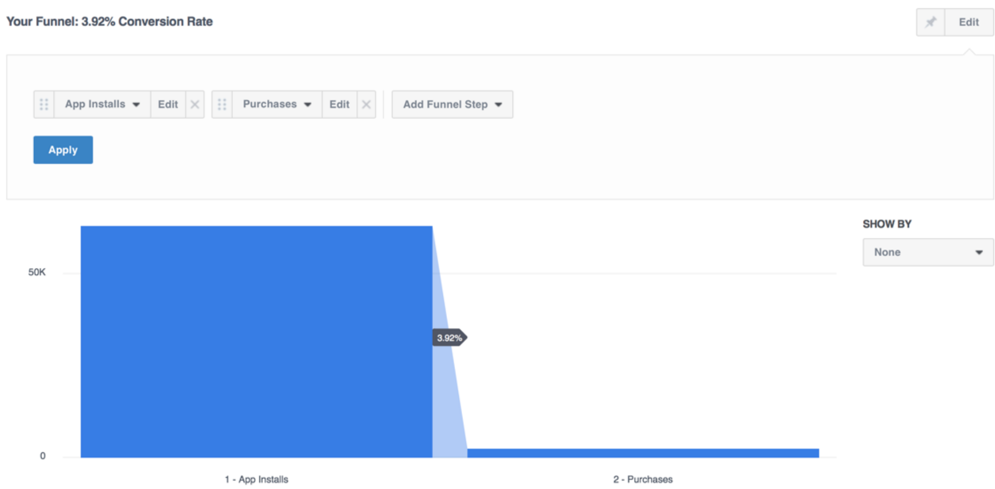



Figure 3: Conversion view in Facebook Analytics.

of content. For this reason the dashboard mostly displays aggregated frequencies and averages, either as snap-shots or over time, and uses a combination of bar-charts, pie-charts, and line-graphs to visualize this data.

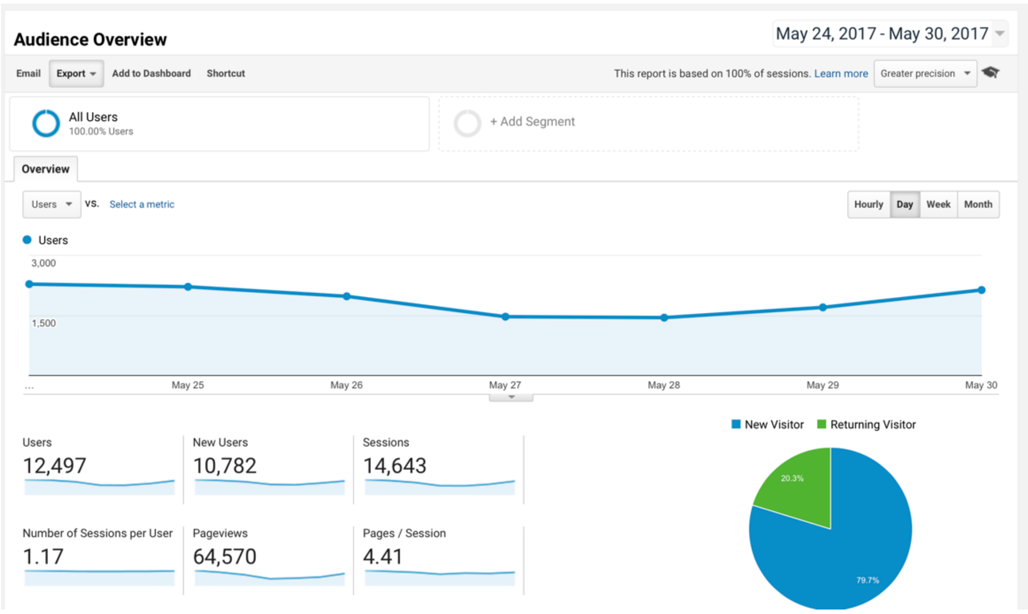
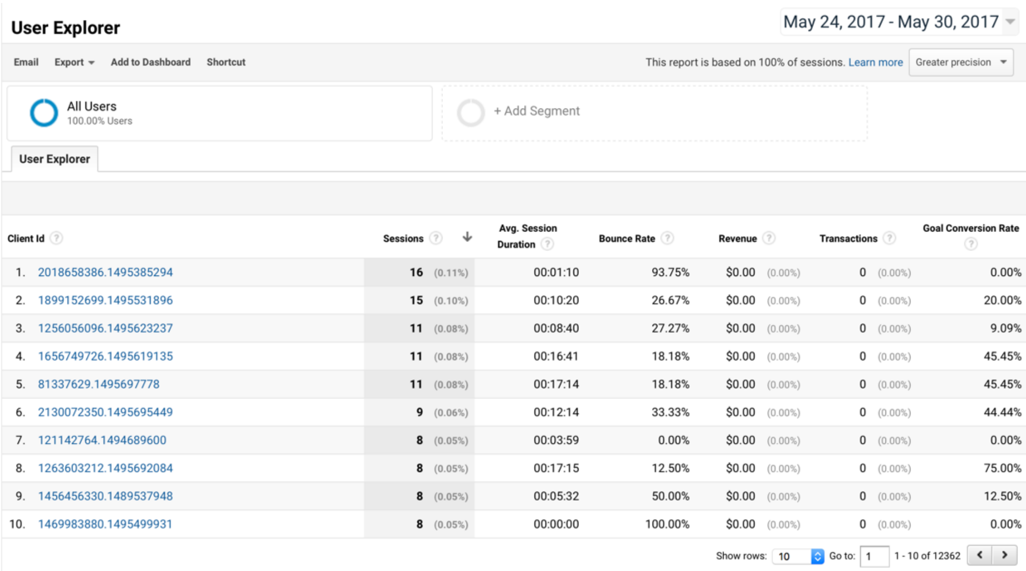



Figure 4: Longitudinal view in Google Analytics.

Other data is communicated through structured tables, again subject to the same kinds of filtering.

The service also offers user-flow analysis that allows for tracking users between specific tracking points and allow for diagnostics in terms of when users leave a digital content universe, typically a website.

In general, Google Analytics uses visualization with an emphasis on time-line visualizations with




| | Client Id | Sessions | | Avg. Session Duration | Bounce Rate | Revenue | | Transactions | | Goal Conversion Rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 2018658386.1495385294 | 16 | (0.11%) | 00:01:10 | 93.75% | $0.00 | (0.00%) | 0 | (0.00%) | 0.00% |
| 2. | 1899152699.1495531896 | 15 | (0.10%) | 00:10:20 | 26.67% | $0.00 | (0.00%) | 0 | (0.00%) | 20.00% |
| 3. | 1256056096.1495623237 | 11 | (0.08%) | 00:08:40 | 27.27% | $0.00 | (0.00%) | 0 | (0.00%) | 9.09% |
| 4. | 1656749726.1495619135 | 11 | (0.08%) | 00:16:41 | 18.18% | $0.00 | (0.00%) | 0 | (0.00%) | 45.45% |
| 5. | 81337629.1495697778 | 11 | (0.08%) | 00:17:14 | 18.18% | $0.00 | (0.00%) | 0 | (0.00%) | 45.45% |
| 6. | 2130072350.1495695449 | 9 | (0.06%) | 00:12:14 | 33.33% | $0.00 | (0.00%) | 0 | (0.00%) | 44.44% |
| 7. | 121142764.1494689600 | 8 | (0.05%) | 00:03:59 | 0.00% | $0.00 | (0.00%) | 0 | (0.00%) | 0.00% |
| 8. | 1263603212.1495692084 | 8 | (0.05%) | 00:17:15 | 12.50% | $0.00 | (0.00%) | 0 | (0.00%) | 75.00% |
| 9. | 1456456330.1489537948 | 8 | (0.05%) | 00:05:32 | 50.00% | $0.00 | (0.00%) | 0 | (0.00%) | 12.50% |
| 10. | 1469983880.1495499931 | 8 | (0.05%) | 00:00:00 | 100.00% | $0.00 | (0.00%) | 0 | (0.00%) | 0.00% |

Figure 5: Segmentable table view in Google Analytics.

the addition of snapshots of information shown as bar charts, pie charts, or representations on geographical maps.

The user flow diagram is the exception to this rule and is notable in the context of digital games and digital learning environments as this kind of visualization shows how users in the aggregate experience content serially, moving from one piece of content to the next. This is related to the concept of the travel-path, described in D1.1 (3), which will be visited later in this deliverable.

### 2.1.4 Trends in General Analytics Solutions

Pulling together observations from the three described General Analytics solutions we see that most of these offerings are centered on tracking numbers of users over time, breaking down users into segments from the data collected or from meta-data available about the users, and tracking key events such as the order of visited content or conversion behavior. Segments are typically defined from singular variables drawn from either behavior observed from event tracking or meta-data. The services are not focused on automatically clustering or otherwise segmenting users based on multidimensional models, possibly because different domains require different clustering methods and different tuning of these methods. None of the solutions are oriented toward following individual users closely, but rather focus on displaying analytics in the aggregate. Individual users are tracked, and used to generate longitudinal statistics and visualizations such as user flows and funnel views, and statistics describing user engagement and churn, but these are again typically presented in the aggregate. All the frameworks are based on event-driven tracking points, quite likely since this is a straight-forward paradigm for collecting data across very different domains, since the developers implementing the tracking points ensure that tracking points are placed in ways where they correspond to meaningful events of interest.

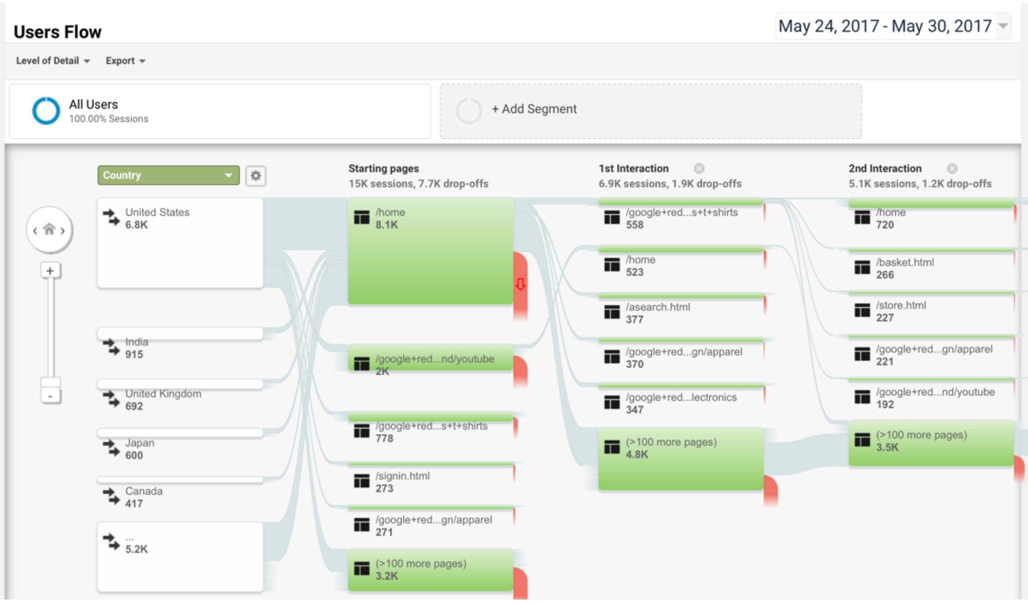



Figure 6: User flow diagram in Google Analytics.

## 2.2 Game Analytics solutions

In addition to the general analytics providers, three of which we describe above, a number of analytics services exist that cater particularly to game developers and game analysts. According to the feedback we received from our network survey, two main services are of interest to game developers: Game Analytics and Unity Analytics.

### 2.2.1 Game Analytics

Arguably one of the first services to cater specifically to game developers, Game Analytics offers a platform tailored for easy integration in the game development process. The service offers plug-ins for a range of game platforms and game engines with easy tracking integration. It also offers benchmarking along key performance indicators between games in a one-to-may relationship. The platform is focused on showing time-line data with a focus on user acquisition, engagement, and churn, allowing analysts to filter data on a range of variables.

The system allows for funnel visualization, akin to that described for Google Analytics, which in the particular case of games can be used to visualize e.g. progression through a game.

The platform allows for segmentation of users based on custom segmentation conditions with reports that are updated continuously.

The system also presents the option of using cohort analysis, where players are grouped by the time they first joined the game. This allows analysts to investigate whether players act differently based on when they started playing the and track e.g. player community responses to updates. The platform also features error tracking through the SDKs for the various platforms, which in turn may allow developers to troubleshoot errors and improve the quality of their game. Another prominent feature of the platform is the ability to track game variables over time, typically those pertaining to player state. This allows to developers to track how players are interacting with the game in aggre-

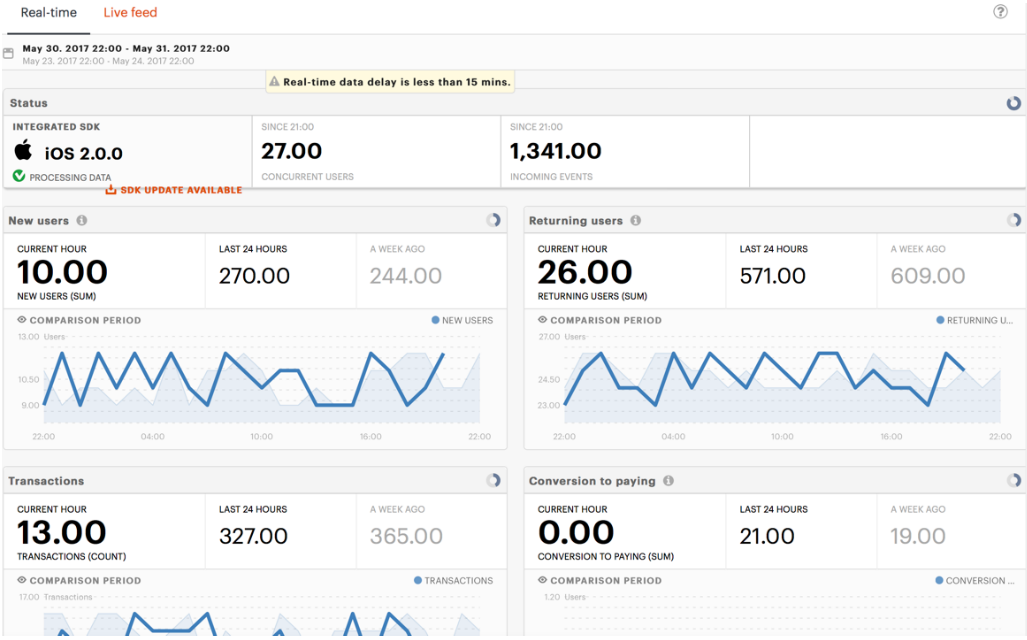



Figure 7: Longitudinal view in Game Analytics.

gate.

As in the case of the general analytics services, Game Analytics is driven by event-driven tracking and a focus on displaying player progression in the aggregate, allowing analysts to filter out particular subsamples and investigate metrics over time within these subsamples. Again, context information about the games providing the data is obtained through the labels which are defined by developers and parsed by human analysts. Most of the visualizations are accomplished through line graphs and tables displaying aggregate metrics attached to events. As such, Game Analytics can be thought of as a specialized, focused offering in line with the general analytics services, but targeting the game market specifically.

### 2.2.2 Unity Analytics

Unity Analytics provides an offering much akin to that provided by Game Analytics, but targeting solely the Unity game engine. In fact, the service is provided as part of the Unity engine license.

The features offered in Unity Analytics roughly match the ones offered in Game Analytics with options for visualizing standard metrics over time, options for segment building, and funnel analysis.

Analysts can define custom events in their game code and track these through Unity Analytics and visualize these in aggregate using the web based interface. Overall, Unity Analytics apply the same visualization strategies as seen above with an emphasis on line and bar charts in combinations with tables to allow for sense-making of the collected data.

### 2.2.3 deltaDNA

Another analytics product that is tailored towards the game industry is deltaDNA (formerly Games Analytics *not* GameAnalytics as described above). deltaDNA offers an large platform that offers

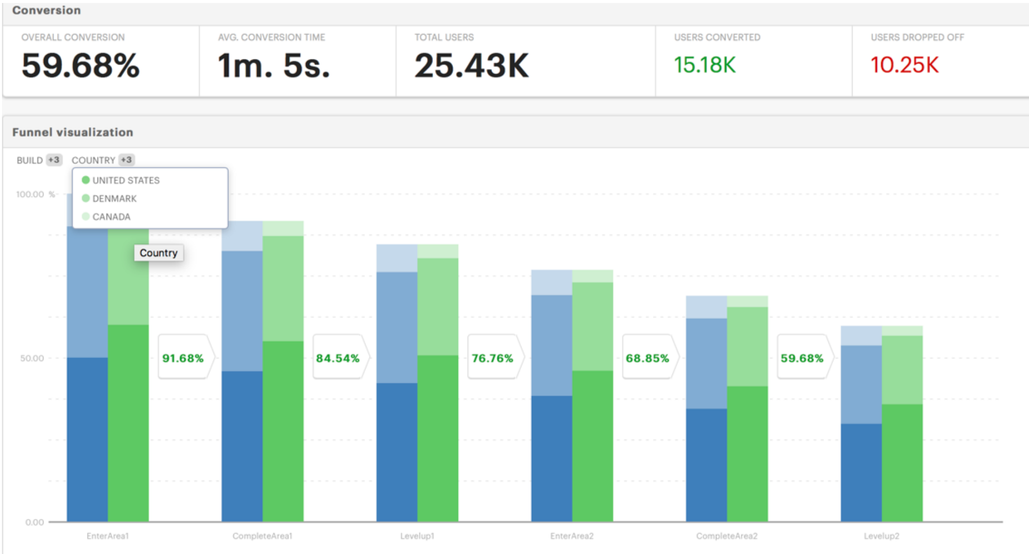
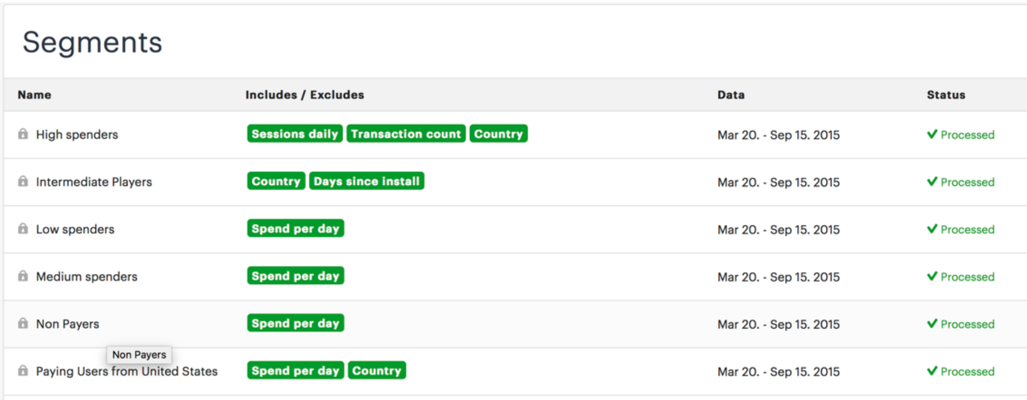



Figure 8: Funnel view in Game Analytics



Segments

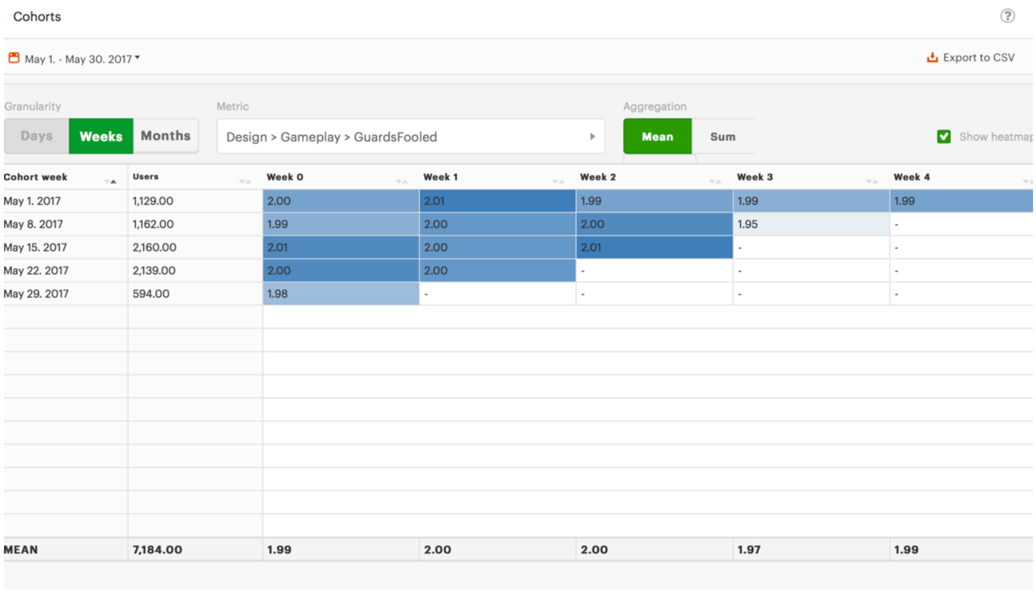
| Name | Includes / Excludes | Data | Status |
|---|---|---|---|
| High spenders | Sessions daily Transaction count Country | Mar 20. - Sep 15. 2015 | Processed |
| Intermediate Players | Country Days since install | Mar 20. - Sep 15. 2015 | Processed |
| Low spenders | Spend per day | Mar 20. - Sep 15. 2015 | Processed |
| Medium spenders | Spend per day | Mar 20. - Sep 15. 2015 | Processed |
| Non Payers | Spend per day | Mar 20. - Sep 15. 2015 | Processed |
| Paying Users from United States | Spend per day Country | Mar 20. - Sep 15. 2015 | Processed |

Figure 9: Segment view in Game Analytics.

besides "Deep data analytics" also "Real-time marketing". However, when referring to "Deep data analytics", one should not automatically assume that this matches the Deep Analytics definition of the ENVISAGE project. For example, many of the deltaDNA's "deep data analytics" features focus on KPIs displayed on a dashboard, benchmarks comparing KPIs to industry standards, or funnels. Besides their pre-configured dashboards, they also provide custom dashboards and direct access to the data. Both options give more flexibility to the customer when setting up their analytics.

As opposed to the "Deep data analytics", deltaDNA's "Real-time marketing" does contain tools and solutions that the ENVISAGE consortium would potentially consider as Deep Analytics. For example, deltaDNA offers A/B tests and balancing. The balancing features allow game developers to adjust parameters of the game, so that an individual user experience is possible for every player [1]. A typical example of a parameter to change is the difficulty level of a game. Clearly, changing the content or difficulty in a game has many counterparts in educational learning and virtual labs. The

[1] See https://www.youtube.com/watch?v=sDvBYzPiDXA for more information.




| Cohort week | Users | Week 0 | Week 1 | Week 2 | Week 3 | Week 4 |
|---|---|---|---|---|---|---|
| May 1. 2017 | 1,129.00 | 2.00 | 2.01 | 1.99 | 1.99 | 1.99 |
| May 8. 2017 | 1,162.00 | 1.99 | 2.00 | 2.00 | 1.95 | - |
| May 15. 2017 | 2,160.00 | 2.01 | 2.00 | 2.01 | - | - |
| May 22. 2017 | 2,139.00 | 2.00 | 2.00 | - | - | - |
| May 29. 2017 | 594.00 | 1.98 | - | - | - | - |
| MEAN | 7,184.00 | 1.99 | 2.00 | 2.00 | 1.97 | 1.99 |

Figure 10: Segmented table view in Game Analytics.

ENVISAGE project is further developing this angle by providing an authoring tool that will allow teachers to actually not only adjust parameters of virtual labs but also certain design aspects. To achieve this, the ENVISAGE consortium require more flexibility than provided by existing solution.

### 2.2.4 Trends in Game Analytics Solutions

Pulling the observations from the two game analytics specific services presented above, we find that these represent specialized instances of general analytics services. The analytic are focused on tracking and visualizing users in the aggregate with a focus on visualizing mostly singular metrics, or sometimes metrics next to one another, or letting the analyst construct new metrics by combining others or defining custom events. Visualization strategies are centered on using line and bar graphs in conjunction with tables to provide analysts with quickly interpretable analytics.

## 2.3 Research Game Analytics Solution: Machinations

A final example of visualization in game analytics that takes an approach different from the examples described above is the Machinations system developed chiefly by researcher Joris Dormans [4, 3]. The Machinations system is event-driven in the same manner as the analytics platforms described above, but also encompasses an abstracted simulation of the game system being analyzed. That is, an analyst wishing to use the Machinations system must first describe the logic of the game at some chosen level of abstraction corresponding to the tracked events, using the Machinations visual flow chart description language. The system is intended to be both a simulation driven analytics tool, deployed before any players actually use the game. Events are simulated used random number generators, statistical distributions, and rules. The flow charts then visualize these events dynamically and how the game's rules respond to these events, for instance when simulating a game's economic system. Once a game is public and players start interacting with the game, these synthetic event

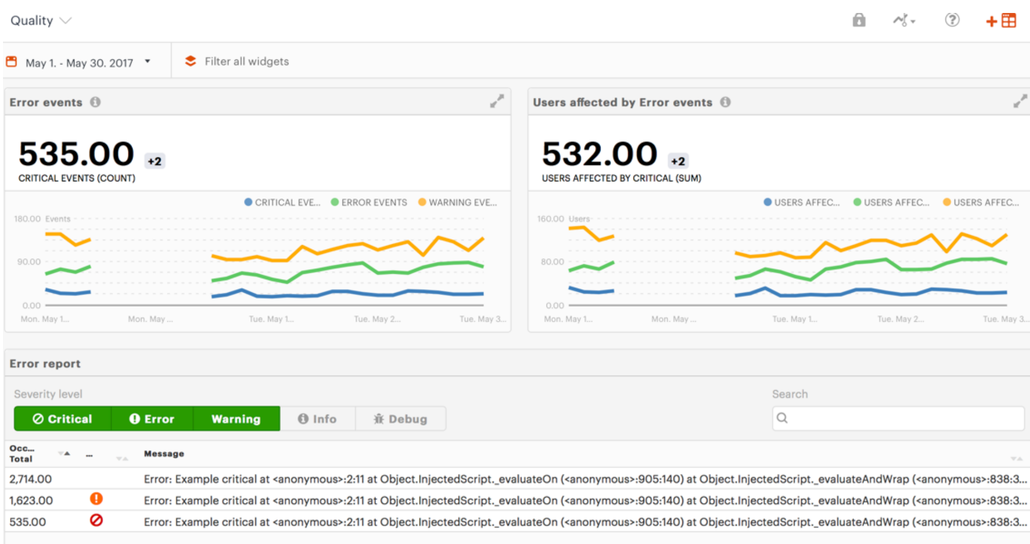



Figure 11: Error view in Game Analytics.

generators can be replaced with actual game events drawn for player, which in turn can drive the interactive, animated flowchart view. This represents a game analytics approach which is much closer to the mechanics and dynamics of a particular game being analyzed, but comes with a larger upfront cost in terms of time investment, since the game must be modeled as a flowchart.

This section presented a brief overview of current approaches in on-line analytics and game analytics, both in terms of general philosophy and visualization strategies. In the following section, we identify particular needs for learning analytics determined in the previous deliverables of the ENVISAGE project and relate these to these existing solutions. We use this to identify which approaches from game analytics may be carried over to learning analytics, and which challenges require novel solutions.

# 3 Requirements for Visualizing Course Progress Reports

To understand what visualization principles and practices can be transferred from analytics in general and game analytics it is helpful to understand what visualization needs have already been identified for ENVISAGE. In this section, we extract visualization requirements from the previous deliverables completed for ENVISAGE and relate these to the approaches identified in Section 2. Specifically, we draw on D1.1 [9], D1.2 [7], D2.1 [6], and D2.2 [8]. Together, these deliverables present a first approach at understanding the interests of teachers using and building virtual labs for teaching, what data would serve these interests well, and the analytical treatments and supporting architecture necessary to serve these interests. Below we extract the requirements or ideals that pertain to the visualization of data drawn from virtual labs, and the technical implementations that follow from them. We group each concern thematically across deliverables in the sections below.

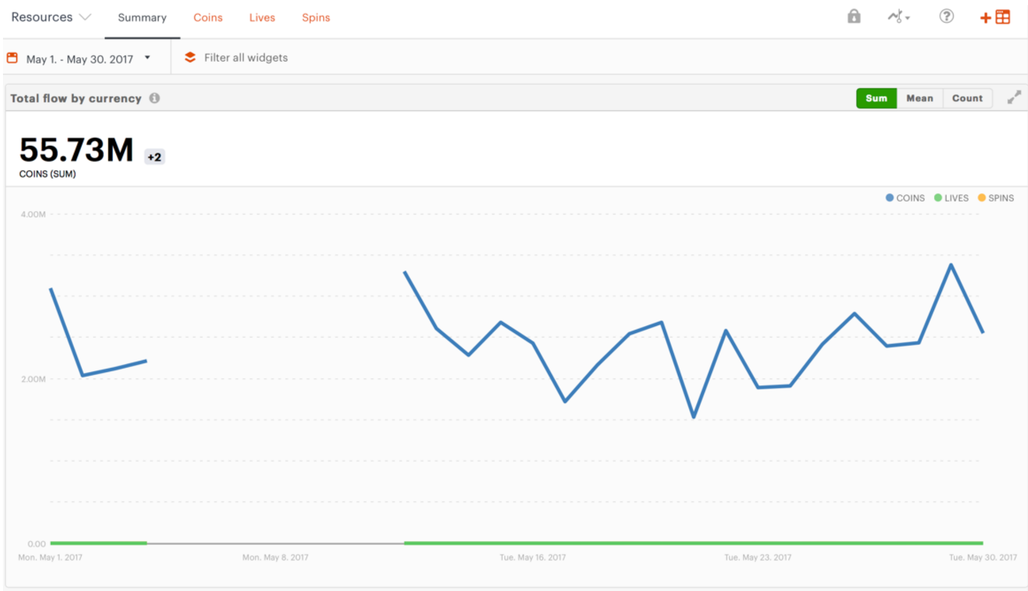



Figure 12: Longitudinal view in Game Analytics.

## 3.1 Considerations related to teacher questions

D1.2 [7] and D2.2 [8] identify a fundamental need for visualizations allowing teachers to visualize learning behavior and indicators as a) individuals or in groups, allow them to b) contrast individuals to one another, groups to one another, or individuals to groups, and c) allow them to make these comparisons either cross-sectional or longitudinally.

### 3.1.1 Selecting individual and or group visualization

Allowing for flexible visualization in groups or at the individual level requires the easy selection of subsamples of data. This creates for the easy definition of filters and segmentations of existing data. Visualization solutions developed through ENVISAGE should support the easy definition of these criteria. Strategies for accomplishing this can be transferred from the filtering and segmentation methods identified in existing (game) analytics solutions.

### 3.1.2 Contrasting individuals and/or groups to one another

To contrast different groups to one another the filtering and segmentation should be used to create multiple instances of the same visualization, but with different data being displayed. This would allow teachers to compare e.g. one student to the rest of the class, to students to one another, or two classes to each other.

### 3.1.3 Allowing for cross-sectional or longitudinal visualization

Given that teachers may be interested in comparing individuals or classes in singular sessions, or track individuals or groups over longer periods of time, ENVISAGE should produce visualizations appropriate for both these use cases. For cross-sectional visualizations, it would be relevant to present

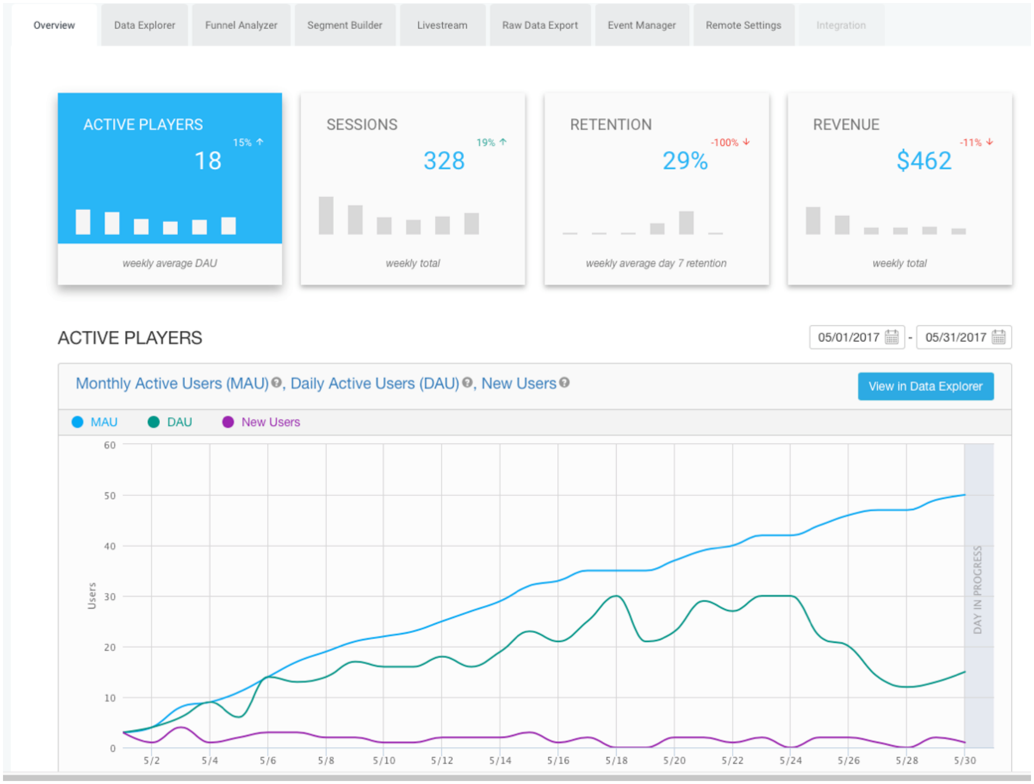



Figure 13: Longitudinal view in Unity Analytics.

values of one or more metrics for both the groups of interest as done in the commercial analytics platforms described in Section 2. We can also consider taking a Machinations-inspired approach and foregrounding the dynamics within the game sessions for each group, exposing how students interact with the mechanics of the learning environment. For longitudinal visualizations, aggregate metric visualization approaches may be imported from the commercial analytics services.

## 3.2 Considerations related to metrics

In this following section we consider the metrics of interest defined in D1.2 [7] and suggest visualization strategies. We consider each metric from the perspective of Stevens's Theory of Scales of Measurement[10].

### 3.2.1 Time-on-task

Time-on-task is fundamentally a quantitative measure with a meaningful zero point and as such qualifies as being measured on a ratio scale. Visualizing this metric for individuals or groups, or individuals and groups can easily be accomplished by line graphs or bar charts that can be scaled linearly to represent the relevant value for the sub-sample in question. Time-on-task may also be range-normalized according to the length of the whole section in which the task was engaged with, presenting a relative measure. Both of these approaches can be relevant to teachers, depending on whether they are interested in understanding absolutely how much time students spend on particular tasks, or whether

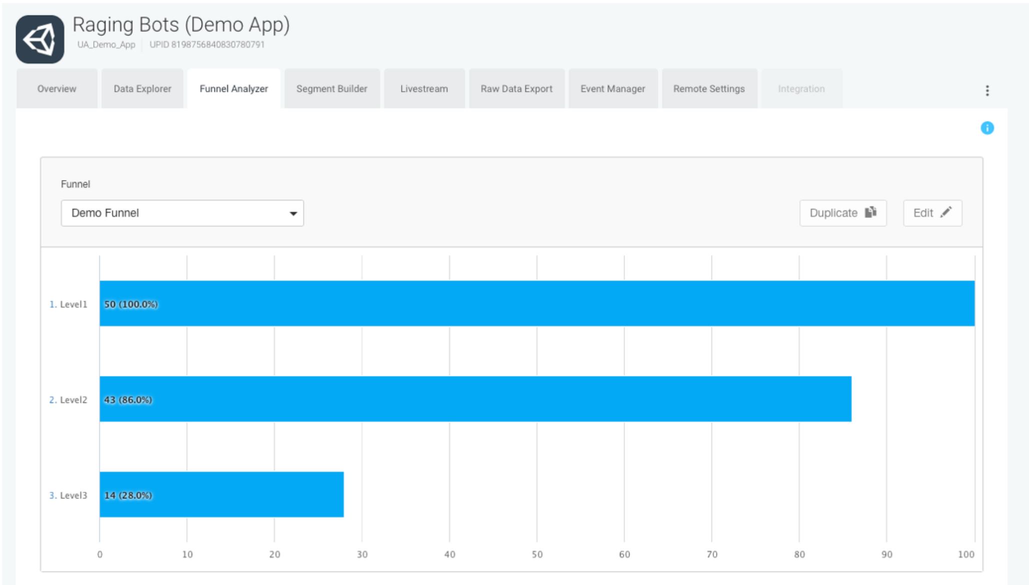



Figure 14: Funnel view in Unity Analytics.

they are more interested in understanding how students spend the time that is available to them, relatively, between the tasks. For this reason, the first ENVISAGE implementation will support both these modes of visualizing this metric.

### 3.2.2 Time-to-completion

Time to completion is also a ratio scale metric, similar to time-on-task. The metric cannot be normalized for an individual session, but could be range-normalized across all the time taken across multiple sessions to provide a relative visualization.

### 3.2.3 Class categorization profile

The class profile is an ordinal metric that divides a class into one of three performance categories: low, moderate or high, on a number of dimensions. The category for the class is calculated based on the aggregate performance of the individual students in the class. The class profile corresponds to a segmentation in the analytics solutions presented in Section 2, but an ordered segmentation rather than a completely nominal (categorical) one. The class profile can be visualized using a color scale, which can be helpful both when measuring the same class over time or when comparing two classes to one another. When data from multiple classes is being visualized the color scale can also be used as a legend coloring data points, adding another dimension to the visualization.

### 3.2.4 Levels of proficiency

Levels of Proficiency is a value calculated based on which percentages of a class population reach which performance categories out of low, moderate, or high. The metric is an interval level measure which is well suited for visualization through line graphs or bar charts.

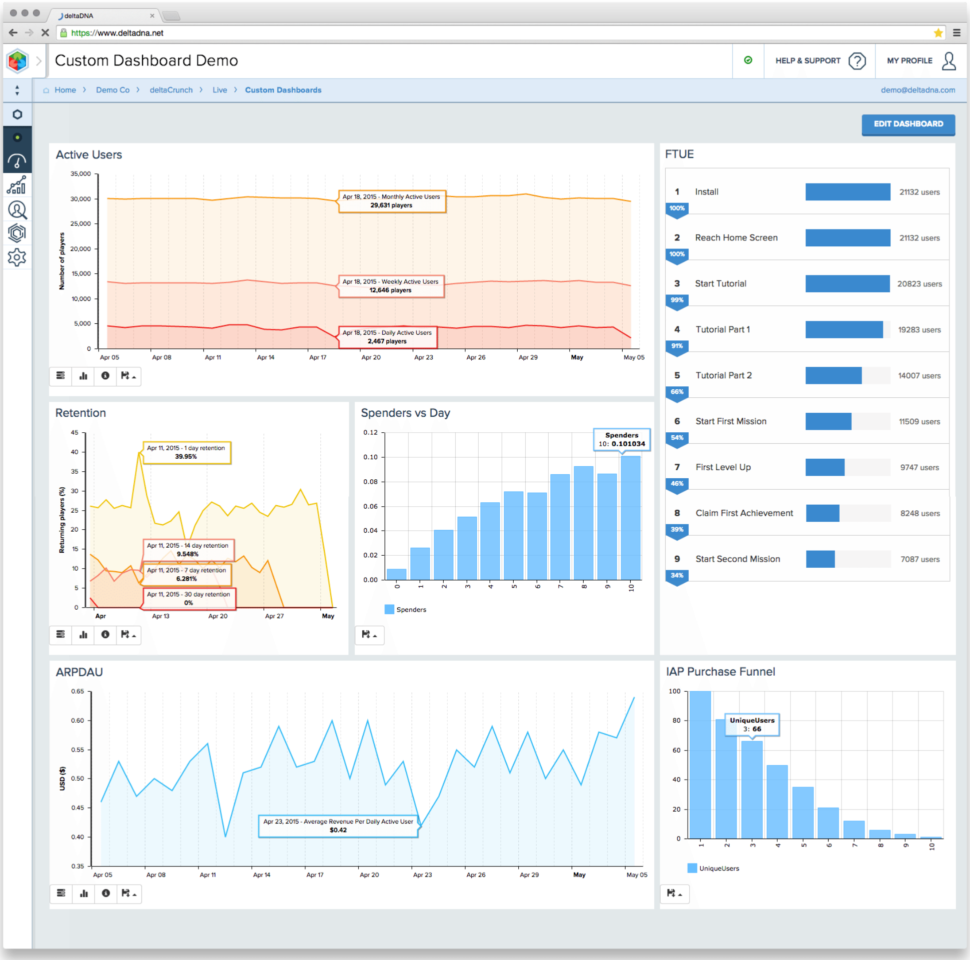



Figure 15: A dashboard from deltaDNA.

### 3.2.5 Mastery index

Mastery index describes how well a student is capable of conforming to some objective measure internal to the learning environment in question. For the Wind Energy Lab e.g. this could be the percent of the time where the student was able to keep the simulation in a correctly powered state. Typically, the mastery index will benefit from being normalized to range from 0 to 1 or 0 to 100, since the index describes the space going from complete incompetence to perfect mastery. For the purposes of ENVISAGE we assume that the mastery is measured linearly or can be transformed into a linear form. Given this assumption, this metric is also an interval level measure and hence suited for visualization through line graphs and bar charts that can express how close a student or groups of students are to achieve perfect mastery and can be used to compare students or groups of students.

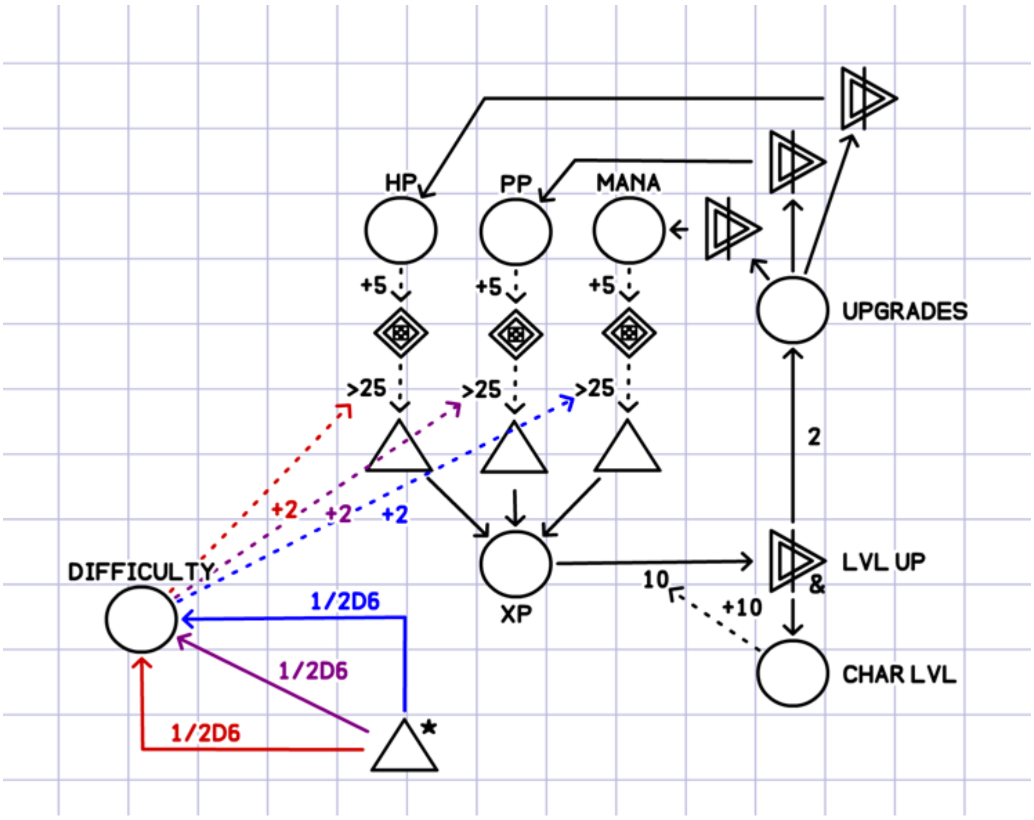



Figure 16: Interactive flowchart and simulation in Machinations.

### 3.2.6 Travel Path

The travel path metric describes how one or more students move through a learning environment i.e. which sequences of actions they took and the corresponding events that were elicited. The travel path consists of a number of sequentially ordered data-points that at the simplest level can contain just labels or at the most complex levels can contain the full-game state as well as meta-data about the student(s), their location, and so forth. As such the travel path can contain multiple levels of measurement and cannot easily be expressed in a compressed format. Rather, the dynamic movements of the students through the learning environment must be described at the level of abstraction defined by the collected events. Appropriate visualization strategies can be drawn from the funnel, user flow, and Machinations-style visualizations described in Section 2. Each of these provide approaches for showing how individuals or groups of individuals move between the events that are possible in an environment.

# 4 Goals of Learning Analytics Visualization

Drawing together the findings from the previous section, a number of high-level goals for learning visualizations that our solution must be capable of addressing become clear. This allows us to define this first iteration of visualization strategies for the ENVISAGE project. Some of these goals are similar to the goals in commercial game analytics and hence visualization methods can most likely be ported directly from this field. Others are specific to learning analytics and require visualization solutions particular to this field.

## 4.1 Cross Sectional and Longitudinal visualization

Learning analytics shares the need for both comparing different segments at specific points in time as well as tracking metrics over the course of time. Methods for accomplishing this are well-developed in commercial game analytics, as described in Section 2 and the described approaches of providing aggregate summaries in the form of pie charts, bar charts, and line graphs can be transferred directly from game analytics to learning analytics.

## 4.2 Small-Scale visualization

Where most of the analytics solutions reviewed in Section 2 focus on aggregating across large numbers of users, except for Machinations, learning analytics sometimes comes with a need for visualizing across a small number of users and possible only a single session or very few sessions [8]. This means that visualization solutions developed for Learning Analytics need to allow the analyst to receive meaningful visualizations at e.g. the class level, or even at the level of an individual user. This is different from the approaches taken in game analytics in general which typically supposes users in the thousands or millions and as a results focuses predominantly on aggregated data.

## 4.3 Unknown Goals

As documented in D1.1 [9] and D2.2 [8]. it is not given that educators will be able to define objective measures of performance ahead of time, when engaging in Learning Analytics. This could be either due to measures of interest arising from studying the data during Learning Analytics or that the processes of interest take place in an environment that is not trackable i.e. in the mind of the student or as a social process in the classroom and not in a process that is observable inside the digital learning environment. We can contrast this with regular analytics where the analyst is often interested in objective measures of user behavior. The examples of game analytics we covered in Section 2 focus on metrics such as engagement (how long do users stay within a digital content universe), app launches, and conversions. These metrics are easily defined, measurable and do not concern themselves with learning or developmental changes in the mind of users. Defining an objective for optimization, thus, is typically easier in commercial analytics and game analytics. We suggest that this motivates developing more flexible, customizable, and process-oriented visualizations for learning analytics than standard game analytics. Since the analysts (i.e. teachers designing digital learning experiences, e.g. in the form of virtual labs) may not be able to directly measure the outcome that they are attempting to optimize for, it becomes increasingly important to facilitate data exploration and render salient how students are moving through the digital learning environment. This is mirrored in the emphasis on the travel path metric in D1.1 [9], D1.2 [7], and D2.2[8].

## 4.4 Clustering for Segment Discovery

For Learning Analytics, the notion of segmenting is useful, much in the same way that it is useful to commercial Game Analytics. Learning Analytics provides several outcome segments that can be calculated objectively and externally, or gained from meta-data about students provided it is possible to identify individual students in the dataset. This could be, for instance, the student's overall performance in the subject that a digital learning environment is designed to teach, represented by e.g. the student's grade in the subject. It could also be the metrics internal to learning analytics, such as e.g. the proficiency level or the mastery index. Nonetheless, teachers or analysts may be interested in mapping these outcome classes to classes that are derived from the collected data, rather from externally defined metrics. For ENVISAGE, it is expected that the Deep Analytics under development will provide these suggested segmentations through processes of clustering. While it is too early to describe how this clustering will work in detail, as this will be determined in a later deliverable, there will be a need for visualizations that support teachers and analyst in making sense of the clustering proposed by Deep Analytics and connecting this to other outcome metrics or objectively defined segmentations. Thus, visualization solutions for both shallow and deep learning analytics will need to flexibly support not only the definition and display of segments, but also the discovery and rejection or confirmation of potential segmentations derived from patterns in the collected interaction data. In the following section, we describe the preliminary visualization strategies that were developed in to satisfy the goals identified here.

# 5 Developed Visualization Strategies

In this section we describe visualization strategies developed in order to accomplish the goals identified in the previous section. It is important to note that the visualizations described here are work-in-progress and will be updated over the course of the project. Most notably, perhaps, they do not take into account visualization strategies for Deep Analytics, as the methods for these are still being developed in forthcoming deliverables.

## 5.1 Basic visualizations: Bar Charts and Line Graphs

Collected here under one heading, the most basic visualizations developed for Learning Analytics are basic aggregate visualizations of singular metrics. These mirror the implementations of visualizations of aggregate metrics that are common-place in commercial game analytics.

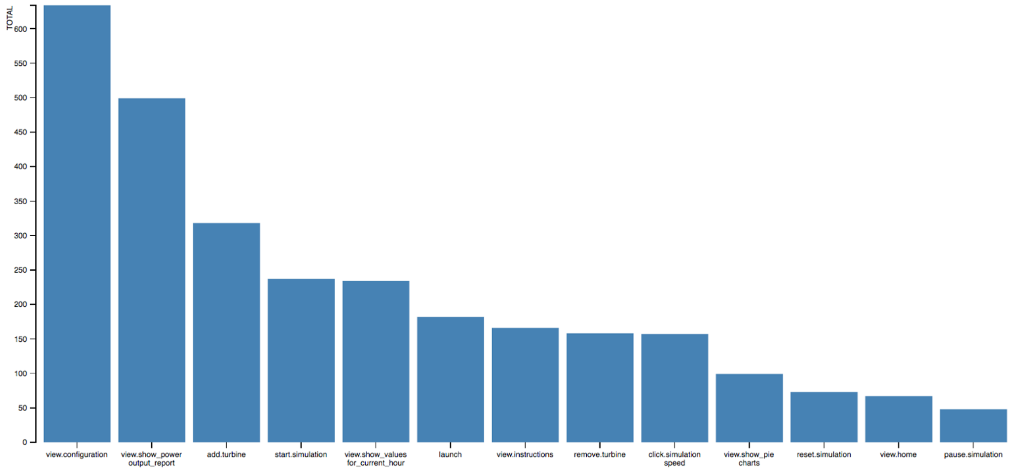



Figure 17: Simple bar charts from the ENVISAGE visualization library.

## 5.2 Force-Directed Graph Visualization

This visualization is focused on visualizing typical travel paths. In the force-directed graph, each node represents a possible event in the virtual lab. The weight of an edge between two nodes visualizes how frequently this transition occurs in the currently selected sample. This visualization is useful for cross-sectional inspection of aggregate behavior in a group of individuals and allows an analyst (typically a teacher) to see which transitions are frequent and which are infrequent. Additionally, placing two or more of these visualizations next to one another, as shown in Figure 18, allows an analyst to quickly compare whether interactions frequencies are similar or different between the groups. If e.g. an analyst wanted to compare the behaviors of two groups of students; one with a high proficiency level and one with a low proficiency level, this visualization would allow for a quick impression of the differences in interactions patterns. While the forced graph visualization provides a quick overview, it is harder to extract detailed patterns from this visualization. This is addressed in the following visualization, the Chord Diagram.

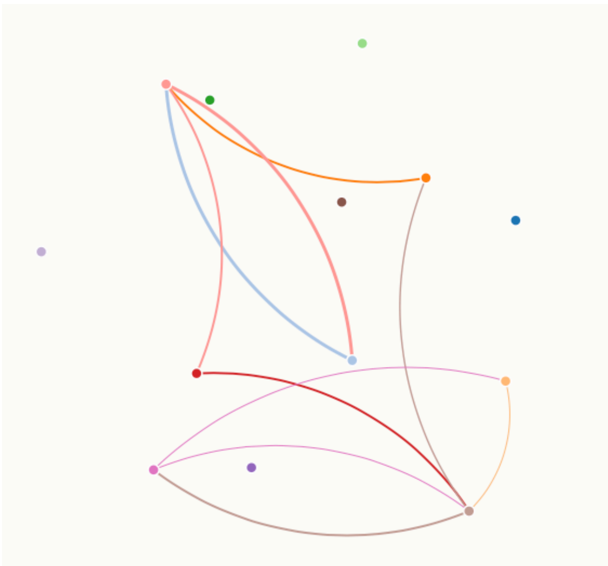


Figure 18: A force-directed graphs of a group of students.

## 5.3 Chord Diagram Visualization

The Chord Diagram also describes transition frequencies between states in the aggregate, allowing an analyst to inspect e.g. the interaction patterns of a class. It draws on principles also observed in the Google Analytics User Flow diagram, described in Section 2. However, many virtual labs including the Wind Energy lab of the ENVISAGE project, which provided the data for these examples, are not strictly linear experiences, but rather systems where the same state may be visited multiple times, at least at certain tracking resolutions. If state or sequence information is removed, the interaction may be described not only as a linear flow of events, but also as transitions in a cycle graph. Here, it becomes interesting to see for each state, which is the most visited next state. The Chord Diagram visualization addresses this by providing the analyst with a view where all transitions from one event to other events are shown as colored flows. Additionally, the implemented visualization allows for filtering the data by hovering over source events, in which case only the flow from the highlighted source event is shown. This visualization collapses time across the digital learning environment in favor of showing transition frequencies in the same way as the force directed graph, but with an emphasis on detail an allowing the analyst to drill down into specific transition patterns. Again, comparing multiple segments may allow an analyst to intuitively make sense of how interactions patterns are different and form hypotheses to be explored on this basis. Collapsing the time dimension into aggregate frequencies does remove some information about the sequences in which students move between events. This is addressed in the following two visualizations: Time-line visualizations.

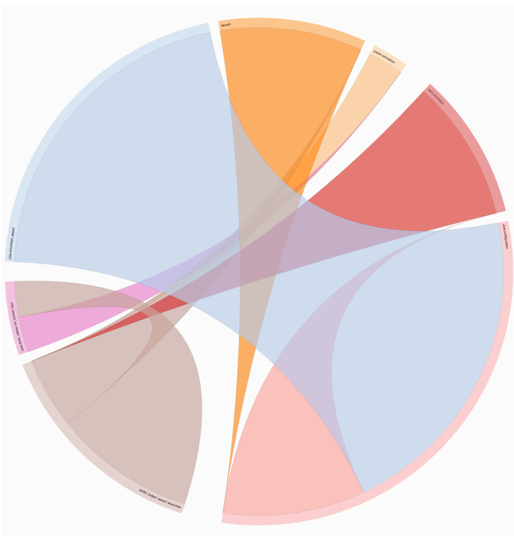


Figure 19: A chord diagram showing next event distribution after the start.simulation event.

## 5.4 Absolute Time-line Visualization

Where the Force-Directed Graph and the Chord Diagram disregard the sequence of user actions in favor of aggregating transition patterns the time-line visualizations developed for ENVISAGE focus on foregrounding the time spent between different events. The time-line visualization leverages the full travel path metric delivered by Shallow Analytics to display which events each individual user experienced and how much time was spent between each event visit. The first visualization taking this approach is the Absolute Time-line visualization. This visualization simply labels the time between events with the name of the event at the start of each period and displays information either for each user or averages across groups of users. This allows an analyst compare individuals or groups in terms of how much time they spend in different parts of the digital learning environment, potentially spurring on insights into reasons for these differences or changes that might move these patterns closer to a desired state. If a high degree of variation is observed in the dataset, e.g. if students spend very different amounts of time overall or on specific tasks, it may become difficult to compare individuals or groups to one another. This problem is addressed in the final visualization strategy developed for this deliverable, the Relative Time-line Visualization.

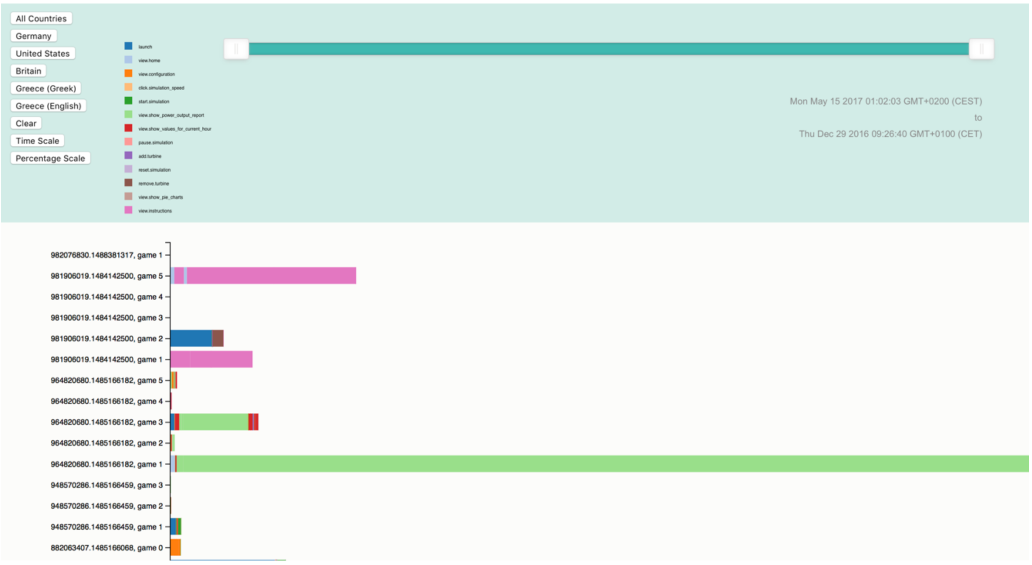



Figure 20: An absolute time-line view of event sequences, i.e. travel paths, in the Wind Energy Lab.

## 5.5 Relative Time-line Visualization

The relative time-line visualization uses the same travel path metrics as the Absolute Time-line Visualization, but normalizes all total session times to the same visual length, and calculates the time spent between events as a relative fraction of this total amount of time. This means that the activities of students are no longer comparable in terms of the total time taken (this perspective is addressed in the Absolute Time-line visualization), but become comparable in terms of where they spend their time, relatively. While this may produce misleading information for datasets with a high degree of variation in terms of the total session length, it may be useful to an analyst when the sessions are relatively similar, but not completely. When working with groups, values may again be averaged before this normalization process. Figure 21 shows this type of visualization for a dataset where a number of individuals are displayed. This visualization additionally sorts the students based on the pattern similarity between their normalized sessions. This is accomplished by drawing on the Levenshtein distance comparison feature included in the Shallow Analytics library: The travel paths of the students, and the time they spent between different events, are quantized into string representation of the patterns. These patterns are then compared in terms of edit distance, using the Levenshtein metric, and sorted according to their distance to one another. This may provide an analyst with a quick sorting of the students or groups of students according to their travel paths, which may in turn by combined with segment information to inspire the analyst to various hypotheses about the differences between these segments.

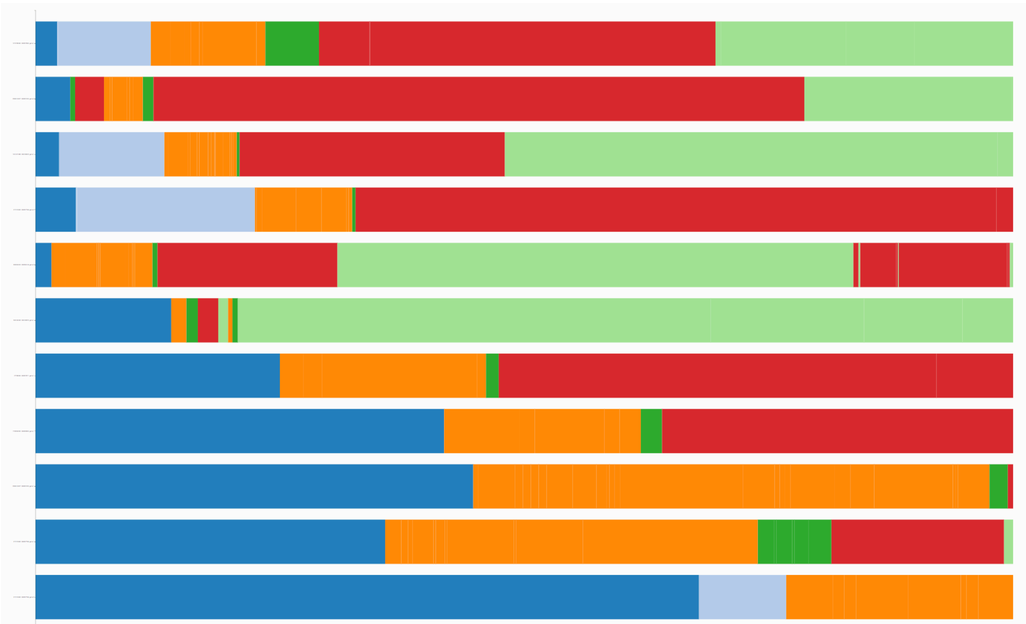


Figure 21: Relative time-line visualization of event sequences, i.e. travel paths, in the Wind Energy Lab for a group of students. Users travel paths are sorted by similarity by measuring the pairwise Levenshtein edit distances. Color legend: Blue: time spent in launch; Light blue: time spent in home view; Orange: time spent in configuration view; Green: time spent in simulation view; Red: time spent looking at current hour values; Light green: time spent viewing power output report.

## 5.6 State-Based Travel Path Visualization

The current tracking resolution of the ENVISAGE Virtual Labs does not support detailed state information in relation to events. This lack of detailed context is a typical property of analytics solutions and hence it is sensible to develop visualization solutions like the ones presented here, that do not presuppose detailed state information but only work with sequences of events. However, as part of advancing the state of the art Learning Analytics, the ENVISAGE project will be implementing detailed state tracking functionality for the newly implemented virtual labs which will also be used to enable

Deep Analytics. Once these are in place, this deliverable will be updated to also include travel-path visualization strategies that take into account the state, and hence the context, of the tracked events, as also described in D2.2 [8].

# 6 Technical implementation

## 6.1 Overall Architecture

The visualizations implemented for ENVISAGE can be found at the following URL:

`https://github.com/Envisage-H2020/Analytics-Visualizations`

As described in D4.1, the ENVISAGE project employs a distributed software architecture where the various tasks in the project are handled by independent modules that communicate REpresentational State Transfer (REST) [5]. . The visualization components developed for ENVISAGE follows these same principles; the visualization software assumes that it may retrieve a dataset, or subsamples of a dataset at URL endpoints provided by the data store solution provided by goedle.io. The received data is then parsed, transformed, and analyzed as necessary, either directly in the visualization implementation or by requesting these transformations and analyses from the implemented Shallow Analytics solution, delivered as part of D2.2 [8]. All JavaScript is executed in the user's browser, as shown in Figure 22 below.

## 6.2 D3 implementation

All the visualizations described and depicted above are implemented using the open source JavaScript visualization library D3 (short for Data Driven Documents). D3 provides a powerful framework for developing visualizations using Scalable Vector Graphics (SVG), Hypertext Markup Language (HTML), Canvas (a feature of HTML), and JavaScript. All modern web browsers are capable of using the technology as are many so-called WebViews (components that embed web browser rendering engines within other software solutions) for game engines or other software applications. This makes D3 a very flexible and general tool for providing the visualizations for envisage. All visualizations described above are implemented in singular HTML files. Each of these HTML files load the D3 visualization JavaScript library and contain the custom JavaScript necessary to generate the visualization embedded inside the HTML file. Stylings necessary for each visualization are defined using Cascading Style Sheets (CSS). Stylings specific to each visualization are also embedded inside each HTML file while stylings shared between the visualizations, currently the styling necessary to display filtering sliders, are stored in shared CSS files. The general solution is displayed in Figure 23 below.

# 7 Conclusion

This deliverable described the background for, the design of, and the implementation of the initial visualization solutions for the ENVISAGE project. The visualization strategies were informed partially by the state of the art in commercial game analytics and partially by the needs identified by previous deliverables over the course of the ENVISAGE project. The result is a demonstrator piece of

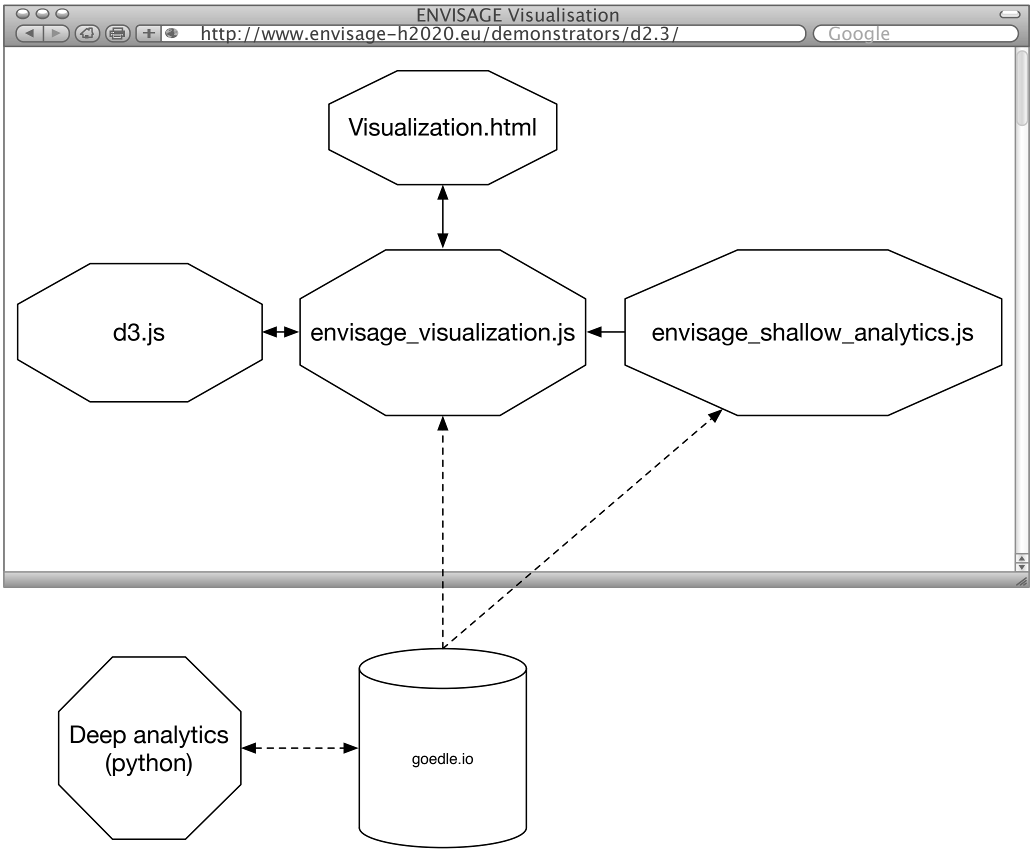



Figure 22: The overall architecture for the visualization part of the ENVISAGE solution. Elements shown in the browser window are calculated and rendered on demand in the user's browser, drawing on the elements show outside the window: the goedle.io data server, and the deep analytics service.

software implemented in JavaScript, leveraging the D3 visualization library, and the initial Shallow Analytics library, delivered as part of D2.2. The demonstrator supports many of the identified needs: the ability to conduct cross-sectional analysis, the ability to filter and select on multiple levels, the ability to overview metrics defined in D1.1 and D1.2, and the ability to explore student travels paths through the virtual labs. Other capabilities will be added to the visualization suite as the ENVISAGE technical platform matures. Future updates of the visualization strategies will be updated to allow for longitudinal analysis as well as integrate information derived from the Deep Analytics implementation, which is underway. They will also expand on the visualization strategies identified and initially implemented in this deliverable.

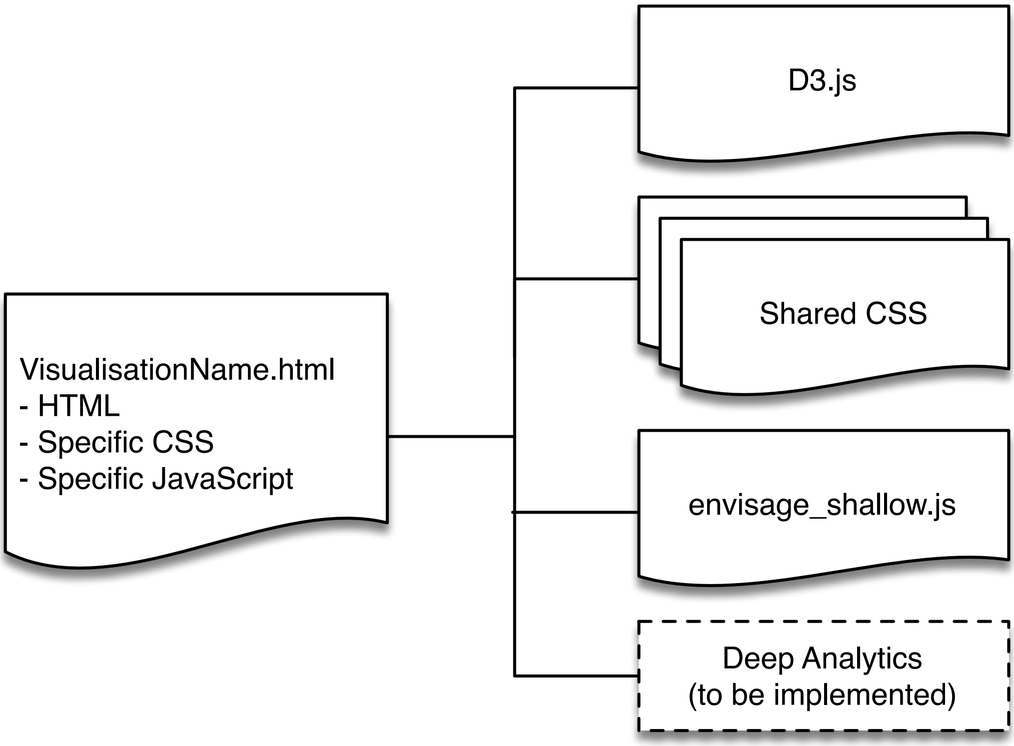



Figure 23: The relations between the documents and libraries developed for this deliverable. The visualization also includes the Deep Analytics service under developed (marked with dotted lines).

# References

[1] App annie.

[2] Google docs.

[3] Ernest Adams and Joris Dormans. The designer's notebook: Machinations, a new way to design game mechanics, 2012.

[4] Joris Dormans. Simulating mechanics to study emergence in games. *Artificial Intelligence in the Game Design Process*, 2(6.2):5–2, 2011.

[5] Roy T Fielding and Richard N Taylor. *Architectural styles and the design of network-based software architectures*. University of California, Irvine Doctoral dissertation, 2000.

[6] Fabian Hadiji and Marc Müller. D2.1 analytics infrastructure installation and data aggregation. In *ENVISAGE*. 2017.

[7] Georgios Mavromanolakis, Pavlos Koulouris, Sofoklis Sotiriou, Giannis Chantas, and Spiros Nikolopoulos. D1.2 data structure and functional requirements. In *ENVISAGE*. 2017.

[8] Benedikte Mikkelsen and Line Ebdrup Thomsen. D2.2 user profiling and behavioral modeling. In *ENVISAGE*. 2017.

[9] Sofoklis Sotiriou, Pavlos Koulouris, and Georgios Mavromanolakis. D1.1 educational scenarios and stakeholder analysis. In *ENVISAGE*. 2017.

[10] Stanley Smith Stevens. On the theory of scales of measurement. 1946.

[11] Edward Rolf Tufte. *The Visual Display of Quantiative Information*. Graphics Press LLC, Cheshire, 2001.