The Derived Data Approach to Support the Construction and Consumption of Explorable Visual Narratives

A thesis submitted to the

University of Dublin, Trinity College

For the degree of

Doctor of Philosophy

Bilal Yousuf

Knowledge and Data Engineering Group

School of Computer Science

Trinity College Dublin

byousuf@tcd.ie

Supervised by Professor Owen Conlan

Declaration

I, the undersigned, declare that this work has not been previously submitted as an exercise for a degree at this or any other University, and that, unless otherwise stated, it is entirely my own work.

Bilal Yousuf

December 2015

Permission to Lend or Copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

Bilal Yousuf

December 2015

Acknowledgements

Firstly I would like to thank my supervisor Professor Owen Conlan for his guidance, support and encouragement over the years. I would also like to thank Dr. Seamus Lawless, Dr. Alex O'Connor and Dr. Theresa Doyle for their time and feedback while I was compiling this thesis.

I would like to express my gratitude to the members of the KDEG research group and the ADAPT centre, both student and staff members who participated in the user trials conducted in this thesis, for their generosity with their time and their invaluable feedback.

I would like to thank my parents for their continuous encouragement and support throughout the entire process. I would also like to thank my brothers and sister for their encouraging words and good humour. Finally, a special thanks to my wife, Arfa, for her endless patience, support and a constant smile.

Abstract

This thesis proposes a novel approach called *the derived data approach* that supports the construction and consumption of explorable visual narratives in Technology Enhanced Learning (TEL). Like many other domains, TEL environments generate valuable data which can be difficult to interpret. As TEL continues to grow in popularity, the challenge of addressing high student dropout rates remains, which is due to a low level of student involvement in the learning process. Learning Analytics aims to motivate learners and enhance their engagement by using the student-logged data generated by TEL environments and also enables educators to monitor students. It uses Information Visualisation to present interactive views using student-logged data that learners and educators can analyse. Information Visualisation research has demonstrated the value of visual narratives in communicating a message, by highlighting facts and making the message more memorable. In addition, visual data exploration can support users in gaining deeper understanding of data. However, in many domains and specifically in TEL, there are deficiencies in the presentation of explorable visual narratives which could enable the analysis of data related to it. In addition, there is limited support offered during the construction of explorable visual narratives, specifically in the automatic generation of sequences which present visualisations that can be applied in such narratives.

The derived data approach supports the creation of explorable visual narratives by automatically generating and presenting sequences of visualisations for a narrative under construction. It also supports the exploration of these narratives by deriving and presenting data related to it, which can be further investigated. A framework consisting of a set of models and components based on this *approach* has been designed and implemented in a technical infrastructure called VisEN (Visual Explorations through Narratives). The implementation of VisEN is tightly coupled to the design requirements to realise this *approach*, hence its evaluations validate the usefulness and effectiveness of this *approach*.

VisEN has been evaluated through a number of user trials, both in a TEL context and as a stand-alone system. In TEL, it was shown that the *approach* supported the majority of ‘improving students’ to engage with course content. As a stand-alone system, the evaluations showed that the *approach* supported both the construction and consumption of explorable visual narratives. VisEN was also used with non-TEL data (Irish economic data), which showed that the *approach* can be effectively used in other domains as long as the data adheres to the same format used in TEL.

Contents

List of Figures

List of Tables

Abbreviations

AIED	Artificial Intelligence in Education
AJAX	Asynchronous JavaScript XML
API	Application Programming Interface
CAM	Contextualised Attention Metadata
CMS	Course Management Systems
CSV	Comma Separated Values
DSL	Domain Specific Language
EDM	Educational Data Mining
HCI	Human Computer Interactions
HTML	Hypertext Markup Language
ITS	Intelligent Tutoring Systems
KML	Keyhole Markup Language
LA	Learning Analytics
LMS	Learning Management System
MOOC	Massive Open Online Course
OLM	Open Learner Model
PLE	Personalised Learning Environment
SQL	Structured Query Language
SUS	System Usability Scale
TEL	Technology Enhanced Learning
URI	Uniform Resource Identifier
VisEN	Visual Explorations through Narratives
XML	Extensible Markup Language

Abbreviations for Evaluation Features

EF1	VisEN provides visual interactions including drilldown, zooming, details-on-demand and filtering to facilitate view manipulations.
EF2	VisEN provides a process to facilitate visual encoding by mapping narrative slice data to a set of appropriate visualisation techniques.
EF3	VisEN supports consumers explore visual narratives by automatically generating derived data explorations of narrative slices and providing links to these, which load separately from the visual narrative, to enable analysis of derived data without disrupting the flow of the narrative.

EF4 VisEN supports authors, who have access to and an understanding of their own or others data, to produce visual narratives, through the automatic generation of derived data slices.

EF5 VisEN tailors the ordering of the derived data explorations, mentioned in EF3, by inferring consumer preferences based on usage of these visualisations.

1 Introduction

1.1 Motivation

As Technology Enhanced Learning (TEL) environments are continuously growing in popularity[1,2], the challenge of addressing poor student engagement and high dropout rates when using these technologies needs to be tackled. In such environments, students' active involvement and engagement with the learning process is lower than traditional classroom settings, and hence they tend to get demotivated, leading to high dropout rates (Koutropoulos et al., 2012). Such environments typically generate large quantities of student-logged data including their interactions with learning resources, assigned activities, discussion forums and course-related quizzes. This data often contains valuable information that can highlight resource usage trends, reveal learner engagement levels and can be used to present peer comparisons. However, in its raw format, it can be quite difficult to interpret and explore by students, educators and other interested stakeholders, who can collectively be referred to as end users. Nonetheless, there is a need for these end users to make sense of this data, to communicate over it and to share it, as it may highlight trends and influence decisions. As a result, mechanisms are required to enable end users to understand and explore this data.

TEL is indicative of many domains, in that it is rich in user data that can be presented through multiple contexts providing several opportunities to enable end users to visually scrutinise and explore the data. Guiding end users through TEL data can highlight important facts, such as learner engagement and resource usage trends, while at the same time presenting an authored message that can be consumed. In addition, supporting end users in exploring this data during consumption can enable them to scrutinise areas of personal interest and obtain a better understanding of the data. Guiding end users through complex data and supporting explorations into it requires mechanisms that enable effective authoring of messages with less prescriptive consumption.

One way of addressing these issues is to provide end users with visual stories that can be navigated and explored to help make sense of such data. These visual stories are constructed by users with knowledge of the data and are supported in the construction process through mechanisms that suggest sequences within the visual story.

---

[1] http://www.forbes.com/sites/michaelhorn/2015/04/23/report-that-says-online-learning-growth-is-slowing-misses-big-picture/
[2] http://elearningindustry.com/elearning-statistics-and-facts-for-2015

Information Visualisation facilitates an effective means to comprehend data as it maps data attributes to visual properties, forming visualisations that support pattern discoveries, communication and understanding of data (Card et al., 1999, Riche et al., 2010). Mapping appropriate visualisations to datasets or segments of data is common in visualisation tools (Gonzalez et al., 2010, Viegas et al., 2007) where a chosen dataset or data segments are automatically visualised. Storytelling in Information Visualisation or visual narratives can be defined as an ordered sequence of steps consisting of visualisations, which are linked or connected to make the communicated message more memorable (Austin, 2011). Visual narratives also provide effective ways of highlighting facts, making points and passing on information (Kosara & Mackinlay, 2013). Research in visual narrative has recently started to gain momentum and has been used in both journalism (Gao et al., 2014, Hullman et al., 2013b, Segel & Heer, 2010) and by visualisation tools[3,4] (Lee et al., 2013). However, visual narrative has not been used to facilitate the exploration of data related to the visual narrative that may support its credibility and offer further insights into it. Visualisation tools currently support users in constructing visual narratives through mechanisms that enable data and filter specifications; however the automatic generation of suggested sequences for the visual narrative are not supported.

Data transformation (Card et al., 1999) is the process of providing derived values and structure for input data. The state of the art in both Information Visualisation and educational systems (that use visualisations to present student data) support data transformations, but these are limited to data parsing, data extraction and providing statistical measures such as aggregates and summaries. Derived data transformations[5] are defined in this thesis as transformed data that is generated from the data source (input data) using predefined mappings to present data related to visualisations in the visual narrative. These transformations can provide statistical measures but they also provide related data views that can help end users to comprehend the data surrounding the visual narrative.

The derived data transformations are manifested as visualisations which can be used to support both the construction and consumption of visual narratives. In the case of consumption, the derived data transformations are appended to the visual narrative, rendering it as an explorable visual narrative. These transformations enable end users to

[3] http://www.gapminder.org/
[4] https://viz.ged-project.de/
[5] Derived data transformations are referred to as derived data slices in the context of visual narrative construction and referred to as derived data explorations in the context of visual narrative consumption and all three terms are synonymous.

investigate the data surrounding the explorable visual narrative and can help them in gaining certain insights that may not be achievable through the original message in the narrative. For example, a visual narrative presenting student performance and consisting of derived data transformations (rendered as explorations) can be used by an educator to assist him/her in comprehending the data surrounding the narrative. The derived data transformations (explorations) can provide the educator with useful insights, such as the time spent by the students against activities or resources used. In the case of visual narrative construction, the derived data transformations suggest subsequent visualisations for the visual narrative by displaying data related to the previous created visualisation.

Visual data exploration supports end users in gaining a better understanding of the data when little is known about it and exploration goals are vague (Keim, 2001). In addition, visual interactions such as selecting, filtering, and navigating support end users in making sense of data through view manipulations (Kosara et al., 2003). In addition, the drilldown interaction technique allows end users to explore details behind the presented visualisations[6], as they render a more detailed view of a specific element within the visualisation that is under examination. Nevertheless, these drilldowns do not necessarily show related data that can include detailed views of specific elements, summaries and derived data explorations of a particular visual narrative.

Data tailoring addresses information overload by representing data customised for the individual which includes filtering out data the end user finds irrelevant (Gauch et al., 2007). As the derived data explorations present data related to the visual narrative (to support end users in gaining insights into the data), tailoring these explorations to the preferences of end users can reduce information overload and present explorations of interest to them.

Addressing the issue of enabling end users to make sense of and explore student-logged data through explorable visual narratives require:

- mechanisms to enable and support individuals (who have some degree of knowledge of the data they have access to, but may not necessarily have any visualisation expertise) to build explorable visual narratives. In this thesis, these individuals are referred to as ‘authors’,
- enabling end users to consume these explorable visual narratives. The end users will be referred to as ‘consumers’ throughout this thesis.

[6] http://spotfire.tibco.com/

This leads to two key challenges; the first of these challenges is to support authors in constructing a story using student-logged data, which is complemented with automatically generated visualisations, resulting in a visual narrative. The visual narratives are further expanded with derived data explorations. Authors are supported in the construction process through suggested sequences, where the derived data slices are used to generate the suggested sequences. The authors have access to student-logged data (through systems they are responsible for or through open and freely available data), and wish to publish it to a target audience in a narrative format that can be easily followed and explored. The visualisations (both the narrative and the derived data slices) in the explorable visual narratives must be appropriate for the data and automatically generated to complement the story.

The second key challenge is to support consumers in navigating and understanding these explorable visual narratives. As mentioned above, exploring the data related to a story may provide end users with a deeper understanding or insight into the message communicated by the explorable visual narrative. Addressing this challenge requires a means to enable consumers to easily navigate and understand the message communicated in the explorable visual narrative, which is complemented by derived data explorations. These explorations are available to consumers as they view the explorable visual narratives and they can customise them by selecting which type of derived data they wish to have included in the explorable visual narrative. Based on the interactions of the consumer, the orderings of the derived data explorations can be tailored to the individual's preferences.

By addressing these two challenges, authors would be assisted in sharing their understandings through a narrative format that is complemented by visualisations, and consumers would be supported in understanding and exploring student-logged data. This would enable the large quantities of TEL generated student-logged data, which contains valuable, sought after information and which is often quite difficult to decipher, to be presented in a format that can be consumed and explored by consumers.

1.2 Research Question

This body of work aims to address the following Research Question:

"Can explorable visual narratives support consumers in understanding, benefiting from and gaining insights into data from the TEL domain; and to what extent can authors be appropriately supported in producing these narratives?"

In the context of this research, authors are users who have access to student-logged data from systems they are responsible for and have an insight of, or have access to and an understanding of others' open data. Authors are not required to have any expertise with visualisations. They may wish to communicate this data in a story format that is complemented by visualisations and that can be understood and further explored by a wide audience of consumers including a target set. For example a university professor may need to provide students who are participating in a personalised online course with live performance and learner engagement data and also highlight performance and engagement issues per individual. His students may wish to explore the time they spent on certain tasks and make comparisons with peers, which may not necessarily be available in the message provided by the professor. Or the professor may need to assess how well the learning needs of his students are being met and may wish to correlate the content they used with their assessments, and communicate his findings to the university and publishers, who in turn may require further details before making a decision.

The Research Question addressed by this thesis consists of two parts:

1. supporting consumers in understanding, benefiting from and gaining insight into TEL data they are interested in through explorable visual narratives,
2. supporting authors in constructing explorable visual narratives.

This thesis describes an innovative *approach*, which aims to address both parts of the Research Question, which are directly aligned to the two challenges discussed in section 1.1 (supporting the construction and consumption of explorable visual narratives). The research includes the implementation of a framework called VisEN (Visual Exploration through Narratives) which addresses both parts of the Research Question to support this *approach*. VisEN supports authors to construct explorable visual narratives (hereafter referred to as visual narratives), and addresses the second part of the Research Question as follows:

- by generating visualisations for author specified data,
- by suggesting sequences for the visual narrative through derived data slices,
- by using these derived data slices (transformations) to append explorations to the visual narrative using appropriate visualisations.

VisEN addresses the first part of the Research Question by supporting consumers in:

- navigating and analysing visual narratives,
- exploring the visual narratives through the derived data explorations,

- tailoring the derived data explorations according to consumer preferences.

Both sets of users need to have rudimentary computing skills, such as Internet browsing, while authors also require basic Microsoft Excel or CSV skills. The main difference between authors and consumers is that the former have some degree of knowledge about the dataset that they are using to build the visual narrative.

1.3 Research Objectives

There are four research objectives (stemming from the Research Question) that this thesis aims to address:

1. Research objective one is to analyse the state of the art to identify how well the two challenges discussed in section 1.1 are currently addressed:
   - Challenge One: Supporting users (authors) to communicate student-logged data in a format that is easily consumable.
   - Challenge Two: Enabling users (consumers) to understand student-logged data and gain deeper insights into it by providing mechanisms to facilitate explorations within the same dataset.

   This analysis will examine the methodologies used by systems in the Information Visualisation domain to support the access and communication of data. It will also discuss best practices in Information Visualisation to support users in making sense of data through visual interactions and analyse the level of exploration supported by visualisation systems. Findings from the analysis of these systems will be used to create a list of important tasks and features that visualisation systems should support to help users in making sense of data. Relevant tasks and features from this list will be used to analyse the TEL state of the art systems which present visualisations using student-logged data in order to support users in making sense of data.
2. Research objective two is to use the analysis of the state of the art (best practices and limitations) to define an *approach* to address both parts of the Research Question (supporting consumers in understanding, benefiting and gaining insights into student-logged data, and supporting authors to create explorable visual narratives using this data). The visual narratives will consist of sections known as narrative slices and each narrative slice will consist of a description, a visualisation and a set of explorations. These explorations will be presented through visualisations showing data related to each narrative slice.

3. Research objective three is to design and implement a framework (VisEN – Visual Explorations through Narratives) that will realise the approach defined in research objective two. Since the vast majority of student-logged data generated by online learning systems is numeric and stored in tabular formats, such as task completion times, performance, number of resources used and discussion forum activity (Govaerts et al., 2012, Jovanovic et al., 2008, Charleer et al., 2014), the *approach* will be implemented using tabular and numeric data.
4. Research objective four is to evaluate the *approach* defined by objective two with authentic users through realistic use cases to assess the extent to which VisEN supports the construction and consumption of visual narratives (addressing both parts of the Research Question). This includes using the *approach* to support the engagement and performance of undergraduate students, by assisting them in understanding their logged data through visual narratives. It also includes using the *approach* to provide educators with visual narratives presenting student learning behaviour. A secondary aim of this objective is to evaluate the implementation of the *approach* on tabular and numeric data from a non-TEL domain in order to determine the effectiveness of the *approach* in a different domain.

1.4 Contribution

This thesis proposes an innovative *approach* called the *derived data approach* that incorporates derived data transformations (synonym for derived data slices and derived data explorations) into the process of constructing and consuming visual narratives from student-logged data. The primary motivation of this *approach* is to use derived data explorations to support consumers in exploring visual narratives and gaining deeper understandings of the student-logged data. The secondary motivation is to use derived data slices to support authors in creating visual narratives.

The major contribution of this thesis is *the derived data approach to the construction and consumption of explorable visual narratives in TEL (the derived data approach)* and its accompanying technologies and models that support it. It contributes to the state of the art by supporting the construction and consumption of visual narratives. The minor contribution of this thesis is the VisEN framework to support this *approach.* A proof of concept prototype was first developed to acquire quantitative and qualitative feedback for users constructing narrative slices and exploring inter-connected visualisations. Version 1 of VisEN automatically generated appropriate visualisations for narrative slice data. Version 2 of

VisEN was used by the AMAS Personalised Learning Environment (PLE[7]) to provide individualised visual narratives (presenting learners' course engagement breakdown, task durations and peer comparisons) to a total of 233 students pursuing undergraduate degrees across two academic years. The AMAS PLE was used by the learners (third and final year Computer Science and Computer Engineering students) as part of their course work for the Information Management and Data Engineering module run in Trinity College Dublin. Version 2 of the framework addressed the first part of the Research Question or challenge two (supporting the consumption of visual narratives). Version 3 of VisEN was used to produce visual narratives with the support of derived data slices. It was also used to support the consumption visual narratives through derived data explorations. Version 3 of the framework addressed both parts of the Research Question or both challenges one and two (supporting the construction and consumption of visual narratives).

This work has contributed to eight conference, work shop and journal publications in the area of TEL. The details of these publications are outlined in Chapter 7.

1.5 Methodology

A review of the state of the art was undertaken in the area of Information Visualisation to identify best practices and limitations in the construction and consumption of visual narratives and to support consumers in making sense of data. From this review a set of recommendations were made regarding important tasks and features visualisation systems should support. Relevant recommendations from this set were used to analyse educational systems using visualisation to support students in making sense of their logged data and assisting educators in viewing student learning patterns. The analysis of the state of the art in Information Visualisation and educational systems (trends and limitations) greatly influenced the definition of the *approach* and the design of the framework that realised this *approach*. The implementation of the VisEN framework closely adhered to the design requirements and was developed in phases (proof of concept prototype followed by three versions) which gradually integrated all of the design requirements.

A proof of concept prototype was first developed allowing participants of a user trial to take on the role of authors and construct narrative slices using a connected data source. It also enabled participants to take on the role of consumers and view, navigate and explore the AMAS PLE student-logged data presented through inter-connected visualisations and hand-

[7] A PLE is a learner-centric educational platform used by students for self-directed learning (www.educause.edu)

coded explorations. The goal of the prototype was to present a partial implementation of the *approach* to participants familiar with student and educator needs in TEL and attain feedback using a questionnaire to determine suitability and effectiveness of this *approach*. The response to the proof of concept was largely positive with participants favouring navigating and exploring data through inter-connected visualisations. This response indicated that a framework supporting the construction and consumption of visual narratives in TEL would be valuable.

Following the positive feedback from usage of the proof of concept prototype, version 1 of VisEN was implemented which generated and ranked visualisations for specified data segments. This version implemented the design requirements to generate and rank a set of appropriate visualisations for specified data. This was evaluated by ten experienced visualisation users and their feedback was used to determine how comprehensible, accurate and appropriate the visualisations were for the data segments. This feedback was gathered using a questionnaire.

Prior to the completion of version 2 of VisEN, the professor conducting the Information Management and Data Engineering module was shown a visual narrative with automatically generated visualisations and derived data explorations. The aim of this preview session was to assess the suitability of such visual narratives individualised for his students. Feedback from the professor was gathered through an interview, and was used to complete the implementation of version 2 of the VisEN framework. This version included an interface for visual narrative presentations and functionality for the derived data explorations. Sequenced visualisations presented in a narrative format and derived data explorations in the form of visualisations linked to the visual narrative were supported in this version. It was deployed to the AMAS PLE and evaluated during the months October – December 2013 and October – December 2014 by two separate groups of students (108 and 125 third and fourth year computer science and computer engineering students respectively). These two sets of user trials consisted of students using the AMAS PLE to conduct coursework, and their usage of their individualised visual narratives were used for the evaluation. Students were also asked to complete a post-course questionnaire regarding their experience with the AMAS PLE, which consisted of six statements related to their individualised visual narratives. The responses to these six statements were also used for the evaluations of the first part of the Research Question (consumption of visual narratives).

Version 3 of VisEN was implemented to include the functionality of narrative construction supported through derived data slices, thereby completing the full implementation of the *approach*. This version was evaluated by 40 users who constructed visual narratives from two data sources (including a data source from a non-TEL domain) and who viewed and explored a previously constructed visual narrative. The data used for this evaluation consisted of task completion times, system interactions, questionnaire responses and structured interviews.

A total of four user experiments were devised for the proof of concept and three versions of VisEN to evaluate *the derived data approach*. The proof of concept prototype and version 1 of the framework implemented aspects of the *approach* and participant feedback from user trials was used to improve implemented features. Versions 2 and 3 of VisEN addressed the two parts of the Research Question outlined in section 1.2. Specifically, version 2 addressed the first part and version 3 addressed the second part of the Research Question.

The research methodology adopted by this thesis was a combination of quantitative and qualitative research methods, which included the analysis of participant logged data, questionnaire responses and interviews. The quantitative method used in this thesis included the analysis of questionnaire responses of participants evaluating the proof of concept, versions 1, 2 (AMAS PLE students) and 3 of the framework. The quantitative method was also used to analyse student visual narrative usage data (version 2) and participant timings in constructing narrative slices (version 3). The remaining analysis in this thesis used the qualitative method which focused on participant interviews, including the Information Management and Data Engineering module course professor interview prior to the completion of version 2 of the framework and its deployment to the AMAS PLE. The evaluation of version 3 of the framework also consisted qualitative evaluations in the shape of structured interviews with 40 participants who used VisEN to construct and consume visual narratives.

1.6 Thesis Overview

Chapter 2 reviews the state of the art in Information Visualisation to identify important visual analysis tasks and features to support authors in sharing and communicating data in a format that consumers can explore and making sense of. It describes best practices in the areas of data access, data transformations, visual encodings, visual interactions and visual guidance. It highlights the limited focus of data transformations (primarily focused towards statistical measures) and discusses the lack of support for derived data transformations that can scaffold both the construction and consumption of visual narratives. The chapter concludes with a list of recommendations that visualisation systems should implement to support authors in constructing visual narratives that can be consumed and explored.

Chapter 3 uses the relevant recommendations from Chapter 2 to analyse the state of the art in TEL systems that present visualisations using student-logged data to learners and/or educators. It identifies best practices including support for peer comparisons, visual interactions and the presentation of live student data. It also highlights limitations which include the lack of visual narratives in TEL and limited focus of data transformations which concentrate only on statistical measures and do not include the presentation of related data. Both Chapters 2 and 3 address research objective one (analysing the extent to which authors are supported in communicating a message using their data through visualisations and consumers are assisted in understanding and exploring this data).

Chapter 4 uses the influences from the Chapters 2 and 3 to define the *derived data approach* and design a framework to realise this *approach*. It provides an overview of the *approach* and identifies a list of design requirements. The chapter consists of detailed descriptions of all of the components of the VisEN framework which amongst others comprises of the Derived Data Visualisation Model (responsible for dynamically generating derived data transformations) and the View Generation Model (responsible for generating visualisations).

Chapter 5 discusses the implementation details for the sub-components, engine and models of the VisEN framework that adhere to the design requirements to realise the *approach*. The chapter includes discussions on the features supported by the implementation of the framework and highlights the technologies and third party libraries used. Chapters 4 and 5 address research objectives two and three (defining, designing and implementing the *approach*).

Chapter 6 addresses research objective four (evaluating the *approach* using authentic users through realistic use cases) by evaluating the effectiveness of the *derived data approach* through the analysis of the results of four user-based experiments. The evaluation consists of quantitative and qualitative studies.

Chapter 7 discusses the conclusions of this thesis by describing the research objectives and achievements, the contributions made to the state of the art and identifies future work and directions for this research.

2 Information Visualisation: Making Sense of Digital Data

The amount of data that has been either created or replicated to date (2015) has almost reached eight trillion gigabytes and has been projected to reach a total of 40 trillion gigabytes by the year 2020[8]. The vast majority of this data has been generated from user interactions, including those with social media and devices connected to the internet, in addition to data published by enterprises and government bodies. This data can quite often be difficult to interpret by ordinary people as it can be overwhelming and obscure. However, there is a need for individuals to make sense of this data, to explore it, to communicate over this data and to share it, as it often contains valuable information that can highlight trends and influence decisions. The need to understand and explore this data may result from requirements of individuals to report findings, or from requirements of analysing data where a user is a stakeholder and intends to take some particular action. As mentioned in Chapter 1, Information Visualisation maps data attributes to visual properties and forms visualisations which can offer simpler ways for users to understand and consume data (Card et al., 1999). In addition Visual Analytics aims to support users in the process of understanding data (Thomas & Cook, 2005). This chapter describes the state of the art in Information Visualisation, by identifying important visual analysis tasks and features to support users in making sense of data (research objective one). It includes a discussion on best practices and limitations in visual exploration and guidance and concludes with a list of features that visualisation systems should support in order to assist one set of users in constructing visual narratives and another set of users in understanding and exploring data.

2.1 Introduction

Information Visualisation taxonomies, reference models, and visual navigation patterns are used in this chapter to highlight important transformations, interaction techniques and models required for visual analysis. A number of taxonomies (Heer & Schneiderman, 2012, Yi et al., 2007, Chi, 2000) outline visual analysis tasks that support the data exploration process and briefly discuss implementations of these. The reference models (Card et al., 1999, Chi, 2000) highlight various abstractions, transformations and mappings applied to data as it is visually analysed by users. The visual navigation patterns (Schneiderman, 1996, Van Ham & Perer, 2009) outline tasks that support users in manipulating views. This chapter discusses best practices in these areas by describing the features supported and their implementations by the state of the art visualisation systems. These areas include data access

[8] http://www.nytimes.com/interactive/2013/06/19/business/Big-Data-Will-Get-Bigger.html

and data transformation, visual encodings, visual interactions, visual guidance and personalisation practices in Information Visualisation. This chapter provides a deeper analysis into these systems than that found in the taxonomies and models referenced above, by comparing and contrasting implementations, highlighting best practices, discussing evaluation results (where available) and identifying limitations.

The remainder of this chapter is structured as follows: section 2.2 discusses data access and data transformation; section 2.3 addresses visual encodings; section 2.4 describes visual interactions; section 2.5 discusses visual guidance and section 2.6 outlines personalisation directions in Information Visualisation.

2.2 Data Access and Data Transformation

Visualisation systems access data from several sources including single and multiple databases, spreadsheets and other file formats. However, this data may require transformations or derivations to alter values and its underlying structure to present new attributes or related data. This section discusses and analyses data access and data transformation research in Information Visualisation.

2.2.1 Data Access

A large number of visualisation systems (Tang et al., 2004, North, 2000, Woodruff et al., 2001, Derntl et al., 2012, Mazza & Dimitrova, 2007) support visualising data accessed from relational data models. Generally this is from a single specified data source, for example, Rivet (Tang et al., 2004) uses three layers to support data access. The first layer is the "ordered relational data model" where data sources use an abstraction to appear equivalent to the users. The second layer is a "named data sources and a data repository" to simplify data retrieval by users. The third layer is an "XML file specification" with information about a data source. Derntl et al. describe an architecture that allows specific users (widget creators) to create widget-based visualisations (Derntl et al., 2012). These visualisations are created by selecting a database connection from pre-configured database connection strings, constructing a query and defining the visualisation technique to be used to display the query results.

In some cases, such as Tableau[9], *née* Polaris (Stolte et al. 2002), data can be accessed simultaneously from multiple data sources. Tableau uses data blending (Morton et al., 2012)

---

[9] http://www.tableau.com

to create visualisations using data from multiple sources. Data blending consists of three steps[10]:

- Data Acquisition (gathering data from the various specified sources),
- Joining Data (combining it into a dataset),
- Data Cleansing (removing unwanted data).

Visualisation systems (Gonzalez et al., 2010, Viegas et al., 2007) also support users uploading data, either from file systems or entering it through web-based interfaces, which then gets visualised. In the case of systems such as Tableau and Tibco Spotfire[11] users can upload data from the file system in addition to connecting to data sources. A system called Many Eyes (Viegas et al., 2007) uses a table based data model consisting of same-length named columns supporting numeric or text type data. Google Fusion Tables (Gonzalez et al., 2010) allows users to upload their data and explore it through appropriate visualisations proposed by the tool (based on the data types present). It supports data uploaded from sources such as CSV files, Spreadsheets and KML (Keyhole Markup Language).

Analysis

Relational models are very prevalent, offering access to existing datasets and allowing designers to decouple the visualisation system from the data (Tang et al., 2004). Users are also interested in viewing and querying related data from multiple data sources (Morton et al., 2012), and it is common to find related data across multiple data sources instead of residing in one dataset. Supporting user-uploaded data through files or web-based interfaces is also a popular mechanism to specify data. This allows users to upload data that is free from missing or erroneous values and allows data to be structured in a fashion that supports the process of visual encodings. However, such architectures introduce overheads where users are required to locate and structure the data according to specifications. This discussion highlights the importance of supporting data from both relational data and uploading data from user files, and it can be seen that commercial visual analytics tools such as Tableau and Tibco Spotfire support both formats. In the TEL domain, student-logged data is usually stored in databases. However learner reports[12], including grades, are usually formatted for CVS and MS Excel files. Hence supporting both relational models and various file formats for accessing data can be considered important.

---

[10] http://www.datawatch.com/what-is-data-blending/
[11] http://spotfire.tibco.com/
[12] http://www.data.gov/

2.2.2 Data Transformation

Data transformation (also referred to as Derive or Connect) is the process of providing derived values and derived structure for the connected data source (Card et al., 1999), in addition to handling data errors and missing values. Data transformations are applied using source data, metadata and database schema including data types, keys and relationships. An example of derived values can refer to applying statistical measures to the data and derived structure can refer to sorting and classifying the data. These derivations can support various views of the data such as counts and aggregates. The derivations can also present attributes related to those shown in the visualisations and can assist users in gaining a deeper understanding of the data through exposure to related data. For example, a visualisation presenting resources used by a student for a selected course task can include transformations that display related data rendered through visualisations, such as the student's performance against the task or resources used by top performing students for the same task. Heer and Schneiderman highlight the importance of such transformations in the data analysis process by noting that: "*visual analytics tools should include facilities for deriving new data from input data*" (Heer & Schneiderman, 2012).

Although data transformations can be used to generate several types of related attributes from input data, they have often been limited to data parsing, data extraction and providing statistical measures, summaries and data aggregations. For example, Tibco Spotfire enables users to apply several configurations to the data such as dynamic aggregation, presenting various statistical measures of the data including sum, average, minimum and maximum. Polaris (Stolte et al., 2002) describes multiscale visualisation using data cubes (a common method for database abstractions) and supports data transformations by categorising data as dimensions and measures and provides summaries including aggregates of data. InfoZoom (Spenke & Beilken, 2000, Spenke, 2001, Spenke & Beilken, 2003) automatically computes data aggregation attributes using functions similar to a spreadsheet program, including sum, count, minimum and average and these attributes are available during the exploration process. OnionGraph (Shi et al., 2014), a visualisation interface supporting hierarchical focus plus context visualisations for networks, supports overview and low level analysis, semantic aggregation for data abstraction and exploration across heterogeneous networks.

Xiao et al. introduce a network visualisation system (Xiao et al., 2006) that is used to visually analyse previously conducted analysis. The system creates a set of events which are derivatives of clauses used in logical models, and uses these derivatives to enable users to view the dataset at a higher level, showing aggregations of the dataset. Mahoney presents a network visualisation showing university study patterns between modules and degrees awarded for the student body across seven academic years (Mahoney, 2013). The system provides three statistical analysis graphs to support exploration by showing data related to the original view. The first statistical analysis graph presents the mean amount of modules shared between degree awards. The second graph shows the density of the graph by calculating the mean amount of connections for each degree award as a proportion of the maximum connections available. The final graph presents the modularity of the degree awards. Data transformations used in TitleBars (Hearst 1995) and Information Mural (Jerding & Stasko, 1998) involve parsing data into feature vectors to support the computation of data intersections (TitleBars) and facilitate dynamic value filtering (Information Mural). Matkovic et al. use a coordinated visualisation system to present an intensive care unit dataset consisting of logged entries of patient data including statistical summaries and differences of the selected data (Matkovic et al., 2012).

Analysis

Data transformations support the data analysis process, by presenting various data attributes derived from the input data, which highlights their importance in supporting users in making sense of data. Various implementations of data transformations have been supported by several visualisation systems. However, it can be seen from the above review of the state of the art that these implementations have been predominantly focused on and limited to data extraction, data parsing and providing statistical measures, summaries and aggregations of input data. The dynamic generation of transformed data that is related to the data shown to users through visualisations has not been implemented. This related data can be generated using derived data transformations which are defined in this thesis and which use preconfigured mappings to present data related to the previous visualisation. For example, if a user is interested in exploring certain data shown through a visualisation (such as specific nodes), related data views can be dynamically generated by applying derived data transformations to the data within the same dataset. The related data can be generated by examining the input data used to generate the visualisation, including constraints, mappings and data relationships, and transforming this input data through derived data transformations. These transformations generate related data views by manipulating input data and constraints. For example, in TEL, if a learner is analysing a visualisation showing his/her course engagement breakdown, derived data transformations such as the students' resource usage, class course engagement breakdowns, top students' course engagement breakdowns can be generated and made available to the learner. Such dynamic transformations are not supported by the state of the art.

2.3 Visual Encoding

Visual encoding is the process of mapping data to either suitable graphical marks that are used to form a view, or to suitable visualisation techniques (a predefined view, such as bar charts, line charts, etc.). Graphical marks are areas, lines and points used to encode information through positional (one, two or three dimensions), retinal (size, colour, texture, shape, etc.) and temporal properties (Bertin, 1983). This section discusses 1) systems that map data to graphical marks to build visualisations and 2) systems that map data to a predefined visualisation technique, and compares and contrasts both methods.

2.3.1 Mapping Data to Graphical Marks

A number of visualisation systems (Livny et al., 1997, Tang et al., 2004, Stolte et al., 2002, Mackinlay, 1986) map data to graphical marks, which are combined to create a visualisation. For example, DEVise (Livny et al., 1997) consists of source data tables (TData) and graphical representations (GData). The records in the TData tables are mapped to visual symbols to create GData table entries consisting of visual attributes, such as colour, orientation, axes, size and shape, and the mapping process uses the TData schema. Rivet (Tang et al., 2004) uses an architecture which maps nominal values to colour and quantitative values to size and defines encodings which map certain data fields to certain visual variables, such as axes and colour.

Tableau uses VizQL, a formal language for describing visualisations (Hanrahan 2006). Users drag and drop data fields onto a visual canvas provided by Tableau to generate VizQL statements, which build visualisations. Tableau categorises and partitions database fields into scale and role classifications. The scale classification categorises the field as ordinal or quantitative and this classification is used for visual representation. Quantitative fields can be represented as axes and ordinal fields can be shown as header or classes. The role classification partitions fields into dimensions and measures. The graphic generation is comprised of three components, which include table configuration, graphic type within a pane, and visual encodings. The table configuration component uses database fields (ordinal and quantitative) to create expressions. The expressions consist of operands (database fields) and three operators (+, x and /). These expressions are used to form the clauses to create table axes. The graphic type is chosen based on the axes, which is determined within the table configuration, and a mark type chosen by the user. The mark types consist of text, line, Gantt bar, rectangle, glyph, polygon, circle and image. Finally the visual mappings component uses the selected mark to encode the relevant database fields.

A Presentation Tool (ATP) (Mackinlay 1986) uses compositional algebra consisting of graphical languages and operators to generate designs. It requires application designers to supply data, which is analysed for structural properties of inputs, such as qualitative, ordinal, numerical and nominal values when synthesising designs, which are produced as an abstract image description. ATP uses graphical objects, including points, lines, and areas as sentences of graphical languages (collection of tuples), and semantics, to encode arrangements of graphical representations. The graphical sentence specifies the location of the graphical object, which can be used to determine the height and width of the object (expressiveness). ATP also considers the perceptions of an individual viewing the image and encodes ordering, size and colour into the graphical design process based on the input data. The composition algebra is used to generate designs by using a set of primitive graphical languages and composing design by merging parts that encode the same information. The synthesis algorithm consists of three processes: 1) partitioning the set of relations until a match is found with a primitive language; 2) selecting of candidate designs for each partition; and 3) combining the designs using composition operators.

Analysis

Both more recent and historical visualisation systems have used, and are currently using, techniques to map data to graphical marks. Expressions, consisting of operands and operators and compositional algebra have been used effectively by Tableau and APT respectively to map data to graphical marks. These types of mappings offer several advantages such as increased generalisation, extensibility and flexibility as new mappings can easily be added (Tang et al., 2004). In addition various types of datasets can use existing mappings. However, the literature suggests that over-generalisation, which is a feature that can result from these types of mappings, requires additional work to create appropriate encodings for specific visualisations.

2.3.2 Mapping Data to Visualisation Techniques

A growing number of visualisation systems (Viegas et al., 2007, Gonzalez et al., 2010, Derntl et al., 2012) map data to visualisation techniques which are either part of the system or taken from the widely available range of existing visualisation libraries. For example, Many Eyes (Viegas et al., 2007) automatically renders user data through commonly used visualisation techniques. It uses a table-based data model consisting of same-length named columns where each column can either have numeric or textual data. The user-inputted data can also be interpreted by the system as unstructured (coming from freeform text entered onto the form), or it can be interpreted as a table if entered as tab-delimited. For tabular data, the system uses heuristics to determine a visualisation technique and it allows users to reverse this decision to select a different technique. Each dataset has metadata associated with it, some aspects of which are provided by the user and other aspects are automatically determined by the system. Many Eyes consists of over a dozen visualisation techniques, each of which has a predefined schema that specifies the data requirements. The schemas consist of mandatory and optional type slots, such as textual slots, multiple textual slots, numeric slots, multiple numeric slots and unstructured slots. The typed slots are matched against the typed columns to determine a visualisation technique to match the dataset.

Google Fusion Tables (Gonzalez et al., 2010) uses the Google Visualisation API[13] to source the visualisation techniques. The data types in the input data are compared to the types needed for each visualisation technique to determine the appropriate set of techniques to be used. The set of visualisation techniques used by the infrastructure discussed by Derntl et al. (Derntl et al., 2012) are the chart-based techniques sourced from the Google Visualisation API. Gretl[14] supports econometric analysis through a number of visualisation techniques, including line charts, box plots and scatter plots. The user is required to provide formatted data and select the technique to render it. Microsoft Excel[15] supports data analysis by automatically generating visualisation techniques for user specified data, with users selecting the chart type from a number of available techniques.

Analysis

Some of the more recent visualisation systems map data to supported visualisation techniques or make use of available visualisation libraries and map data directly to appropriate visualisation techniques. Using these libraries offers the advantage of reducing the coding effort required to build the visual encoding component. Nevertheless, an understanding of the characteristics and affordances of each of the visualisation techniques used needs to be developed and integrated into the visual encoding component to allow the correct visualisations to be selected.

This section has addressed the two methods used for visual encoding and highlighted their advantages and limitations. In comparing and contrasting both practices, it can be seen that

[13] http://developers.google.com/chart/
[14] http://gretl.sourceforge.net/
[15] Microsoft Office 2013

mapping data to graphical marks is more extensible and flexible than mapping data to visualisation techniques as it can handle a broader range of mappings. However, it requires a greater coding effort as it cannot take advantage of existing visualisation libraries. In the case of mapping data to visualisation techniques, a myriad of literature (Dias et al. 2012, Chi 2000, Carr et al. 1987, Heer et al. 2010, Graham & Kennedy, 2010), including evaluation papers exist that can support designers in encoding characteristics for visualisation techniques. In summary, both practices serve their individual purposes quite well and depending on the system requirements one can be chosen over the other.

2.4 Visual Interactions

A primary aim of Information Visualisations is to support pattern discovery, communication and understanding of data (Riche et al., 2010). Visual interactions support individuals in understanding data better through mechanisms and techniques that enable them to analyse, manipulate and explore visual representations (Kosara et al., 2003). Yi et al (Yi et al., 2007) and Heer and Schneiderman (Heer & Schneiderman, 2012) define popular visual interaction categories that enable view manipulations. These include *select*, *explore*, *elaborate*, *filter*, *navigate* and *multiple and coordinated views*. This section presents best practices amongst visualisation systems using these visual interaction techniques to support users in manipulating views to make sense of data. Other visual interactions including zoom, panning, details-on-demand and drilldown are included within the discussions of this section.

2.4.1 Select

The *select* interaction technique is supported by most visualisations systems and allows users to click and highlight items of interest within a view. Heer et al. introduce a visualisation system supporting a higher selection criteria (compared to selecting a set of items), allowing users to select a region within the views to analyse details (Heer et al., 2008). The selections are modelled using SQL-like queries, which are generated by users through visual interactions by selecting an item, a range (by dragging) or an attribute from the legend.

Figure 2.1 shows murder cases of Hispanic victims in a selected region of Los Angeles in 2007. The system also supports query relaxation techniques, which generalise a selection made by users to include additional related items. Clicking on a node in the scatterplot in Figure 2.1 queries its details, but multiple clicks adds related attributes to the query to expand it with day, week and month.

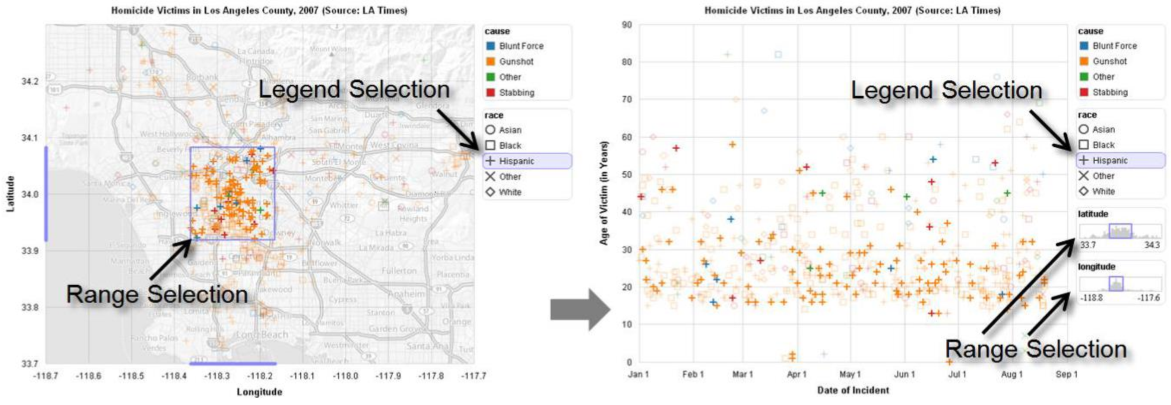

Figure 2.1 Supporting Higher Selection Criteria

Table Lens (Rao & Card, 1994) implements the focus plus content technique allowing users to select elements or regions of interest within the visualisation. This presents a zoomed view maintaining the surrounding context by distorting the layout, dividing the space appropriately for the cells in the focus space and dividing for cells in the context space (shown in Figure 2.2).

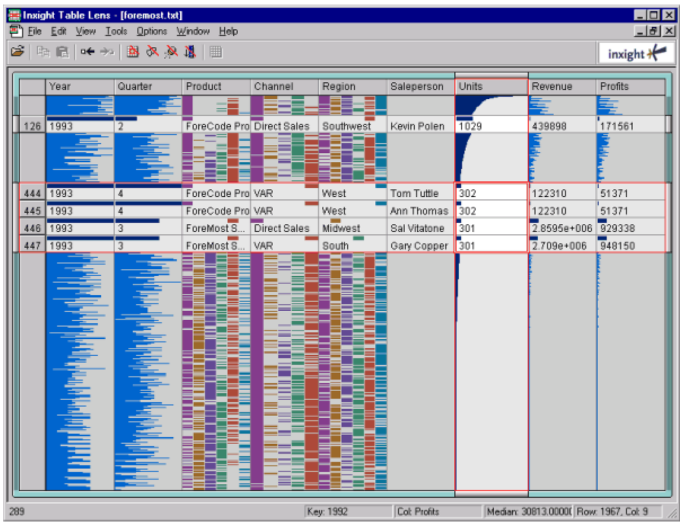

Figure 2.2 Table Lens zoomed view

Dust & Magnet (Yi et al., 2005) allows users to view and interact with multivariate data to extract information of interest by presenting data points as "dust" particles and variables as "magnets". The tool allows users to select and drag magnets within the visualisation which cause the dust particles to readjust their positions. Changing the position of magnets enables users to identify data attributes with respect to the magnets in the visualisation.

Analysis

The review above covers a summary of three examples of how the *select* interaction technique has been implemented. The evaluation of the visualisation system discussed by

Heer et al. supporting selection ranges and query relaxation techniques (Heer et al., 2008), found that participants using the system benefited from the selection criteria. The evaluation of Dust and Magnet (Yi et al., 2005) highlighted how users were able to select several dust particles of interest and view their details in textual format. Table Lens uses *select* to support the analysis of large quantities of data present in a table. The *select* interaction technique has been implemented by most visualisation systems and is commonly used by users interacting with interfaces. This interaction technique supports users in establishing data interests, specifying ranges and boundaries within visualisations and hence implementations of *select* are very important for systems supporting visualisations.

2.4.2 Explore and Elaborate

Explore presents a different part of the data, usually not shown due to the size of the data, screen or cognitive limitations. Panning and direct-walk techniques are commonly used as a form of *explore* where a user scrolls across a view (panning) or moves the view position through selections or hyperlinks (direct-walk). The *elaborate* technique allows users to change the representation of the visualisation from an overview to a detailed view through techniques such as drilldown. The details-on-demand (revealing element details) technique also falls into this category.

Vizster (Heer & Boyd, 2005) has been used to present interactive social networks using node-link representations and supports a number of interaction techniques including *explore* and *elaborate*. Double-clicking a node causes it to expand revealing the network for the selected member and double-clicking it a second time causes it to contract. This allows users to drilldown to view networks of individual members in the context of the entire data.

Elzen and Wijk introduce a novel visual exploration technique: small multiples visual exploration (Elzen & Wijk, 2013), consisting of visualisations rendering the data for analysis and small multiples (Tufte & Graves-Morris, 1983) to support exploration. This enables users to select parameters, which load a set of small multiples, allowing users to compare alternative visualisations of the chosen parameter and select one of these views (the most interesting) to enlarge and explore. Lee and Grinstein introduce a portion of the Exbase (Lee & Grinstein, 1995) architecture which describes a visual database exploration model. It supports visual exploration of the database by enabling users to alter queries and visualisation transform components. Altering query and visualisation components are supported at a range of levels, where a query could include restrictions on conditions to change the target and enable exploration of different portions of the database or changing the visual display settings to a new visualisation.

Amongst several other interaction techniques, Tibco Spotfire supports drilldown, where items in the visualisation can be selected to view details through a different visualisation as shown in Figure 2.3. These visualisations are linked and the drilldown view shows a visualisation of the details-on-demand data. The tool supports linking of several visualisations in this manner and allows users to drilldown into various parts of the data.

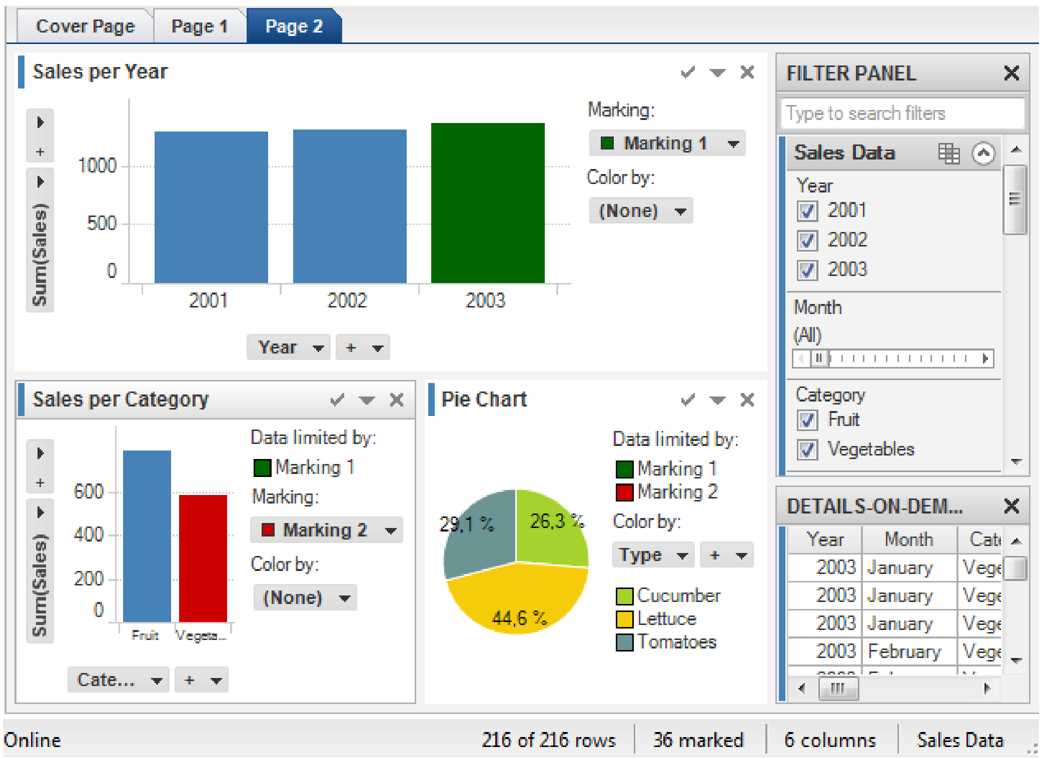

Figure 2.3 Visual Explorations in Tibco Spotfire[16]

Analysis

The *explore* and *elaborate* interaction techniques enable users to view data that is not rendered, using the panning, direct-walk, drilldown and details-on-demand interaction techniques, which support the data exploration process. Three implementations of *explore* and *elaborate* are discussed above that incorporate drilldown and details-on-demand (Tibco Spotfire and Vizster), presenting detailed views, and facilitating data comparisons through small multiples (Elzen & Wijk, 2013). The evaluation of the visualisation system supporting small multiples exploration (Elzen & Wijk, 2013) found that small multiples did not improve

[16] http://spotfire.tibco.com/

task execution times, nor reduce the number of errors made by users. Nevertheless, it found that users required fewer steps to answer questions and less time to explore the data.

The *explore* and *elaborate* interaction techniques address user needs to view data that is beyond the capacity of a visualisation. They support the data exploration process by panning across views, drilling down to view items of interest and viewing details that can support users in making sense of data.

2.4.3 Filter

The *filter* technique allows users to specify the data to be shown in the visualisation by supplying a range or a condition. Ranges can be specified through sliders, search boxes, hierarchies or checkboxes made available through interfaces.

Tibco Spotfire and Tableau support the drag and drop operations, enabling users to select data to analyse and specify filters. Tibco Spotfire includes users controls, such as checkboxes, radio buttons and numeric range sliders that can be used to apply filters to the visualisations. It also supports the sorting of data within visualisations through properties to enable users to find patterns in the data. Vizster (Heer & Boyd, 2005) supports data filtering by keyword search of profile attributes through a search box provided by the interface. Resulting search nodes are highlighted and the remaining nodes are greyed-out. A second mode of filtering is supported through checkboxes provided next to attributes such as age, gender or number of friends in the profile panel. Clicking on one of these checkboxes causes Vizster to remove member images from the network visualisation and shows the filtered data.

Riche et al. discuss interactive legends (Riche et al., 2010) as controls that support interaction techniques such as selecting and filtering data that are dynamically updated as users visually explore the dataset. The interactive legends have references to nominal, categorical, ordinal and numerical data types presented by Bertin (Bertin, 1983). For categorical and ordinal data, the design supports click, shift-click and control-click to filter items, select ranges and toggle items.

NameVoyager[17] (Wattenberg & Kriss, 2006) presents visualisations presenting the most common baby names since 1900 in the US. The data is visualised through interactive stacked graphs, where the x-axis shows the year and the y-axis presents numbers to show the usage frequency of names. Users can filter the dataset by selecting either the Boys or Girls radio

[17] http://www.babynamewizard.com/voyager

button and can type letters into the name search box. As a user types in letters, at every keystroke, NameVoyager immediately updates the visualisation, filtering the data for the names that start with the letters entered. Figure 2.4 shows popularity of names starting with the letter ‘A’ for both boys and girls for over a century. The darker colours show the most popular names used in recent times.

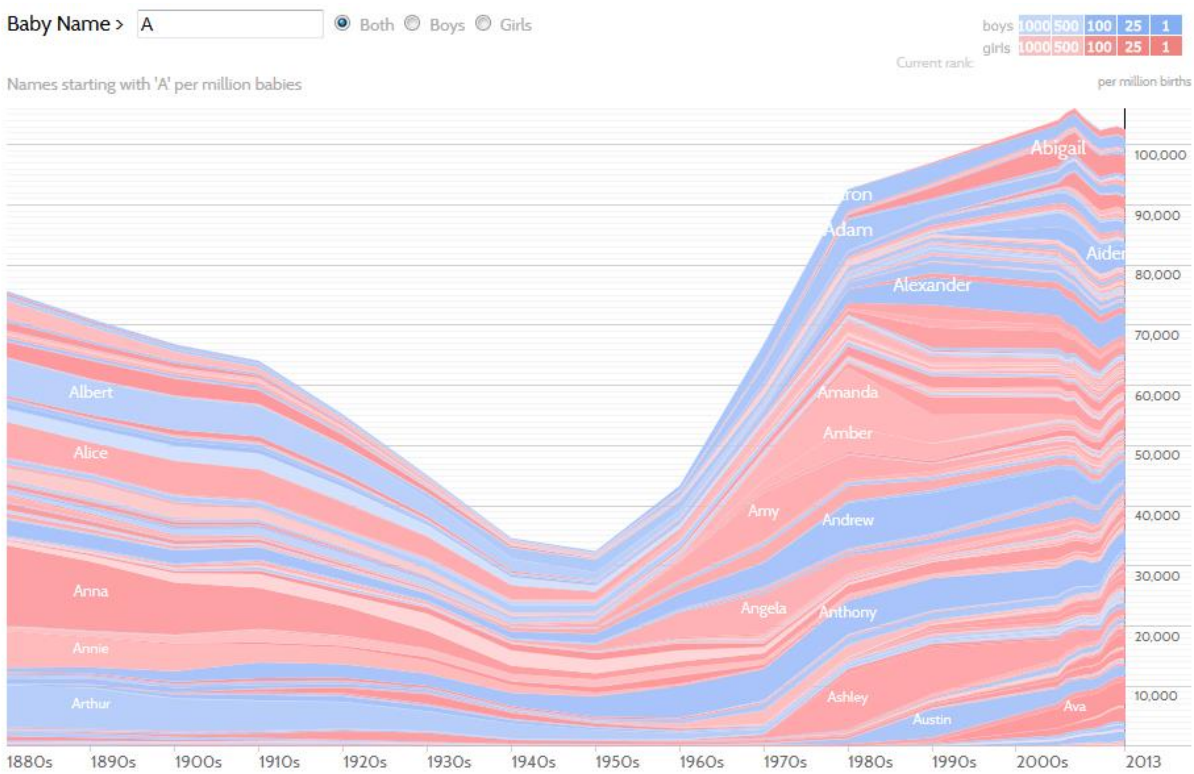

Figure 2.4 Name usage frequency in NameVoyager

Analysis

The *filter* interaction technique enables users to specify data attributes of interest to narrow the range of data that can be shown through a visualisation. In the case of NameVoyager, it is used to reduce the results displayed in the visualisation through user specifications. In the case of Tibco Spotfire, *filter* is used to specify data conditions when building visualisations and it is also used to manipulate existing views to show data of interest through checkboxes, radio buttons and numeric range sliders.

Riche et al. discuss an evaluation that focused on whether interactive legends (Riche et al., 2010) enhance understanding of data attributes and opacity mappings, and if these legends are slower to use for filtering than widgets. The evaluation consisted of two controlled experiments which were designed to compare users’ understandings and interactions between interactive legends and standard widgets. The measure to determine if using interactive legends was slower than widgets recorded the time taken by participants to complete two tasks. In the first task, participants were asked to filter values by a specified opacity and in the second task, they were asked to filter opacity for one or more values. The comparisons between the results of both experiments (interactive legends and standard widgets) showed that level of understanding amongst participants improved and tasks were conducted twice as fast with interactive legends.

Data filtering is fundamental for visual analysis as users are not usually interested in the entire dataset presented and more often require mechanisms to specify their interests or certain data attributes. Hence it is important that the filter interaction technique is supported by visualisation systems to help users in making sense of data.

2.4.4 Navigate

The visual information seeking mantra (Shneiderman, 1996) describes a commonly used navigation pattern which specifies "*Overview first, zoom and filter, then details on demand*". This pattern enables users to first have a broad view of the data before viewing details of areas of interest. However, at times, an overview may not be useful due to the size of the dataset and the "*Search, show context, expand on demand*" (Van Ham & Perer, 2009) navigation pattern can be more appropriate.

InfoZoom (Spenke & Beilken, 2000, Spenke, 2001, Spenke & Beilken, 2003) displays attribute relationships in databases presenting all the data objects in one view by using techniques to compress neighbouring cells with the same values into one cell. Double-clicking on marked cells zooms into the details of the attributes, removing other attributes at the higher level and offering a drilldown view of the clicked cell. Users can drilldown further by clicking on cells that appear in the zoomed-in views to analyse the data in more detail. Vizster supports navigation through the zoomed-in views, which appear by selecting a region within the display and the system draws higher resolution images or photographs for the updated view. In addition, focus plus context views are also supported by Vizster during zooming and distortion is used to increase the size of the nodes in focus while other nodes are greyed-out. DateLens (Bederson et al., 2004) provides users with a focus plus context calendar, supporting navigation through scrolling and magnifies selected elements (focal regions).

NameVoyager implements Schneiderman's information seeking mantra to facilitate user interactions to analyse names usage frequency. It first provides an overview of all the data in the database, where each distinct name is shown as a stripe. Clicking or double clicking on a stripe allows users to navigate usage frequency, statistics, and popularity during a specified time period, in addition to historical information about the name.

Analysis

Navigate supports the visual analysis process by allowing users to either start with a broad view of the dataset, or through a narrower view, and then apply a number of visual interaction techniques to explore other areas of the data. Four implementations of *navigate* are discussed above using overview, zoom and focus plus context. Implementations of focus plus context methods (DateLens and Table Lens) aid navigation by presenting overviews and detailed views. Users are presented with a visual overview of the data and they can zoom into areas of interest, while retaining the context of the surrounding data. An evaluation of Vizster (using over 100 users), highlighted that the participants found zooming to view details regarding user relationships and experience particularly useful.

Navigate incorporates important visual interaction techniques discussed in this chapter including: filter and details-on-demand, which have already been established as important to the process of making sense of data. In addition, it addresses user needs to view various parts of the dataset through broad and narrowing searches and supports zooming.

2.4.5 Multiple and Coordinated Views

Multiple and Coordinated Views consists of a single window with multiple synchronised visualisations controlled by coordinated mechanisms that support synchronised scrolling and drilldown.

Improvise (Weaver, 2004) provides an environment to construct and analyse coordinated interactive visualisations using relational data, and showing multiple views on the screen. It implements a "shared-object coordination mechanism" consisting of direct coordination (live properties) and indirect coordination (coordinated queries), which allow users to specify interactions and dependencies regarding behaviour and appearance of views. Some of the coordinated views provided by Improvise include synchronised scrolling, overview and detail, nested views and brushing (selecting data in one visualisation highlights corresponding data in other visualisations).

Snap-Together (North, 2000) allows users to dynamically connect visualisation tools to construct coordinated visualisations that can be navigated and explored without the need to program code. It uses the relational data model, and the database schema is utilised by users to create coordinated visualisations through join relationships. Snap-Together provides several interaction techniques including brushing and linking (selecting elements in one view, which highlights corresponding elements in other views), overview and detail, drilldown, synchronised scrolling and details-on-demand. The brushing and linking technique allows users to select an item in one view, which causes the corresponding item to be highlighted in another visualisation. The overview and detail technique supports selecting an item in the overview visualisation and causing the detail view to navigate to the selected item details. The drilldown technique allows users to click on an item and view summaries and details of that item, such as aggregates. Synchronised scrolling enables two views to show corresponding items and details-on-demand loads details of items in a textual view.

Analysis

Multiple and Coordinated Views has been shown to be effective in displaying large amounts of data through several perspectives on one screen (Weaver, 2004). Snap-Together was evaluated to determine the usability and usefulness of coordinated interfaces using 24 participants (North, 2000). Six participants were asked to create three coordinated interfaces and the remaining participants were required to perform nine browsing tasks that increased in difficulty. The results showed that users found the snapped views more beneficial than independent views and were not distracted by the coordination. This evaluation highlighted some benefits that multiple and coordinated views can offer. In general coordinated views support a design that can help users make sense of data and enable designers to avoid building interfaces cluttered with data. However, a greater cognitive effort is required during the analysis process (Heer & Schneiderman, 2012) as viewing various aspects of complex datasets can overwhelm novice users (Krishnamoorthy & Northm 2005), including experienced analysts. Hence, decisions to incorporate multiple and coordinated views into designs should include a careful consideration of the target audience (level of competence with visual analysis), and cognitive load required to process the data.

2.5 Guidance

The main focus of Information Visualisation has been to assist users in comprehending data by facilitating analysis, enabling patterns to be identified and supporting data explorations. More recently, an increasing interest in storytelling in Information Visualisation, commonly known as narrative visualisation or visual narratives (Segel & Heer, 2010) has emerged to support the visual analysis process. A visual narrative can be defined as an ordered sequence of visualisation steps that include textual descriptions and focus on data (Kosara & Mackinlay, 2013). Visual Narratives provide effective ways of passing on information and make the communicated message more memorable (Austin, 2011). From the ongoing research in visual narratives, Hullman and Diakopoulos define a taxonomy of presentation manipulations, which include rhetorical techniques to support communications and framing data presented in narrative visualisations (Hullman & Diakopoulos, 2011). Hullman et al. discuss how automatic sequencing (Hullman et al., 2013a) between visualisations in narratives can be approached by algorithmically identifying effective sequences considering single attribute changes between transitioning visualisations. They discuss how automatic sequencing could support the visual authoring process, specifically in cases where authors lack training in visualisation design.

In recent years, interactive visual narratives have started to be incorporated into online journalism. Segel and Heer discuss the design strategies for visual narratives mainly in journalism, consisting of genre, visual narrative tactics and narrative structure tactics (Segel & Heer, 2010). Seven genres (annotated chart, slide show, magazine style, flow chart, partitioned poster, comic strip and file/video animation) are described which can be combined to produce a visual story. Visual narrative tactics refers to the structure of the narrative, directing users' attention to items of importance in the narrative and transitioning views without disorientating users. Narrative structure tactics refers to ordering a path that users follow through the narrative, supporting interactions and a facilitation mechanism to communicate details presented in the views.

Visual interactions are also beneficial for visual narratives to support exploration of the data presented. However, a balance between interactions (which can distract from the message of the narrative) and focus needs to be met (Kosara & Mackinlay 2013). Recent workshops[18,19] have found that visual narrative supporting interactions help users explore data and specifically assist in the comprehensibility and credibility of the message. Nonethless, these interactions can reduce the level of guidance offered through the data (Ma et al., 2012) as they can distract the viewer from the original message. Segel and Heer (Segel & Heer, 2010) discuss a balance between author-driven and reader-driven stories, which are increasingly found in visual narratives. Three common methods of discovering such a balance can be found in online journalism (Martini Glass Structure, Interactive Slideshow and Drill Down Story). Martini Glass Structure initially provides a linear visual narrative that allows user interactions once the intended narrative is complete. The Interactive Slideshow format (used

[18] http://flowingdata.com/2011/11/04/telling-stories-with-data-visweek-2011/
[19] http://flowingdata.com/2010/11/11/telling-stories-with-data-a-visweek-2010-workshop/

by a system called Gapminder[20] and also by The New York Times[21]), supports interactions centred within each slide, such as details-on-demand. The Drill Down Story structure displays the theme and enables users to select an instance of interest to drilldown and view details of the data.

This section discusses the use of narrative visualisations in state of the art systems, specifically focusing on the level of support provided for constructing and consuming visual narratives. It also includes the taxonomy of presentation manipulations, automatic sequencing, design strategies and the balance between author and reader-driven visual narrative considerations within the analysis of the systems.

2.5.1 Constructing Visual Narratives

This section describes a number of systems supporting the process of constructing visual narratives. In particular, it analyses the degree of support offered to authors during the narrative construction phase. In addition, it discusses evaluation results presented by each system (publications), to assess the impact the system had on both authors and consumers.

Ellipsis (Satyanarayan & Heer, 2014) supports the construction of visual narratives by allowing users to import existing visualisations into its design environment. The visualisations are decoupled from the narrative, enabling widgets to be used to control and manipulate views. Parameters, consisting of name-value pairs are set for imported visualisations, which specify the characteristics of the visualisation, such as filter ranges, chart width and height. These parameters can be bound to widgets, such as buttons and enable users to change the visual elements displayed. The process of constructing a visual narrative in Ellipsis involves importing an existing visualisation and gaining access to the data, the visualisation techniques, and the parameters using an embedded JavaScript Domain Specific Language (DSL) API. The API can be used to manipulate the parameters and data to tell the story, which consists of scenes of annotated visualisations. Once imported, the interface renders the visualisation, as shown in Figure 2.5 and presents the parameters at the bottom of the interface.

[20] http://www. gapminder.org/downloads/human-development-trends-2005/

[21] http://www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html

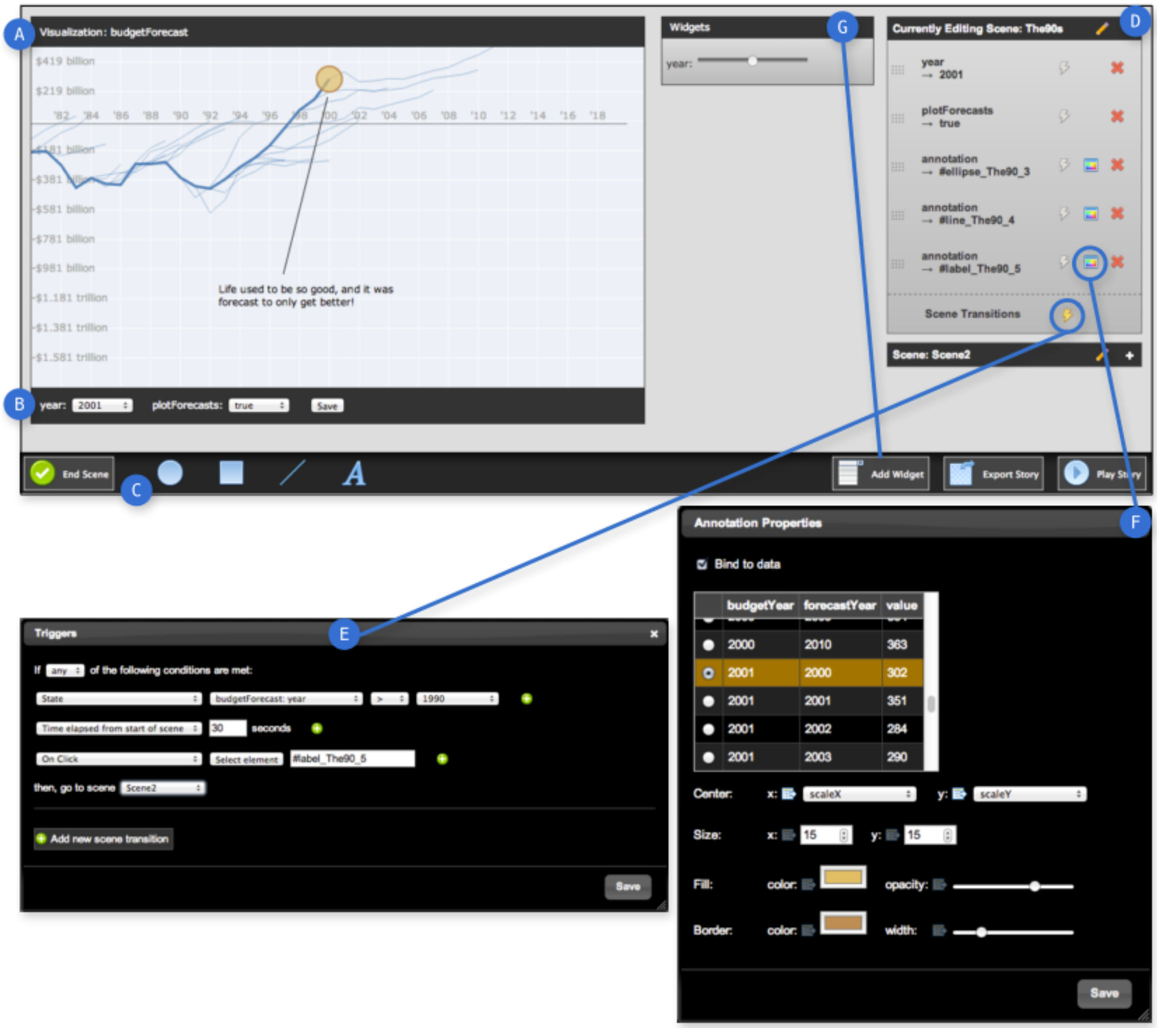

Figure 2.5 Ellipsis Narrative Authoring

Satyanarayan and Heer discuss the evaluation of Ellipsis to determine how well the tool met needs (Satyanarayan & Heer, 2014) of journalists. The participants (eight experienced data-driven story-telling journalists) were given training and walked through an example using the tool. They were then asked to create a short story, were encouraged to think-aloud and were interviewed following the completion of the tasks. All participants were able to build the visual narratives (consisting of 2-3 scenes) with minimal guidance and were able to include annotations, and scene transitions. Half of the participants commented that the tool would be useful for collaborative design, prototyping narratives and sharing with colleagues. Two participants compared the separation between the narrative creation process and visualisations, to reporters in newsrooms and developers. A number of shortcomings were noted including issues with story planning and annotation creation.

GeoTime (Eccles et al., 2008) presents the connectedness of information and events over time through an interactive three-dimensional space to support the analysis of data using a single view. GeoTime Stories is a prototype built on GeoTime, supporting pattern detection in data, narrative construction through textual additions and story collaboration. The prototype provides a template to assist the authoring of descriptions for the visualisation. The tool allows authors to capture visualisation states (snapshots of the analysis process) and these are presented as clickable thumbnails offering exploration points for users viewing the narrative. The snapshot scene is stored as an image with a unique reference, which can be used to load the same data state. The prototype supports collaborative authoring and feedback of a narrative through a HTML text editor. Coloured bars are used to define text sections, such as team input, versioning, tracking and feedback and can be filtered by authors/analysts to include and exclude sections from the narrative.

GeoTime Stories was evaluated during a period of six months through analysts who assessed the system using data that included vehicle tracking, billing data, phone call and payment transactions and provided feedback (Eccles et al., 2008). Test data loaded into GeoTime was presented to the analysts, including pattern detection and visual narratives. The response from the users mainly focused on the pattern search functionality, which was received very positively, with analysts noting the level of support it provided.

Tableau version 8.2 introduced Story Points as a mechanism to create visual narratives allowing several visualisations to be connected and presented as a narrative, as opposed to having them all cluttered in a single dashboard. Data can be accessed from multiple sources, including databases and files (MS Excel and text files), which can be used by an author to construct a visual narrative and highlight important trends that can support user insights. Tableau visual narratives consist of a series of worksheets that are presented in a slideshow format, where users can click on links at the top of the interface to navigate the story. Tableau supports the dragging and dropping of data fields onto specified shelves on a visual canvas to build individual visualisations and worksheets. The worksheets can then be connected together and textual descriptions added to explain the data and form the story. Tableau supports several interaction techniques such as applying filters, selecting elements, zooming and drilldown. The filter implementation allows viewers to specify the data attributes to be shown within an individual worksheet of the visual narrative.

SketchStory (Lee et al., 2013) is a data-enabled digital whiteboard that allows presenters to sketch a visual narrative using pre-specified charts, which get invoked and drawn through gestures (Figure 2.6). For example drawing an icon such as ‘L’ creates an x and y axis and drawing an open ended circle creates a pie chart. SketchStory uses the corner-finding algorithm (Cheema et al., 2012) to convert a stroke drawn on the canvas into a polyline, which is matched by the tool to create an appropriate visualisation. Once an appropriate visualisation is matched (bar, pie, tally and line charts, scatterplots and maps are supported), it is generated using the data from the data files.

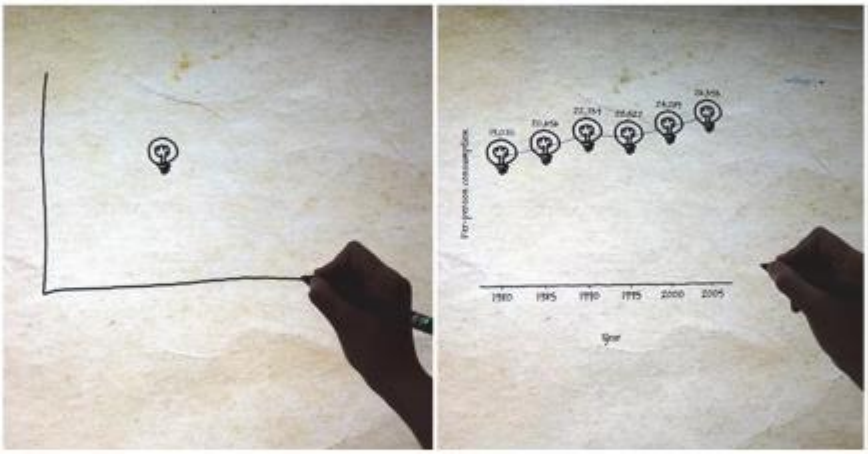

Figure 2.6 Visual Authoring in SketchStory

Several charts can be sketched on the canvas and these can follow a sequence to narrate a visual story. The visualisations are coordinated and the tool supports visual interaction by enabling the presenter to tap visual keywords, which are part of the visualisations in the presentation, and this action causes other charts on the canvas to update according to the selection. The visual narratives are stored in XML files containing the visualisation techniques, data used and the sequencing between the charts.

SketchStory was evaluated to assess its usability through a user trial which included six participants who were asked to construct a visual narrative using global energy consumption data and complete a questionnaire regarding their experience with the tool (Lee et al., 2013). The participants liked using the tool and gave it an average score of 5.5 out of 7, ease of learning was scored at 5.3 and ease of use was scored at 4.7. Some of the issues arising from this study suggested that first time users may struggle with creating content and constructing the narrative at the same time. In addition some participants were concerned about their sketch quality and others suggested that as the charts are moved, annotations should also follow.

Analysis

From the systems discussed above that support the authoring of visual narratives, the evaluation of Ellipsis showed that non-technical users can build narratives using the tool without much guidance. Ellipses supports data access by enabling narrative authors to import both visualisations and the associated data. The adopted design strategy consists of annotated visualisations presented as scenes within a visual narrative. Ellipsis supports interaction and exploration of visual narratives within the individual scenes. Although Ellipses supports non-technical users' in authoring visual narratives, it does not facilitate automatic sequencing between scenes and users are required to manually author the individual scenes. Ellipsis was evaluated (Satyanarayan & Heer, 2014) using budget forecast data and users were asked to create a visual narratives consisting of two to three scenes. As the user created scenes (for example, country-wide unemployment percentage), they were not presented with dynamically generated scenes consisting of related data (for example, regions in the country with the highest unemployment levels, etc.), which can be generated through derived data transformations. Such scenes may support authors in constructing visual narratives through suggestions for next scenes (in this case).

GeoTime Stories supports user detection of patterns within data. It assists the authoring process through templates that can be used to provide textual descriptions to highlight interesting insights. GeoTime Stories supports visual interactions through snapshots that present visualisations and data states that were saved during the process of analysis carried out by authors. These offer exploration points to users, which can be loaded and the various data states can be viewed to support insight. These snapshots need to be saved by authors as they compose their stories', however, they can decide not to save any intermediary snapshots, hence reducing the level of exploration.

Tableau 8.2 provides functionality to author visual narratives using multiple data sources. The visual narratives consist of connected worksheets presenting the data through a slideshow format. Narrative authors are supported in the construction of visual narratives through the drag and drop operations allowing data attributes to be selected and filtered to build visualisations. However, authors are required to create individual worksheets as Tableau does not facilitate automatic sequencing within the narrative. For example, if a user creates a visualisations for his/her narrative, the dynamic generation of sequences showing subsequent visualisations (that can be included in the narrative) may support him/her in constructing the story. Tableau was discussed in section 2.5.1, which highlighted that the Tableau visualisations supported several interaction techniques that can assist users in making valuable insights. Specifically, users can filter datasets to view various segments of the supplied data which lends well to reader-driven stories. Nevertheless, literature (Ma et al., 2012) suggests that this may distract viewers from the original message within the visual narrative as they could move away from the data in the actual story and explore other aspects of the dataset.

SketchStory supports data accessed from files that are used to build visualisations that are mapped to strokes sketched on the digital whiteboard canvas. Its evaluation showed that participants were able to effectively author visual narratives and found it easy to use and learn. Similar to Ellipsis and Tableau, authors using SketchStory are required to construct each visualisation as automatic sequencing is not supported. SketchStory enables authors to construct multiple coordinated charts that support interactions across the visualisations. The evaluation also highlighted that the participants found the stories were very engaging, specifically the supported interactive features.

From this analysis, it can be seen that there is a focus on supporting authors without technical expertise to construct interactive visual narratives. However, there is little support offered to narrative authors in terms of creating the actual story, i.e. none of these systems dynamically suggest sequences for the visual narrative. The start of this section (2.5) discussed some of the benefits that automatic sequencing can offer authors (Hullman et al., 2013a), highlighting its importance to the process of constructing visual narratives. This thesis defines derived data transformations which can be used to dynamically generate visualisations showing related data, and these can be used for automatic sequencing. The review of the systems in this section included examples of the benefits of related data views for creating sequences. Such sequences can be automatically generated using derived data transformations of both the authored visualisation and the dataset. These can be of great value to non-expert designers, who may have domain expertise. In summary, it can be concluded that it is important to support the authoring process, which has been done to a degree by the state of the art systems; nonetheless, designs should also consider implementing facilities to dynamically generate visualisation sequences.

2.5.2 Consuming Visual Narratives

Section 2.5.1 focused on the support offered to authors in building visual narratives and briefly discussed features that also support the analysis of a constructed visual story. This section focuses on tasks and techniques that support consumers in exploring visual narratives to understand the message communicated.

Gapmider[22] has been used to present visual narratives of global health (Figure 2.7) and development trends through interactive charts. It enables users to create motion charts, that can be overlaid on a map and which animate changes in data over time. Motions charts are used to present multidimensional data over temporal intervals and support interactions and user controls (Battista & Cheng, 2011). Node sizes and colour coding in the visualisations are used to help communicate the data. It supports several interactions including filtering data by country and geographical region, selecting countries and tracing changes over times, zooming to view nodes of interest and sizing nodes according to preferred indicators.

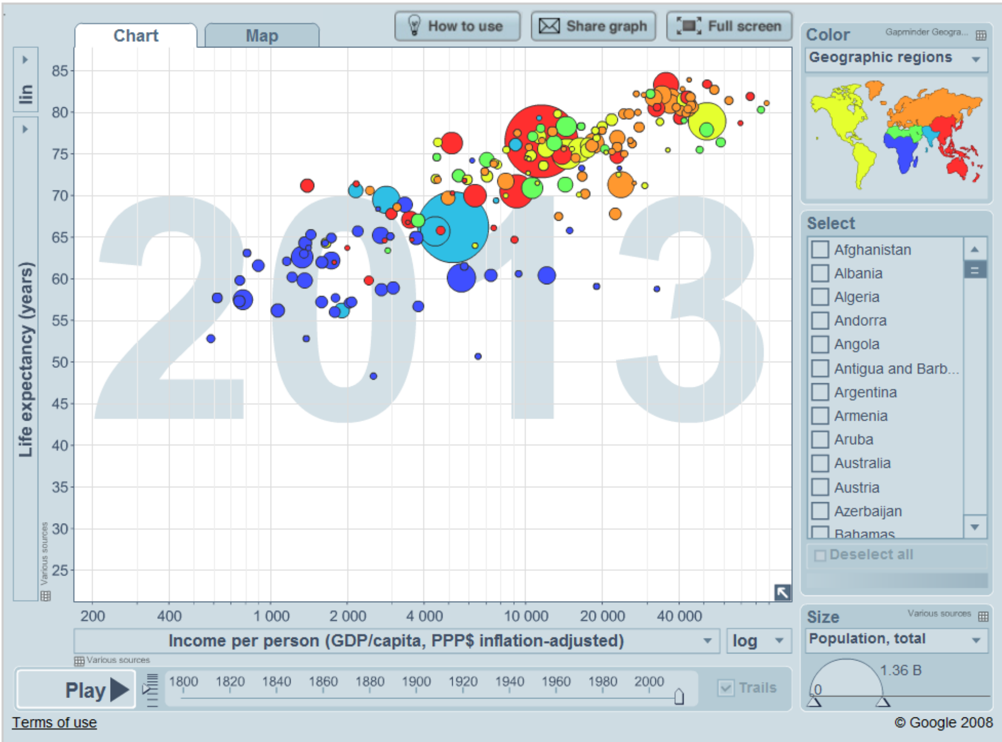

Figure 2.7 Gapminder presenting world health data

Gapminder has been effectively used to communicate a story or message by sequencing transitions using visualisations and thus gradually building views from simple to more complex ones. The interactive stories or presentations may be viewed on a web browser and may be run as a chart or a map view. The stories are presented as animated slide shows where each slide consists of sections that can be navigated using a progress bar or the start and back buttons. The interface contains a play button which shows animations of the data changing over time. Gapminder has been used in a statistics courses at Iowa State University (Le, 2013) where students designed a project using Gapminder, which enabled them to explore statistical data. The student projects and feedback highlighted that it was easy to use and an

[22] http://www. gapminder.org/downloads/human-development-trends-2005/

enjoyable way to learn statistics. Gapminder has also been used by schools to support teachers in expanding student knowledge through interactive presentations of world data.

Sense.us (Heer et al., 2007) is a data analysis website providing visualisations of US census data since 1850 and supports graphical annotations, users' comments, saved bookmark trails and doubly-linked discussions. The website consists of two panels: the left panel presenting an interactive visualisation displayed in a Java applet and the right panel displaying users' comments and bookmarked views saved by the users.

The system supports a concept called doubly-linked discussion, which allows users to enter comments about a view which is saved and is then available to other users to click causing a bookmarked state to load showing the view seen by the author of the comment. The website enables users to save views as bookmarks (thumbnails are used to represent bookmarks) and construct a trail of saved bookmarks. These bookmarks are used to support visual comparisons and visual narratives through a bookmark trail widget, where users can drag and drop bookmarks onto a text area, and text can be associated with each bookmark by the user. The widget allows users to read the text and load the visualisation thumbnail and clicking it loads the view.

The system was evaluated using IBM employees and 18 researchers (Heer et al., 2007). It was deployed on the IBM corporate intranet for a period of three weeks consisting of eight visualisations and publicised to employees through emails. In total, across the studies, 258 comments were created by the participants and it was found that most of these involved data analysis by discussing trends observed in the visualisations. It was also found that users were interested in patterns across visualisations and explored across multiple views in a narrative format. The study found that most of the researchers did not use the bookmark trails but from the participants that did (IBM employees and some researchers), multiple remarks showed its usefulness. The participants suggested that more storytelling features such as presenting all comments without removing older linked ones, should be added.

Online Journalism is increasingly using interactive visual narratives to communicate messages. The New York Times presented a visual narrative on carbon emissions treaties and the possible impact at the Copenhagen climate conference in December 2009, and made it available on its news website[23]. The visual narrative consisted of three tabs, titled Global Emissions, Lessons from Kyoto and Possible Impact. Each tab contained a set of slides that

[23] http://nyti.ms/sFYztk

users could navigate by clicking on the number representing each slide or the next button as shown in Figure 2.8. The visualisations consisted of interactive maps with bubbles (size and colour highlighting specific points), colour-coded regions, and line charts representing temporal data. The slides included captions for all the visualisations, describing the data presented.

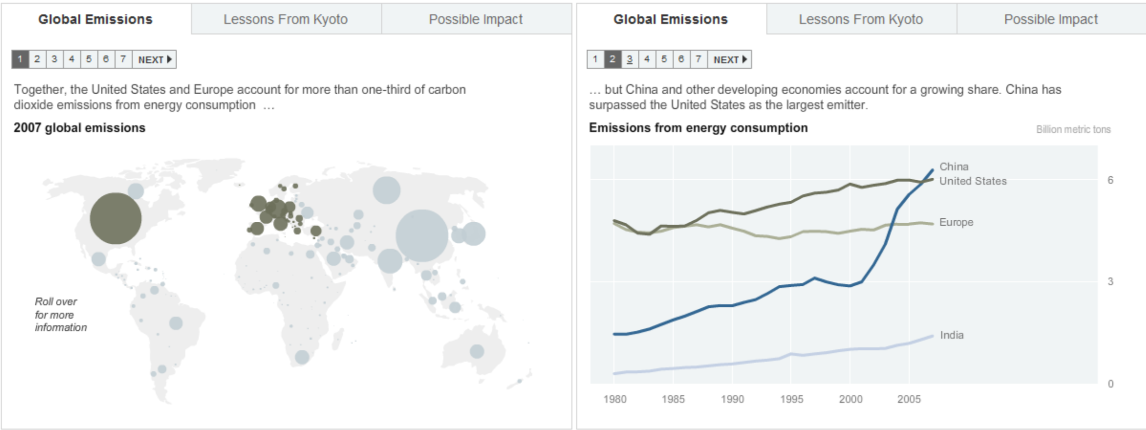

Figure 2.8 The New York Times visual narrative on carbon emissions

The visual narrative described a complex topic showing countries that have implemented the agreed protocol to reduce emissions and the ways these regions believed it should be done. The narrative supported visual interactions, enabling users to hover over bubbles and along the lines to view details such as emissions per region and by time. The "Possible Impact" tab consisted of map visualisations that displayed photographs of possible impact.

In 2009, The Financial Times[24] presented an article with a visual narrative comprising of four user selectable tabs presenting the effectiveness of Provincial Reconstruction Teams (PRTs) implementing development projects in Afghanistan. The tabs displayed maps of the country divided by provinces presenting development data on nation-building (cost per province), security (level of insurgent activity) and counter-narcotics (opium cultivation levels). The visualisations were interactive and each tab provided a selectable list of NATO members. Clicking a country highlighted the province where that nation had troops working on one of the development projects. In addition to highlighting the province, a textual description appeared below the map describing the activities that the selected country was involved in. The visual narrative also provided details-on-demand for each NATO member nation via a "PRT INFO" button. The counter-narcotics tab provided a timeline slider that allowed readers to compare opium cultivation levels from 2005-2009.

[24] http://www.ft.com/cms/s/0/663b649e-b7e6-11de-8ca9-00144feab49a.html

In 2010, The New York Times published a visual narrative showing US budget forecast[25] following the same slide show method it used in 2009 to present the global carbon emissions narrative. The visual narrative used line charts to present budget surplus and deficits from 1980 and included estimates until 2020.

Analysis

From the review of the systems discussed above, it can be seen that implementing annotated charts, slideshow strategies and animations are being used to communicate a message through interactive visualisations. In addition, these systems generally balance both author-driven and reader-driven techniques within the visual narratives by supporting visual interactions and, at the same time, ensuring the intended message is communicated. However, there is a limitation in the data explorations that users can conduct, especially around viewing data related to the visual narrative. Such exploration can be generated using derived data transformations (discussed in section 2.2.2) and can support users in getting a better understanding of the data and the communicated message.

Gapminder has used animated transitions of visualisations to present real world case studies to get people interested in the data. Gapminder supports explorations of the data it supplies, such as the health and development data, however, it does not allow users to load or import their own data. Hence it cannot be used as a tool to communicate a message using data other than that which is currently supported. Gapminder supports several visual interaction techniques; however, it does not support transformations that present related data views to allow users to explore connected data. In the example of world health data shown through Gapminder (Figure 2.7), which presents a visual narrative of income per person versus life expectancy, derived data transformations could present most common causes of death by age categories. This would enable interested users to explore the visual narrative beyond the message provided by the original story.

Sense.us provides several interactive visualisation techniques to represents US census data. It includes functionality supporting graphical bookmark trails, which allows users to save thumbnails of bookmarked views and build a visual narrative, highlighting trends in the data that may be viewed by other users. Users can hover over these thumbnails and load each visualisation of the narrative and read the comments associated with the visualisation. However they are unable to explore the data related to the information presented in the view. For example, if a bookmark from the trail presents a visualisation showing US occupations

[25] http://www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html

since 1850, users may be interested in exploring outliers in the view through links showing visualisations of data related to these outliers.

From the overview of visual narratives used in online journalism, it can be seen that annotated charts and interactive slideshow design strategies have been effectively used to present stories. Such strategies provide users with a simple mechanism to navigate visual narratives. Nevertheless, from the review it can be seen that the visual narratives supported a limited range of interactions through details-on-demand, highlighting values and greyed-out values. In addition, readers were not able to explore data related to the visual narrative, such as viewing a breakdown of carbon emissions from Europe by country, presented by The New York Times (Figure 2.8). These derived data transformations enable readers to explore the narrative and view data beyond the communicated message.

From the analysis, it can be seen that there is a lack of support for consumers in exploring data related to the visual narratives. However, it is important to note that such explorations raise the challenge of ensuring that a balance is still maintained between author-driven and reader-driven techniques. The Martini Glass Structure, Interactive Slideshow and Drill Down Story techniques (Segel & Heer, 2010) have been used in visual narratives to provide a balance between author-driven versus reader-driven narratives. The Martini Glass Structure supports user interactions once the intended message has been communicated. The Interactive Slideshow technique has been used in the visual narratives of The New York Times and The Financial Times. The Drill Down technique presents the visual narrative and enables users to drilldown into individual views to uncover details, and this technique has greater emphasis on the reader-driven method. A way to address the author-driven versus reader-driven balance when incorporating derived data explorations into a narrative is to use a combination of the Interactive Slideshow and Drill Down techniques. With the combined technique, the intended message can be presented using the Interactive Slideshow and the explorations can be displayed through popup windows which can be analysed by a user before returning back to the main narrative.

Section 2.2.2 and 2.4 highlighted the importance of view manipulations and derived data explorations, and hence it can be concluded that both of these features should be supported in visual narratives, but importantly measures must be taken to ensure the intended message is not lost through visual interactions.

2.6 Personalised Visualisations

Personalisation in Information Visualisations has been used to support users in making sense of data through a number of techniques. These include eye gaze processing, biomedical sensors, highlighting elements within visualisations to match text read by users, detecting confusion and providing adaptive interventions, adapting views based on users' visualisation preferences and emotion-triggered adaptation of visualisations. The early stage evidence from studies has been encouraging showing that eye gaze processing can identify individuals' cognitive abilities (Toker et al., 2013). Steichen et al. discuss the use of eye gaze to personalise visualisations through inference of cognitive abilities (such as awareness speed), which can help in determining the different visualisation types to offer to various users (Steichen et al., 2012, 2013). Conati et al. discuss using eye gaze to determine when a visualisation change is required, by analysing the time spent looking at certain elements within the view, locating the area yet to be looked at by the users, and examining pupil dilation (Conati et al., 2011).

Burns et al. (Burns et al., 2012, 2013) introduce a system which infers intended messages from visualisations by adapting the view to the individual user. It includes changing the type of graphic to display the same information and to facilitate the presentation of the data through the individual's preferred graph type thereby supporting the user's strengths. It also suggests using a number of measures to adapt the graphics within the same graph type by adapting the proximity of elements within the view, altering salience, shifting scales and reordering elements. Carenini et al. discus tracking what the user reads of textual content and dynamically adapting the visualisation(s) to highlight elements within the view that correspond to the text currently being read (Carenini et al., 2013). Cernea et al. evaluate the use of emotion triggered visual adaptation to assist users during periods of boredom and frustration while using visualisations (Cernea et al., 2013). The evaluation was conducted using a wearable brain-computer interface headset, which detects brain signals including frustration and emotion. The results showed that adaptation techniques such as highlighting and hints (less intrusive) were most useful in improving users' emotional states. Nazemi et al. (Nazemi et al., 2011) and Gotz et al. (Gotz et al., 2009) present systems that monitor user interaction patterns to determine if visualisations are suitable, and infer users' tasks.

Analysis

Research in personalised visualisations has focused on a number of research directions from using headsets to detect frustration, to eye gaze processing and adapting views for the individual users. Early indications have shown encouraging results for systems supporting personalised visualisations (Cernea et al., 2013, Toker et al., 2013). For example, identifying individuals' cognitive abilities through eye gaze processing and adapting visualisations to highlight data have shown some promising results. Nevertheless, personalised visualisation is an emerging area and the research is quite diverse. Further work is required before solid conclusions can be drawn about the various directions that are currently being investigated. Hence, as some of the directions are in early stages, they require further research before conclusions can be drawn about whether they should be included in designs to support users in making sense of data.

2.7 Summary

This chapter analysed the state of the art in visualisation systems to identify important visual analysis tasks and features that visualisation systems should support in order to promote user understanding, manipulation and exploration of complex digitised data. The analysis found:

- **Data Access**: Digital data is stored across several source types, and the most prevalent of these sources include relational databases and file formats. The state of the art visualisation systems support data access from a range of source types including databases, Excel, CSV and KML (Keyhole Markup Language). In some cases, systems such as Tableau, Tibco Spotfire and Google Fusion Tables support data from a variety of these sources.
- **Data Transformations**: The review has shown that visualisation systems support data transformation through the extraction and parsing of data, generation of statistical measures, and through the creation of summaries and aggregations of input data. However, there is no support for data transformations that dynamically generate views presenting data related to the elements shown within visualisations. For example if a student is analysing a visualisation presenting his/her performance data, then related data views can display data related to the student (such as engagement, resource usage) or performance (such as peers at the same performance level).
- **Visual Encoding**: The state of art visualisation systems support visual encoding through the mapping of data to visual marks or to visualisation techniques. Both forms of mappings are popular and both have advantages and disadvantages. The former is more extensible and flexible as it can handle a broader range of mappings but it requires a greater coding effort. The latter takes advantage of existing visualisation libraries, nonetheless, an understanding of the characteristics and affordances of each of the visualisation techniques is required.

- **Visual Interactions**: The analysis highlighted the importance of visual interaction techniques in supporting users to make sense of data. It singled out a number of interaction techniques commonly used by the state of the art, which include select, explore and elaborate, filter, navigate and multiple and coordinated views.
- **Guidance**: The review highlighted the importance of guidance through visual narratives and identified limitations in both supporting the construction and consumption of visual narratives, which can be addressed through derived data transformations.

From the analysis, a number of recommendations were made that were deemed to be important in the process of supporting users to make sense of data. These suggested that visualisation systems should:

1. enable users to load datasets through a number of data sources including databases and file formats,
2. include derived data transformations to enable users to view explorations, including statistical measures and related data,
3. facilitate Visual Encodings to enable data to be mapped to graphical marks or visualisation techniques,
4. support popular interaction techniques including select, filter, explore, elaborate and navigate,
5. supporting authors in constructing visual narratives that can be used to guide end users through a message. Assist viewers consume visual narratives through interactive visualisations that enable data explorations, while ensuring that a balance between the author-driven and reader-driven techniques is maintained.

These recommendations will be used in Chapter 3 to analyse the state of the art systems in TEL that use visualisations.

3 State of the Art: Visualisations in Technology Enhanced Learning

Chapter 2 analysed the state of the art in Information Visualisation to identify important visual analysis tasks and features that should be supported by visualisation systems, in order to promote user understanding, manipulation and exploration of complex data. This chapter discusses how Information Visualisation has been used in the TEL domain, specifically focusing on Learning Analytics, and uses the visual analysis tasks and features identified in Chapter 2 to analyse these systems (research objective one). The tasks and features identified in Chapter 2 that are relevant to the TEL domain include visual interactions, data transformations and guidance. The tasks and features also include data access and visual encodings which are important for standalone visualisation systems that enable users to import and visualise data. However, features that support data access in TEL are not so important, as educational systems would typically use a small number of data sources to store student-logged data and these sources can be connected to directly during deployment. In addition, visual encodings are also not important as the visualisations to be used to render the student data are also known in advance, so features to map data to visualisations or graphical marks are not required. Hence both data access and visual encodings are not considered in the discussion on the analysis the systems below.

3.1 Introduction

In TEL, visualisations have been used to support learning content, learning process representations, and strategies for managing and organising learning concepts (Gomez et al., 2010). Research on the use of Information Visualisation in the TEL domain has largely focused on three areas, which comprise of Open Learner Modelling (OLM), Educational Data Mining (EDM) and Learning Analytics (LA).

- **OLM** consists of student models that are updated as the learner studies, to present current understandings and beliefs. It is a model of "*knowledge, difficulties and misconceptions of the individual*" (Bull, 2004).
- **EDM** can be defined as "*developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in*" (International Educational Data Mining Society[26]).
- **LA** can be defined as "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising*

[26] http://www.educationaldatamining.org/

learning and the environments in which it occurs" (Learning Analytics Research Society[27]).

OLM visualises learner models to support reflection and allows students to participate in the construction or modification of their personal model. It uses student responses to questions, number of attempts, and task response times, to build models of student competencies and levels of understanding. Visualisations are used to present student models showing learner competencies and levels of understanding, supporting reflection, assessment and monitoring (Bull & Kay, 2013). The visualisations are inferred from learner interactions and have been shown to guide learning, help improve the performance of weaker students, and motivate stronger learners (Bull & Kay, 2007, Mitrovic & Martin, 2007). Studies have shown OLM to increase learner awareness and improvements in reflection, motivation and engagement (Mitrovic & Martin, 2007).

EDM generally does not focus on the entire system, but instead analyses data components and their relationships. It uses algorithms from machine-learning, data mining and statistics to detect patterns and predict outcomes that can benefit learners. It also concentrates on automated discovery and adaptation through Intelligent Tutoring Systems. EDM algorithms use student-logged data and enable learners to reflect on their learning processes and offer feedback with suggestions and recommendations to improve learning. In addition the algorithms raise awareness amongst educators regarding the learning paths and processes of their students and enable teaching reflections (Merceron & Yacef, 2005a). Data mining can be used to visualise and explore online learning data to determine patterns of sequences and generate feedback for educators (Talavera & Gaudioso, 2004). Similar to data mining, EDM consists of four steps, which include collecting data (from online learning environments), pre-processing data (data transformations and formatting), application of mining algorithms and interpreting and deploying results (Romero, 2008).

LA generally focusses on analysing the data from an entire TEL system to inform and generate awareness in learners and educators. It includes systematic interventions, social comparisons and sense-making models. Visualisations have been used to present student activity data to either educators (to enable them to understand student progress), or to students (to promote change in behaviour) or to both educators and students. Wang et al. introduced "academic analytics" and suggested that student success could be determined through the analysis of online learning activities (Wang et al., 2001). Morris et al. and

[27] http://www.solaresearch.org/mission/about/

Macfadyen and Dawson presented a positive relationship between the discussion activity level of students and academic performance (Morris et al., 2005, Macfadyen & Dawson, 2010).

Aside from OLM, EDM and LA, Intelligent Tutoring Systems (ITS) and Artificial Intelligence in Education (AIED) focus on learner modelling, maximising learning through feedback and motivating learners. ITS aims to simulate a human tutor's behaviour and provides learners with exercises. The responses are used together with the learner model to personalise feedback (Ahuja & Sille, 2013). ITS consists of four models comprising of the domain model (assesses learners' performance), the student model (records learner activity), the tutor model (governs feedback for learners) and the user interface model (applies graphics, windows, pointers, menus and text to interact with students) (Sharma et al., 2014). In general, ITS focusses on student learning through fuzzy linguistic modelling, cognitive load, gaming, students' emotional state and machine learning. AIED involves modelling students and educators while considering social and collaborative characteristics to support learning. AIED systems have generally focused on using virtual gaming environments, cognitive tutors to monitor and guide students, adaptive courses, interactive hints to guide learners to solutions, and novel user interfaces which include eye tracking, gesture and speech recognition (Underwood et al. 2013). Although ITS and AIED both share many of the objectives of OLM, EDM and LA, they do not have a strong focus on Information Visualisation and hence these research directions are not within the scope of this thesis.

The remainder of this chapter briefly discusses the role of self-reflection and social comparison in online learning. It also briefly discusses the use of Information Visualisation in OLM and EDM and the impact it has on learners. The aim of the research described in this thesis is to define, design and implement an *approach* that supports the construction and consumption of visual narratives in TEL, to address the Research Question. The objectives of this thesis include the realisation and validation of this *approach* in the TEL domain by supporting students to enhance online learning engagement and performance, through the use of visual narratives to render their personal activity data. While sections 3.3 and 3.4 briefly discuss OLM and EDM, these areas are not covered in detail as their focus is on modelling student beliefs and understandings (OLM) and predicting student performance using data mining (EDM), which is beyond the scope of this thesis. The main focus of this chapter is on the analysis of the state of the art systems using Information Visualisations in the LA domain. This chapter uses the relevant visual analysis tasks and features that were identified in Chapter 2 for Information Visualisation systems to analyse the state of the art.

3.2 Self-Reflection and Social Comparison

Personal logged data generated through learning systems can be used to support self-reflection amongst learners and promote behavioural awareness (Carver & Scheider, 2001) and discussion (Endsley, 1997). Zimmerman describes self-reflection as a combination of self-judgement and self-reaction (Zimmerman, 2002). Self-judgement involves comparing one's performance against previous performance or a standard performance, and assessing one's beliefs regarding success and failure. Self-reaction involves the level of self-satisfaction impacting learner motivation, and the positive affect of performance. Several learning environments, such as SAM (Govaerts et al., 2012), CAM (Santos et al., 2012a), ALAS-KA (Ruipérez-Valiente et al., 2014) and StepUp (Santos et al., 2013a) present student activity data through visualisations to promote self-monitoring and reflection.

Social comparison theory (Festinger, 1954) describes self-improvement, self-evaluation and self-enhancement as motives for individuals to compare their performance or achievements with others. Nevertheless, there is a stronger tendency amongst individuals to compare achievements with targets of a similar category. Studies (Shepherd et al., 1995, Dijkstra et al., 2008) have shown that upward comparisons lead to improved performances amongst learners. Visualisations presenting social comparisons have been used across these three areas (self-improvement, self-evaluation and self-enhancement) enabling learners to compare their course activities with peers. Social comparison can also support learner motivations and engagement by allowing students to compare their own progress with peers (Linton & Schaefer, 2000, Vassileva & Sun, 2007, Barolli et al., 2006, Méndez et al., 2006, Vassileva, 2008) and these have been presented in both LA and OLM through visualisations.

3.3 Open Learner Models

Progressor (Hsiao et al., 2013) and Progressor+ (Hsiao & Brusilovsky, 2012) present visualisations of both learner model and social comparisons showing student progress and comparisons. Progressor+ supports two levels of comparison, the first is at a macro level, which allows the learner to view his/her own model and see thumbnails of peers' models. The second level of comparison is at a micro level, which shows the details of a selected peer's model. OLMlets (Bull et al., 2009) facilitates independent learning and assessment through visualisations by providing a student with a skills meter, a ranked list and a textual summary of his/her knowledge level. Reflect (Kay & Fekete, 2007) shows students their progress together with feedback to support the reflection process through self-assessment. KnowViz (Falakmasir et al., 2012) presents visualisations showing learners' attempts to master course topics and knowledge gained, including peer comparisons. MyExperiences

(Kump et al., 2012) uses visualisations to allow students' to access and alter their individual knowledge level within APOSDLE[28] and understand the automated suggestions and recommendations. Narcissus (Upton & Kay, 2009) provides visual representations of a group model, allowing individual learners to analyse their own progress and the progress of their team. CALMsystem (Kerly et al., 2008) presents learner performance and self-assessments which change according the learner's current knowledge. Competency Map (Grann & Bushway, 2014) presents learner competencies against course assignments. GVIS (Grapple Visualisation Infrastructure Service) (Mazzola & Mazza, 2010) represents open learner models by extracting and using data from different sources and adapts the visualisations through a conditional construct to present learner behaviour. Albert et al. (Albert et al., 2010) present visualisations techniques used by the GRAPPLE Research project to show the learner model, which captures the knowledge user model variable. It presents this variable in addition to the class average and expected goal of the learner.

The objective of this thesis is to address the Research Question by defining, designing and implementing an *approach* that supports the construction and consumption of visual narratives in online learning environments, and by evaluating of this *approach* through authentic users and realistic use cases. These use cases include supporting the engagement of students through visual narratives that use student-logged data. OLM systems focus on building student models presenting learners' competencies, current understandings and beliefs to support reflection and monitoring. As the focus of this thesis is to present visual narratives using student-logged data and not on competencies, current understandings or beliefs, OLM is not within the scope of this research.

3.4 Educational Data Mining

Romero et al. introduce a tool that can be integrated into Moodle[29] to allow educators to data mine learner logs (Romero et al., 2013). An educator can select student activities and decide how to have the data visualised, and save the data to an MS Excel file for mining purposes. Pedraza-perez et al. present a data-mining tool for course educators (Perez et al., 2010), who have the option to analyse this data through a number of visualisations and can mine the data by selecting a mining method (Classification, Regression or Association). Romero et al. describe a tool used for mining and visualising students' most visited trails and present these through a network graph visualisation (Romero et al., 2008). Mostoe at al. discuss an EDM tool which uses student events and time intervals to present interactions between learners

---

[28] http://www.aposdle.org/

[29] https://moodle.org/

and educators by visualising summary interactions and facilitating dynamic drilldowns (Merceron & Yacef 2005b). The DRAL tool (Zafra et al., 2013) aims to identify patterns and relationships between learner performance and work they completed for interested educators. TADA-Ed (Merceron & Yacef 2005b) integrates visualisations and data mining of student-logged data from online learning environments to permit educators to analyse study patterns and progress. Student Success System (S3) (Essa & Ayad, 2012a, Essa & Ayad, 2012b) is an early warning system designed for course educators to identify students who are at risk, understand why a student is at risk, design interventions and assess the success of the intervention. Signals (Arnold, 2010, Arnold & Pistilli, 2012) provides early warnings to students of potential problems by highlighting learner efforts and performance using green, yellow and red indicators. Student Inspector (Zinn & Scheuer, 2007) was developed to analyse e-learning systems' log data and generate visualisations using this data for educators to explore learner performance, topic coverage and misconceptions.

EDM systems focus on supporting learners by detecting patterns and predicting outcomes using algorithms from machine-learning, data mining and statistics. The research objectives of this thesis focus on the definition, design, implementation of an *approach* to support the construction and consumption of visual narratives in the TEL using student-logged data. It does not focus on using machine-learning and data mining algorithms to present predictions to educators. Hence, similar to OLM, EDM falls outside of the scope of this thesis.

3.5 Learning Analytics

Online learning environments, for example massive open online courses (MOOCs), typically can have a very large corpus of students participating in a course, which makes it very difficult for educators to observe learning behaviours and consider student preferences. In addition, with distant learning environments, student involvement in the learning process is lower so they consequently tend to get demotivated, leading to high dropout rates (Koutropoulos et al., 2012). LA enables educators to monitor student behaviour, observe performance and provide students with appropriate feedback when required, or even alter course material to prioritise topics that students are struggling with. A survey of the empirical research in LA between the years of 2011 and 2014 (Nistor et. al., 2015) highlighted that most studies used student-logged data, and it outlined that the use of visualisations was the most popular form of providing educators with instructional support. LA also aims to motivate learners and enhance engagement by presenting progress and peer comparisons (Duval, 2011). However, the design of LA systems need careful consideration regarding the collection and presentation of data as these can be misinterpreted by students and significantly affect motivation levels and performance (Lonn et al., 2015). The remainder of this chapter describes and analyses the state of the art in LA systems that use visualisations to support learners and/or support educators in monitoring student learning behaviour.

Fifteen educational systems and prototypes in the area of TEL, such as SAM (Govaerts et al., 2012), CAM (Santos et al., 2012a), Moodog (Zhang & Almeroth, 2010) and CourseVis (Mazza & Dimitrova, 2007) are examined below. More educational systems (including proposed systems) presenting visualisations using student-logged data are discussed in appendix A. The majority of the systems discussed below provide visualisations to both learners and educators and support interactive visualisation techniques and data transformations. A portion of these systems including CourseVis, GISMO (Mazza & Butturi, 2007), SNAPP (Bakharia & Dawson, 2011) and LOCO-Analyst (Jovanovic et al., 2008) focus on only supporting educators in understanding student behaviour and engagement through visualisations.

There are a number of lessons that can be learnt from the visualisations used, degree of interactivity, exploration and transformations supported by these fifteen systems. This chapter describes and analyses these systems by highlighting best practices and discussing limitations. Section 3.6 describes and collectively analyses five systems that use visualisation to exclusively support educators in monitoring student learning. Section 3.8 discusses and individually analyses ten systems that support students and in most cases both students and educators in making sense of student-logged data through visualisations. These systems are individually analysed as they are more aligned to the research objectives of the *approach*, which aims to support both students and educators (consumers) in making sense of student-logged data.

3.6 Supporting Educators in monitoring student learning

This section discusses and analyses learning systems that support educators in monitoring student learning patterns through interactive visualisations.

3.6.1 Instructional Intelligence System

The Instructional Intelligence System (Thille & Smith, 2011) provides educators with a dashboard communicating student progress and learning at a class level in an attempt to deduce real time instructions. The visualisations on the dashboard present an analysis of students’ interactions with the learning objectives and show class measurements, enabling the educator to adjust or prioritise topics to teach. The dashboard shows the percentage of completed activities, responses to open-ended questions, checkpoints and quizzes. The visualisations are made up of horizontal bar charts and colour-coded bars with nodes, as shown in Figure 3.1. The colour-coded bars classify student learning and the accompanying nodes (one node per student) indicate the learning for individual class students. This visualisation highlights the students who are excelling in their course work and those who are experiencing issues. The bar chart shows how the class is performing by sub-objective, and selecting a coloured bar presents the estimated learning by students and class progress by sub-objectives.

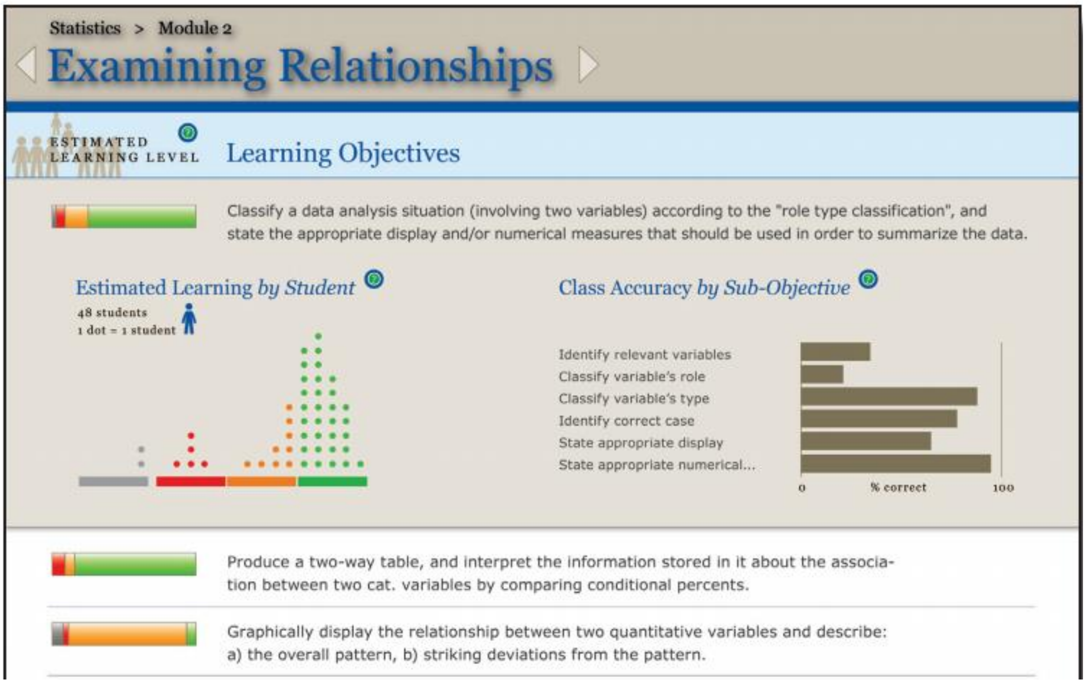

Figure 3.1 Instructional Intelligence System

3.6.2 CourseVis

CourseVis (Mazza & Dimitrova, 2007, Mazza, 2006) tracks student data from a Web-based Course Management Systems (CMS) and provides visualisations that support educators in gaining an understanding of the social, cognitive and behavioural aspects of distance education students. It extends the functionality of CMS by mapping relevant data to 2D and 3D visualisations. An example visualisation offered by CourseVis to view the social aspects of learners is a 3D scatter plot as shown in Figure 3.2, which represents student discussions throughout the course. Visual interactions (including brushing, zooming and rotating) are supported to enable educators to manipulate views.

The CourseVis architecture follows the reference model described by Card et al. (Card et al., 1999), importing student data from the CMS, applying transformations and visual mappings to it and providing visualisations of the processed data. The architecture consists of a Raw

Data Repository, a Domain Model Repository and a Data Processing Procedure. The Raw Data Repository contains student interactions data from the CMS, and the Domain Model Repository contains concepts which are associated with group exercises and quizzes provided by the educators. The Data Processing Procedure consists of four stages which adhere to the reference model: Selection of relevant input data, Data transformation, Visualisation mapping and View transformation. The visual mapping and view transformation stages use OpenDX[30], which supports image manipulation by educators.

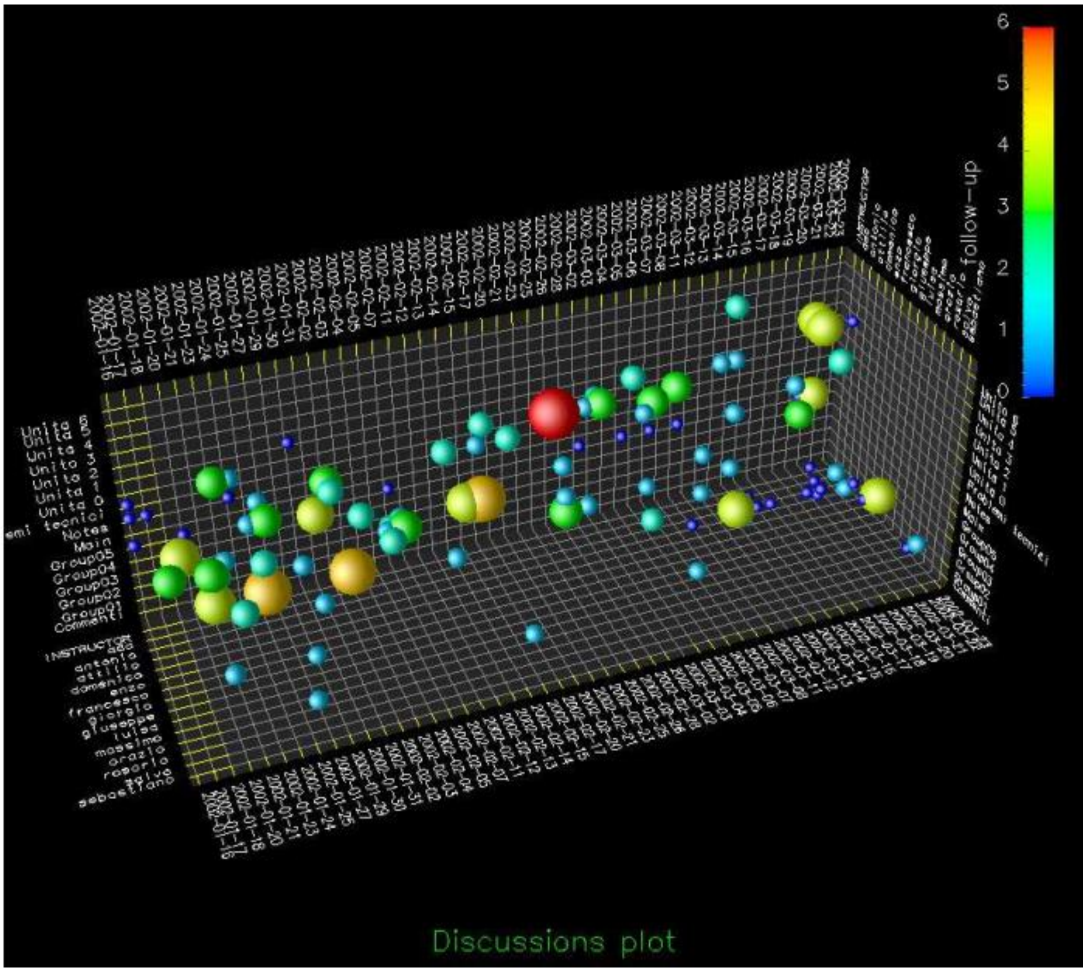

Figure 3.2 CourseVis Scatterplot

CourseVis was evaluated (Mazza & Dimitrova, 2007) through a Focus group, an experimental study and semi-structured interviews to assess suitability, effectiveness, efficiency and usefulness of the tool. The aim of the focus group was to evaluate the usefulness of the visualisations and it was found that all the participants would like to use the scatterplot visualisation to detect student issues in their courses, although the discussion scatterplot, shown in Figure 3.2 was overwhelming in 3D, but usable in 2D. The second

[30] http://www.opendx.org/

evaluation, experimental study, consisted of six instructors, divided into two groups, the first group using a WebCT only and the second group using a WEBCT with CourseVis. A java online course with data from 2002 was used by both groups and the participants were given the same questions to answer pertaining to the data. From the results, it was found that the participants from the first group were unable to answer some questions and were generally slower than the second group of participants, who used the WebCT with CourseVis and hence could use the visualisations to help answer the questions. The second group of participants generally had more accurate responses.

3.6.3 GISMO

GISMO (Mazza & Butturi, 2007) was developed to visualise student-logged data gathered by the Moodle Learning Management System (LMS) to support educators in monitoring student learning behaviour. LMSs[31] are course-centric platforms that support the delivery of learning content and resources to students and administers that track and report on e-learning courses. Some of the data collected by Moodle includes resource interactions by learners, daily activity and pages visited. GISMO presents this logged data using an improved version of the visualisations that educators found useful in CourseVis and was developed using the same architecture (used by CourseVis). GISMO includes an additional component, the Data Exporter, which gathers student-tracking data exported from Moodle and enters it into a MySQL database.

The GISMO visualisation interface consists of three panels: one to display the visualisations, another to select data and filters and the final one to show a timescale. When an educator logs in, a visualisation showing an overview of individual student accesses with the LMS are presented, as shown in Figure 3.3. This enables educators to identify study trends, access peaks, and the most active learners. A visualisation is provided to enable educators to analyse student logs by forum discussion activity and identify the students initiating threads, posting and reading messages. A third visualisation allows educators to analyse the students' usage of learning resources, identifying the resources accessed by learners, when they were accessed, and the number of times the resources were used.

[31] http://www.educause.edu/

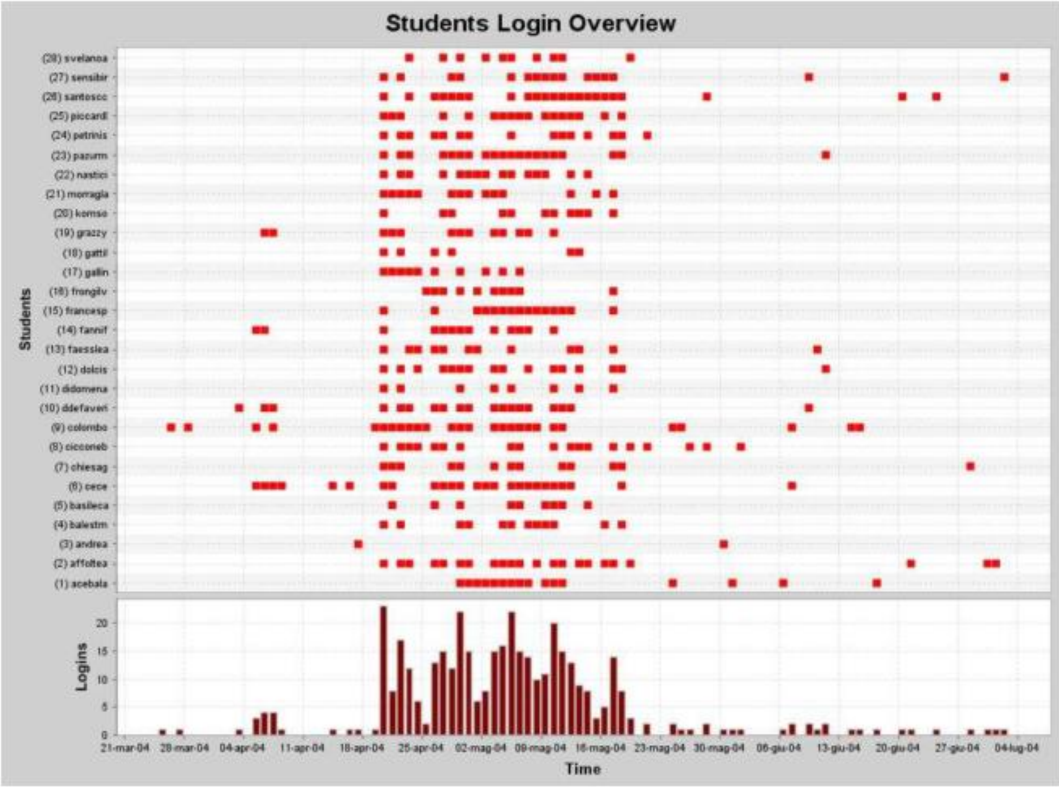

Figure 3.3 Student Access logs in GISMO

GISMO was used in an Instructional Design course (Mazza & Milani, 2004), which was mainly a project-based course run for 16 weeks with students assigned to different groups. Instructors used GISMO every three weeks to analyse learner activities and the overview visualisation enabled them to identify students who had a small number of logins. GISMO did not enable the instructors to analyse the quality of the messages posted by the students. However, instructors could analyse student participation, as the visualisations displayed the number of threads initiated, and number of messages posted and read by individual students. This study showed that GISMO effectively supported instructors in attaining a detailed understanding of student participation and usage patterns using the visualisations.

3.6.4 SNAPP

Social Networks Adapting Pedagogical Practice (SNAPP) (Bakharia & Dawson, 2011) presents student participation and interactions with LMS discussion forums to allow educators to analyse this data through sociogram and ego-network visualisations. Sociogram visualisations are graphs showing links between entities and ego-networks are node link graphs with a focal node, as shown in Figure 3.4. The visualisations use post-reply data from the forum to establish the relationship between participants. Forum posts are stored in a database with a title, description, author and date, and SQL queries are executed to retrieve data which is visualised.

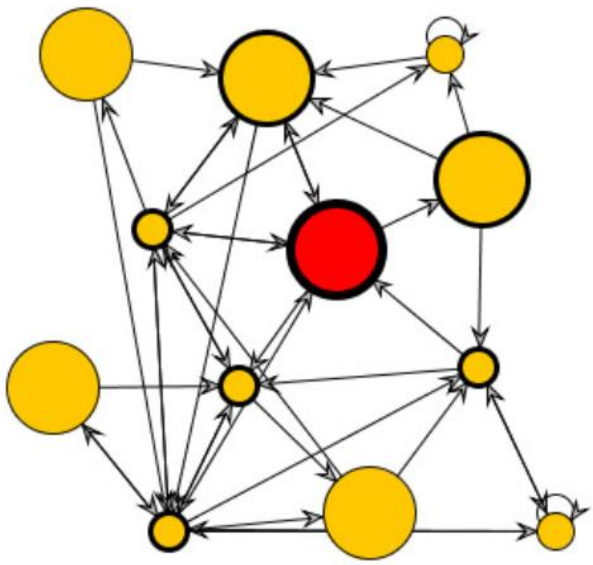

Figure 3.4 SNAPP ego-network visualisation

The aim of the tool is to enable educators to identify participant engagement and intervene when learners show signs of frustration, such as instances of isolated nodes where learners have created posts but there are no responses. The sociogram visualisation can be filtered by learner relationship strengths, by date and by a threshold value, which can allow educators to analyse interactions of various learner categories. Filtering by a threshold shows nodes above or below a specified value and helps identify engaging and non-engaging learners. Filtering by date allows the educators to view student interactions at various time intervals throughout the duration of the course. Clicking a node highlights all the connected nodes allowing the educator to easily see the relationship strength of the selected participant, which is especially useful when there are large numbers of nodes. In addition, SNAPP provides an animation of the changes in the network, which can support educators to identify the emergence of patterns amongst learner interactions.

3.6.5 LOCO-Analyst

LOCO-Analyst (Jovanovic et al., 2008) provides feedback to help educators understand student learning activities, usage and social interactions in online learning environments through tracking data. The feedback is based on Learning Object Context (student interactions with learning content), which abstracts useful tracking data from e-learning systems.

Educators can view a visual breakdown of the scores achieved by students, showing learner performance through a 3D bar chart, shown in Figure 3.5, where bars are coloured to indicate students with top, above average, below average and poor scores. The view consists of two panels, one showing the visualisation and the other (below it) displaying rows of the registered students and their score on the selected activity. The lower panel also presents the overall average score achieved by students on previous learning activities. Educators can explore student interactions with the learning environment by selecting any row to load a dialog containing visualisations presenting student interactions with forms and learning content. In the interactions with the learning content view, a scatter plot visualisation is used to display the amount of time spent by a learner on the various assigned tasks and the days that the student worked on the tasks. Each topic is represented by a colour and shape on the chart and users can interact with the visualisation to zoom in and out to view node details.

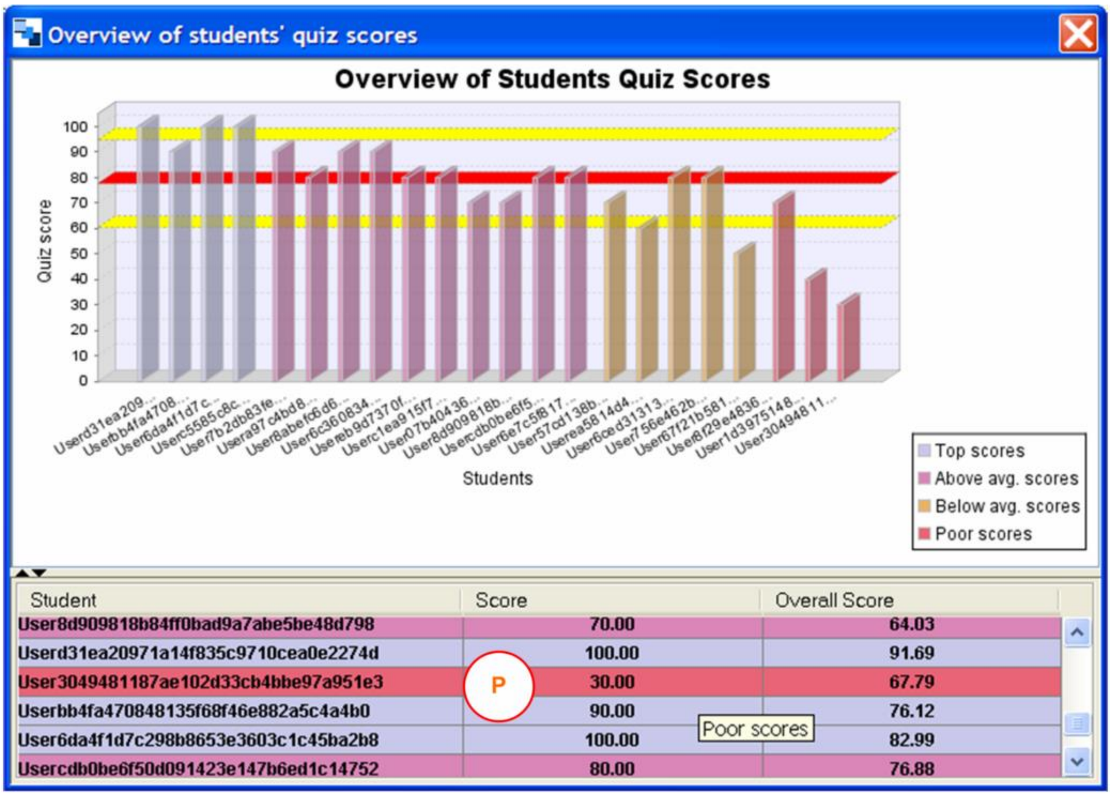

Figure 3.5 LOCO-Analyst Student performance view

LOCO-Analyst was evaluated (Jovanovic et al., 2007) using 18 experienced educators from three universities (University of Saskatchewan, Simon Fraser University and the University of Belgrade), who were asked to test the tool and complete a questionnaire. The questionnaire focused on 1) comparing the feedback provided by LOCO-Analyst with that provided by other systems they used; 2) evaluating the functionalities of the tool and 3) assessing the user interface. The majority of participants found the feedback provided by LOCO-Analyst easier to use than other systems they had employed in the past as it filtered out unnecessary details. The functionality of the system was also found to be effective by the majority of the participants as they acknowledged that relevant details about student interactions were communicated. The user interface evaluation received very positive responses, with the vast majority of the participants acknowledging it to be very good or good.

3.6.6 Analysis

This section has discussed how visualisations have been used to support educators in monitoring student learning patterns and analysing their progress. It included evaluation findings from three of the systems which in general highlighted the benefits that the participants evaluating CourseVis and LOCO-Analyst and the educators using GISMO experienced when using the visualisations. From the discussion above, a trend can be seen with the learning systems focusing on presenting visualisations of student data, with the learner at the centre of the analysis and not the learning content or the tasks or activities. This is an important trend as it can assist educators in focusing on the learning patterns and performance of their students.

Chapter 2 highlighted a number of important visual analysis tasks and features that support individuals in making sense of data including visual interactions, data transformations and guidance, which are relevant to the context of the systems discussed in this section. The paragraphs below analyse the extent to which these tasks and features have been implemented in the LA systems discussed above, and the level of support they offer to educators to help them make sense of the student data.

3.6.6.1 Visual Interactions

Chapter 2 discussed systems supporting interactive visualisations which included query relaxation techniques (Heer et al., 2008), small multiple explorations (Elzen & Wijk, 2013) and interactive legends (Riche et al., 2010). The evaluations highlighted that users benefited from these features, such as spending less time exploring data and executing tasks. Although the five systems discussed in this section do not support the same level of visual interaction capabilities, it can be seen that they provided educators with interactive visualisations to monitor student learning by manipulating views to attain a better understanding of the data. This process included:

- identifying students progressing, and those that were struggling (Instructional Intelligence Dashboard) to comprehend the material,
- understanding the social, cognitive and behavioural aspects of students in online learning environments (CourseVis),
- analysing student learning behaviour (GISMO, LOCO-Analysts and SNAPP).

The Instructional Intelligence System's dashboard visually presents learning objectives and allows educators to manipulate views by selecting bars to view details. Selecting a bar shows the estimated learning by students and class progress for the sub-objectives. CourseVis presents visualisations showing interactions between students and educators, performance of students by course topic and quizzes, and individual student progress. It supports visual interactions such as brushing, zooming and rotating to enable educators to manipulate views and make sense of the student data. GISMO presents visualisations of student accesses (logins, resource and discussion forum) that support selecting and filtering data. An individual student can be selected and filtered to allow educators to identify learning behaviours and issues that learners may be experiencing. SNAPP uses sociogram visualisation to present forum discussions and relationships between learners. It supports selecting and filtering learner relationship strengths, by date and by threshold values, which enables educators to analyse and understand learner behaviours. LOCO-Analyst provides visualisations to support educators in analysing student learning activities, usage and social interactions. The tool supports the drilldown interaction technique, loading a dialog with visualisations showing student interactions in forms, and interactions with the learning content. It also supports zooming of individual task topics (represented by shape and colour on a chart) to view details.

A number of interaction techniques have been implemented by the systems discussed above including select, filter, zoom and drilldown. These have been used to enable users to manipulate views to analyse student data; for example, participants from the CourseVis user trial appreciated the flexibility supported by the visualisation to explore complex data. The evaluations of GISMO and LOCO-Analyst also highlighted user satisfaction with the interfaces and the level of support offered to identify and understand usage patterns.

3.6.6.2 Data Transformations

Chapter 2 highlighted the importance of data transformation and introduced a number of visualisation systems (Spenke & Beilken, 2003, Xiao et al., 2006, Shi et al., 2014) outside of TEL that support transformations but which are limited to data parsing, extraction and statistical measures. Similar statistical measures are evident in GISMO (aggregations of accesses) and LOCO-Analyst (aggregations of scores). However, they do not provide educators with transformations of the data to support derived data explorations. The literature suggests that visual analytics tools should support transformations to present derived views using input data (Heer & Schneiderman, 2012). Derived data transformations have been defined in this thesis to include statistical measures but can also to include data that is derived from the input data (learner data) and which is related to the data presented through the visualisations that the educators are analysing.

3.6.6.3 Guidance

Chapter 2 highlighted that narrative visualisations (ordered sequence of visualisation steps with textual descriptions focusing on data) are being increasingly used in Information Visualisation to support the visual analysis process (Segel & Heer, 2010). They are effective in passing on information and assist in making the message more memorable (Austin 2011). Although the student data generated by the systems discussed above can be presented through visual narratives, none of these systems support them. LOCO-Analyst supports drilldown exploration through a connected visualisation that is presented in a popup window but does not provide the necessary descriptions of the sequenced views to classify it as a visual narrative. The lack of support for visual narratives in education systems presenting visualisation to educators is maybe due to the fact that an author is required to construct a narrative, which in general may not be available in such scenarios.

3.7 Supporting Students in monitoring their progress

Research objectives three and four defined in section 1.3 of this thesis outlined the implementation of the *derived data approach* and its evaluation, by supporting the engagement and performance of undergraduate third level students, through visual narratives using student-logged data. To this effect, this section discusses the state of the art learning systems that support students in monitoring and understanding their progress. These systems enable learners to view and interact with visualisations displaying their learning activities including resources used, time spent on tasks and peer comparisons. It is important to note that there are some indications in the literature (Burleson et al., 2005) which highlight that some students do not find peer comparisons useful. Nonetheless, research (Vassileva & Sun, 2007, Linton & Schaefer, 2000) also suggests that learners can be motivated to engage when comparing their own progress with peers.

The majority of the systems discussed below use student-logged data, including activity traces, task durations, resource usage and discussion forum interactions as data for the visualisations. This section highlights best practices, describes evaluation findings and analyses the degree to which visual interactions, data transformations and guidance (identified as important tasks and features for visual analysis in Chapter 2 that are relevant in TEL) are supported. The majority of systems discussed below also support educators in visually analysing student learning behaviour and this aspect is also discussed.

3.7.1 SAM

Student Activity Meter (SAM) (Govaerts et al., 2010, 2011, 2012) can be described as a multi-panel non-narrative tool supporting visual interactions for both learners and educators. It has been deployed in a Personalised Learning Environment (PLE) on the Responsive Open Learning Environments[32] (ROLE) project. Learners can monitor what they have been working on and the time they spent on activities and view peer comparisons through the visualisations.

Figure 3.6 SAM User Interface

The user interface consists of three panels as shown in Figure 3.6; the first panel shows a number of visualisations separated by tabs. The first of these visualisations is a line chart with each line representing a student and the red line representing the logged-in student. The lines represent the total time spent by learners on course activities by date. The visualisation allows learners to compare the time spent on course activities with peers. The second tab contains a parallel coordinates visualisation with four axes representing the time spent on the course, the number of documents used, the average document usage time and the average time students used the PLE. Each polyline represents a student and the logged-in student is shown in red. The third tab contains a bar chart with data categorised by the time spent and the number of documents used. The green bar is the category that the logged-in student belongs to. The tool allows users to drilldown by selecting a bar to view details. The second pane presents statistical measures, including time spent and document usage figures for the logged-in learner, and comparison figures for a student selected through the visualisations.

[32] http://www.role-project.eu/

The third pane graphically presents recommended documents to the learner based on the most popular documents amongst learners and the ones with the greatest usage time.

The first iteration of SAM was evaluated (Govaerts et al., 2012) using twelve students from a User Interface course. The aim of the trial was to assess usability and user satisfaction and consisted of two sessions. The first session involved a task-based interview, where the ‘think-aloud protocol’ (Van Someren et al., 1994) was used, which required students to think-aloud as they used the interface and explain what they understood from the visualisations. This session was used to measure user errors, learnability and efficiency with the interface. It consisted of seven students and in general the participants could follow the visualisations with the exception of some explanations required for the parallel coordinates and line chart visualisations. The second session evaluated user satisfaction using the System Usability Scale (SUS) (Brooke, 1996), which was run after one month of usage and consisted of 9 student participants. The average SUS score was 73.33 with a standard deviation of 9.35, which is an above average[33] score for usability.

SAM also supports educators in identifying patterns in student learning behaviour and in monitoring the activities through visualisations that are taking the most time. The second and third iterations of SAM introduced a help button to assist first time users (iteration 2) and enabled the rearrangement of the parallel coordinate axes to improve data comparisons (iteration 3). The second iteration was evaluated (Govaerts et al., 2012). using twenty educators (teachers and teaching assistants). Moodle logs were used as evaluation data and participants were required to complete an online survey covering teaching issues and whether SAM could address these. The third iteration was deployed to an online course with 270 registered learners and evaluated using the same survey as the second iteration with twelve educators (Govaerts et al., 2012). The results of both qualitative studies highlighted that delivering feedback to learners and resource usage awareness were important issues for educators and that the SAM visualisations were useful in raising teacher awareness.

The fourth iteration enhanced the line chart visualisation by facilitating zooming (on the vertical axes), filtering by time spent and searching via a text box. Filtering, searching and axes removal functionality was added to the Parallel Coordinates visualisation. The evaluation consisted of data gathered from a Systems Architecture course. Eleven educators were involved in the course and were given six tasks to complete in a one-hour session (Govaerts et al., 2012). The tasks involved them making findings using parallel coordinates,

[33] http://www.measuringu.com/sus.php

identifying ‘strong’ and ‘at risk’ learners and understanding how time how learners spent their time. The results showed that all the participants were able to see how the students were spending time on the course, detect patterns and issues using the parallel coordinates, and identify the strong and at risk students.

Analysis

SAM provides visualisations to learners supporting self-reflection and peer comparisons. The visualisations also allow educators to analyse how students are spending time and support identification of students at risk so interventions can take place. Chapter 2 discussed the role of visual interaction techniques in supporting users in making sense of data and highlighted successful evaluations of systems, for example, supporting data filtering and zooming (Heer & Boyd, 2005). The visualisations available in SAM present current learner data, and both students and educators can interact with them to view past learning activity patterns. The tool supports a number of interaction techniques, including select, filter (supports searching and removal of axes from the parallel coordinates) and drilldown. Four evaluations were conducted and the results showed that the interactive visualisations were found to be very useful by learners. Educators also found them quite beneficial and used them to identify strong and weak students, detect learning patterns and gain insights. This again highlighted the importance of the various interactive visualisations.

In Chapter 2, the concept of data transformations were discussed, which are used to generate data related to input data. Several systems (Spenke & Beilken, 2003, Xiao et al., 2006, Shi et al., 2014) in domains other than TEL support data transformations, however these are limited to statistical measures, including aggregation, and averages. SAM has a dedicated panel displaying statistical measures that include average, maximum and minimum time spent and resources used by learners. However, data transformations which generate data related to that presented through the visualisations that the learner or educator is analysing are not supported. For example, the line chart in Figure 3.6 displays all the students and the times they have spent on their activities. This can be filtered by the total time spent or it can be filtered to show a single student. A derived data transformation could present a visualisation displaying the time spent by a related category of students (based on the search option selected), or the resource usage by the filtered student.

Chapter 2 also introduced user guidance through visual narratives, consisting of sequenced visualisation with textual description focusing on data. The evaluations of systems such as

Ellipsis (Satyanarayan & Heer, 2014), SketchStory (Lee et al., 2013) and Gapminder[34] discussed highlighted the benefits of visual narratives. The goal of visual narratives is to support users in understanding the message and making it more memorable (Austin 2011). SAM consists of multi-panel visualisations and consists of some connected visualisations manifested through the drilldown interactions, however it does not support visual narratives of the data.

3.7.2 CAM Dashboard

The CAM Dashboard (Santos et al., 2011, 2012) consists of goal-oriented visualisations for students to reflect on the time spent on assigned activities, and view comparisons with fellow learners regarding progress made towards goals. The dashboard also allows students to communicate with teachers for support. The visualisations present student behaviour using the various tools required by the course including the Eclipse IDE and MS Word.

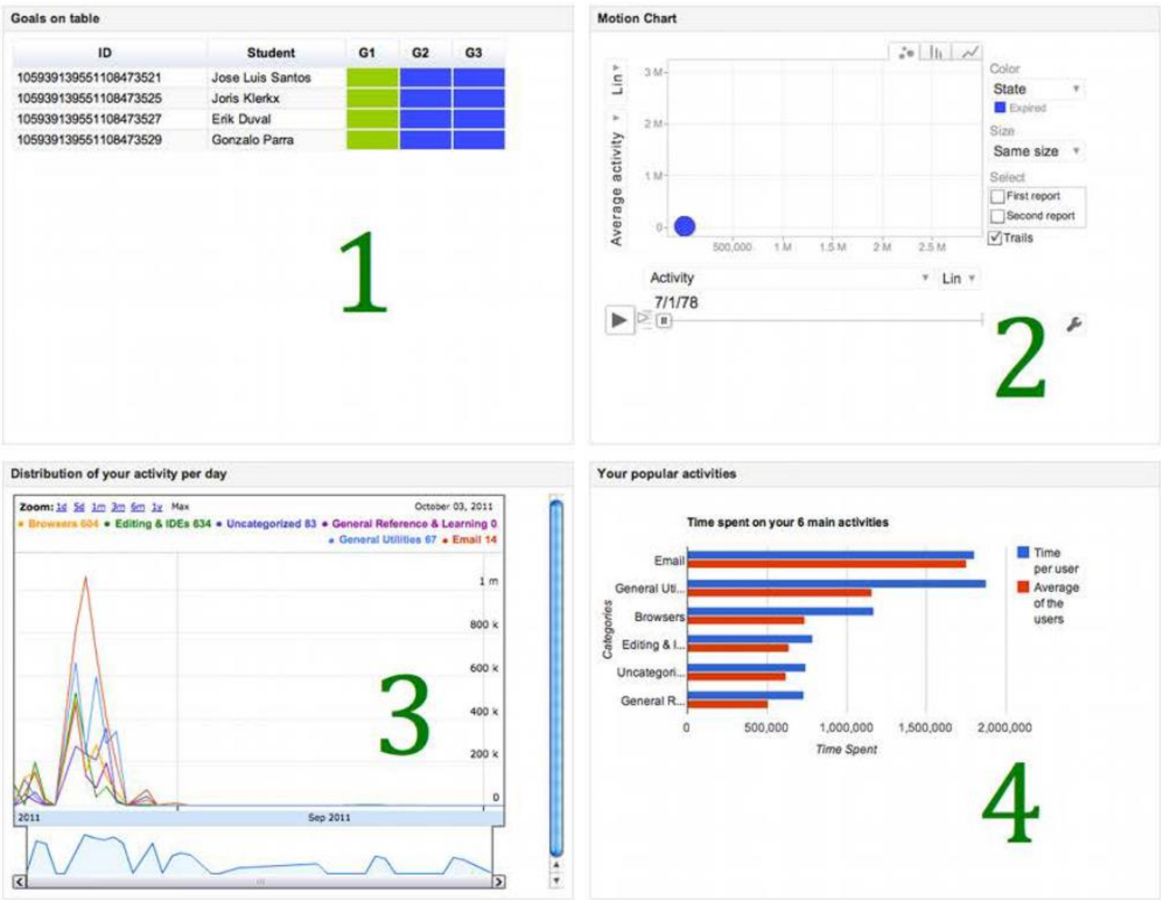

Figure 3.7 CAM Dashboard

The dashboard, shown in Figure 3.7 presents a number of visualisations showing an overview of the status of a student's goals, time per goal, and time spent or number of events

[34] http://www. gapminder.org/downloads/human-development-trends-2005/

per tool and day. It also presents visualisations showing the time spent on an activity by the student compared to the average time spent by peers. The tool supports visual exploration through coordinated visualisations. For example, if a student selects a day, he/she will be shown a drilldown view of data through a different visualisation on the dashboard.

The dashboard uses OpenSocial[35] widgets to present the visualisations, sourced from the Google Chart library, which allows communication between widgets and facilitates coordination between visualisations. RescueTime[36] and Rabbit Eclipse plug-in[37] were used to gather data from the various tools used by the students, which tracked time spent by learners on applications, documents and websites. Two working releases are reported which consist of line charts, bar charts and motion chart visualisations as shown in Figure 3.7.

The first working release of the dashboard was evaluated with 36 students in an engineering undergraduate course after a demonstration of the application (Santos et al., 2012a). Participants were asked to complete a questionnaire which involved them ranking the widgets and rating the usefulness, usability, effectiveness and their satisfaction with the dashboard. The analysis of the questionnaire responses showed that students generally found the visualisations useful and helped them to achieve their goals. Students found the visualisations that supported peer comparisons to be very useful as they could monitor how other students were working through the course. The bar chart visualisations, which presented learners with their own time and a comparison with average student time for activities, received the highest rating.

The second working release was evaluated using a real course setting (Santos et al., 2012a), which analysed data generated from 36 students using the dashboard. The tasks involved implementation of web applications and use case design during four laboratory sessions. The evaluation was conducted with a subgroup from the first working release evaluation participants, and consisted of ten students. The students were given a SUS questionnaire to complete and they were also asked to rate the visualisations, score the usefulness of the dashboard, and rate their satisfaction with it. The SUS score from this evaluation was 72, which is an above average rating, and again the bar charts were rated as the most useful visualisations available on the dashboard.

[35] Open Social API, http://code.google.com/apis/opensocial/
[36] RescueTime, http://rescuetime.com/
[37] Rabbit Eclipse plug-in, http://code.google.com/p/rabbit-eclipse/

Analysis

The CAM Dashboard supports self-reflection amongst learners and enables them to view their progress and compare it with peers through interactive and explorable visualisations. The dashboard consists of interactive visualisations supporting select, filter, zoom and drilldown and multiple coordinated views. The drilldown interaction technique enables students to explore data details by selecting a bar representing a weekday to view a further visualisation showing activity details for the day. This is similar to the functionality offered through Tibco Spotfire which was discussed in Chapter 2. The CAM Dashboard visualisations present real-time logged data to the learners and supports visual interactions to enable them to look back and make comparisons with other students. The two trials found that the participating students found the interactive visualisation beneficial in achieving their goals, and particularly found the bar chart displaying peer comparisons of the time spent on various activities the most useful. As observed in the analysis of SAM, the findings highlighted the importance of the various interactive visualisations.

From the descriptions of the CAM dashboard, it can be seen that it presents views of student activity times (averages). However, it does not provide learners with visualisations that present derived data transformations showing the data related to the visualisations on the dashboard. For example, if a top-performing learner is analysing a visualisation that is displaying the time he/she has spent on certain resources, derived data transformations could present visualisations showing the most popular resources amongst the class students or resources used by other top-performing students.

Although the CAM dashboard consists of coordinated visualisations and connected views supporting drilldown, visual narratives consisting of sequenced visualisations with textual descriptions are not supported.

3.7.3 Moodog

Moodog (Zhang & Almeroth, 2010, Zhang et al., 2007) is a Moodle plug-in that uses the Course Management System's logged data to visualise students' learning activities for both learners and educators. Moodog modifies the Moodle home page with a visual overlay showing coloured bars that indicate the percentage of students that have viewed a resource. The plug-in aims to facilitate insights into student interactions with course material and supports learners through peer comparisons, where individual students may compare their progress with others. Moodog provides visualisations displaying student resources and time-based statistics, including the activities started and completed, the number of times activities were accessed by students and the resources that have not yet been viewed. It also presents comparisons between students such as the number of times course material was viewed, the number of sessions per students, the time spent on the course and forum activity. In addition to the visualisations for the learners, Moodog provides aggregated statistical reports and reminds students to view resources that they have not accessed to date.

The Moodog architecture consists of an interface, and three components: a filtering component, data collection and calculation component and a Unix Cron process component. The interface presents the statistical results to both learners and educators. Some of the views contain learner identities (for educators to view) and the interface ensures the identities are not visible to other learners. The collection and calculation component extracts logged data and course metadata from Moodle tables and runs a number of calculations on the logs. Finally the Cron process component runs every hour and stores the statistical data, used for the visualisation. This helps to improve the response times to users' requests.

Moodog places a visualisation of coloured bars (green and red) for each course activity or resource over the Moodle interface. The green indicates the percentage of students who have viewed the resource and the red presents those who are yet to view the resource. More detailed visualisation reports showing student-based statistics such as total views, number of sessions, time spent online and viewed resources per learner are also available, as shown in Figure 3.8. These statistics also include threads initiated by the learner and the total number of follow-up posts. A second report based on overall course resource statistics visualises the number of unique students who viewed a resource and the total view counts per resource.

Figure 3.8 Moodog Student Activity

Moodog was used to analyse an Introduction to Computer Communication Networks course with an of enrolment 38 students in 2006 (Zhang et al., 2007). The analysis showed that student forum interactions positively correlated to course performance. The study also analysed resource usage, which showed that not all of the users viewed every resource.

Analysis

Moodog provides a visual overlay for Moodle presenting resource views (coloured bars presenting resource popularity) amongst learners for both educators and students to analyse. The plug-in provides a number of statistical measures of the data that are visually presented showing student activity data and resource usage counts. However, the plug-in does not support visual interactions and hence does not allow students to explore the data. For example, students and educators are unable to select and filter for data and resources. They cannot drilldown to explore the statistical reports of interest to analyse data of interest, such as the times when resources were accessed. Chapter 2 discussed the role of visual interaction in supporting users in making sense of digital data, which was highlighted through a number of evaluations (Heer et al., 2008, Yi et al., 2005).

Moodog provides visual representations of statistical reports presenting activity and resource usage, which can be considered as data transformations of the logged data. This is similar to SAM, CAM and a number of systems in domains other than TEL (Spenke & Beilken, 2003, Xiao et al., 2006, Shi et al., 2014). However, derived data transformations presenting views of data related the visualisations being analysed is not supported. For example, the overlay visualisation on the Moodle home page presents resource view percentages. A student or educator is unable to visually explore related data to this, such as the number students that accessed the resource, when it was accessed, and the duration it was accessed for.

Literature suggests that visual narratives support the visual analysis process (Segel & Heer, 2010) and provide effective ways of passing on information (Austin 2011). It can be seen from the description that the plug-in does not support visual narratives as the overlay and the statistical report are stand-alone visualisations.

3.7.4 MACE Zeitgeist

The Zeitgeist application (Schmitz et al., 2009) was developed as part of the MACE system (Stefaner et al., 2007), a set of integrated learning repositories that include Contextualised Attention Metadata (CAM) (Wolpers et al., 2007). Zeitgeist is used to show learning activities available through MACE by presenting resources used by students, the usefulness of these resources and study paths taken by learners.

The Zeitgeist application consists of a dashboard, shown in Figure 3.9, presenting a usage summary, a visualisation of the usage history and a daily content viewing report of individual students' MACE-related learning activities. The aim of Zeitgeist is to support self-monitoring and reflection amongst learners by enabling them to view their learning paths, their activities over time and the history of their learning development. The usage summary section provides the number of resources viewed, downloaded, bookmarked and tagged, which can be filtered by month and year. The usage history section provides a time-line chart visualisation, plotting lines for the number of resources viewed, downloaded, bookmarked and tagged by day of a selected week and year. The daily content history section provides a tabular view of the resources and the action taken on them, such as those viewed, downloaded, bookmarked or tagged and the time of the action.

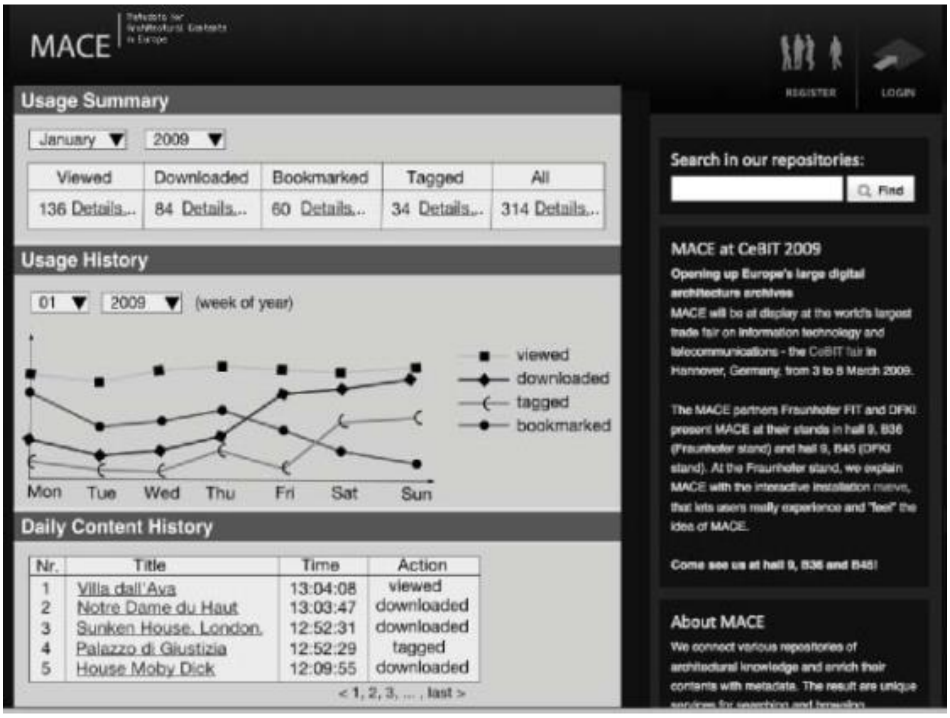

Figure 3.9 Zeitgeist Dashboard

The dashboard supports learner interactions by enabling students to filter usage history by date. In addition, it accumulates learning activities of individual students from the MACE repositories and presents overviews of this data to learners to support peer comparisons of learning behaviour and usage. The dashboard also gathers individuals' metadata, which allows students to view comparisons with similar learners, namely those students with similar usage profiles.

Analysis

Zeitgeist provides learners with a dashboard showing usage statistics and visualising usage history through a line chart and enables students to compare learning activity with peers. The dashboard presents current student activity data and enables learners to look back at historical actions. In addition, the dashboard focusses on the individual learner and presents the content viewed and activities completed for the student. A learner can analyse his/her usage history over a specified period of time through the line chart and apply filters to view data at different timings and view usage details. However, other interaction techniques discussed in Chapter 2 such as zooming, details-on-demand and drilldown which have been shown to be beneficial in supporting users in understanding data are not supported.

The dashboard provides a usage summary view, however it does not provide visualisations that are directly related to the data the student is analysing (derived data transformations). The MACE-related learning activities data may be used to produce visualisations through derived data transformations. For example, if a student is viewing his/her usage history for a specified week, a derived data transformation could show usage of peers with similar profiles during the same week.

Similar to other the tools described in this chapter, Zeitgeist does not use visual narratives to present the MACE-related learning activities.

3.7.5 LARAe

LARAe consists of a dashboard for students that has been used in an Open User Interfaces course, with 38 engineering students, visualising learner actions to support progress and awareness (Charleer et al., 2014). The tool also enables educators to view learner outcomes and progress, and support, interventions. Learners are split into groups of three and are required to report on activities using Twitter, blog posts and comments. LARAe allows students and educators to select a learner and module to view and analyse.

The dashboard shows activities represented by circles which are sorted by student groups and activity type. Student groups are represented as rows and activity types are represented by columns. Colour gradients are used to indicate how old the activity is. Users can choose an activity to focus on and view the context by selecting it and viewing the associated thread as shown in Figure 3.10B and Figure 3.10C, which are considered as drilldown views. The numbers in the activity circles indicate the size of the thread and at a glance educators can see the activities with low numbers of comments so they can intervene if necessary. Selecting or highlighting an activity circle allows the learner to view the distribution of comments throughout the class, read through others posts and become aware of the most active groups.

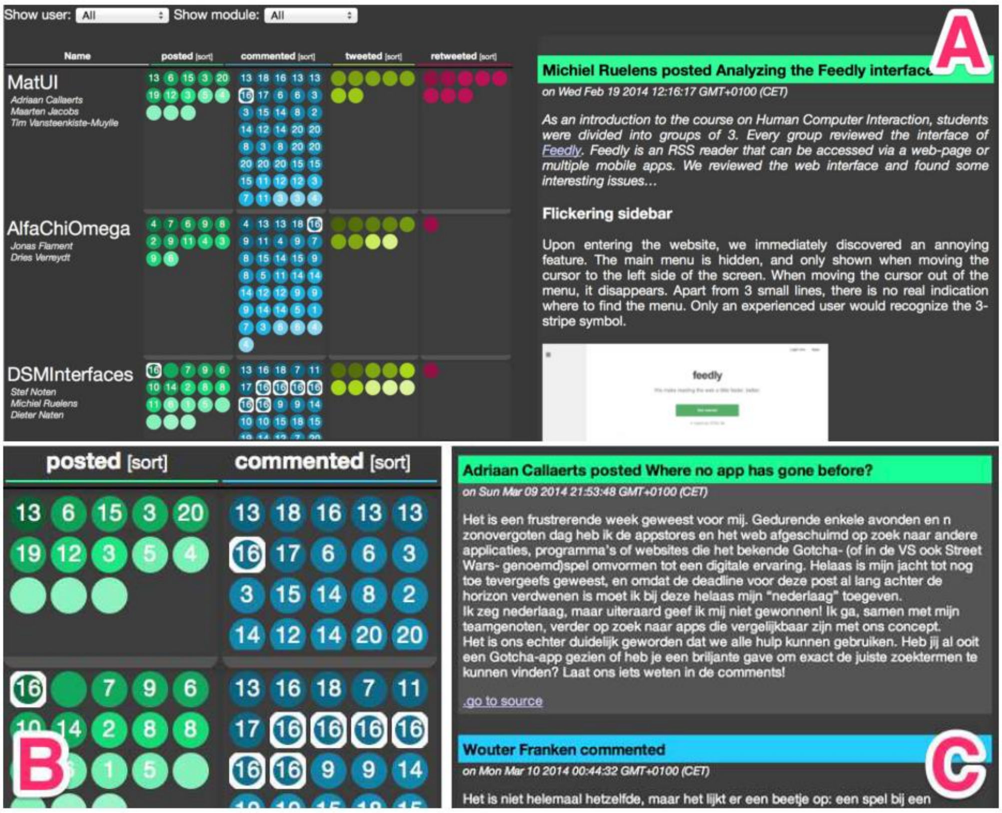

Figure 3.10 LARAe Dashboard

Analysis

LARAe presents a dashboard to enable learners to raise awareness regarding activity content and context, and supports peer comparisons of forum posts. The dashboard can also be used by educators to monitor student progress and it supports interventions. The dashboard uses bespoke visualisations displaying live student forum activity data that can be explored to analyse thread posts and comments. It focusses on learning content filtered by students or modules. The dashboard provides interactive visualisations enabling student and module filtering, and presents thread details through drilldown views when activity circles are selected.

The dashboard consists of circles which represent activities and contains a number that is an aggregation of the total posts or comments associated with it. This aggregation can be considered a data transformation using the student activity data. LARAe does not support the generation of derived data transformations of the data analysed through the visualisations, which may provide further insights. For example, a derived data transformation for the visualisation presented in Figure 3.10A can show the time spent on activities and number of sessions per activity or on all activities.

The dashboard also does not support narrative visualisations, which can help support the data analysis process for both learners and educators (Segel & Heer, 2010).

3.7.6 eMUSE

The eMUSE (empowering MashUps for Social E-learning) platform (Popescu, 2012, Popescu & Cioiu, 2011) uses a mashup of content from social media websites to support learners' access and track learning content, and provides visualisations to educators to monitor student activity and student grades. The platform provides learners with an integrated learning space to access a number of web tools, some of which include Twitter, Blogger and YouTube. It also provides a bar, a line and a pie chart visualisation to students, as shown in Figure 3.11, presenting learner activity and peer comparisons and presents the student's score based on criteria defined by the educator. The platform provides the educators with visualisations showing the overall course activity and individual students' participation. Educators can also use the tool to define scoring criteria and use it as a control panel to manage enrolment and grading.

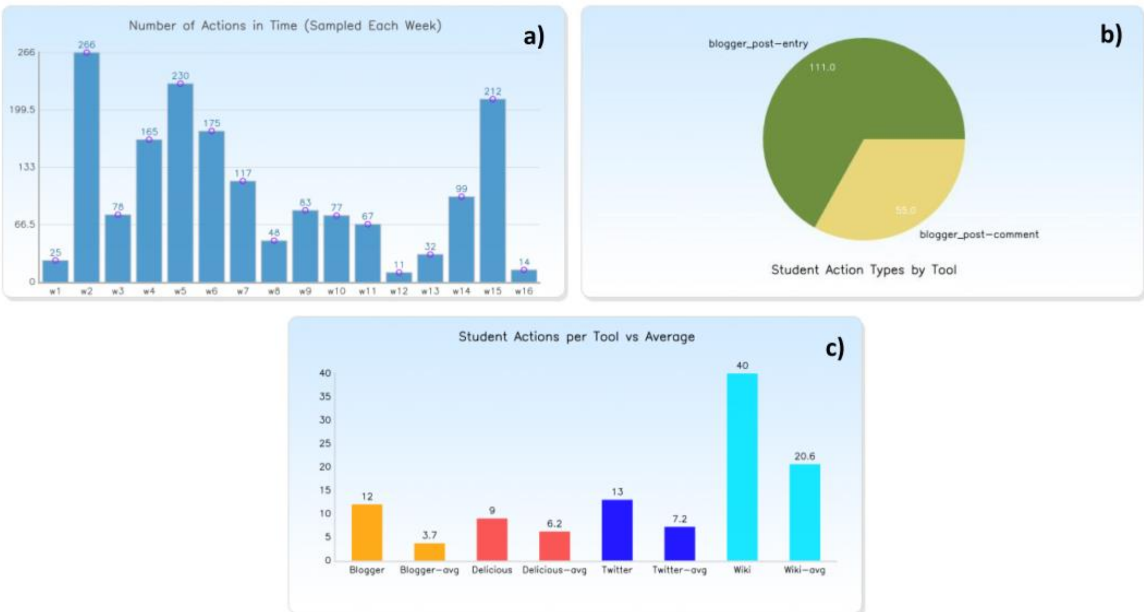

Figure 3.11 eMUSE Visualisations

The eMUSE platform was used for project-based tasks in an undergraduate Web Applications Design course for one semester with 45 registered students, split into eleven teams, at the University of Craiova, Romania (Popescu & Cioiu, 2011). The teams were required to meet weekly and use eMUSE for their project. The evaluation of the platform involved student log analysis and feedback. The log analysis suggested that project and course involvement can indicate level of performance, however, only the data for the best student was shown to present this. The second part of the analysis evaluated learner feedback through a post-study questionnaire, which focused on the usefulness of eMUSE, the extent it motivated the students and whether they would consider using it in the future. The results were very encouraging with most students positively commenting on increased motivation using the platform and the usefulness of comparing progress with peers. Nevertheless, students suggested that the visualisations did not highlight the quality of contributions.

Analysis

The eMUSE platform provides a mashup consisting of learning related Web 2.0 tools accessible through a single point and visualisations showing student contributions and actions and activity times. The platform records learner interactions with a number of commonly used web tools, and provides visualisations presenting statistical measures of interactions, and overall time spent using these tools. It supports peer comparisons and the evaluation revealed that students found this feature to be beneficial. However, the platform does not support visual interactions, such as filter, zoom or drilldown to enable students to manipulate views of explore the data. Evaluations focusing on visual interactions (Riche et al., 2010, Heer et al., 2008, Heer & Boyd, 2005), discussed in Chapter 2, highlighted the benefits of interaction techniques in supporting the data analysis process.

The three visualisations shown in Figure 3.11 present action aggregations and averages from the student activity data. However visualisations, showing data related to the student activity counts presented in the visualisations, are not supported by the platform. For example, Figure 3.11a shows the number of total actions by the users with all the web tools per week. A derived data transformation could visualise a breakdown of time per web tool per week. In addition, the platform does not support narrative visualisations for the activity data, which may support learners in interpreting the data better through a guided message.

3.7.7 ALAS-KA

Khan Academy[38] provides visualisations through a dashboard to learners, to reflect on their performance, and to educators, to enable them to view learner progress and the time they have spent on different resources. ALAS-KA (Muñoz-Merino et al., 2013, Ruipérez-Valiente et al., 2014) is a Khan Academy plug-in that processes raw learner data to extract information at a higher level through a set of metrics and presents it through visualisations

[38] https://www.khanacademy.org/

to learners. The metrics consist of five modules, which comprise of: 1) total use of the platform; 2) correct progress on the platform; 3) time distribution of the use of the platform; 4) gamification habits; and 5) exercise solving habits. The metric calculations use student activities, which sometimes include both video and exercises or only exercises:

- the *total use* of the platform metric uses attempts, time and efficiency in completing activities (where efficiency is reduced if an activity is repeated). In addition the time spent using optional items in the learning environment is also included,
- the *correct progress* metric uses the total number of correct exercises, number of proficiencies attained and efficiency of correct progress (exercises solved correctly versus time to solve),
- the *time distribution* metric uses the time intervals during which activities are worked on and how constant learners are with their activities during the study period,
- the *gamification habits* metric measures badges awarded to learners to determine the motivational effects of gamification,
- *exercise solving habits* focusses on learner behaviours when attempting exercises, to determine if they follow a recommended learning path.

The visualisations consist of two groups (class and individual). The class visualisations, shown in Figure 3.12, consist of a set of pie charts per metric (selectable via a dropdown on the interface), presenting a distribution overview for the entire class, or a selected group of students. The pie charts consist of colour quadrants with legends presenting a high level view of how students performed per metric. The individual student visualisations allow learners to reflect on their performances. These visualisations are presented by metric and consist of two bar charts, where a student's interactions are presented over time or activity and his/her performance compared against the mean class performance. The objective of the visualisations are to present the metrics and support interactions with the plug-in.

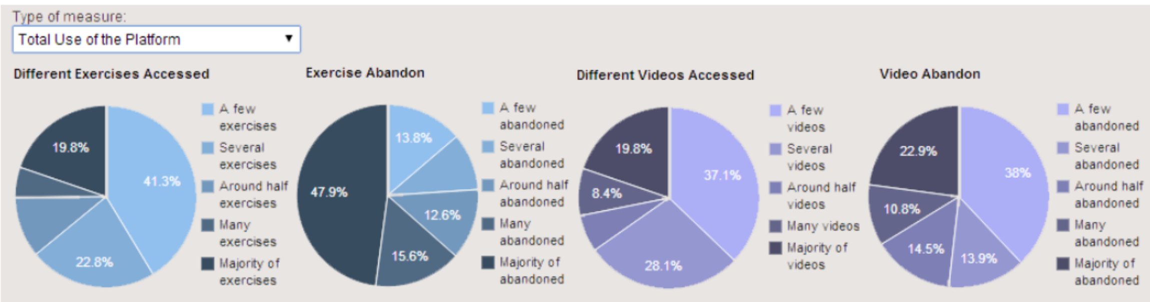

Figure 3.12 ALAS-KA Class Visualisations

Analysis

ALAS-KA uses five metrics to visually present student activity traces to learners to support self-reflection and to educators to enable them to monitor progress. However, there are limitations in the support offered to students through the visualisations, such as visual interactions and peer comparisons. Students can choose the metric they would like to analyse but the visualisation presenting the metrics does not support visual interactions. Literature has highlighted the benefits users gain from visual interaction techniques, specifically when analysing data (Riche et al., 2010, Heer et al., 2008, Heer & Boyd, 2005). In addition, the tool supports limited comparisons by allowing students to compare their usage against the class average for each metric but does not allow comparisons with individual learners. While there are some indications (Burleson et al., 2005) showing that students can at times be averse to peer comparisons, literature (Vassileva & Sun, 2007, Linton & Schaefer, 2000) suggests that they can support learner motivations and engagement by allowing students to compare their own progress with peers.

ALAS-KA presents class averages that are shown next to the learner's data. However, the plug-in does not support transformations that present data related to the student's activity data shown through visualisations. In addition visual narratives presenting the student activity and calculated metric are also not supported by the plug-in.

3.7.8 TrAVis

TrAVis (May et al., 2011) provides learners and educators with visualisations showing student activity data on Computer Mediated Communications such as discussion forums and blogs. The aim of TrAVis is to enable students to analyse and monitor their activities and reflect on their interactions on communication forums by analysing their participation rates and their learning porgress. TrAVis visualises student tracking data at four levels of communication activities, including aggregation (threads started, messages posted), discussion (number of messages posted and content communicated), cooperation (goal-oriented group activities) and collaboration (time-oriented group activities).

TrAVis provides a visualisation called "Time Machine" to students and educators showing all the activities of the learners, which can be interacted with to reveal message meta-data including timestamp, forum and message title. Learners can view peers who can also read the same message; each peer who read the message is represented by a coloured sphere, which has a size proportional to the time spent reading the message and the colour indicates how much of it has been read. Using this visualisation, learners can monitor their own activities as well as compare their discussion interactions with peers. TrAVis also provides radar graphs as shown in Figure 3.13 to enable learners to monitor their interactions at the four communication levels mentioned above. For example, at the aggregation level, the radar graph represents the magnitude of the number of threads started, messages posted, replied and quoted and connection frequency. Leaners can compare their interactions at any of these four levels with peers.

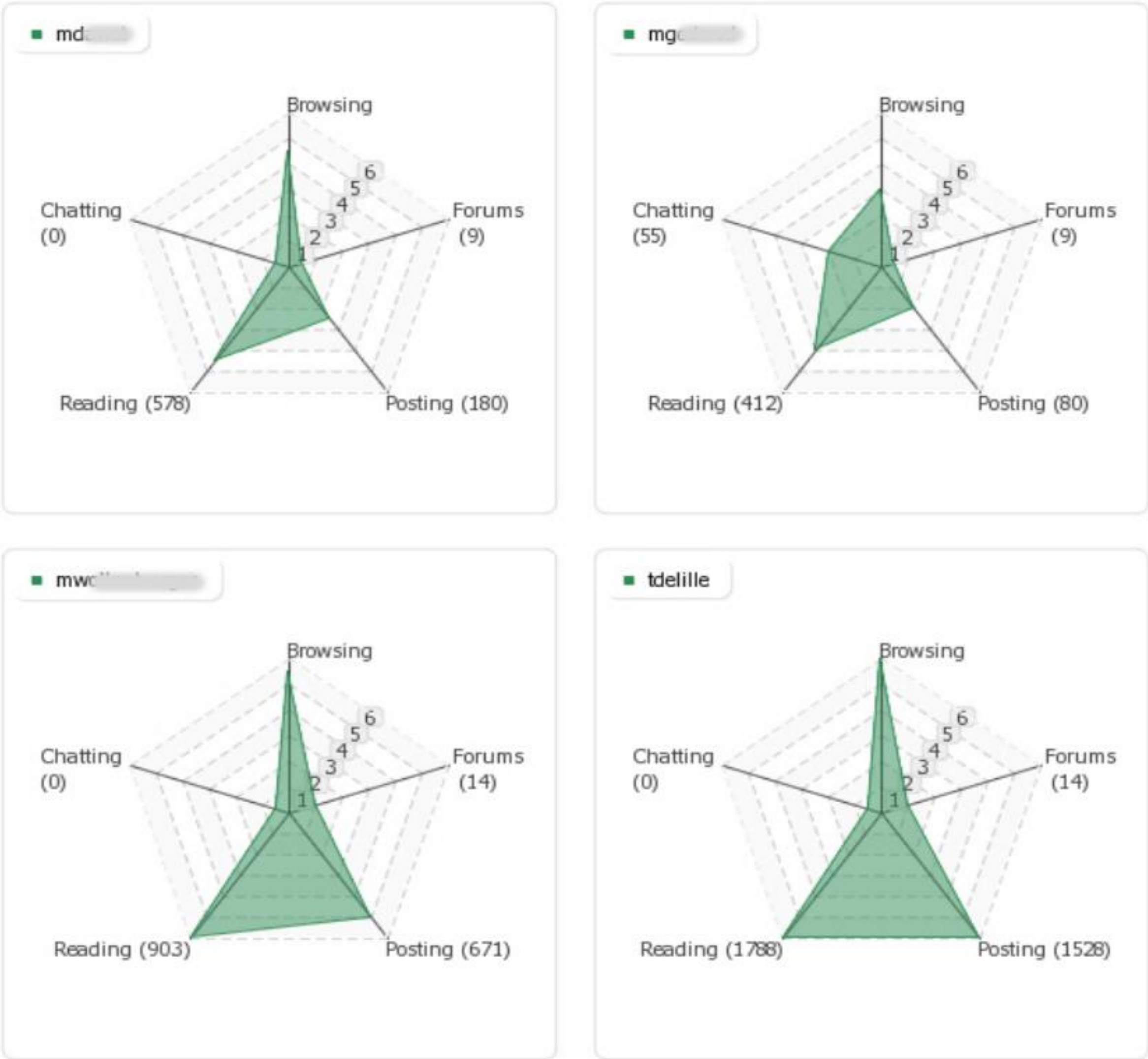

Figure 3.13 TrAVis Radar graphs showing peer comparisons

In 2009 TrAVis was evaluated at the University of Grenoble during a three month period with thirteen students and three tutors from the French as a Foreign Language masters course (May et al., 2011). The aim of the evaluation was to assess the functionality of TrAVis and the impact it had. The students were placed in groups and tutors used TrAVis to view student interactions and communications. The learners used TrAVis to monitor their activities and view peer comparisons. During the three months, a total of 330 threads were created with

2804 messages exchanged in 28 discussion forums. A user questionnaire completed by seven participants (six students and one tutor) highlighted that the system scored well in terms of usability but participants took time in getting used to the functionality of TrAVis. The evaluation showed that radar graphs enabled tutors to gauge the engagement and motivation levels of learners.

Analysis

TrAVis provides learners and educators with two sets of visualisations presenting user message metadata and supporting peer comparisons. The "Time Machine" visualisation presents real-time monitoring of student activities and allows learners and educators to look back at how the student has been engaging with the learning activities over the duration of the course. This bespoke visualisation supports interactions by providing details for selected items, however, the radar graphs do not support visual interactions.

The radar graphs presents aggregations of the discussion forum activity traces, however, derived data transformations displaying data related to forum activities are not presented. For example, a derived data transformation may present a visualisation presenting data related to the radar graphs, such as a breakdown of times when students were most active with the forums. In addition, visual narratives, presenting student activity traces with the forum discussions, are not presented to support the data analysis process through sequenced visualisations accompanied by textual descriptions.

3.7.9 Navi dashboard

The Navi dashboard (Santos et al., 2013b, Charleer et al., 2013) aims to improve learner motivations in online learning courses by presenting visualisations of learner data, peer comparisons using badges, social visualisations (Hsiao & Brusilovsky, 2012) and activity streams (Olson et al., 2012). Badges (Goligoski, 2012) have been used to reward students for activities (based on how they were conducted) and achievements, and have been shown to increase learner motivation (Domínguez et al., 2013). Activity Streams (Olson et al., 2012) have been used to collect all activities from the various course systems to be presented within a dashboard.

Navi presents gold, silver and bronze badges to learners to reflect achievement. The badges have icons to give them context within the course and a description explaining the activities to be completed to achieve each badge. The dashboard supports details-on-demand views for each badge presenting the students that have been awarded with them. It also supports a drilldown view displaying how badges were acquired. The drilldown view is presented through a time-series line chart, where each line represents a badge and when it was awarded, presented by a circle, as shown in Figure 3.14. Educators can also view the line chart to analyse student progress and determine where the learners find the material challenging.

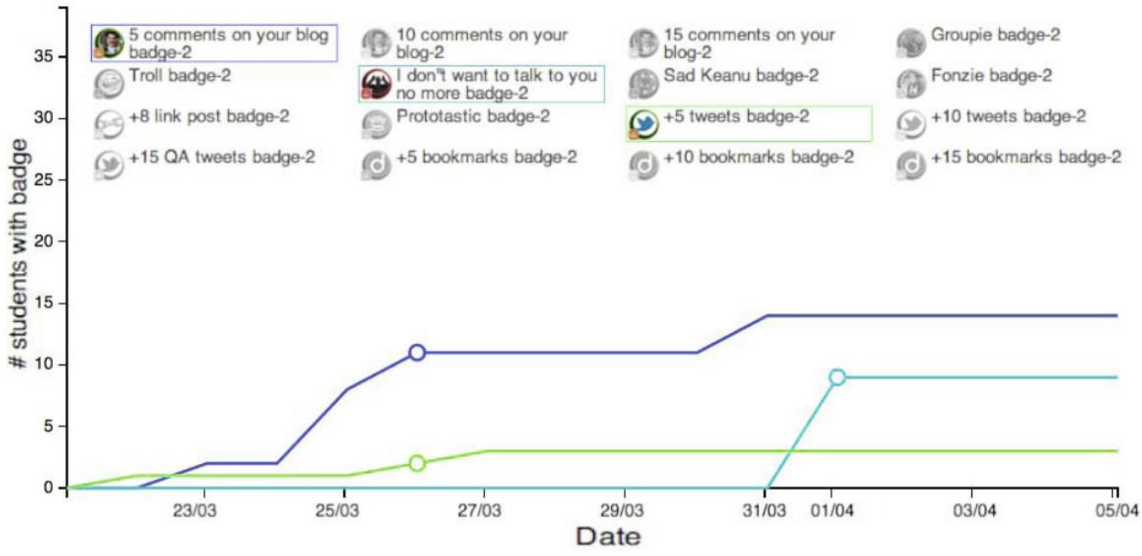

Figure 3.14 Navi Dashboard: Badges awarded over times

The Navi dashboard was evaluated by analysing the logged data and questionnaire responses of 26 engineering students partaking in a user interfaces course. The course used to evaluate the Navi dashboard involved students blogging, tweeting and bookmarking activities, which all contained activity action, author, date and artefacts (Charleer et al., 2013). The questionnaire focused on the frequency of use, the usability, and the usefulness of the dashboard and the badges. The log analysis showed that the dashboard had high usage during face-to-face sessions with educators. The user feedback suggested that the dashboard was rarely used to reflect on activities, regardless, the learners perceived the badges to motivate them to post tweets and comments and to read others' blogs.

Analysis

Navi aims to motivate learners by awarding gold, silver and bronze badges visualised through a dashboard. It supports visual interactions including drilldown views and supports social comparisons. It enables both learners and educators to view data in real time and interact with the system to explore when the badges were awarded. The social comparisons allow learners to view the badges awarded to peers.

The dashboard presents an aggregation of students that have been awarded each badge, however transformations visualising related student activity data are not provided. For example, a derived data transformation may be used to present a visualisation showing the time spent by individual learners, who have achieved the badge associated with a learning activity. Such a visualisation may be linked to the time-series line chart, shown in Figure 3.14 to present related data to support further explorations. Similar to the other tools discussed in this chapter, the Navi dashboard does not support visual narratives to present the learner data and badge awards.

3.7.10 StepUp

The StepUp dashboard (Santos et al., 2013a) provides visualisations of learner data to assist students in the learning process and to promote reflection. Prior to the development of SetUp, 56 students from three courses were involved in three brainstorming sessions to highlight learning issues and students' needs. The participating learners were from a Multimedia course, a problem solving and design course and Masters degree students. The session highlighted 34 student issues, seven of which were used as areas of focus, in the design of StepUp. The issues addressed included: uninvolved group members, group communication, time distribution, alerts, student motivation, balance between social activities and studies, and awareness of resources.

Figure 3.15 StepUp Dashboard

The implementation of StepUp that followed, visualises the time students spent on course content, resource usage and social media interactions. Learners are clustered on the dashboard by named groups which they have been assigned to, as shown in Figure 3.15. Cells corresponding to group members show the number of posts made by the students. The dashboard also displays the number of posts, number of comments, the total number of hours spent on the course and the total number of tweets made by individual students, to support peer comparisons amongst learners. The dashboard contains sparklines (Tufte, 2006) showing the overall activity per learner. Students can interact with the dashboard and view a breakdown of their activity by selecting the sparklines to load bar chart visualisations presenting the time they have spent on different activities over the week, and the distribution of posts during the week.

StepUp was used and evaluated during the three courses by the 56 students that were involved in the brainstorming sessions described above (Santos et al., 2013a). The evaluation consisted of a survey that focused on learning issues and how the tool addressed issues raised during the brainstorming session. The responses regarding how well the tool addressed learner issues were quite mixed. Students did not believe StepUp raised their awareness of resources and tools, nor was it perceived to have assisted with peer comparisons. Learners were also unsure if they were motivated by the visualisations. However, all the students believed the StepUp visualisations enabled them to reflect on how they were performing in the course and the Masters degree students believed the tool helped them analyse how they spent their time. Overall the findings showed that StepUp was more effective for learners assigned to groups and working through a topic than learners working individually on different topics. The evaluation also analysed students' logged data to extract page views and visit durations over time. The evaluation of earlier work (Santos et al., 2012b) focused on the usefulness and usability of the tool, and showed that course activity times were useful for peer comparisons.

Analysis

The StepUp evaluation highlighted that the dashboard supports self-reflection amongst learners allowing them to analyse how their time was distributed. The dashboard visualises student posts and comments, supports interaction techniques and facilitates peer comparisons. It presents real-time data and learners can interact with the sparklines to explore drilldown views showing the time spent on different activities and the distribution of posts over the weeks of the course. It also allows learners to view activities details by selecting the sparklines.

The StepUp dashboard presents automatically populating bar chart visualisations when students interact with the sparklines, displaying activity times and post distributions. However, similar to the systems mentioned above, this is also limited to statistical measures. Derived data transformations, consisting of visualisations presenting data related to the input data and views shown via the dashboard are not supported. For instance, interacting with a sparkline for a top performing student may generate a visualisation, using derived data transformations, showing the resources used by similar performing learners.

The StepUp dashboard does not provide visual narratives of the activity data and post distribution to learners, which could be used to help make the messages more memorable.

3.8 Discussion

Chapter 2 analysed the state of the art in Information Visualisation and discussed tasks and features that were identified as important in supporting the data analysis process and assisting users in making sense of data. Three of these, as discussed at the start of this chapter, are relevant to the TEL domain, namely, data transformations, visual interactions and guidance through visual narratives. These were used in this chapter to analyse the state of the art in LA systems providing visualisations to students and educators. This chapter has described and analysed the state of the art educational systems that use student-logged data to present visualisations to both learners and educators, discussing the features supported and highlighting limitations. This description and analysis aided the identification of best practices, trends and limitations, which are discussed in this section.

3.8.1 Best Practices and Trends

Chapter 2 introduced visual interactions and discussed how it supports the data analysis process and assists users in making sense of data. From the descriptions and analysis of the educational systems discussed in this chapter, it can be seen that the majority support visual interactions to varying degrees. The implemented interaction techniques include select, zoom, filter, details-on-demand and drilldown views, and the evaluations published for these systems highlighted the positive impact the interactive visualisations had on the data analysis process.

A number of trends supported by the systems discussed in this chapter can be identified, including social comparisons, viewing real-time data and looking back at previous learner data, the visualisation techniques used, and the types of data they focused on. The majority of systems discussed above, presenting visualisations to learners support social comparisons using students' personal logged data. Peer comparisons can support learner motivation and engagement by allowing students to compare their own progress with peers (Vassileva & Sun, 2007, Linton & Schaefer, 2000). While there are some indications (Burleson et al., 2005) highlighting that students may at times be averse to peer comparisons, studies have shown that upward social comparisons can help improve student performances (Dijkstra et al., 2008, Shepherd et al., 1995). In addition, the majority of systems presented real-time data and enabled learners and educators to interact with the visualisation to view activity data from other periods during the course. The visualisation techniques supported by the systems mainly consist of common techniques, such as bar charts, line charts, pie charts, and scatterplots, nevertheless, there were some exceptions, such as parallel coordinates, network graphs and bespoke visualisations (Travis and LARAe). These visualisations focused on presenting learner data such as resource usage, tasks times and the level of student progress with learning activities.

3.8.2 Limitations

Chapter 2 introduced data transformations and highlighted the importance of visual analytics tools supporting transformations of input data (Heer & Schneiderman, 2012). The analysis sections in this chapter highlighted that the majority of systems described support some degree of data transformations, however, these are limited to statistical measures, such as averages and aggregates. What is clear from the analysis is that these systems do not support derived data transformations of the input data (learner logged data), which maybe generated using input data and the data which is used in the visualisation that the student or educator is analysing. These transformations present an opportunity for consumers to explore data beyond the affordances of visual interactions techniques. Chapter 2 highlighted the importance of such transformations. Derived data transformations present visual explorations to users beyond statistical measures. These transformations can enable the students or educators to view data that is related to the visualisations and enable them to explore data beyond the original message communicated.

A large number of the systems discussed in this chapter use a dashboard to provide visualisations to both learners and educators, some of which support coordinated visualisations. Encouraging results have been highlighted that show that both students and educators can benefit from such formats and learners can improve their academic results through individualised dashboards and visualisations (Thille & Smith, 2011, Arnold, 2010). Dashboards aim to provide students with insights into their learning activities to promote a positive outcome in behaviour (Verbert at al., 2013), however, they offer limited guidance to users (Harrer, 2015). Chapter 2 introduced visual narratives which consist of sequenced visualisations with textual descriptions which help guide consumers and make the message more memorable (Austin 2011). In addition, they provide effective ways of highlighting facts, making points and passing on information (Kosara & Mackinlay 2013). From the descriptions and analysis of the systems described above, none of these support visual narratives as a means to present student data.

3.8.3 Summary

Table 3-1 provides a summary of the features supported, best practices and limitations of the fifteen learning systems discussed in this chapter.

Name	Visualisation user	Visualisations Supported	Visual Interactions	Personalised Views	Social Comparison	Data Transformations	Visualisation Format	Evaluations
SAM	students, educators	bar and line charts, parallel coordinates	drilldown, filter, search	ISD	PC	SM	multiple panels	questionnaire (efficiency and usability)
CAM Dashboard	students	bar, moon and line charts	coordinated views, drilldown, zoom, filter	ISD	PC	SM	dashboard	questionnaire (usefulness, usability and effectiveness)
Moodog Plug-in	students, educators	coloured bars	-	ISD	PC	SM	overlay and stand-alone	Moodle interactions analysis
MACE Zeitgeist	students	line chart	filter	ISD	PC	SM	dashboard	-
LARAe	students, educators	bespoke visualisation	drilldown	ISD	PC	SM	dashboard	-
eMUSe	students, educators	bar and pie chart	-	ISD	PC	SM	standalone views	log analysis and learner feedback (motivations)
ALAS-KA	students, educators	bar and pie charts	-	ISD	class average	SM	individual views in an interface	log analysis (identifying correlations)

TrAVis	students, educators	bespoke visualisation, radar graphs	selecting bars	ISD	group comparisons	SM	dashboard	usability questionnaire
Navi Dashboard	students, educators	badges and time-series line chart	details-on-demand and drilldown	ISD	PC	SM	dashboard	questionnaire (usefulness and usability), log analysis
StepUp	students, educators	sparklines, bar charts	drilldown	ISD	PC	SM	dashboard	Survey and SUS
Inst. Intelligence System	educators	bar chart, colour-coded bar with nodes	selecting bars	-	-	-	dashboard	-
CourseVis	educators	scatterplot, line chart, bespoke visualisation	rotating views, brushing and zooming	-	-	SM	OpenDx	focus group, experimental study ad interviews
GISMO	educators	bar chart, scatterplot	filter	-	-	SM	login, time and filter views	case study
SNAPP	educator	sociogram and ego-network	filter, node selection	-	-	-	standalone	-
LOCO-Analyst	educator	bar chart and bespoke visualisation	drilldown	-	-	SM	multiple panels	questionnaire (18 experienced educators)

Table 3-1 System Summaries (ISD: Individual Student Data; PC: Peer Comparisons; SM: Statistical Measures)

The columns in Table 3-1 address a number of important features, based on identified trends, best practices and limitations. The types of visual interactions supported by the systems are also summarised, and it can be seen that most of the learning systems provide a degree of support to manipulate views. The table also shows that the visualisations are individualised for the users, and it can be seen that this is the case if the target users are students, their personal logged data is used. It can be seen from the social comparison column that almost all of the systems presenting visualisations to students support peer comparison. The data transformation column shows that the majority of systems support transformations by presenting various statistical measures for the learner data. However, derived data transformations are not supported. The penultimate column mentions the presentation type, highlighting that the browser panels and dashboard are amongst the most popular but also shows that visual narratives are not supported by any of the systems. The final column specifies the type of evaluation conducted, which include logged data analysis and questionnaires.

The evaluations of the systems discussed in this chapter highlight the positive impact the visualisations had on learners. Literature (Li et al., 2010) suggests a positive behavioural change if learners become aware of their data through visualisations, reflect upon it by making sense of what is needed to improve, and driving impact by taking the necessary steps.

From the analysis of the state of the art systems in LA using both evaluation findings from these systems and the important features and tasks to support visual analysis introduced in Chapter 2, the best practices and trends have been identified. In addition, gaps have been identified in the state of the art, namely the limitations that have been discussed in this chapter. Using these gaps, best practices and trends, including those identified in Chapter 2, a list of requirements can be drawn up to support the construction of visual narratives that can be consumed to enable educators to monitor student behaviour and performance, and help motivate learners to enhance their engagement with learning systems. These requirements include the following:

- providing interactive visualisations to support view manipulations,
- presenting learners using personal logged data and supporting peer comparisons,
- presenting real-time data, which includes resource usage, tasks times and progress, and supporting users to look back through logged data.
- supporting data transformation that present statistical measures and related data views,

- guiding students and educators through data to help communicate the intended message and to minimise misinterpretations of data.

These requirements will be used in the design chapter to guide the definition, design and implementation of the *derived data approach*.

4 Design

This chapter defines and provides an overview of *the derived data approach* (developed to satisfy the Research Question) to visual narrative construction and consumption. This definition is followed by a detailed description of the design of a novel framework, called VisEN (Visual Exploration through Narratives), that realises this *approach*. Both the definition and the design details address research objective two (definition of the *approach*) and three (design of a framework to realise the *approach*) outlined in section 1.3 of Chapter 1. Chapter 2 identified important Information Visualisation tasks and features that support users in making sense of digital data. Chapter 3 used the tasks and features that were relevant to TEL to analyse the state of the art LA systems that present visualisations using student-logged data to educators and learners. From this analysis, a set of requirements were derived that will be used in this chapter to guide the definition and design.

4.1 Introduction

The analysis of the state of the art in educational systems in LA discussed in Chapter 3 highlighted best practices and trends in using Information Visualisations. It also identified gaps or limitations within the state of the art. The definition of the *approach* and design of the VisEN framework use the best practices identified in the state of the art and include mechanisms to address the gaps. This chapter first defines the *approach* using the influences (best practices and limitations) from the state of the art and then provides an overview of the *approach*. The framework requirements are then defined and its description is provided. This is followed by detailed design specifications of the various components, including its sub-components and models supported by the framework. Finally, the scope of the framework is discussed.

4.2 Approach Definition

The Research Question addressed by this thesis is: *Can explorable visual narratives support consumers in understanding, benefiting from and gaining insights into data from the TEL domain; and to what extent can authors be appropriately supported in producing these narratives?"* Research objective one focused on analysing the state of the art in Information Visualisation, to identify the extent to which authors are supported in communicating student-logged data, in a format that is easily consumable. It also focused on analysing the extent to which consumers are supported in understanding student-logged data and gaining deeper insights into it. The analysis of the state of the art in Information Visualisation identified a list of important tasks and features that should be supported by visualisation systems (Section 2.7, page 43). Relevant tasks and features were used to analyse of the state of the art in TEL, specifically in LA, and influences from both sets of analyses were used to define the *derived data approach* to address the Research Question.

The analysis of the state of the art in Information Visualisation (Chapter 2) discussed data access and visual encodings to support the communication of data. It identified best practices and compared and contrasted the mapping of data to graphical marks against its mapping to visualisation techniques. Based on the analysis of both data access and visual encoding, it is important for the *approach* to support the use of data which is accessible from both databases and file systems, and to support one of the two visual encoding techniques discussed (Section 2.3, pages 17-20). With the multitude of visualisation libraries available, it was decided to map data to visualisation techniques. Using these libraries also reduces the coding effort as visualisations do not need to be hand coded.

Visual interactions consist of mechanisms and techniques that support users in understanding data by enabling them to analyse and manipulate visual representations (Kosara et al., 2003). The review of the state of the art in Information Visualisation, discussed in Chapter 2, highlighted the role of visual interactions in supporting users in making sense of data. It also discussed a number of evaluations (Yi et al., 2005, Heer & Boyd, 2005, Heer et al., 2008) which showed that participants in trials benefited from the supported visual interactions, and in general, these interactions assisted them in understanding the data. The analysis of the state of the art in TEL, discussed in Chapter 3, showed that the majority of educational systems, presenting visualisations using student-logged data to educators or students, support some level of visual interactions to enable users to manipulate views. Evaluations (Govaerts et al., 2012, Santos et al., 2012) discussed in Chapter 3 also described how students and educators benefited from the support offered through visual interactions. Hence it is important that the *approach* supports, at minimum, the popular visual interaction techniques supported within TEL, which include select, filter, drilldown views, details-on-demand and zoom (Section 3.8.1, page 80).

Chapter 2 discussed Information Visualisation reference models (Card et al., 1999, Chi, 2000) which include data transformations that provide derived values and derived structure for input data (Card et al., 1999). The discussion highlighted that data transformations are popular (Spenke & Beilken, 2003, Shi et al., 2014, Xiao et al., 2006), but they are limited to providing data parsing, extraction and statistical measures. A similar trend was found in TEL where the majority of educational systems discussed in Chapter 3 did not support transformations beyond statistical measures. Examples were presented in both chapters showing how data related to the input data beyond statistical measures may support educators and students to in exploring data sources. For example, if a learner is analysing a visualisation showing his/her resource usage timings, derived data explorations, including the resource usage times of peers and top performing students, can be generated dynamically and made available to the learner. This can allow a student to explore data beyond the view that is presented through the visualisations. Hence it is important for the *approach* to both support transformations that present statistical measures but also go beyond these to dynamically generate and present derived data explorations.

Chapter 2 discussed the emergence of visual narrative in Information Visualisation and highlighted how these can support the process of communicating a message. Visual narratives can be defined as a set of sequenced visualisations containing textual descriptions that focus on data (Kosara & Mackinlay 2013). They are important as they provide an effective way of passing on information and make the communicated message more memorable (Austin 2011). Chapter 2 described a number of systems that support the construction and consumption of visual narratives. It was noted that there was limited support for authors constructing visual narratives, particularly in suggesting visualisation sequences. Chapter 2 highlighted that authors could be supported in visual narrative construction through automatic sequencing (Hullman et al., 2013a), and these sequences could be generated using derived data slices.

In addition, limitations in supporting the exploration of visual narratives were also noted, such as viewing data related to the visual narrative. It was highlighted that derived data explorations could address this gap by presenting visualisations that consumers could use to explore the dataset used to construct the visual narrative. The analysis of the state of the art in Chapter 3 showed that visual narratives are currently not supported in TEL. However, as visual narratives provide an effective way of passing on information, coupled with the successful evaluations from the Information Visualisation domain, it was decided that the *approach* should support visual narratives that contain derived data explorations. This is important in TEL as consumers (students, educators and other interested stakeholders) could navigate through and interact with an intended message delivered in a structured format and explore it through related data to gain deeper insights.

As mentioned in Chapter 1, derived data transformations is the general term for derived data slices and derived data explorations and all three terms are synonymous. In the context of narrative construction, derived data transformations are referred to as derived data slices which support the construction of the visual narrative. In the context of narrative consumption, derived data transformations are referred to as derived data explorations which support the consumption of the visual narrative.

Chapter 2 discussed personalisation in Information Visualisation, which focused on eye gauge tracking, adapting visualisations to user preferences, highlighting visual elements corresponding to text the user is reading, and detecting frustration. Chapter 3 highlighted that students' personal logged data is used within visualisations to individualise the message. Based on these findings, the *approach* allows authors to construct visual narratives that can be individualised (using a learner's personal logged data) to each consumer once the narrative is completed. In addition the *approach* adapts the ordering of the presentation of the explorations to consumers based on their preferences and exploration patterns.

The guiding influences (data access, visual encodings and visual interactions) and limitations (related data presented to support the construction and consumption of visual narratives) identified in the state of the art, and discussed in this section, have led to the defining of *the derived data approach to the construction and consumption of explorable visual narratives in TEL*.

4.3 Approach Overview

The *approach* supports of a set of users, known as authors who have some level of knowledge of one or more data sources, in communicating the data in an explorable visual narrative format. Authors construct visual narratives using a data source of interest. A visual narrative in the *approach* is composed of narrative slices, which are the individual components or sequences of the narrative. Authors construct narrative slices by specifying data from the chosen data source and providing textual descriptions. To enable authors to construct narrative slices, the *approach* provides a component, called the Narrative Builder that allows authors to connect to a data source and specify data to communicate, accompanied by textual descriptions. Each narrative slice has an accompanying visualisation which renders the data specified by the author. A set of appropriate visualisations for the specified data is generated, using another component supplied by the *approach*, called the Visualisation Engine, which consists of pre-defined visualisation schemas. The *approach* consists of a model, called the Derived Data Model, which generates derived data slices using the data related to the previously created narrative slice, and these are presented as visualisations to support authors in creating further narrative slices.

The *approach* supports a set of users, known as consumers, in navigating through, interacting with and exploring the constructed visual narratives. It provides a component called the Visual Narrative Explorer which enables consumers to select a visual narrative to view, and specify the level of exploration. Consumers can navigate through the visual narrative, interact with the visualisations and explore data related to the narrative slices through visualisations generated by the Derived Data Model (derived data explorations). The explorations are also tailored to the consumer based on usage patterns. Figure 4.1 provides a high level view of the *approach*.

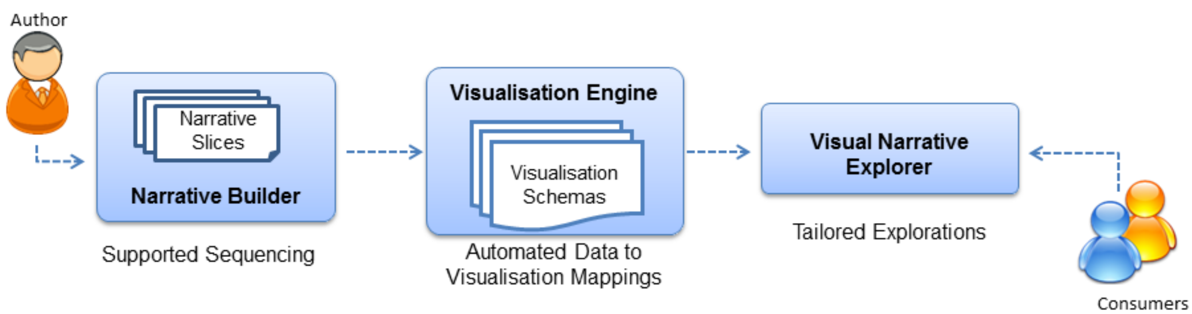

Figure 4.1 Approach Overview

4.4 Defining Framework Requirements

This thesis addresses the *approach* of supporting authors (individuals with access to data that they have an understanding of) to effectively produce visual narratives that can be navigated through, interacted with, and explored by consumers. This section outlines both author and consumer requirements based upon 1) best practices and 2) addressing limitations identified within the state of the art. These are then used to define a list of framework requirements, which are outlined at the end of this section.

4.4.1 Supporting Visual Narrative Authoring

To support the visual narrative construction process, the following author requirements need to be addressed:

- mechanisms to enable authors to access data, including connecting to a database or uploading tabular data from the file system,
- mechanisms to allow authors to select data of interest to include in a narrative and to specify filters,
- appropriate visualisation techniques offered to authors once they have specified data, to present views of the data they have selected. Authors can then choose one visualisation technique from the set to use in the narrative slice,
- presenting visualisations to authors of data related to the constructed narrative slice through derived data explorations, and allowing authors to decide which ones, if any, to allow consumers to view for their exploration,
- presenting possible sequences of narrative slices to authors through derived data slices.

To address these requirements, the framework needs to support:

- connection to data sources and tabular data files,
- generation of a set of appropriate visualisations through author specified data and filters,
- suggestions of potential narrative slices through the generation of derived data slices, presented in the form of visualisations.

4.4.2 Supporting Visual Narrative Navigation and Exploration

To support the navigation, interaction, exploration, and understanding of data in the visual narratives, the following consumer requirements need to be addressed:

- empowerment of consumers to view, navigate and interact with visual narratives with minimal disruption to the story flow,
- facilitation of the exploration of the data of the visual narrative to gain a deeper understanding of the communicated message,
- assistance of consumers in exploring the visual narrative by inferring user preferences and ordering explorations accordingly.

To address these requirements the framework needs to support:

1. exploration of visual narratives through derived data explorations of narrative slices while minimising the disruption to the flow of the story,
2. tailoring the order of the derived data explorations using statistics of the consumer's usage of explorations to infer preference,
3. supporting view manipulations of narrative slices through interaction techniques.

4.4.3 Framework Requirements

Based upon author and consumer requirements outlined above, which incorporate the best practices and limitations identified and discussed in Chapter 3, the following seven design requirements have been defined for the VisEN framework. The framework is required to:

1. support authors, who have access to and an understanding of some data, to produce visual narratives, by automatically generating derived data slices (based on narrative slices they have constructed) which are presented as visualisations. These visualisations are suggestions that can be included in the visual narrative,
2. support consumers in exploring visual narratives by automatically generating derived data explorations of narrative slices and providing links to these visualisations. These links should load separately from the narrative, to enable data exploration without disrupting the flow of the visual narrative,
3. tailor the ordering of the derived data exploration visualisations, mentioned in point 2 above, by inferring consumer preferences based on the usage of these views,
4. provide mechanisms to support access to tabular data sources with numeric data,
5. provide a process to facilitate visual encoding by mapping narrative slice data to an appropriate set of visualisation techniques,
6. provide operations to enable the authoring, and mechanisms to piece narrative slices together to produce a visual narrative,
7. support consumers manipulate narrative slice views through appropriate interaction techniques.

Requirements 1 - 3 are novel and requirements 4 – 7 are best practices outlined by the state of the art analysis.

4.5 Framework Description

This section describes how the requirements outlined above led to the design of the VisEN framework, discussing the functionality of the various components and models. The discussion first focuses on the narrative construction design followed by the visual narrative navigation and exploration design.

4.5.1 Visual Narrative Construction Design

Figure 4.2 shows the component and engine that address the narrative construction requirements presented in section 4.4.3. The component responsible for data access and narrative authoring is called the Narrative Builder. The mapping of data to appropriate visualisations is handled by an engine called the Visualisation Engine.

At step one in Figure 4.2, the author connects to a data source of choice, which can include databases or tabular data files. Authors connect to a database by specifying connection parameters or selecting preconfigured data source parameters, which are stored in configuration files. Similarly, tabular data files can be uploaded by authors. The Data Connection sub-component of the Narrative Builder presents the metadata (table names and columns) to authors allowing the structure of the data to be analysed.

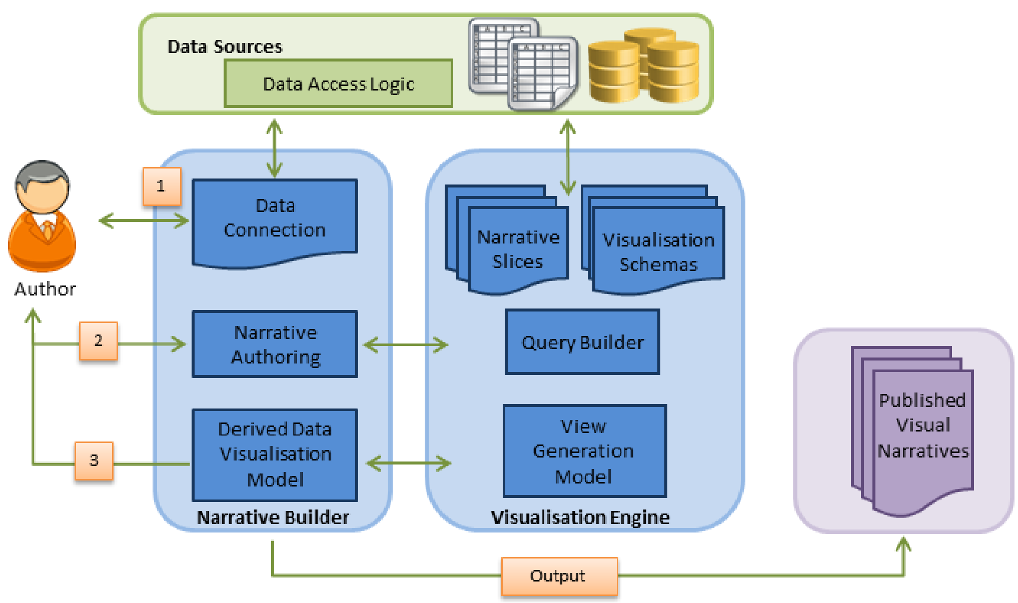

Figure 4.2 Visual Narrative Construction design

One of the design requirements of the framework is to provide mechanisms to support authors in constructing visual narratives. At step two in Figure 4.2, the author uses the Narrative Authoring sub-component to construct individual narrative slices. The constructed narrative slices consist of metadata, data fields and filters, all of which are used by the Visualisation Engine to select a set of appropriate visualisation techniques to render the data. The set of visualisation techniques may consist of more than one technique, in which case the author must select one when saving the narrative slice. Once all the narrative slices are created, the author connects them in a preferred order using the Narrative Authoring sub-component. In step three of Figure 4.2, the author is presented with a set of visualisations showing data related to the previous narrative slice through derived data transformations. The author can analyse these and use some of the suggestions to construct further narrative slices and others can be specified as explorations for the visual narrative (made visible to consumers as explorations). The author can specify which derived data explorations to include and exclude from the set to present as explorations to consumers. By default, all of the visualisations in the set are made available to consumers.

4.5.2 Visual Narrative Navigation and Exploration Design

Figure 4.3 shows the component and the model that addresses the visual narrative navigation and exploration requirements presented in section 4.4.3. The component responsible for supporting consumer exploration is called the Visual Narrative Explorer, which consists of the Visual Narrative Viewer and the Tailoring Engine sub-components. The Visual Narrative Explorer communicates with the Derived Data Visualisation Model to present derived data explorations for the visual narrative that can be explored by consumers to gain insights. Only the derived data explorations specified by the author are presented, but, all of them are shown if none of them were specified by the author.

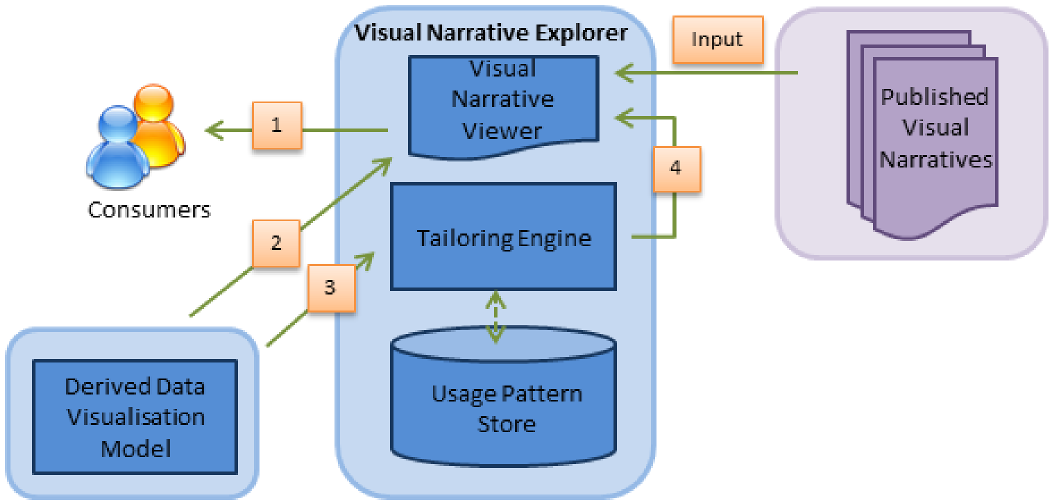

Figure 4.3 Visual Narrative Exploration design

One of the key and novel requirements of the framework is to show visualisations presenting derived data explorations of a constructed narrative slice to consumers, which can be used to enable them to explore the visual narrative. At step two of Figure 4.3, the Derived Data Visualisation Model generates derived data objects for each narrative slice of the selected visual narrative using the preferences specified by the consumer and the derived data explorations that the author has included. These are made available to the consumer through the Visual Narrative Viewer. One of the requirements is to support the exploration of visual narratives but also to minimise the disruption to the story flow. Hence the derived data explorations are not shown as part of the visual narrative, but as separate links within the interface (derived data explorations links).

At step three and four of Figure 4.3, the Tailoring Engine examines the consumers' specified derived data exploration settings, the derived data explorations from the Derived data Visualisation Model, and the usage pattern history in the Usage Pattern Store sub-component. It then orders the derived data exploration links accordingly.

4.5.3 Summary of Framework Components and Models

Table 4-1 presents a summary of the functionality of the framework's sub-components and models. The table specifies whether the sub-component or model is related to visual narrative construction and/or consumption.

Name		Functionality	Narrative Construction	Narrative Consumption
Narrative Builder	Data Connection	Stores database connections. Enables tabular data file upload and handles database connections. Presents connected source metadata to authors.	Yes	
	Narrative Authoring	Facilitates the construction of visual narratives using metadata from the connected source and displays visualisations generated by the Visualisation Engine.	Yes	
	Derived Data Visualisation Model	Creates derived data objects using narrative slices constructed by authors and predefined mappings.	Yes	Yes
Visualisation Engine	Visualisation Schemas	Contains schemas describing the affordances and characteristics of visualisation techniques. It contains the mandatory and optional data requirements for visualisation techniques.	Yes	Yes
	Query Builder	Creates and executes database queries using metadata supplied in the narrative slices and or derived data objects.	Yes	Yes
	View Generation Model	Generates a set of appropriate visualisations from imported libraries using metadata, formatted query results and visualisation schemas.	Yes	Yes
Visual Narrative Explorer	Visual Narrative Viewer	Presents published visual narratives to consumers and supports navigation and interactions across the narrative slices.		Yes
	Tailoring Engine	Infers consumer preferences of derived data explorations based on the data supplied by the Usage Pattern Store and the Derived Data Visualisation Model.		Yes
	Usage Pattern Store	Stores consumers' derived data exploration usage patterns and session history.		Yes

Table 4-1 Summary of functionality of framework sub-components and models

4.6 Overview of VisEN framework

Figure 4.4 shows an overview of the VisEN framework including its interfaces, sub-components and models. Based on 1) the author and consumer requirements (based on influences from the state of the art), and 2) the resulting requirements and functionality of the framework, the following architecture has been created for the VisEN framework.

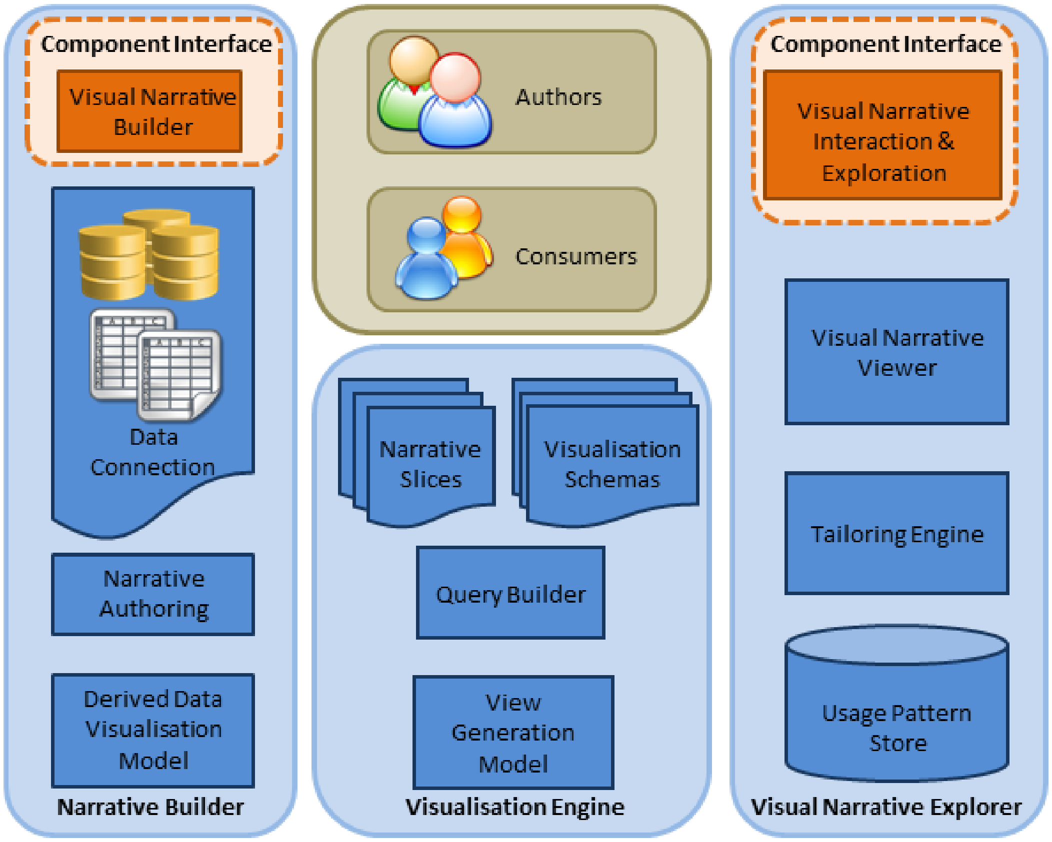

Figure 4.4 VisEN Framework

The subsequent sections 4.7, 4.8 and 4.9 describe the design details of the sub-components and models of the Narrative Builder, Visualisation Engine and Visual Narrative components of the VisEN framework. These descriptions include design specifications for data source access and the component interfaces that support the functionality.

4.7 Narrative Builder

This section describes the Narrative Builder, focusing on the design of the Data Connection and Narrative Authoring sub-components and the design of the Derived Data Visualisation Model.

4.7.1 Data Connection

The objective of the Data Connection sub-component is to support authors in uploading or connecting to a data source and formatting its tables and columns. Table 4-1 outlined the functionality of the Data Connection sub-component, which included connecting to data sources and presenting metadata. The requirements of the sub-component are to:

- provide a list of databases that the author can connect to,

- allow authors to connect to databases not on the list by supplying connection parameters,
- provide an extensible design to support various types of data models,
- support the upload of data from other formats available on the authors file system,
- format metadata from the source data to support the construction process.

Figure 4.5 presents the steps involved in connecting to data sources and outputting formatted metadata. At step one of Figure 4.5, authors are offered two options to connect to data sources. The first option enables authors to select a source file (tabular data) from their file directory through the Visual Narrative Builder interface by directly entering the path of a file or using the file lookup option provided by the interface. The second option allows authors to connect to a database using an existing connection parameter by selecting the name of the database from a list of saved database connections. The interface presents a list of names of saved database connections (previously used by the same author) and any of these can be used for connection. If a database is not available through this interface, the author can specify connection parameters by entering the database URI, name, username and password, which are then saved for future use. As mentioned in requirement four in section 4.4.3, the framework supports tabular and numeric data and the Data Connection sub-component ensures the data is numeric by examining the field types. For database connections, the column type is examined, however for tabular data files, the individual fields are examined. In addition, the source data must provide metadata for each column. This is not an issue with databases, however, tabular data files are examined to ensure the first column consists of metadata describing the data contents including table names and columns.

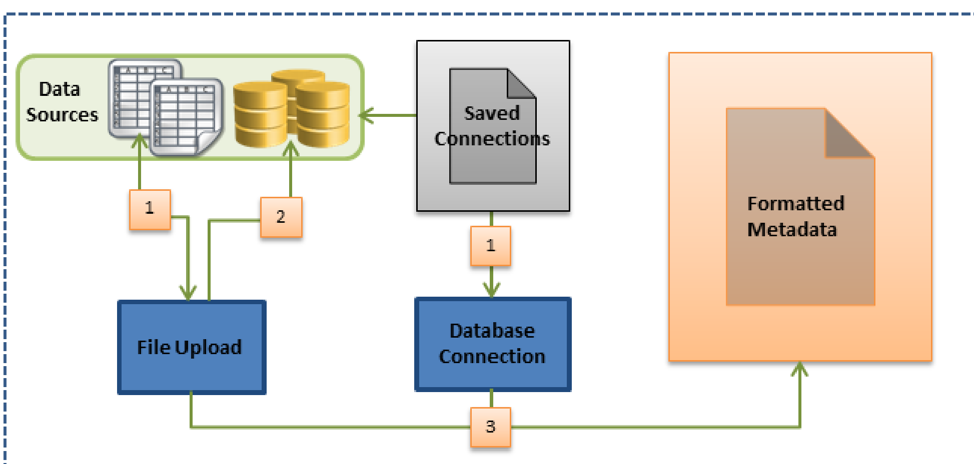

Figure 4.5 Data Connection design

If the author uploads source data from the file system, a database is created using the Database Server running on the same machine hosting the VisEN Framework, and the data is stored in it as indicated by step two. Once a new database is created, the data is stored in tables with columns matching the data types. The name of the database corresponds to the name of the file. Saving the file data in a database is required by the Visualisation Engine (described in section 4.6), which runs queries against the data source to fetch and validate query results for visualisations.

Finally, the metadata obtained through step one is formatted by applying mappings to database table names and columns. The mappings aim to address the presentation of the metadata by adjusting letter casing and replacing underscores with spaces. This formatting is only applied to database tables and not to files uploaded from the file system, as it can be assumed that file headings are already formatted for users.

4.7.2 Narrative Authoring

The objective of the Narrative Authoring sub-component is to support authors in constructing narrative slices and assembling visual narratives. Table 4-1 outlined the functionality of the Narrative Authoring sub-component, which includes selecting data and applying filters to build narrative slices and assembling visual narratives. The requirements of the sub-component are to:

- present the formatted data supplied by the Data Connection sub-component through an interface that allows the author to analyse the structure of the source data,
- allow authors to create narrative slices by specifying data fields and filters using the formatted metadata (which generates a set of appropriate visualisations by passing the data to the Visualisation Engine) and enter a description of the data,
- enable Authors to accept one of the visualisation techniques for a narrative slice returned in the set by the Visualisation Engine,
- support the construction of narrative slices by showing authors derived data slices for the narrative slices constructed by invoking the Derived Data Visualisation Model,
- enable authors to decide which of the derived data explorations are made available to consumer,
- allow authors to assemble visual narratives by connecting the saved narrative slices and enable them to be published.

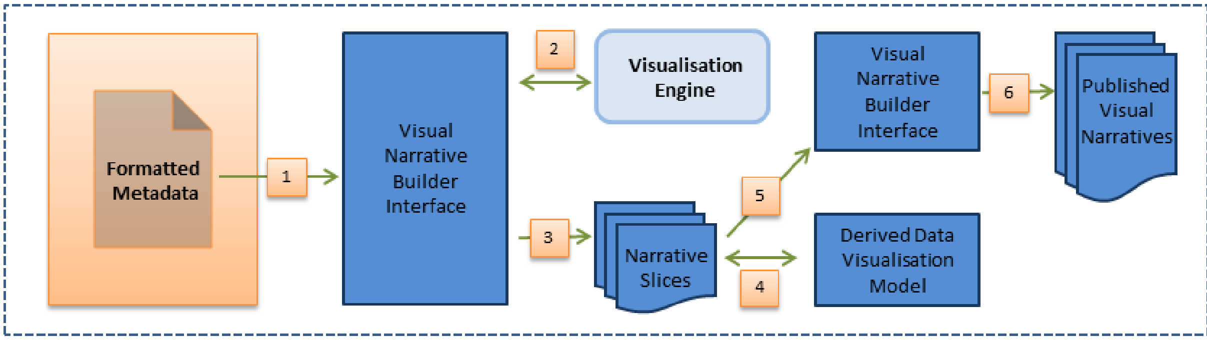

Figure 4.6 Narrative Authoring design

Figure 4.6 presents the steps involved in constructing narrative slices and assembling and publishing visual narratives. The Visual Narrative Builder Interface uses the formatted metadata (containing table names and columns) as indicated by step one in Figure 4.6 and presents it to authors. This interface allows authors to select column names and specify filters using selecting and linking operations provided by the interface. Filtering data fields allows authors to set data ranges, exclude and include certain values, etc. As authors specify data selections and/or filters, the Narrative Authoring sub-component invokes the Visualisation Engine, which returns a set of appropriate visualisation techniques. All of the visualisation techniques returned by the Visualisation Engine are presented to the author through the Visual Narrative Builder Interface, where he/she chooses one to use in the narrative slice by selecting it. The interface also provides fields to enter a description and a title for the narrative slice. The description is usually a summary of the analysis of the data presented by the visualisation technique. Finally the narrative slice can be saved as indicated by step three.

Step four shows the invocation of the Derived Data Visualisation Model, which returns a set of derived data transformations for a constructed narrative. These are presented as visualisations to the author through the Visual Narrative Builder Interface. The Derived Data Visualisation Model can be invoked once a narrative slices has been constructed. The authors can specify which of the derived data explorations to supply to consumers with the narrative slice through this interface by selecting the ones to include. If none are selected, then by default, all of the derived data explorations are made available to the consumers. The derived data slices offer suggestions to authors for further narrative slices, where they can use the interface to select a visualisation and save it as a narrative slice by supplying it with a description and saving it.

The Visual Narrative Builder Interface displays all of the author's saved narrative slices (step five) and allows him/her to specify which to include in the visual narrative by identifying and selecting the titles. Selected narrative slice titles get linked, where a narrative slice links to the previous one (if available) and the subsequent one (if available). Authors can adjust the ordering by moving the narrative slice titles and prioritising the order, which re-establishes the links between them. Finally the interface allows authors to enter a title for a narrative slice and publish it (step 6), making it available to view by consumers.

4.7.3 Derived Data Visualisation Model

The objective of the Derived Data Visualisation Model is to produce data objects related to narrative slices that can be used to generate visualisations to support the construction and exploration of visual narratives. Table 4-1 outlined the functionality of the Derived Data Visualisation Model, which uses the data fields, filter values and metadata from narrative slices and applies template based mappings to them to generate derived data objects. The requirements of the model are to:

- provide a set of mappings that can be applied to author selected data and filter fields in narrative slices to generate derived data objects,
- support the addition of new mappings at run time,
- support the lookup of appropriate transformations for narrative slices,
- apply appropriate mappings to narrative slices to create derived data objects,
- invoke the Visualisation Engine passing the derived data objects to generate derived data transformations.

The Derived Data Visualisation Model uses narrative slice data fields, filters and metadata to create derived data objects that are sent to the Visualisation Engine to generate derived data transformations. Derived data objects are created by extracting data fields, filters and metadata from a narrative slice and applying appropriate mappings to it. The Derived Data Visualisation Model consists of several preconfigured mappings, some of which are applied to generate statistical data such as maximum, minimum, average, etc. using the data fields in narrative slices to create derived data objects. For example, if a narrative slice presents the learning resources used by a student during the first month of the semester, the model will apply statistical mappings to the data fields in the narrative slice and create derived data objects presenting the resources with the maximum and minimum usage, and the average resource usage by the learner.

The Derived Data Visualisation Model also creates derived data objects containing data beyond statistical measures. For example the derived data objects can consist of new filter fields which are created by altering the time period from the above example and presenting the learner's resource usage over the entire course duration or altering the filter and creating related data showing the best students' resource usage. These alterations are handled through mappings. The framework design facilitates the addition of mappings at runtime, which are included in the lookup. The Derived Data Visualisation Model mappings adhere to the following structure:

- are encoded as tuples on the server,
- have conditions that must be satisfied by the data in the narrative slice for the mapping to be applied (for example, if the narrative slice data does not contain any filters, then the mappings handling filter data are not used),
- consist of tuples containing a key value pairs (if the narrative slice data fields or filters match the key, then the mapping is applied).

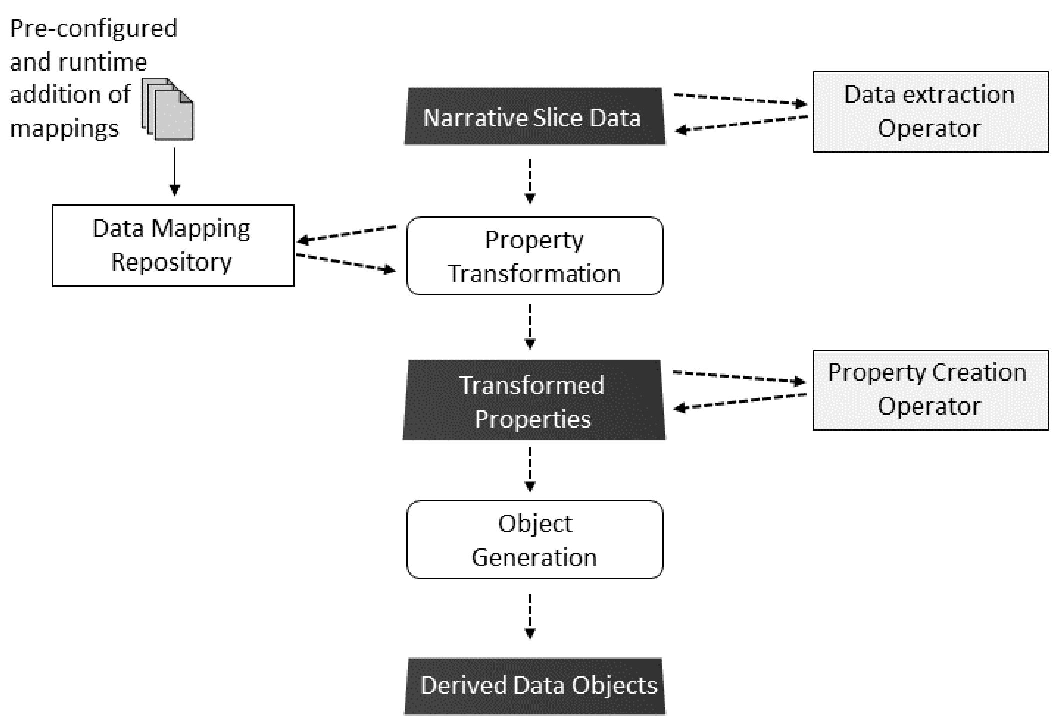

Figure 4.7 Derived Data Visualisation Model

Figure 4.7 shows the design of the Derived Data Visualisation Model, consisting of three data stages (Data, Transformed Properties and Derived data objects) and the two processing steps (Property Transformation and Object Generation) to create the derived data objects. Below are the descriptions of the data stages (two of which contain operators for extracting and setting values) and the processing steps:

- **Data**: the data includes metadata, data field values and filter field values from a narrative slice. This stage consists of the Data Extraction Operator, which extracts the metadata, data field and filter field values,
- **Property Transformation**: this processing step iterates through the set of mappings from the Data Mappings Repository returning a list of the mappings that can be applied to the narrative slice. For example, if the narrative slice consists of filters, then all the mappings handling filters are returned in the list,
- **Transformed Properties:** mappings from the returned list are used by the Property Creation Operator to create new properties of data fields, filter field values, and metadata, such as the titles and legends. Properties are created by applying the mappings to the narrative slice data fields, filters and metadata,
- **Object Generation**: an object is created for each returned mapping using the corresponding set of properties and metadata,
- **Derived Data Objects**: the set of objects (derived data objects) are the outputs of the model.

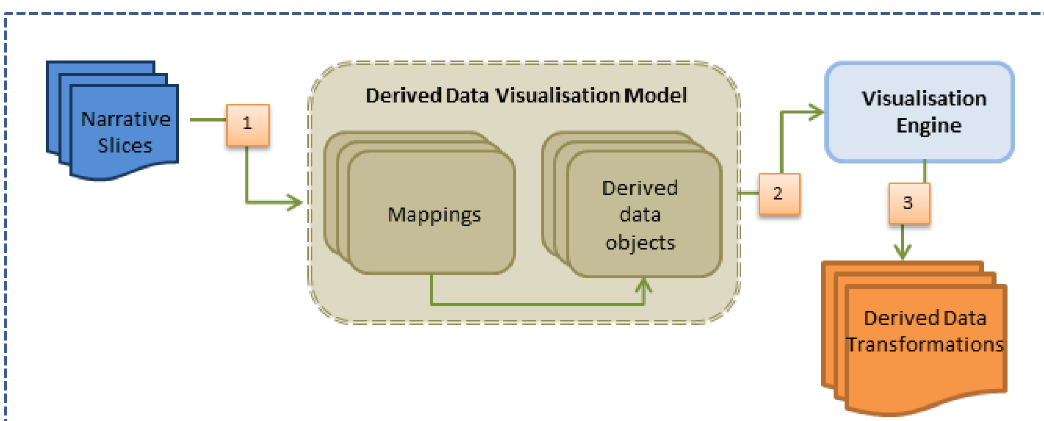

Figure 4.8 Generating Derived data transformations

Figure 4.8 presents the steps involved in generating a set of derived data transformations. The Derived Data Visualisation Model uses the data fields and filters specified in narrative slices to lookup a set of appropriate mappings. In step one, appropriate mappings are sought, each of which is used to create a derived data object. Several mappings can be matched for each narrative slice leading to the generation of multiple derived data objects. The derived data objects are passed to the Visualisation Engine, as shown in step two, which returns a visualisation technique for each object, called a derived data transformations, as shown in step three.

4.8 Visualisation Engine

Table 4-1 outlined the functionality of the sub-components and the View Generation Model of the Visualisation Engine, which use data, filters and metadata from a narrative slice to generate appropriate visualisations rendering the data. It also generates appropriate visualisations for the derived data objects. The following requirements are supported by the Visualisation Engine:

- uses visualisation techniques from imported libraries that can render numeric data,
- supports the addition of new visualisation techniques at run time,
- supports the encoding of affordances and characteristics of visualisation techniques in visualisation schemas and enables new encodings to be added at run time,
- extracts data fields and filter values from narrative slices and derived data objects to build queries that are run against the connected data source,
- uses the metadata from narrative slices and derived data objects, and the query results returned from the data source, to determine appropriate visualisation techniques,
- generates a set of visualisation techniques to render the narrative slice data and a visualisation technique to present derived data object properties.

Figure 4.9 shows the sub-components of the Visualisation Engine and its View Generation Model, and presents the steps it takes to generate visualisations. The design of the sub-components and the View Generation Model are discussed below.

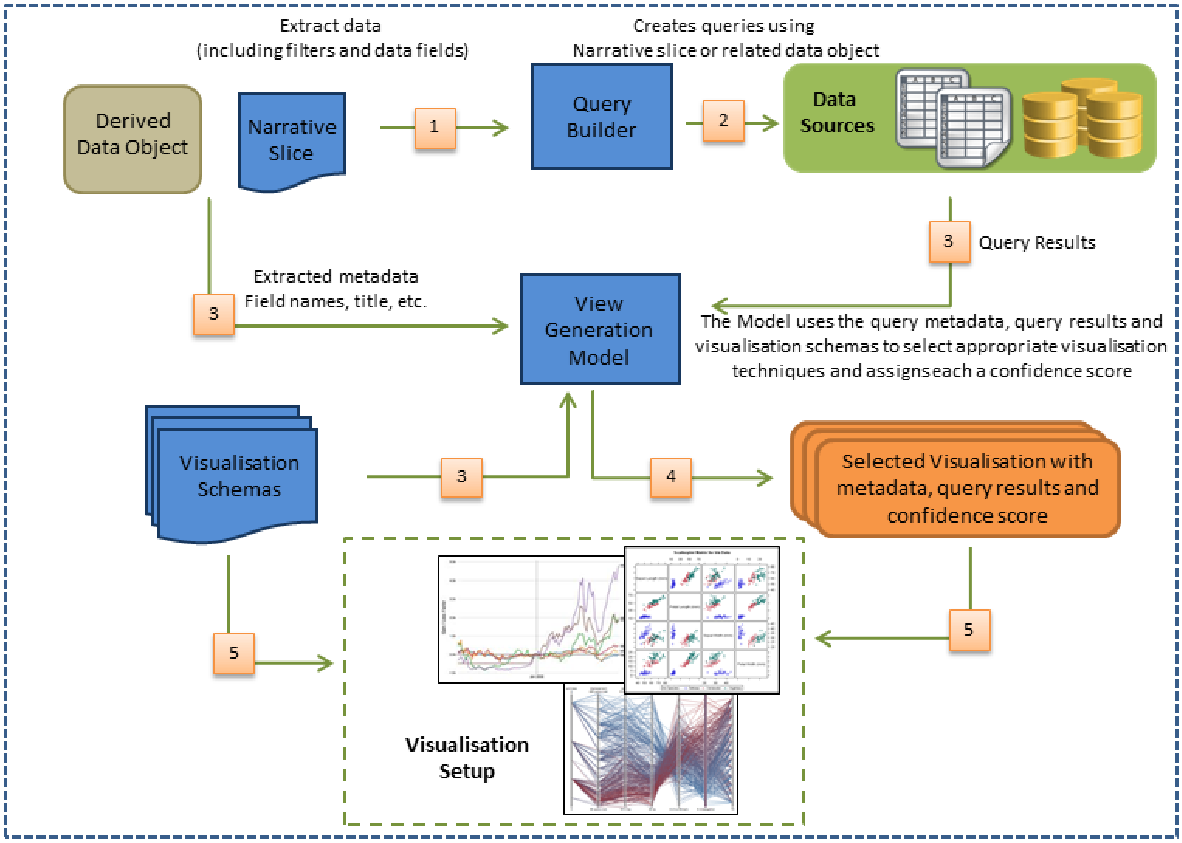

Figure 4.9 Visualisation Engine design

4.8.1 Visualisation Schemas

Visualisation schemas describe the characteristics and affordances of visualisation techniques and are used by the View Generation Model to find appropriate visualisations for data specified in the narrative slices and the derived data objects. Each visualisation technique supported by the framework must consist of a visualisation schema for it to be considered in the process to find appropriate visualisation techniques. Visualisation schemas are files stored on the server that describe visualisation technique requirements. The data requirements are specified as tuples consisting of key value pairs and consist of mandatory and optional data requirements. An example of a mandatory requirement key value pair can be 'data type' as the key and 'numeric' as the value. The number of elements that can be presented by the visualisation technique and the number of data series per technique are also amongst the mandatory data requirements. A data series represents values associated with one entity; for example, the engagement breakdown of a student per learning activity. A second data series could represent the average engagement breakdown of class members per learning activity. Both these data series can be shown through a single visualisation technique.

The schemas have a hierarchical structure and are divided into optional and mandatory data requirements, as separate objects. The mandatory data requirements must all be satisfied for a visualisation technique to be included in the set of techniques that can be used to render the data. The mandatory data requirements also specify the structure of the data file that is used by the visualisation technique to render the data. This specifies the format of titles, legends, data series names, and data series values. The optional requirements are used to assign a confidence score to the chosen visualisation technique in the event that multiple techniques can appropriately render the data. The confidence score assigned to a visualisation technique is determined by the number of optional requirements that are satisfied. The selection of visualisation techniques and the assigning of confidence scores is conducted by the View Generation Model. Table 4-2 presents the keys of the schema tuples including an explanation of the key, value type and whether the data requirement is mandatory (m) or optional (o). The suffix associated with some of the optional requirements is a number that is used to calculate the confidence score for a selected visualisation technique.

Key (Explanation)	**Value type**	**Required**
visualisation (name of the visualisation technique)	string	m
data type supported (quantitative, ordinal, nominal)	string	m
data range block object (combination of number of series and points)	-	m
→ data series (number of entities)	numeric	m
→ max data points (for the number of above entities)	numeric	m
→ min data points (for the number of above entities)	numeric	m
→ ideal data point (the number of points best suited for the series)	numeric	o<suffix>
coordinates per point	numeric	m
ideal coordinates (coordinates per point best suited)	numeric	o<suffix>
ideal data series (number of series best suited)	numeric	o<suffix>
ideal data points (number of points best suited)	numeric	o<suffix>
number of axes	numeric	o
x-axis (if required)	boolean	o

y-axis (if required)	boolean	o
data structure object (object describing the data format)	-	m
→ series name (entity name)	string	o<suffix>
→ data (format of the data)	string	m
→ colour (colour coding entities)	string	o
title (if required)	boolean	o
subtitle (if required)	boolean	o
legend (if required)	boolean	o
axis title (if required)	boolean	o

Table 4-2 Visualisation Schema Keys

The structure of the visualisation schemas follow a hierarchy with nested elements (denoted by arrows in Table 4-2). The data range block object in the visualisation schemas, shown in Table 4-2, allows several series and data point combinations to be encoded. For example, a visualisation technique can support several entities, each with its associated data points. However, as the number of entities increase, the number of data points that can be displayed are reduced. The data structure is an object describing the data format required by the visualisation technique, including series name, colour and data values format (e.g. comma separated values, arrays, etc.).

The data range values (including minimum and maximum data points), ideal data series and ideal data points can be configured for specific deployments, screen resolutions and capabilities of devices and can be adjusted at both deploy and run time.

The Visualisation Engine consists of a number of predefined visualisation schemas and a template that can be used to add new visualisation techniques to the framework from imported libraries. The template consists of optional and mandatory data requirement objects and sample tuples that can be used to create a visualisation schema. Adding new visualisation techniques and schemas does not require code updates. The view Generation Model only uses the requirements configured in the schemas to determine matches and does not require other coded rules. Hence, once a visualisation schema is created, it is automatically picked up at run time and does not require code updates.

4.8.2 Query Builder

The objective of the Query Builder is to use the data fields and the filter values supplied by the narrative slice or the derived data object to build queries that are run against the connected data source. The requirements of the Query Builder are to extract data fields, and filter values and conditions from narrative slices, and derived data objects, and to build and execute queries that conform to the standard of the data source used.

The data fields and the filter values from narrative slices are specified by authors via the Narrative Builder Interface; and the data fields and the filter values in the derived data objects are properties set by the Derived Data Visualisation Model. Narrative slice and derived data objects must consist of at least one data field. The Query Builder first extracts all the select data fields from the narrative slice or the derived data object, as shown in step one of Figure 4.9, and uses these to create a query statement. The order of the fields in the statement follow the same order used by the author in constructing the narrative slice or the order in the derived data object. If the narrative slice or derived data object consists of filter fields, then these are extracted and used to append filters to the query statement.

The generated query statement is executed against the connected data source as indicated by step two in Figure 4.9, and the result-set is passed to the View Generation Model.

4.8.3 View Generation Model

The objective of the View Generation Model is to generate a set of ranked visualisations using: 1) the imported libraries; 2) the data in the narrative slice or a derived data object; and 3) its corresponding query results. The methodology adopted by the View Generation Model follows that presented by the Information Visualisation Data State Reference Model (Chi, 2000). Step three in Figure 4.9 shows the model's inputs and outputs. Figure 4.10 shows the design of the View Generation Model, consisting of three data and view stages (Data, Visualisation Abstraction and Ranked Visualisation Set) and the two processing steps (Data Mapping and View Mapping) to generate a set of visualisations.

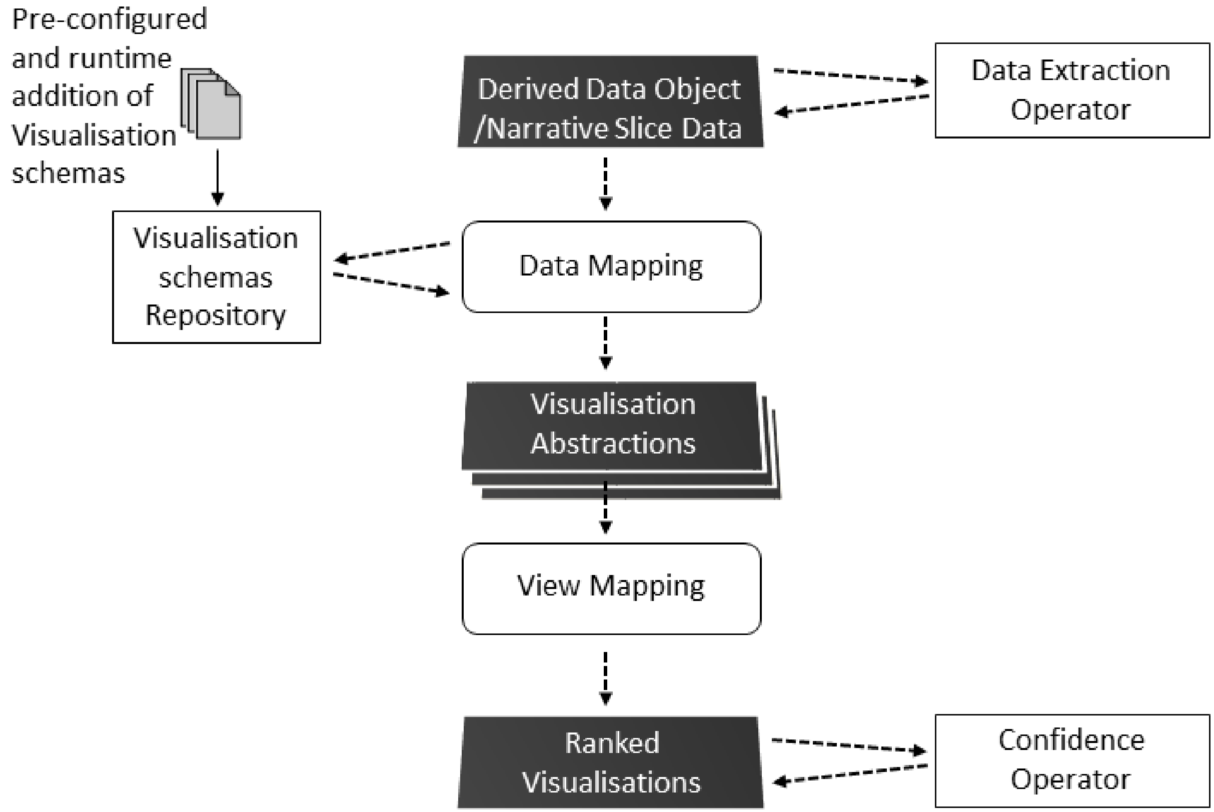

Figure 4.10 View Generation Model

The data and view stages and the processing steps are discussed below:

- **Data**: this stage consists of the Data Extraction Operator, which extracts the metadata, data field and filter field values from the narrative slice or the derived data object,
- **Data Mapping**: this processing step iterates through all the visualisation schemas (one per visualisation technique) and formats the query results according to each schema to determine if all the mandatory fields' requirements are satisfied. The visualisation techniques that can be used to display the data as a result of this step are marked,
- **Visualisation Abstraction**: this stage presents data in various formats so it can be visualised using any of the marked visualisation techniques,
- **View Mapping**: this processing step creates data objects adhering to the structure described in the visualisation schemas for the marked visualisation techniques, which are then invoked through the imported visualisation libraries,
- **Ranked Visualisations**: this stage consists of the Confidence Operator, which ranks the marked visualisation techniques according to how well the optional requirements described in the visualisation schema are met by the data.

4.8.3.1 Data Mapping

The Data Mapping processing step formats query results from the Query Builder, and the extracted data and metadata of the narrative slice or derived data object, to determine a set of visualisation techniques for the data. Visualisation schemas specify the format that the data must adhere to so it can be visualised, which can vary per technique. If the 'data type supported' key value in a schema matches the data type, the data is formatted according to 'coordinates per point' key value if possible. Following the data formatting, the processing step checks if any of the remaining mandatory requirements are satisfied in any of the schemas. This includes checking if the data range blocks in any of the schemas are satisfied. All the schemas that have the mandatory data requirements met are marked.

4.8.3.2 View Mapping

The View Mapping processing step uses the marked visualisation techniques from the Data Mapping processing step. As mentioned in section 4.6.1, other than the data requirement conditions, the visualisation schemas also consist of the structure that the data and metadata must adhere to so it can be processed by the imported libraries to produce the visualisation techniques. Table 4.2 included the data structure object, which describes the format of the data. The View Mapping processing step creates objects for the chosen visualisation techniques and sets the data properties according to the specification in visualisation schemas. It uses the metadata, such as title and data field names, to set the title and series name properties, and sets the data properties by formatting the result set accordingly.

4.8.3.3 Confidence Operator

For the marked visual schemas (satisfying the mandatory field requirement checks), the Confidence Operator analyses the optional field requirements, which include checks to determine how well the data meets the ideal display condition of the visualisation technique. Some of the optional requirements have a numeric suffix, as shown in Table 4.2, which are used to determine the confidence level of the visualisation techniques. The suffix with higher values maps to the higher confidence levels. The Confidence Operator determines the confidence level by adding the suffix values for each marked visualisation technique and ranks them according to the total suffix value. The confidence levels are low confidence, medium confidence and high confidence and the total suffix score is used to rank them accordingly (techniques with the highest totals fall into the high confidence category, followed by medium and then low). For cases where only one or two visualisation techniques can render the data, then the high confidence, or high confidence and medium confidence scores, are used respectively.

Finally, the View Generation Model creates a list of ranked visualisation objects (according to confidence scores) with each object containing the name of the visualisation technique that has been marked, its confidence score, formatted data and metadata including field names and titles. These parameters are used to generate visualisations using the imported libraries.

4.9 Visual Narrative Explorer

Table 4-1 outlined the functionality of the Visual Narrative Viewer sub-component, the Tailoring Engine, and the Usage Pattern Store, which use the published visual narratives and the Derived Data Visualisation Model to support consumers view, navigate, explore, and understand the communicated message. The following requirements are supported by the Visual Narrative explorer:

- present published visual narratives to consumers and support viewing, navigation and visual interactions,
- support consumer exploration by presenting derived data explorations,
- tailor the ordering of derived data explorations by inferring consumer preferences,
- minimise disruption to the flow of the story by separating the analysis of the derived data explorations from the visual narrative by presenting the derived data explorations as links (derived data explorations links) beside each narrative slice.

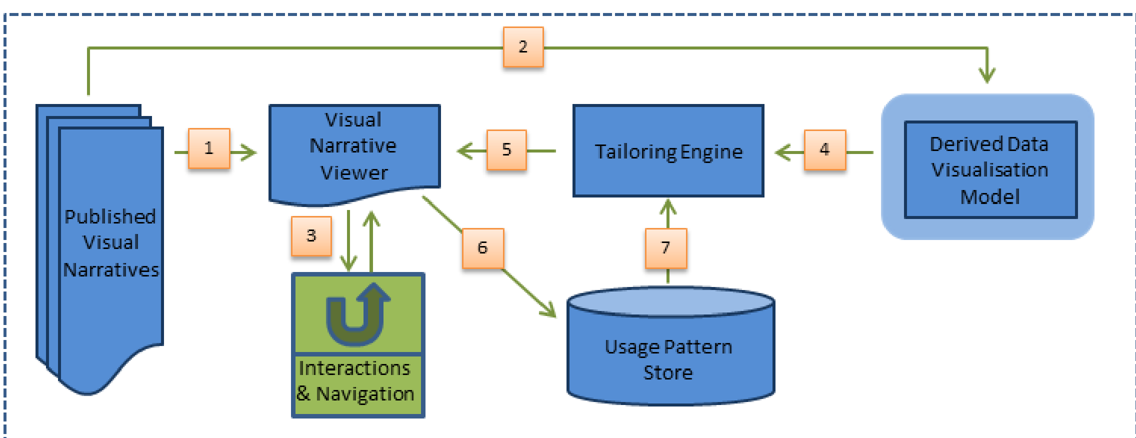

Figure 4.11 Visual Narrative Explorer design

4.9.1 Visual Narrative Viewer

The objective of the Visual Narrative Viewer is to enable consumers to view, navigate, interact with and explore published visual narratives. It consists of an interface presenting the titles of published visual narratives that can be selected by the consumer to load and navigate. The interface also presents the data used by the derived data explorations to the consumers to enable them to specify which explorations they are interested in. For example, if a published visual narrative presents course engagement and learning resource usage for a student, the derived data exploration options that can be selected by the consumer can include engagement and resource usage. Selecting course engagement as the exploration option will include derived data exploration of engagement, such as engagement breakdown per activity, engagement comparisons between class students and top engagement levels.

Selecting a visual narrative through the Visual Narrative Viewer causes all the narrative slices to appear individually within the interface, allowing the consumer to navigate across them to view the entire visual narrative, reading the descriptions and interacting with the visualisations (step three in Figure 4.11). The visual interactions include details-on-demand revealing element values, data series filtering to include or remove a series of values and zooming to view details of clustered values. The visual interactions come as part of the sourced visualisation techniques and are not included in the design of the Visual Narrative Viewer.

The Derived Data Visualisation Model (discussed in section 4.7.3), constructs derived data objects that are visualised by the Visualisation Engine. Once a visual narrative is selected (step one of Figure 4.11) by the consumer, the Visual Narrative Viewer extracts the narrative slices and presents links to the derived data explorations. It does this by calling the Derived Data Visualisation Model passing all the narrative slices (step two of Figure 4.11), which returns a set of derived data explorations (via the Visualisation Engine) for each narrative slice. Consumers can explore the derived data explorations by selecting the corresponding links to load the visualisations separately to the visual narrative. The consumer can analyse the derived data explorations individually and then return to the visual narrative.

4.9.2 Usage Pattern Store

The objective of the Usage Pattern Store is to monitor consumer activities and record the interaction patterns of derived data exploration usage. The Usage Pattern Store monitors consumer interactions with the derived data and the visualisations corresponding derived data exploration by logging events related to both the derived data and the corresponding visualisations. It is important that both data and visualisation logs are recorded to handle scenarios where certain data, for example a course engagement narrative may consist of derived data explorations showing the engagement level of a student and a category of students. In this scenario, there are two derived data explorations for course engagement data, resulting in two sets of logs for the visualisations and one set of logs for the data (course engagement). In this scenario, the Usage Pattern Store will record interactions for the course engagement data and the interactions with the visualisations. The events recorded include, the time spent viewing the data and visualisations, the number of times visualisations using the data were visited and the number of interactions carried out by the consumer against these visualisations. Consider an example where a narrative slice consists of a visualisation displaying a student's learning resources usage patterns. In this example, the derived data explorations can present:

- Visualisation 1: the student's course engagement patterns,
- Visualisation 2: the student's completed learning activities,
- Visualisation 3: all students' learning resource usage patterns,
- Visualisation 4: the best students' learning resource usage patterns.

In this scenario, the Usage Pattern Store will monitor and record activities against course engagement, learning activities, and learning resource data. Interactions, viewing times and the number of visits for visualisations 3 and 4 in the example are logged against the learning resource data and the two visualisations. Similarly interactions, viewing times and the number of visits for visualisations 1 and 2 are logged against course engagement and learning activities respectively and the two visualisations.

The Usage Pattern Store records data using maps with the string representation of the data and visualisation names as the keys and interactions, viewing times, and the number of visits as properties to objects mapping to the keys.

4.9.3 Tailoring Engine

The objective of the Tailoring Engine is to order the derived data explorations links presented with each narrative slice by inferring consumer preferences. The engine uses the three logs (interactions, viewing times and the number of visualisation revisits) per data and visualisations corresponding to each derived data exploration recorded by the Usage Pattern Store, to order the derived data exploration links accordingly. Consumer preferences are determined by assigning weightings to the three recorded logs and calculating user interests. The three recorded logs are assigned weightings as follows:

1. the number of exploration revisits has the highest weighting assigned to it as these repeat visits indicate that the consumer is interested in the related data,
2. visual interactions present consumer interest in the data, but this is not as strong as revisiting an exploration, hence the weighting is lower,
3. viewing times also indicates consumer interest in the data, however, an exploration can be loaded but the consumer's attention may be elsewhere and this needs to be taken into consideration, hence the weighting assigned to this log is the lowest.

Both the exploration and data scores are considered in the process to determine the ordering of the links, as this indicates the data that is of most interest to the consumer, and the data including any filters or configurations set by Derived Data Model. The Tailoring Engine iterates through the list of derived data exploration links, extracting the data and filter fields and applies these weightings. The derived data exploration links are displayed on the Visual Narrative Interaction and Explorer interface and their ordering corresponds to the calculated score. The derived data exploration link with the highest score is presented at the start of the list followed by the remaining links in descending score order.

4.10 Scope of the Framework

The focus of VisEN is to support the construction of visual narratives in TEL that can be navigated and explored by consumers. The framework provides authors with an interface to construct visual narratives and supplies consumers with an interface to navigate and explore these visual narratives. The design of these interfaces focusses on the construction and consumption of visual narratives as opposed to Human Computer Interactions (HCI). Hence the state of the art, design and evaluation chapters do not focus on HCI, which is beyond the scope of this research.

The VisEN framework has been developed for the TEL domain supporting the construction of visual narratives using student-logged data. Usually log data consists of numeric data within this domain (resource usage figures, activity times, course access statistics and discussion forum activity) and is stored in tabular format. Hence VisEN supports the construction and exploration of visual narratives using numeric and tabular data, which can be sourced from databases or uploaded from the author's file system. In addition, the framework requires tabular file sources to have column headers, which are used for metadata. Although data can be sourced from databases, VisEN does not support narrative construction from multiple data sources at the same time.

Chapter 2 discussed two best practices for visual encoding: mapping data to visual structures and mapping data to visualisation techniques. VisEN implements the latter allowing it to take advantage of available visualisation techniques from JavaScript visualisation libraries and hence reducing the coding effort required to build visualisations. However, for the automatic mapping of data to visualisation techniques, the latter method requires developers to encode schemas for visualisation techniques, describing its affordances and characteristics. The design of VisEN supports the addition of new visualisation techniques, but requires that schemas are created for these. The design of VisEN supports several visualisation techniques and the framework caters for the addition of new techniques when required. However, new visualisation libraries added at run time are not supported as these have specific data formatting requirements.

VisEN supports a number of interaction techniques described in Chapters 2 and 3 that have been used in the TEL domain, including drilldown, zoom, filter and details-on-demand. In addition it supports data transformations (derived data transformations) to facilitate explorations and assists the visual narrative construction process. There are a number of other interaction techniques that are not supported by the framework, such as multiple coordinated views and focus plus context, for example. Although these have been shown to be important and beneficial to the process of making sense of digital data, they do not necessarily lend well to the process of supporting the construction and consumption through derived data transformations of visual narratives, and hence are beyond the scope of this thesis.

VisEN individualises visual narratives to consumers, such as students viewing narratives which use their personal logged data. It also tailors the derived data explorations based upon user preferences and usage history. Chapter 2 discussed the directions adopted by the state of the art in personalised visualisations, namely, eye gaze tracking, adapting visualisation techniques to user preferences, highlighting visual elements corresponding to text that the user is reading, and detecting boredom and frustration. These research directions are beyond the scope of the objectives of this thesis and hence are not considered within the design of the framework.

Finally, as VisEN has been implemented for the TEL domain, it enables and supports educators to construct visual narratives and facilitates the exploration of these visual narratives by students. It has been developed as a standalone framework that requires source data from the learning environment and provides interfaces to construct, navigate and explore visual narratives by respective users. One of the objectives of the framework is to offer students a visual narrative that can help them understand their performance and engagement, such as resource usage, activity times, course access statistics and engagement with course activities and comparisons with peers. However, student learning pedagogies are beyond the scope of this research. In addition, generating open learner models and the use of data mining and machine translation algorithms to predict student performance are also beyond the scope of this thesis.

4.11 Summary

This chapter defined *the data derivative approach to the construction and consumption of explorable visual narratives in TEL* (*derived data approach*) and described the design of the VisEN framework to support it. The analysis from the state of the art (best practices and limitations) guided the definition of the *approach*. The author and consumer requirements were outlined based upon 1) these best practices and 2) addressing the identified limitations from the state of the art. These were used to define the design requirements for the VisEN framework to realise the *approach*. This was followed by an outline of the functionality and detailed design descriptions of the components, including the sub-components and models of the VisEN framework. Finally, the scope of the design of the VisEN framework was also discussed. The design details of the sub-components, engines and models will be used as the basis of the technical implementation of the VisEN framework, which is discussed in Chapter 5.

5 Implementation

Chapter 4 defined *the derived data approach* (to address the Research Question of this thesis) and outlined the design required and provided the design specifications of the VisEN framework to realise this *approach*. In addition, it included detailed descriptions of the design of the various sub-components, engines and models of the framework. The sections of this chapter provide detailed implementation descriptions for the sub-components, engines and models of VisEN which adhered to the design specifications in order to fulfil the framework requirements.

5.1 Introduction

Chapter 4 introduced and described the design of the Narrative Builder, Visualisation Engine and the Visual Narrative Explorer components of the VisEN framework. The design specifications for the three components included design details for data source access and the service interfaces that support the functionality. This chapter first outlines the features supported by these components and the technologies used; it then describes the implementation details of the sub-components, engines, models and interfaces required by the framework. Finally, the workflows for visual narrative construction and consumption are presented. By following the design requirements and specifications detailed in Chapter 4, the implementation of the VisEN framework makes the *approach* a technical reality.

5.2 VisEN Features

The Narrative Builder, Visualisation Engine and the Visual Narrative Explorer are the core components of VisEN and, based on the design requirements outlined in section 4.4.3 of the design chapter, their implementations support the following features:

1. enabling authors to connect to tabular data sources (databases and tabular source files) by specifying connection parameters or a file to load,
2. formatting and presenting the connected source data through a web browser displaying table names and column names,
3. enabling authors to drag and drop column names onto a canvas, specify filters if required, and provide narrative descriptions for the data,
4. mapping author selected data and filters to a set of supported and appropriate visualisation techniques and pre-selecting the technique that is best suited to the data (with the highest confidence score). Enabling authors to select other visualisation techniques from the set to represent the narrative slice if more than one is returned,

5. supporting authors in creating visual narratives by generating and presenting derived data slices constructed using the data from the previously created narrative slice,
6. enabling consumers to navigate through visual narratives and manipulate views to support them in understanding the data through interaction techniques,
7. supporting consumers in exploring the dataset through derived data explorations linked to each narrative slice,
8. inferring consumer preferences of the derived data explorations by monitoring their usage and ordering the explorations accordingly,

5.3 Technologies Employed

VisEN has been implemented using Java and the client-server three-tier architecture, consisting of the Presentation, Application and Data tiers. It currently supports data sources from MySQL databases, Comma Separated Values (CSV) and Microsoft Excel file formats. The Apache POI[39] library is used to parse MS Excel files. The framework's interfaces are web-based and implemented using HTML5, JavaScript, AJAX and the jQuery library. The combination of HTML5, JavaScript, AJAX and the jQuery library were chosen to enable the framework to use JavaScript visualisation libraries. The visualisation techniques are sourced from JavaScript libraries that currently include D3[40], D3 plus[41] and Highcharts[42]. These were chosen to provide a range of visualisations that were not all available through one library. Supporting multiple stacks also enables the framework to be extended to support a larger range of visualisation techniques as they become available.

The interfaces use the jQuery Accordion widget to display the data source metadata (formatted table and column names) and use the jQuery Draggable and Droppable widgets to support the construction of visual narratives. JSON objects (using the Google GSON[43] API) are employed to encode the narratives slices and describe the characteristics and affordances of visualisation techniques (visualisation schemas). The implementation of the interface used by VisEN adhere to the web usability principles outlined by Nielsen (Nielsen, 1999).

[39] https://poi.apache.org/
[40] http://d3js.org/
[41] http://d3plus.org/
[42] http://www.highcharts.com/
[43] https://github.com/google/gson

VisEN can be extended to support data from other sources and to encode narratives slices and visualisation schemas using other technologies such as XML, without impacting the design described in Chapter 4. In addition the architecture can also support visualisation techniques sourced from other JavaScript libraries.

5.4 Narrative Builder

The objective of the Narrative Builder is to support authors in constructing visual narratives through the Data Connection and Narrative Authoring sub-components, and through the Derived Data Visualisation Model. Chapter 4 introduced the Narrative Builder and described the design of its sub-components and its model. The design of these included specifications for the component interface (Visual Narrative Builder Interface), which together with the sub-components and model supports the narrative construction process. This section discusses the implementation details for the two sub-components (Data Connection and Narrative Authoring), the Derived Data Visualisation Model, and the Visual Narrative Builder Interface, conforming to the design requirements outlined in section 4.7.

5.4.1 Data Connection

Section 4.7.1 discussed the Data Connection sub-component, outlined its requirements, and described its design specifications. The implementation of the Data Connection sub-component supports data sourced from MySQL databases, CSV and MS Excel file formats. It enables authors to specify MySQL connection parameters, including the database URI, the database name, a user name and a password, as shown in Figure 5.1. Alternatively, authors can specify a path to a CSV or MS Excel file to upload. The supplied parameters are used to connect to or upload the data source. The Data Connection sub-component has been implemented using Java Servlets and Java classes; and the Visual Narrative Builder Interface, has been implemented using Java Server Pages, HTML5, JavaScript and JQuery provides supporting functionality.

Figure 5.1 Data Source Specification and Connection

The first step in constructing a visual narrative requires authors to specify the location of the data source. The implementation details of the Data Connection sub-component developed through the Presentation, Application and Data tiers of the framework are described below. The implementation details of the Visual Narrative Builder Interface that are relevant to the Data Connection sub-component are included.

5.4.1.1 Presentation tier

The Visual Narrative Builder Interface implementation details to support the Data Connection sub-component are described in Table 5-1, which are encoded in the dataconnector.jsp and the dataconnector.js files. These files handle the authors' inputs and invoke the Application tier of the framework to establish data source connections.

JSP/JS	**Implementation description**
dataconnector.jsp	• Four text fields: the database URI, the database name, a user name and a password parameters are encoded using html text input types and another field for the CSV or MS Excel file path is provided. The input fields have an id and the values set by the authors are extracted using jQuery in dataconnect.js. • The "Browse file" button, shown in Figure 5.1, enables authors to select a file by triggering the jQuery call (#browse-file) in dataconnector.js. • The "Connect" button establishes a database connection by triggering the jQuery call (#establish-connection) in dataconnect.js, providing the database or parsing the CSV or Excel file.
dataconnector.js	• Two jQuery functions are encoded: $(#establish-connection) and $(#browse-file). • **$(#establish-connection)**: A call to the NarrativeBuilder Servlet doGet method is triggered using AJAX passing in the connection parameters (database or CSV/ MS Excel file name). • **$(#browse-file)**: A dialog window to lookup the CSV or MS Excel file is opened. The selected file path is used by $(#establish-connection).

Table 5-1 Presentation tier implementation details

5.4.1.2 Application tier

The Application tier classes for the Data Connection sub-component, described in Table 5-2, consist of the NarrativeBuilder, DataProcessing, XLSReader and CSVReader classes. These classes handle the formatting of the connection parameters and the returned metadata (table/sheet names and column headers) from the data tier.

Class	Method implementation description
NarrativeBuilder	• **doGet(HttpRequest, HttpResponse)** A databaseQuery object is created setting the four database connection parameters (for database connections) and invokes the DataProcessing.executeQuery() method passing the databaseQuery object. The returned data is formatted to replace underscores with spaces and capitalises the starting letters of table and column names. • For MS Excel and CSV file connections, the methods XLSReader.readXLSData() and CSVReader.readCSVData() are called respectively passing in the file path. The returned metadata is stored in the HTTP session.
DataProcessing	• **executeQuery(databaseQuery)** This method invokes the data tier by instantiating the DataConnector class and calling the getTableDescription() method passing the databaseQuery object. The resulting dataset is returned to the NarrativeBuilder class as a JSON object.
XLSReader	• **readXLSData(path, connectionData)** This method uses the XSSFWorkbook, XSSFSheet, XSSFRow, Row and Cell classes from Apache POI to extract MS Excel data, including the sheet names and the first row to represent table names and column headers respectively. It writes it to the default MySQL data source hosted on the server by calling writeXLSData() passing extracted data and the default connection data. • **writeXLSData(extractedData, connectionData)** A database is created and populated using the extracted data by instantiating the DataConnector class in the data tier and calling the createDatabase() and populateDatabase() methods.

CSVReader	• **readCSVData(path, connectionData)** This method uses Java's input and output streams to extract the data values and column headers from the CSV file. • **writeCSVData (connectionData)** A database is created using the file name and populated using the extracted data by instantiating the DataConnector in the data tier and calling the createDatabase() and populateDatabase() methods.

Table 5-2 Data Connection Application tier implementation details

5.4.1.3 Data tier

The Data tier implementation for the Data Connection sub-component functionality consists of the DataConnector class. Its methods, described in Table 5-3, handle the querying of data, and the creation and population of databases.

Class	Method implementation description
DataConnector	• **getTableDescription(databaseQuery)** This method uses the parameters in the databaseQuery object to establish a connection with the specified database. • It executes the "show all" sql query, which returns the table and column names. These fields are extracted and written to a JSON object, which is returned. • **createDatabase(connectionData)** This method uses the connection parameter to establish a connection with the default MySQL data source. • A query to create a database using the parameter is constructed and executed. • **populateDatabase(extractedData)** This method uses the database name, the table names and data from the parameter object to populate the created database. • Prior to creating the data, it determines the data types using the data values. • Create database table statements are constructed and executed followed by insert statements for the data.

Table 5-3 Data Connection Data tier implementation details

5.4.2 Narrative Authoring

Section 4.7.2 discussed the design of the Narrative Authoring sub-component, outlining its objectives and features. Its implementation supports authors as follows:

- to view the connected data source metadata (tables/sheets and column names) through a browser,

- to drag and drop columns onto select and filter panels (and specify filter conditions and values) to generate appropriate visualisations and construct narrative slices.

The visualisations are displayed through the Visual Narrative Builder Interface which provides text fields that allow authors to specify narrative slice titles and textual descriptions for the data. Figure 5.2 shows a screenshot of the interface presenting the narrative slice authoring process.

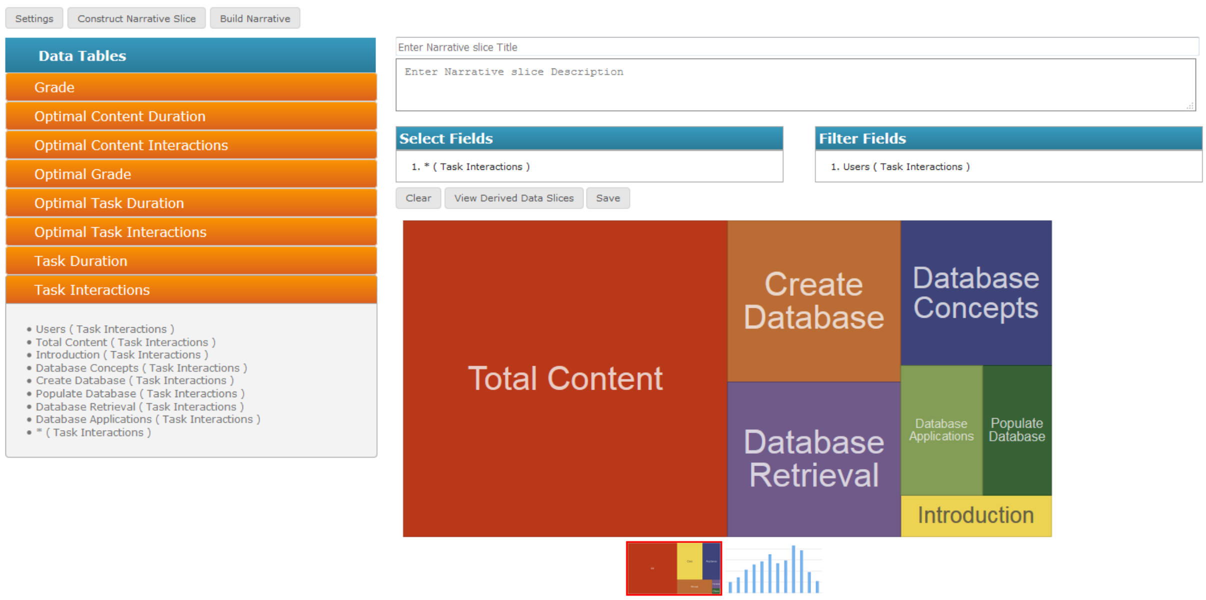

Figure 5.2 Narrative slice construction

Upon connecting to a data source, the next step of the visual narrative construction process requires authors to specify the data fields to be use in the narrative slices, set filters (if required) and supply textual descriptions for the data. The Visualisation Engine is invoked when data fields are selected or when filters are saved; this returns a set of visualisations to be displayed on the canvas. The visualisation with the highest confidence score is displayed (default); however authors can choose other visualisations from the set to be used in the narrative slice.

Once narrative slices have been constructed, the final step in the narrative construction process is to sequence them forming a visual narrative and supplying a title, as shown in Figure 5.3. The interface lists the constructed narrative slices and enables authors to view their description and the associated visualisation before sequencing them. The interface allows authors to publish visual narratives and make them available to consumers to navigate and explore. Two videos, used by the participants of Experiment 4, discussed and referenced in Chapter 6, present the construction of narrative slices and assembly an explorable visual narrative.

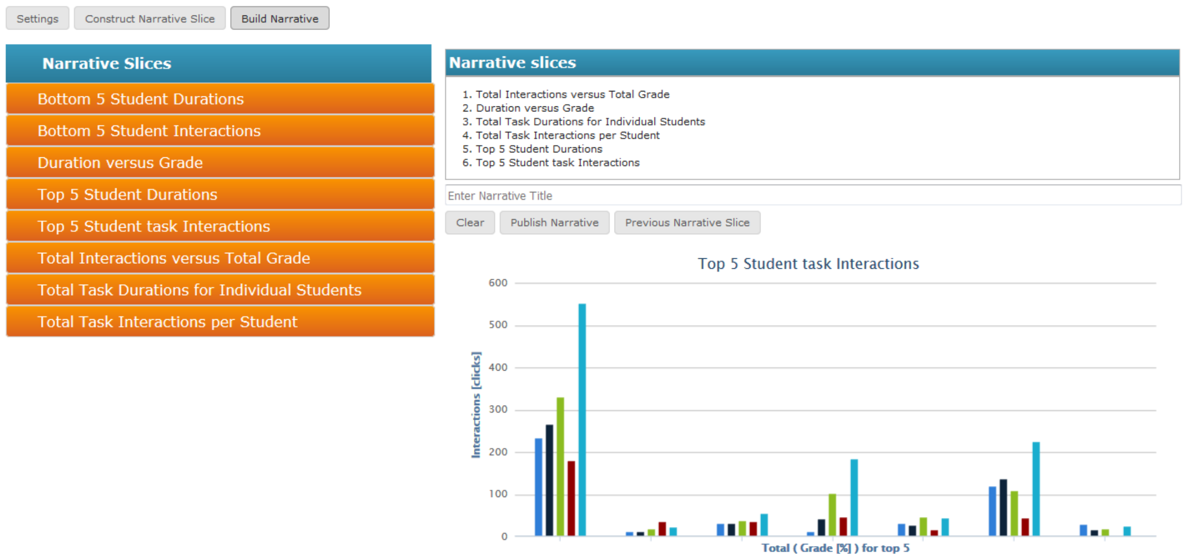

Figure 5.3 Assembly of a visual narrative

5.4.2.1 Presentation tier

The Visual Narrative Builder Interface implementation details for narrative construction are encoded in the narrativebuilder.jsp, narrativebuilder.js and the narrativeslicesaving.js files and are described in Table 5-4. These files encode the “construct narrative slice” button, which presents the data source metadata using the jQuery Accordion, Draggable and Droppable widgets to support the dragging and dropping of data elements for narrative slice construction. The interface provides a canvas to display the generated visualisations and presents a list of saved narrative slices, which can be sequenced to build a visual narrative.

JSP/JS	Description
narrativebuilder.jsp	• **Constructing Narrative Slices:** The “construct narrative slice” button invokes a jQuery call to the buildTableAccordion() function in narrativeslicebuilder.js, which fetches the metadata and builds an accordion table with draggable elements. • The interface uses the jQuery Draggable widget to build select and filter panels that accept elements dropped into them from the accordion table. • The visualiseSelection() function in narrativebuilder.js is invoked when data elements are dragged and dropped onto the select panel, which triggers calls to the Visualisation Engine to return a set of visualisations for the data (Figure 5.2).

narrativebuilder.jsp	• **Creating Filters**: loadFilterPopup() in Narrativebuilder.js is called using jQuery when data elements are dragged and dropped onto the filter panel, which loads a dialog box to set filter conditions. Once these conditions are saved, visualiseSelection() in Narrativebuilder.js is called to get a set of visualisations. • The data displayed in the filter dialog panel include equality and relational operators and clause conditions such as *and*, *or*, *orderby* and *between*, which are set by the authors. In addition the loadFilterPopup() fetches all the data values for the dropped elements using the connected data source. These values are displayed in the dialog panel for the authors to set the data values. • The save button implementation runs checks against the select panel data, title and description fields ensuring they are populated and calls the save() function in narrativeslicesaving.js. • **Constructing Visual Narratives:** The build narrative button implementation invokes the loadNarrativeBuilder() function in the narrativebuilder.js file to get the saved narrative slices and displays them using the jQuery Accordion widget. • The interface provides a panel implementing the jQuery Droppable widget where narrative slice titles can be dropped and sequenced (Figure 5.3). • The “publish narrative” button implementation runs checks against the title and ensures at least one narrative slice has been included in the narrative, and then calls the publishNarrative() function in narrativeslicesaving.js.
Narrativebuilder.js	• **buildTableAccordion()**: This function calls the NarrativeBuilder.doGet() method to fetch the metadata saved in the HTTP session for the connected data source. • The jQuery Accordion widget is used to display the metadata and the select and filter panels are created using the jQuery Droppable widget. • **visualiseSelection()**: This function calls the NarrativeBuilder.doGet() method passing the selected data and filter fields to invoke the Visualisation Engine. • The returned visualisations are displayed through the Narrative Builder Interface.

Narrativebuilder.js	• **loadFilterPopup()**: This function calls the NarrativeBuilder.doGet() method passing the data field that has been dropped onto the filter panel to get a list of associated values. It then calls visualiseSelection() once the filter data is saved. • **loadNarrativeBuilder()**: This function calls NarrativeBuilder.doGet() method to get the saved narrative slices.
narrativeslicesaving.js	• **save()**: This function calls the NarrativeSlice.doPost() method passing the data fields, filter data and conditions, title, description and the visualisation reference of the narrative slice, to save it to the HTTP session. • **publishNarrative()**: This function calls the NarrativeBuilder.doPost() method to save the narrative slice references, their sequences and title of a visual narrative to file and makes it available to consumers.

Table 5-4 Narrative Authoring Presentation tier implementation details

5.4.2.2 Application tier

The Application tier for the Narrative Authoring sub-component implementation, described in Table 5-5, consists of the NarrativeBuilder class. The implementation details handle writing the data source metadata to the response stream, calling the Visualisation Engine and saving narrative slices and visual narratives to file.

Class Method	Description
NarrativeBuilder. doGet(HttpRequest, HttpResponse)	• **Retrieving metadata:** The connected data source metadata (table or sheet names and column header) is extracted from the HTTP session and set in a JSON object using the GSON library and written to the Servlet response stream. The JSON object consists of the table or sheet names and corresponding arrays containing the column headers. • **Retrieving saved narrative slices:** A call is made to DataFileReader. getDataSlices(), which returns a list of JSON objects (one per narrative slice) and this is written to the response stream.

NarrativeBuilder. doGet(HttpRequest, HttpResponse)	• **Retrieving data values for the filter panel:** A call is made to DataConnector. getDataValues() passing the fields dropped onto the filter panel, the specified filter conditions, and the data connection parameters saved in the HTTP session. • **Calling the Visualisation Engine**: A narrativeSliceData object is created setting the title, data fields, filter values and conditions and sent to the Visualisation Engine. The returned set of visualisations are written to the response stream to be displayed through the Narrative Builder Interface.
NarrativeBuilder. doPost(HttpRequest, HttpResponse)	• **Saving a narrative slice**: The doPost() method reads the narrative slice data from the HTTP request and sets the title, description, selected data fields, filter data and visualisation technique into a narrativeSliceData object. • The narrative slice data is written to file by calling the DataFileReader. setDataSlice method passing the narrativeSliceData object. • **Publish a visual narrative**: A narrativeSequence object is created and the title, narrative slices references and the sequences are set. A call is made to DataFileReader. publishNarrative() passing the narrativeSequence object.

Table 5-5 Narrative Authoring Application tier implementation details

5.4.2.3 Data tier

The Data tier implementation of the Narrative Authuring sub-component consists of the DataConnector, DataFileReader, and the DataFileWriter classes, described in Table 5-6. These classes are responsible for establishing a MySQL database connection and creating and running a select query, reading narrative slice data from file, and writing narrative slice data and visual narrative data to file.

Class	Method implementation description
DataConnector	• **getDataValues(filterField, connectParam):** This method uses the connection Parameters to establish a connection with the data source. A query is created using the fields in the filterfield parameter, which is executed against the connected data source and result set is returned in a list.

DataFileReader	• **getNarrativeSlices():** This method reads the files containing the narrative slices data on the server. • It parses the data from each file into JSON objects setting the title, description, field data, filter data and the visualisation technique and returns a list of JSON objects.
DataFileWriter	• **setNarrativeSlice(narrativeSliceData)** This method creates a file using the title of the narrative slice and the timestamp and saves it to the server. • It then writes the narrative slice title, description, data fields, filters and the chosen visualisation technique to the file. • **publishNarrative (narrativeSequence**) A file is created on the server and the title of the visual narrative, and its narrative slice references and sequences are written to it.

Table 5-6 Narrative Authoring Data tier implementation details

5.4.3 Derived Data Visualisation Model

Section 4.7.3 discussed the design of the Derived Data Visualisation Model, outlining its objectives and features. Its implementation supports the construction of visual narratives by presenting derived data slices of constructed narrative slices that can be used within the visual narrative. The implementation also supports the exploration of published visual narratives through derived data explorations. The Visual Narrative Builder Interface consists of the View Derived Slices button, which appears when a visualisation is generated and provides authors with visualisations of related data through a popup window. Several derived data slices can be generated for each narrative slice data, which are all made available to the narrative author.

The derived data transformations applied to narrative slice data are based on pre-defined template mappings, as highlighted in Table 5-7 below. The model uses the fields dropped onto the select and filter panels in addition to the filter conditions and filter values specified. The current implementation of the model supports:

- **Model Relationship Transformations** that map TEL metrics, such as engagement, grade and duration,
- **Associate Transformations** that map fields and filter conditions to related items, for example, *greater than* and *greater than or equals* is mapped to associated data such as the data with the highest values,

- **Symmetric Data Transformations** that invert filter conditions, for example *greater than* is mapped to *less than*,
- **Opposite Group Transformations** that map fields and filter conditions to opposite but related items, for example, *greater than* and *greater than or equal* is mapped to data with the lowest values,
- **Summary Details Transformations** that map data fields presenting summary views to detailed views, for example, aggregations are mapped to breakdowns and vice versa,
- **User Group Transformations** that map individuals to associated groups and vice versa, for example user group views are mapped to a breakdown of individuals.

	Field/condition	**Mapping**		**Field/condition**	**Mapping**
Model Relationship Transformations	Engagement	Duration	**Symmetric data Transformations**	>	<
	Engagement	Grade		>=	<=
	Duration	Engagement		<	>
	Duration	Grade		<=	>=
	Grade	Engagement		Bottom	Top
	Grade	Duration		Top	Bottom
Associate Transformations	>/>=	Top	**Group/User Transformations**		
	</<=	Bottom		Group	Top
	<=/</>/>=	Group		Group	Bottom
	Optimal	Top/Group		User	Group
	Bottom	Group		Group	Breakdown
	Top	Group			
Opposite group Transformations			**Summary/Detail Transformations**		
	>/>=	Bottom		Aggregation	Breakdown
	</<=	Top		Breakdown	Aggregation

Table 5-7 Implemented Transformations

Table 5-7 presents the mappings that are currently supported by the Derived Data Model and these are used to generate the derived data transformations to support both construction and consumption of visual narratives.

5.4.3.1 Presentation tier

The Visual Narrative Builder Interface implementation details supporting the Derived Data Visualisation Model are described in Table 5-8, which are encoded in the narrativebuilder.jsp and the explorationpath.js files. These files handle the calls to the Application layer to invoke the model to generate derived data transformations. The interface displays visualisations of these transformations in popup windows and supports navigation between them.

JSP/JS	Description
narrativebuilder.jsp, explorationpath.js	• Narrativebuilder.jsp presents the “View Derived Data Slices” button which triggers the #load-exploration-point-builder jQuery function in the explorationpath.js file to extract the data fields, filter conditions and values from the select and filter panels. • A call is made to the NarrativeBuilder.doGet() method passing a generateDerivatives string and the narrative slice data fields and filter settings. • The call to the Servlet returns a set of derived data slices which are presented through a popup window, with “view next visualisation” and “view previous visualisation” buttons to navigate through the views.

Table 5-8 Derived Data Visualisation Model Presentation tier implementation details

5.4.3.2 Application tier

The Application tier for the Derived Data Visualisation Model implementation, described in Table 5-9, consists of the NarrativeBuilder class and the DerivativeDataGenerator super class. It also consists of six sub classes responsible for implementing the transformations outlined in Table 5-7 above.

Class Method	Description
NarrativeBuilder. doGet()	• **Generate Derivatives:** This method calls the DerivativeDataGenerator super class passing the fields from the select panel and filter data set in a narrativeSliceData Object, which returns a list of DerivedData objects. • The list of DerivedData objects are then passed to the Visualisation Engine, and the visualisations with the highest confidence scores for each DerivedData object are used to render the derived data slices.
DerivativeDataGenerator Super class	• The DerivativeDataGenerator class extracts the individual data fields and filter conditions and values from the narrativeSliceData object. • Six sub classes, one for each of the transformations described in Table 5-7 are implemented and these are invoked passing the relevant extracted data. • These classes return transformation properties if the mapping conditions are satisfied. • For each mapping, a DerivedData Object is created using the narrativeSliceData object data and the transform properties. A list of DerivedData objects is returned.
Derived Data Visualisation Model Sub classes	• The model's sub classes include: ▪ ModelRelationshipMappings, ▪ SymmetricDataMappings, ▪ SummaryDetailsMappings, ▪ GroupUSerMappings, ▪ AssociateMappings, ▪ OppositeGroupMappings. • These sub classes implement the transformations outlined in Table 5-7 by applying mappings to data fields, filter values and filter conditions and return an updated narrativeSliceData object.

Table 5-9 Derived Data Visualisation Model Application tier implementation details

5.4.4 Analysis

Chapter 5 outlined five design requirements for the Data Connection sub-component (section 4.7.1), which included:

1. providing a list of databases that have been previously connected to by the narrative author.
2. enabling authors to supply database connection parameters.
3. supporting data sourced from other formats available on the narrative author's file system.
4. an extensible architecture to support different data models.
5. formatting data source metadata to support the narrative authoring process.

With the exception of feature 1, the remaining four have been implemented through an architecture that can be extended to support data sourced from formats other than MySQL, CSV and Excel. Feature was included as a design requirement as it made it more convenient for authors regularly using VisEN to construct visual narratives. However, in the context of this thesis, the user trial did not require participant to regularly use VisEN, hence this feature does not assist in answering the Research Question and was left out.

Six design requirements were outlined in section 4.7.2 for the Narrative Authoring sub-component, which included:

1. presenting the formatted metadata from the connected data source and allowing the narrative author to view the structure of the source data,
2. enabling authors to specify data fields and filters using the formatted metadata and entering descriptions for the selected data,
3. enabling Authors to select one of the visualisation techniques or use the default for a narrative slice returned in the set by the Visualisation Engine,
4. presenting visualisations of derived data transformations to authors by invoking the Derived Data Visualisation Model to support the construction of the visual narrative,
5. enabling authors to decide which of the derived data transformations are made available to consumer for explorations,
6. supporting the sequencing of narrative slices to build a visual narratives that can be published.

From the implementation details presented in this section, features 1, 2, 3, 4 and 6 have been implemented. Although the implementation details of feature 4 was not discussed above, it is covered in the next section. Design feature 5 has not been implemented and all the visualisations of the derived data slices are made available to the consumers as derived data explorations. This feature was intentionally excluded to evaluate whether user trial participants would prefer authors to decide which derived data visualisations should be included in the narrative or whether to make all of them available to the consumers. The results of this evaluation is presented in the findings of Experiment 4 in Chapter 6.

A further five design requirements were outlined for the Derived Data Visualisation Model in section 4.7.3, which included:

1. supporting a set of derived data transformations that are used to map data fields and filter conditions and values in narrative slices,
2. enabling the addition of mappings for the Derived Data Visualisation Model at run time,
3. supporting the lookup of appropriate transformations for narrative slice field and filter properties,
4. applying mappings to narrative slice field and filter properties to create derived data objects,
5. calling the Visualisation Engine to generate a visualisation for each derived data object.

The implementation details presented in this section show that features 1, 3, 4 and 5 have been implemented. Design feature 2 has not been implemented as the transformations are defined within the code and not in property files. The inclusion of mappings at run time (in additions to those already encoded in the framework) were not required to address the Research Question, hence design feature 2 was not implemented.

5.5 Visualisation Engine

The objective of the Visualisation Engine is to generate visualisations for narrative slice data and derived data transformations using imported JavaScript libraries. It consists of the Visualisation Schema and the Query Builder sub-components and the View Generation Model. Chapter 4 introduced the Visualisation Engine and described the design of its sub-components and its model. This section discusses the implementation details of the two sub-components and the model, which adhere to the design requirements outlined in section 4.8.

5.5.1 Visualisation Schema

Section 4.8.1 discussed the design of the visualisation schemas, which are used to characterise visualisation techniques using key value tuples. The design described the structure of the schemas highlighting the tuples marked as mandatory and optional and explained the keys and value types. The implementation of the Visualisation Engine currently supports ten visualisation techniques, outlined in Table 5-10, that are generated using the D3, the D3 Plus and the Highcharts JavaScript libraries.

Visualisation Techniques	Bar Chart	Pie Chart
	Line Chart	Scatterplot
	Parallel Coordinates	Gauge
	Bubble Chart	Stacked Area Chart
	Stacked columns	Treemap

Table 5-10 Supported Visualisation Techniques

The visualisation schemas are encoded as JSON objects and stored as files on the server. Ten visualisation schemas have been encoded, one for each of the supported visualisation techniques. Literature (Dias et al. 2012, Chi 2000, Carr et al. 1987, Heer et al. 2010, Graham & Kennedy, 2010) was used to support the encoding of characteristics for visualisation techniques. Figure 5.4 presents the visualisation schema of the scatterplot visualisation technique.


```
{
 "visualization technique":"scatterplot",

 "data ranges": [
    {
        "data series":"1",
        "minimum points per series": "10",
        "maximum points per series": "100",
    },
    {
        "data series":"2",
        "minimum points per series": "10",
        "maximum points per series": "100",
    }
 ]

 "maximum coordinates per point": "2",
 "minimum coordinates per point": "2",
 "number of coordinates": "2"
 "first coordinates type": "quantitative",
 "second coordinates type": "quantitative",
 "data categories required": true,
 "maximum data categories": "10",
 "minimum data categories": "1",

 "number of axes":"2",
 "yaxis data": "true",
 "xaxis data": "true"
 "xaxis title": "true",
 "yaxis title": "true",
 "data type supported x-axis":"quantitative",
 "data type supported y-axis":"quantitative",
 "technique source":"javascript",
 "title": "yes",
 "subtitle": "yes",
}
```


Figure 5.4 Sample Schema

From Figure 5.4, the scatterplot visualisation schema shows that it can handle two sets of data with a range of 10-100 points per data set (data range). It can only handle data with two coordinates, requires x and y axes, and requires quantitative data. The visualisation source indicates that the technique is sourced from is a JavaScript library.

5.5.2 Query Builder

The objective of the Query Builder sub-component is to extract data fields, filter conditions and filter values from narrative slice data and from derived data objects to generate and run queries against the connected data source.

5.5.2.1 Application tier

The Query Builder consists of the BuildQuery Java class which contains methods that return segments of a query which are combined to form the overall query. Table 5-11 presents the implementation details of the Application tier of the Query Builder.

Method	Description
buildSelectStatement(narrativeSliceData, derivedData)	• This method constructs and returns the select segment of the query, handling data from both the narrative slices and the derived data objects. • For both objects, the data fields are extracted and appended to a Select string. If either object has aggregate settings such as sum, count, etc. then these are used in the statement.
checkForJoins(narrativeSliceData, derivedData)	• If the data fields are from two or more tables, calls are made to the getPrimaryKeys(), getForeignKeys() and getMatchingColumns() methods in the QueryExecution class in the Data tier passing the name of each table. • For primary keys, foreign keys and columns, a check is run to ensure the keys or columns match. If a match is found, a query segment is created using the matching keys or columns, and this query segment is returned.
buildWhereClause(narrativeSliceData, derivedData)	• Filter data, conditions and values are extracted from both parameter objects. • Authors can specify a number of conditions including *and*, *or*, *between* and *order by*. They can also specify equality and relational operators. The following methods encoded in the Application tier are called that handle each condition: addAndCondition(), addOrCondition(), buildBetweenClause(), buildOrderBy Clause(). The relevant method is called passing the field data and which return a query segments. Equality and relational operators are handled by the addOperators(),in the Application tier, by passing the field data, values and conditions which return a query segment.

buildQuery()	• The buildSelectStatement(), buildWhereClause() and checkForJoins() methods return query segments that are used to build the SQL query. • The query segments are ordered and appended to form a query that conforms to the SQL syntax by parsing through the returned query segments. • The QueryExecution.executeQuery() method in the Data tier is called to run the query and the results are passed to the View Generation Model.

Table 5-11 Query Builder Application tier implementation details

5.5.2.2 Data tier

The Data tier class used by the Query Builder sub-component consists of the QueryExecution class and its methods handle the execution of SQL queries. The QueryExecution class also handles the primary key, foreign key and matching column header lookups. The implementation description is provided in Table 5-12.

Method	Method implementation description
executeQuery (sqlQuery, connParams)	• This method uses the saved connection parameters to establish and initialise a database connection. • The query is executed and the results returned.
getPrimaryKeys (table), getMatchingColumns(table), getForeignKeys(table)	• The getPrimaryKeys() method uses the DatabsebaseMetaData class from the Java SQL package to retrieve a set of primary keys using Resultset.getString(“PKCOLUMN_NAME”). • Similarly the getForeignKeys() method uses the DatabsebaseMetaData class to lookup a set of foreign keys. • The getMatchingColumns() method runs a “show all” query and “describe” query for each table to determine matching columns.

Table 5-12 Query Builder Data tier implementation details

5.5.3 View Generation Model

The objective of the View Generation Model is to use the data from the narrative slice or the derived data object and the imported libraries to generate a set of ranked visualisations. The model consists of four super classes (VisualisationOverviewBuilder, VisualisationMatcher, TechniqueMatcher and VisualisationPresentation) and to ten sub classes that represent the supported visualisation techniques.

5.5.3.1 Application tier

The classes of the View Generation Model are all implemented in the Application tier of the architecture. These classes generate a ranked set of visualisations that can be used to render the data, which is formatted for the individual visualisation techniques in the set. Table 5-13 discusses the implementation details.

Class	Method implementation description
VisualisationOverviewBuilder	• **getMetaData(narrativeSliceData, relatedData):** Query results, data fields and the title are extracted from the narrative slice or the derived data object and set in a visualisationMetaData object. Visualisation schema references are read from the server and set in the same object. VisualisationMatcher.getVisualisationList() uses this data to determine appropriate visualisation techniques to render the data. • **getDisplayData(visualisationMetaData):** The sub classes of the matched visualisations are called, with each returning an object containing the data results, axes and legends formatted for the visualisation technique. These objects are added to a list and returned in a JSON object.
VisualisationMatcher	• **getVisualisationList(visualisationMetaData)**: For each schema, all the mandatory fields are matched against the query results and the narrative slice or derived data object meta data, including data ranges, data type, and coordinates per point. These matches are handled making calls to the getDataRanges(), getDataTypes() and getCoordsPerPoint() methods in the sub classes. These methods return data ranges, data types and the number of coordinates in the result. • The visualisationMetaData object is updated with a set of visualisation techniques that are suitable to render the narrative slice or the derived data object data.

TechniqueMatcher	• **determineTechniqueConfidenceScore(visualisationMetaData):** This method sums up the suffix values of the optional schema properties in the visualisation MetaData object. • It uses the set of the suitable visualisation techniques for the ranking. The visualisations with the greatest total suffix sum is ranked with the high confidence level, followed by checks for medium and low confidence levels.
Ten Subclasses	• The ten sub classes include: ▪ BarChartVisualisationTechnique ▪ BubbleChartVisualisationTechnique ▪ GaugeVisualisationTechnique ▪ LineChartVisualisationTechnique ▪ ParallelCoordinatesVisualisationTechnique ▪ PieVisualisationTechnique ▪ ScatterplotVisualisationTechnique ▪ StackedAreaVisualisationTechnique ▪ StackedColumnVisualisationTechnique ▪ TreeMapVisualisationTechnique • All of these classes call the VisualisationPresentation.getVisPresentationData() method, which returns the presentation format for the visualisation technique. This method is used to create a JSON object with the visualisation data (technique, data, axes, legends), which is used to generate the visualisation technique.
VisualisationPresentation	• **getVisPresentationData(metadata, queryResults)**: The axes titles (if applicable), legends, details-on-demand and data categories are generated using this method. • Axes titles and legends are generated using the metadata in the narrative slice or the derived data object. • Details-on-demand data is determined using query results and the field names. A string is generated and formatted to present both the results and the field names. • Data categories are determined by the number of data series in the query results.

Table 5-13 View Generation Model Implementation Details

5.5.4 Analysis

The design requirements of the Query Builder include extracting data fields, filter values and filter conditions from narrative slices and derived data objects, and then generating and executing queries that conform to the standard of the data source used. To this extent the implementation described above adheres fully to these requirements developed for a MySQL database.

Section 4.8 of the design chapter of this thesis outlined six design requirements for the View Generation Model, which included:

1. using visualisation techniques from imported libraries to render numeric data,
2. supporting the addition of new visualisation techniques at run time,
3. enabling the encoding of characteristics and affordances of visualisation techniques in visualisation schemas and supporting new encodings to be added at run time,
4. using data fields, filter conditions and filter values from narrative slices and derived data objects to build queries to run against the connected data source,
5. using the query results from step 4 together with the metadata from narrative slices and derived data objects to determine appropriate visualisation techniques,
6. generating a set of visualisations to render narrative slice data and derived data objects.

The implementation uses visualisation techniques sourced from JavaScript libraries and creates JSON data structures that adhere to the requirements of the libraries to generate the visualisations. Developers encode the visualisation schemas that specify the characteristics of the visualisation techniques and these are used to 1) determine a set of appropriate visualisations for the data and 2) rank the visualisations. The current implementation supports ten visualisation techniques; however importing other techniques requires the addition of new sub classes (one per technique). The addition of visualisation techniques at run time advances the framework and makes it more suitable for data that may not be numeric. However, for the purpose of this thesis, this feature does not assist in addressing the Research Question, and was not included in the implementation of the framework. Hence five of the six design features are supported by the VisEN framework.

5.6 Visual Narrative Explorer

The requirements of the Visualisation Narrative Explorer are to display published visual narratives and support navigation and visual interactions, between and within the narrative slices. The requirements also include supporting tailored data explorations through derived data explorations. The Visualisation Narrative Explorer consists of the Visual Narrative Viewer, the Usage Pattern Store and the Tailoring Engine sub-components. The design of these sub-components included specifications for the Visual Narrative Interaction and Exploration interface. This interface together with the sub-components supports the visual narrative consumption process. Chapter 4 introduced the Visualisation Narrative Explorer and described its design; this section discusses the implementation details of the three sub-components, which comform to the design requirements outlined in section 4.9.

5.6.1 Visual Narrative Viewer

The Visual Narrative Viewer sub-component uses the Visual Narrative Interaction and Exploration interface to present a list of published visual narrative titles to consumers through a dropdown menu on a browser window. Selecting a visual narrative triggers the interface to display checkboxes showing the data used by the derived data transformations, which consumers can select to specify the explorations that they are interested in. In addition the interface provides a slider to select the level of exploration that the consumer would like to set. The interface includes a “View Narrative” button, which loads the selected visual narrative, and supplies the specified level of derived data explorations as links below each narrative slice description. The interface implementation uses tabs to present each narrative slice (title, visualisation, description and derived data exploration links) and these tabs are rendered through a browser window. Figure 5.5 presents the Visual Narrative Interaction and Exploration interface displaying a sample visual narrative.

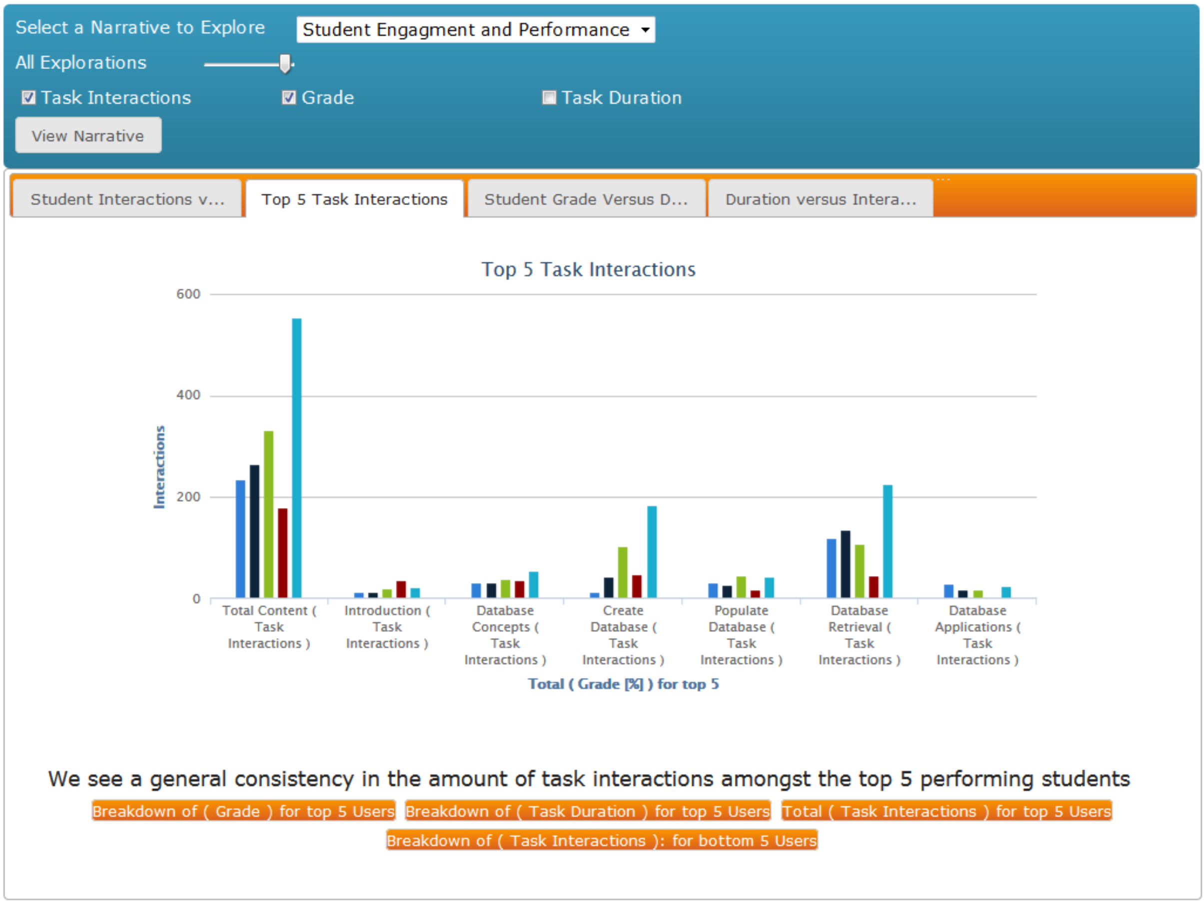

Figure 5.5 Visual Narrative Interaction and Explorer interface

5.6.1.1 Presentation tier

The Visual Narrative Interaction and Exploration interface implementation is encoded in narrativeviewer.jsp and narrativeviewer.js files. The implementation details discussed in Table 5-14. The ten supported visualisation techniques each have a JSP and corresponding JavaScript file to render the visualisation within the interface.

<table>
<tr><th>JSP/JS</th><th>Implementation description</th></tr>
<tr><td>narrativeviewer.jsp</td><td>
The JSP file implements a dropdown menu, which invokes a jQuery call in narrativeviewer.js (#narrative-titles-dropdown) when the interface is loaded.
The derived data explorations consist of data attributes, for example, interactions, durations, and grade in the case of TEL, and these are generated through a jQuery call in narrativeviewer.js (#view-selected-narrative) and rendered via checkboxes.
A slider is encode which can be set to “All”, “Some” or “No Explorations”, which are used to determine the level of explorations the visual narrative will have.
The implementation provides the “View Narrative” button which invokes a jQuery call in narrativeviewer.js (#view-selected-narrative) to call the showTab() function to fetch the visualisations, which are loaded into the tabs as iframes.
The interface presents a set of derived data exploration links at the bottom of the narrative slice tabs using a div that gets populated with the rest of the tab data (title, description and visualisation) through the showTab() function.
The tabs are indexed and their IDs are used to populate the div with ordered derived data exploration links.
</td></tr>
<tr><td>narrativeviewer.js</td><td>
$(#narrative-titles-dropdown) calls the NarrativeViewerServlet.doGet() method passing the selected title to generate the ordered derived data explorations.
$(#view-selected-narrative) iterates through the ordered derived data exploration and sets each data attribute in a checkbox.
showTab() calls NarrativeViewerServlet.doGet() passing the visual narrative title to get the narrative slices and derived data exploration links. For each narrative slice, one of the ten visualisation technique JSP files is loaded into an iframe.
The showTab() function iterates through the list of derived data exploration links and displays them sequentially under the visualisations in the interface.
Clicking a derived data exploration link invokes the NarrativeViewerServlet. doGet() method passing the link index, which returns the derived data exploration. The selected exploration is displayed in a dialog window through an iframe (the relevant visualisation technique JSP file is loaded).
</td></tr>
</table>

Visualisation techniques JSP/JS Files	• The ten visualisation techniques call their corresponding JavaScript functions to generate the visualisation that is rendered within the interface. • The JavaScript function calls the NarrativeViewerServlet.doGet() method passing the visualisation technique name to get the narrative slice data and description.

Table 5-14 Visual Narrative Viewer Presentation tier implementation details

5.6.1.2 Application tier

The NarrativeViewer class implements the Application tier logic of the Visual Narrative Viewer. It makes calls to the Data tier to get the saved visual narratives and the individual narrative slices. It also handles calls to the Derived Data Visualisation Model to generate derived data explorations. The implementation details of the NarrativeViewer class are discussed in Table 5-15.

Class	Method implementation description
NarrativeViewer	• **doGet():** This method calls JsonFileReader.getNarrativeJsonData() to get a list of published visual narrative titles, which are written to the response stream. • **Load selected visual narrative**: JsonFileReader.getNarrativeSlices() is called passing the name of the selected visual narrative. This call returns a JSON object with the narrative slice data, which includes the description of the narrative slice (supplied by the author during construction) and the visualisation technique reference. • **Related Data**: Calls are made to the Derived Data Visualisation Model passing the narrative slice data, which return derivedData objects that are passed to the Visualisation Engine to generate a visualisation for each derivedData object.

Table 5-15 Visual Narrative Viewer Application tier implementation details

5.6.1.3 Data tier

The JsonFileReader class consists of methods to get a list of saved visual narratives and its narrative slices. The implementation details are discussed in Table 5-16.

JsonFileReader	• **getNarrativeJsonData()**: A JSON object with an array of published visual narrative titles are read from the published_narratives file that is saved on the server. • **getNarrativeJsonData(narrativeTitle)**: The visual narrative title is used to fetch the relevant file from the server. • **getNarrativeSlices(narrativeTitle)**: The visual narrative title is used to fetch the relevant file from the server with the narrative slices. A JSON data object is created setting a list of the titles, descriptions, visualisation technique references (each element in this list corresponds to a narrative slice of the visual narrative). The JSON data object is returned to the called method.

Table 5-16 Visual Narrative Viewer Data tier implementation details

5.6.2 Usage Pattern Store

The Usage Pattern Store sub-component records user actions associated with derived data exploration attributes and associated visualisations. As derived data explorations can have various configurations, such as filters for example, and hence can have several visualisations per attribute, actions are recorded separately for data attributes and visualisations. Recorded actions include the number of visits per visualisation and the time spent viewing the data through the visualisation.

The Usage Pattern Store has been implemented in the Application tier of the VisEN architecture and recordings are stored as HTTP session attributes and only available during visual navigation and exploration.

5.6.2.1 Application tier

The Usage Store Pattern sub-component is implemented in the NarrativeViewer and UsageStore Classes (implementation details discussed in Table 5-17). The NarrativeViewer stores usage data in the HTTP session which is available for the duration of the navigation and exploration of visual narratives. The UsageStore class methods update the number of visits and viewing time variables.

Class	Method implementation description
NarrativeViewer	• This class creates HTTP session variables for the derived data exploration visualisation visits and viewing times per related data attribute, setting both at the default value of zero. • For each derived data exploration, calls are made to incrementVisits() and updateViewingTime() in the UsageStore class.
UsageStore	• **incrementVisitsCount(datatype, visualisation):** This method uses the HTTP session variable associated with the datatype and visualisation parameters and increments it by one. • **updateViewingTime(datatype, visualisation, status):** This method is called when a derived data exploration is opened and closed (indicated by the status parameter). It records the time from when the visualisation is opened until it is closed and updates the HTTP session variable associated with the datatype and visualisation parameters by adding the viewing timing to the existing value.

Table 5-17 Usage Pattern Store Application tier implementation details

5.6.3 Tailoring Engine

The Tailoring Engine orders the derived data exploration links according to inferred user preferences. The implementation of the Usage Pattern Store records the number of visits and viewing times for both the derived data exploration attributes and associated visualisations. The Tailoring Engine applies separate weightings to viewing times and visits to calculate the inference. A weighting of ten units is applied to visits and a weighting of two units is applied to viewing times. The viewing time weighting is lower as factors such as consumers leaving a derived data exploration open for some time before or after analysing it are taken into consideration. The following formulae are used to determine an inference score (*iScore*) for derived data exploration links. The formulae below represents one way of calculating the inference score:

$$visScore = (visNumVisits \times 10) + (visViewingTime \times 2)$$

$$dataScore = (dataNumVisits \times 10) + (dataViewingTime \times 2)$$

$$iScore = visScore + dataScore$$

5.6.3.1 Application tier

The NarrativeViewer and the TailoredExploration classes encode the Application tier logic (shown in Table 5-18) of the Tailoring Engine to assign weightings to the derived data explorations and rendering them accordingly.

Class	Method implementation description
NarrativeViewer	• **doGet(Httprequest, HttpResponse)** When a requested derived data exploration visualisation is opened, the TailoredExploration.getLinkAttributes() method is called, which extracts the viewing time and visit count variables from the HTTP session. • When the requested derived data exploration visualisation is closed, the TailoredExploration.setExplorationMetaData() method is called, which updates the viewing time and the visit count variables for the related data and the visualisation. A call is then made to TailoredExploration.tailorExploration(), which re-orders the derived data exploration links and this is manifested when a visual narrative tab is loaded.
TailoredExploration	• **getLinkAttributes(derivedData)**: This method gets the HTTP session variables associated with the derived data exploration attributes and associated visualisations. • **tailorExploration(derivedData)** The derived data objects are re-ordered through the implementation of the *iScore* formula presented above. • **setExplorationMetaData(derivedData):** This method gets viewing time and visit count properties for the derived data objects and updates the values accordingly .

Table 5-18 Usage Pattern Store Application tier implementation details

5.6.4 Analysis

The following four design requirements were outlined for the Visual Narrative Explorer in section 4.9:

1. displaying published visual narratives to consumers and supporting navigation and visual interactions between and within the narratives,
2. presenting derived data exploration to support explorations within the visual narratives,
3. tailoring the ordering of derived data explorations by inferring consumer preferences,

4. separating these explorations from the visual narrative to minimise disruption to the flow of the story.

The four design requirements have been implemented, however the Tailoring Engine sub-component implementation does not include consumer visual interactions in the calculation of the inference score. This is because the visualisations including the interactions supported are sourced from JavaScript libraries and not implemented within the framework, meaning that the visual interactions are not accessible.

5.7 Summary

This chapter discussed the implementation details of the individual framework components including their sub-components, engines and models using the three-tiered architecture. From the details discussed per component, together with the analysis sections, it can be seen that the vast majority of design requirements were implemented by VisEN, highlighting that the implemented framework adheres to the design. The four design requirements that were not implemented, included:

1. providing a list of databases that narrative authors had been previously connected to,
2. enabling authors to decide which of the derived data transformations are made available to consumer for explorations of the derived data transformations,
3. enabling the addition of mappings (used by the Derived Data Model to generate derived data transformations) at run time,
4. supporting the addition of new visualisation techniques at run time.

These four requirements do not directly support the construction and consumption of visual narrative and do not assist in addressing the Research Question. Hence they were not included in the implementation of the VisEN framework. However, these requirements indirectly support the *approach* through the addition of new visualisations loaded at run time and cached data source connections, and these will be included in future releases of the framework.

6 Evaluation

This chapter discusses the experiments conducted as part of addressing the Research Question defined by this thesis (Section 1.2, page 4). In addition, the experiments aimed to evaluate the effectiveness and usability of *the derived data approach* which was developed to satisfy the Research Question. The research discussed in this thesis was evaluated initially through a proof of concept prototype implementation of a visualisation system that supported the authoring of narrative slices and exploring hand-coded, inter-connected and explorable visualisations. The prototype was trialled with users experienced in the development of PLE to attain feedback regarding the suitability and effectiveness of this *approach*. Following the positive response from this trial, the VisEN framework was developed and evaluated through a further three experiments to assess its architectural components and end-to-end usage.

This chapter first outlines the evaluation strategy used in this research and then describes the four experiments conducted, which used a variety of techniques. These techniques included user trials, students using and evaluating the visual narratives which were included in their course work, log analysis of students' usage patterns, questionnaires and interviews. The experiments used VisEN as a standalone framework and also in conjunction with the AMAS (Staikopoulos et al., 2015) PLE to provide students with visual narratives. The chapter also includes analysis of the findings from each of the experiments, and a discussion which ties the findings from these experiments to an assessment of the *approach*. Copies of the four experiment questionnaires and interview questions, where applicable, are presented in appendix B.

6.1 Evaluation Overview

Section 1.3 outlined the research objectives for this thesis, with objective 4 (supporting authors in creating visual narratives that can be followed, analysed and explored by consumers) addressing the evaluation of the *approach* with authentic users through realistic use cases. In order to realise this objective, it was necessary to refine it into features that VisEN could support.

The following design requirements were defined in section 4.4.3:

RQ1 Support authors, who have access to and an understanding of some data, to produce visual narratives, by automatically generating derived data slices (based on narrative slices they have constructed) which are presented as visualisations. These visualisations are suggestions that can be included in the visual narrative.

RQ2 Support consumers in exploring visual narratives by automatically generating derived data explorations of narrative slices and providing links to these visualisations. These links should load separately from the narrative, to enable data exploration without disrupting the flow of the visual narrative.

RQ3 Tailor the ordering of the derived data exploration visualisations, mentioned in RQ2, by inferring consumer preferences based on their usage of the explorations.

RQ4 Provide mechanisms to support access to tabular data sources with numeric data.

RQ5 Provide a process to facilitate visual encoding by mapping narrative slice data to an appropriate set of visualisation techniques.

RQ6 Provide operations to enable the authoring, and mechanisms to piece narrative slices together to produce a visual narrative.

RQ7 Support consumers manipulate narrative slice views through interaction techniques.

Based on research objective four, four experiments were conducted and their evaluations provided evidence that the following set of evaluation features (which closely follow the framework's design requirements) was supported:

EF1 VisEN provides visual interactions including drilldown, zooming, details-on-demand and filtering to facilitate view manipulations (RQ7).

EF2 VisEN provides a process to facilitate visual encoding by mapping narrative slice data to a set of appropriate visualisation techniques (RQ5).

EF3 VisEN supports consumers in exploring visual narratives by automatically generating derived data explorations of narrative slices. It provides links to these explorations, which load separately from the visual narrative, to enable analysis of derived data without disrupting the flow of the narrative (RQ2).

EF4 VisEN supports authors, who have access to and an understanding of their own or others data (numeric data), in producing visual narratives, through the automatic generation of derived data slices (RQ1, RQ4, RQ6).

EF5 VisEN tailors the ordering of the derived data explorations, by inferring consumer preferences based on their recorded usage of these visualisations (RQ3).

The above evaluation features are a decomposition of *the derived data approach* and the evaluations of these features enabled conclusions to be drawn regarding the effectiveness and usability of the *approach*. In addition, the ordering of the evaluation features supported a phased development of the framework, where the later evaluation features (EF3, EF4 and EF5) had the highest priority. The implementation of VisEN, discussed in Chapter 5, closely followed the framework's design requirements, outlined in section 4.4.3, which were directly associated with the *approach*. The experiments and the evaluation findings discussed in this chapter assist in validating these design requirements and the Research Question and research objectives outlined in sections 1.2 and 1.3.

6.2 Evaluation Strategy

To evaluate the thesis based on research objective four and the framework design requirements, the evaluation strategy involved obtaining feedback from key stakeholders at various stages of the development to enable user comments and findings to be incorporated into the design process. Hence there was a need for a number of user-centred trials, focusing on various framework requirements. The stakeholders involved in the experiments consisted of professors, students, researchers and instructors as authors and consumers. In total, this research comprised of four experiments which used a combination of quantitative and qualitative evaluations including the analysis of participant logged data, questionnaire responses and interviews.

Experiment 1 consisted of a user trial, which was conducted through a proof of concept prototype. This prototype consisted of functionality to enable participants to analyse the AMAS (Staikopoulos et al., 2015) student data and construct narrative slices. It also consisted of an interface that presented pre-configured, interactive, inter-connected and explorable visualisations. The purpose of this experiment was to acquire qualitative feedback for constructing narrative slices and exploring inter-connected and explorable visualisations. The feedback from this experiment was used to determine the suitability and effectiveness of the *approach*. The proof of concept prototype was evaluated by a group of users (AMAS team members) experienced in the development of PLE.

As the feedback from the prototype was largely positive (discussed in section 6.3), it was decided to design and implement the first iteration of the VisEN framework, which consisted of the Narrative Builder, Visualisation Engine and Visual Narrative Explorer components. Experiment 2 consisted of two parts: the first part focused on evaluating the View Generation Model of the Visualisation Engine component of the VisEN framework, which automated the generation of visualisations using libraries. It was decided to evaluate the Visualisation Engine first to determine whether the View Generation Model supplied a set of appropriate visualisations for narrative slices. This was important as the mapping of narrative slice data to appropriate visualisations is central to the *approach*. The evaluation gathered feedback from experienced visualisation users, which was used to assess how accurately the Visualisation Engine determined appropriate visualisation techniques for supplied data. From the evaluation, the findings presented development areas for the Visualisation Engine.

The second part of Experiment 2 presented a prototype to the professor of the Information Management and Data Engineering course delivered every academic year in Trinity College Dublin which the AMAS PLE that supports students in learning SQL programming. The aim of this part of the experiment was to gather feedback from the professor regarding a visual narrative presenting student course engagement, activities and submissions. The objective was to allow the professor to view and analyse a visual narrative and provide feedback as to whether such a narrative, individualised for each learner, would be useful to his students during the course. The feedback from both parts of Experiment 2 was used to design and implement the second iteration of VisEN.

Experiment 3 consisted of two parts assessing the second iteration of the VisEN framework. The first part of this experiment used VisEN to provide individualised visual narratives to a total of 233 students enrolled on the Information Management and Data Engineering course during the 2013-2014 and 2014-2015 academic years by deploying it to the AMAS PLE. The visual narratives provided students with an individualised explorable visual narrative of course engagement, activity completion, time spent on resources and peer comparisons. The purpose of this experiment was to promote student engagement with course content by presenting an individualised visual narrative to each learner. The visual narratives were composed using student-logged data and aimed to provide learners with a mechanism to easily understand and explore their logged data and view peer comparisons. The evaluation analysed learner activities during the periods in which the course took place, examining visual narrative visits during poor and good course engagement by learners. At the completion of the course learners completed a questionnaire, which assessed the usability and usefulness of the visual narratives.

The second part of Experiment 3 included a structured interview with the course professor, who had used a VisEN visual narrative showing his students' data at the mid-point of the 2014-2015 academic year, containing students' engagement, completed activities and resource usage. The aim of this part of the experiment was to enable the professor to monitor his students' learning behaviour as they worked through the course activities. It also aimed to determine how useful it was for the course professor to view and analyse a visual narrative showing his students' learning behaviours. Experiment 3 provided evidence of the suitability and the effectiveness of the *approach* for authentic consumers through a realistic use case.

The fourth experiment consisted of user trials, which were conducted using the third iteration of the framework. The purpose of this experiment was to evaluate a full manifestation of the *approach* from the visual narrative construction and consumption perspectives. It focused on evaluating the end-to-end usage of the VisEN framework, including constructing and consuming visual narratives. It evaluated the support offered to authors in constructing visual narratives and the support offered to consumers in exploring visual narratives. The evaluation consisted of log analysis, and feedback from professors, instructors, demonstrators and students from various courses through questionnaires and structured interviews. It provided evidence for the suitability and the effectiveness of the end-to-end process of the *approach* for both authors and consumers.

As VisEN has been developed for the TEL domain, all of the experiments used student-logged data to construct and explore visual narratives. The AMAS student-logged data was the primary source of data, which consisted of thousands of logged entries at the end of each semester. The AMAS logged data was chosen as it was envisioned from the start of the research that the VisEN framework would be deployed to the AMAS PLE to deliver individualised visual narratives to learners as they worked through the course activities. The AMAS student data from the years 2011-2012, 2012-2013, 2013-2014 and 2014-2015 was used for the four experiments. In addition to using the AMAS student logged data, Experiment 4 also used Irish import/export[44] data to determine whether the framework could be used in domains other than TEL.

VisEN was developed in three iterations (excluding the prototype), where each iteration was evaluated through experiments and the feedback was used to guide the development of further iterations. Table 6-1 summarises the experiments conducted within this research.

	Name	Evaluated Features	Evaluation Subjects	Evaluation Method
Experiment 1	proof of concept evaluation	EF1	seven amas team members	user trial and questionnaire
Experiment 2	evaluation of the View Generation Model	EF2	nine experienced visualisation users and one professor	user trial, questionnaire and interview
Experiment 3	evaluation of visual narrative consumption through real world use cases	EF1, EF2, EF3	233 undergraduate students and the course professor	log analysis, questionnaire and interview
Experiment 4	end-to-end evaluation	EF1, EF2, EF3, EF4, EF5	40 instructors, professors and students	user trial, questionnaire and interview

Table 6-1 Thesis Experiments Summary

6.3 Experiment 1: Proof of Concept Evaluation

The first experiment evaluated a proof of concept, which enabled participants[45] to take the role of authors and connect to a MySQL database. It allowed them to analyse the data and drag data fields onto a visual canvas to construct narrative slices. It also enabled them to take on the role of a student and interact with and explore pre-configured, inter-connected and explorable visualisations that supported drilldown, zooming, details-on-demand and filtering.

6.3.1 Experimental Goals

The goal of the first experiment was to evaluate two important user activities: 1) constructing narrative slices and 2) interacting with inter-connected and explorable visualisations. The evaluation was conducted through a user trial, which gathered feedback from professors,

[44] http://wits.worldbank.org/

[45] Age and gender were not considered important in the user trials, however, skill level was an important consideration, especially in experiments 1 and 2.

instructors and teaching assistants, to assess if constructing narrative slices and interacting with inter-connected and explorable visualisations was useable and beneficial for stakeholders in TEL.

6.3.2 Experimental Setup

The first experiment consisted of a user trial conducted with the AMAS project team members (Staikopoulos et al., 2015). The AMAS team members were chosen as they were experienced in the development of PLEs and had a deep understanding of student and instructor needs. Feedback regarding the two user activities was gathered through two questionnaires (one per activity).

All seven AMAS project team members participated in the user trial, which consisted of two user tasks, which used the AMAS PLE student-logged data from the 2011-2012 academic year, stored in a MySQL database. The first task involved querying the AMAS student data source and constructing a narrative slice. The purpose of querying the data source allowed the participants to examine the data structure and values and identify the narrative slice to construct. Each participant was provided with ten minutes of training, which covered how to construct a narrative slice, and was then asked to construct one following a script. The script included screen shots, the steps to be executed and the purpose of each field in the form. Feedback for this task was gathered through a questionnaire that focused on whether the participants believed they could query data sources and construct narrative slices via the prototype interface and whether they considered it user friendly or not.

The second task required participants to attempt to answer eight questions by interacting with two sets of pre-configured, inter-connected and explorable visualisations through a web browser. Inter-connected and explorable visualisations were indicative of what was believed at the time that VisEN would produce. The first set of visualisations included a bar and line chart visualisations, showing the number of students that had completed each of the AMAS SQL course activities (bar chart), and a breakdown of the days taken for each activity to be completed (line chart). The second set of visualisations included a network visualisation and a bar chart. The network visualisation showed the relationships between students who received one or more identical title recommendations from the recommender component of the AMAS SQL course. The bar chart showed the ratings given by students to recommended titles. The visualisations supported a number of interaction techniques, including drilldown, details-on-demand and data filtering. Figure 6.1 shows one of the visualisations from the second set of visualisations.

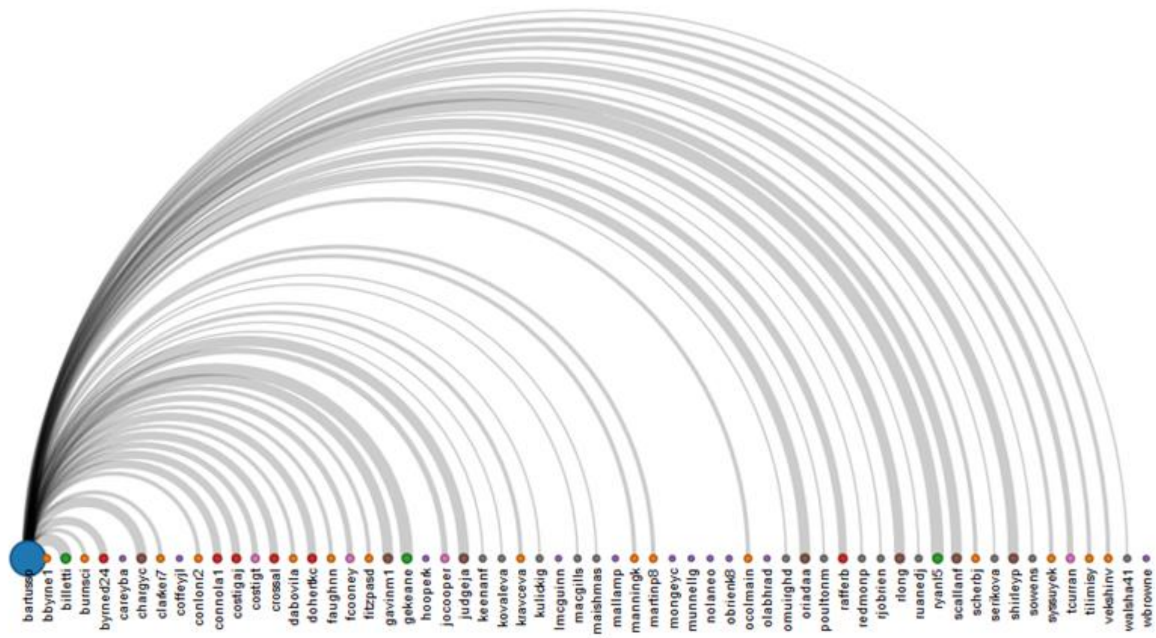

Figure 6.1 Arc diagram showing student relationships regarding recommended titles

Feedback for the second task was gathered through a questionnaire that focused on the usefulness of the inter-connected and explorable visualisations that supported navigation across a set of views which included descriptions of the data presented.

6.3.3 Experimental Results

Each participant was asked to complete the two task-related questionnaires and two usability (SUS) questionnaires (one per task). The SUS questionnaire is presented in appendix C. The data analysis and narrative slice construction questionnaire (from the first task) consisted of three statements and asked users to indicate the level of difficulty in creating narrative slices and comment on the construction process. The visual navigation interface questionnaire (from the second task) consisted of six statements and addressed the experience navigating and interacting with inter-connected and explorable visualisations.

Statements	Strongly Agree count	Agree Count	Disagree Count	Strongly Disagree Count
I found it easy to query the source data through the Data Analysis Interface	2	5	0	0
Once it was demonstrated how to build a data table[46], I had little trouble in building the one required	0	7	0	0
The Data Analysis interface used to build data tables was user friendly	0	6	1	0

Table 6-2 Data analysis and narrative slice construction questionnaire responses

Table 6-2 displays the responses to the Data analysis and Narrative slice construction questionnaire, which shows that the process of constructing narrative slices received largely positive feedback. The questionnaire included a section for further comments for participants to add. The prevalent comments included: providing onscreen guidance, providing help options, adding more clarity to certain parts of the narrative slice creation process, allowing the editing of data fields once saved, and pre-loading the interface with data source fields. Most of these comments focused on making enhancements to the interface and were addressed during the third iteration of the framework. The data table creation interface received a mean SUS score of 67.9 with a standard deviation of 16.0. A SUS score of under 68[47] is considered below average, which indicated that further work was required to improve the usability of the interface.

Statements	Strongly Agree count	Agree Count	Disagree Count	Strongly Disagree Count
It was useful to analyse the progress of user activities through visualisations	5	2	0	0
It was useful to explore the data behind[48] a completed task bar	3	4	0	0
Loading the time-series chart to show the data behind a completed task bar was user friendly	4	3	0	0

[46] In the prototype, narrative slices were referred to as data tables.
[47] Sauro, Jeff. "Measuring usability with the system usability scale (SUS)." (2011).
[48] In the prototype data behind a visualisation referred to derived data explorations.

The visualisations used to show the recommended titles are well connected and easy to explore and interact with	5	2	0	0
Loading the bar chart to show the average ratings for the recommended titles was user friendly	2	5	0	0
Viewing the breakdown of the title ratings was useful	2	5	0	0

Table 6-3 Data analysis and narrative slice construction questionnaire responses

Table 6-3 displays the results of the visual navigation interface questionnaire, which shows a very positive response to navigating and interacting with inter-connected and explorable visualisations. This questionnaire also included a section for further comments for participants to add and the prevalent ones included: increasing the node sizes of the network visualisation and allowing text orientation to be changeable to make nodes easier to read. The interface received a mean SUS score of 90 (above average) with a standard deviation of 9.1, with no users rating it below 70.

6.3.4 Analysis & Summary

Overall, the participants were able to complete the required steps in the user trial. From the results of the user trials, it can be seen that the responses to the concept of narrative slice construction and navigating inter-connected and explorable visualisations had been largely positive. In addition the participants particularly liked interacting with and navigating the visualisations and expressed its usefulness in PLEs where students can view and explore their own progress and view peer comparisons. A number of suggestions were made by the participants regarding narrative slice construction, outlining a number of improvements that can aid the construction process. These comments were taken into consideration in the subsequent design and implementation details for the Narrative Builder interface.

The aim of Experiment 1 was to evaluate narrative slice construction and navigation of inter-connected and explorable visualisations that supported drilldown, zooming, details-on-demand and filtering. To this effect, evaluation features EF1 was fully addressed. The results of Experiment 1 showed that the *approach* is usable and useful and highlighted that a framework which supported the production of visual narratives that could be consumed by consumers would be very valuable in the TEL domain.

6.4 Experiment 2: View Generation Model Evaluation

Following the encouraging outcome from the evaluation of the proof of concept prototype, it was decided to design and implement the first iteration of the VisEN framework. The first component that was developed was the Visualisation Engine to automatically generate appropriate visualisations for narrative slice data, which is central to the *approach*. In addition, evaluating the appropriateness of these visualisations was key to constructing visual narratives. The design and implementation details of the Visualisation Engine component is discussed in Chapters 4 and 5 respectively. The experiment consisted of two parts: part one evaluated the outputs of the Visualisation Engine, generated by the View Generation Model (a sub-component of the Visualisation Engine), through experienced visualisation users; part two involved a professor generating narrative slices and viewing the visualisations produced by the View Generation Model.

6.4.1 Experimental Goals

As discussed in the design and implementation chapters of this thesis, the Visualisation Engine uses visualisation techniques from imported libraries. It maps narrative slice data to appropriate visualisation techniques and generates a ranked set of visualisations for each slice. The goal of part one of the experiment was to assess how well the visualisations generated by the View Generation Model matched the narrative slice data and to determine if the ranking was appropriate. The goal of part two of the experiment was to gather feedback from the professor of the Information Management and Data Engineering Course (delivered every year in Trinity College Dublin), after viewing and analysing a sample visual narrative targeted for his students in the following year. The final goal of this experiment was a technical assessment to determine whether the implementation of the Visualisation Engine and its sub-components provided expected outputs as described by the design of the framework.

6.4.2 Experimental Setup: Part one

Members of the Knowledge and Data Engineering research group in Trinity College Dublin, who had experience analysing data through visualisations were invited to participate in part one of Experiment 2. It was important that the participants were experienced visualisation users as they could provide useful feedback regarding the accuracy of the outputs of the Visualisation Engine and offer suggestions about other visualisations that could be used. In total nine participants responded to the call who consisted of post-doctoral researchers and PhD students who had experience working with visualisations through their own research. Prior to the user trial, five narrative slices consisting of data and metadata but without visualisations were created. The 2012-2013 student data (from 88 learners) gathered by the AMAS PLE was used to construct the narratives slices.

During the user trial, the participants were presented with descriptions from five narrative slices through a web browser. As each narrative slice was presented, a call to the Visualisation Engine was made passing the data and metadata of the narrative slice and references to a a set of generated visualisation techniques (populated with the narrative slice data) together with confidence rankings were presented through a dropdown menu in the browser. The View Generation Model ranked the generated visualisations in the high, medium or low confidence score categories. The generated visualisations were ordered in the dropdown menu in descending order of confidence and indicated the confidence score next to the type. The number of visualisations generated per narrative slice ranged from one to three per dropdown depending on how many could appropriately render the data. Participants were asked to analyse each of the visualisations per narrative slice. Figure 6.2 shows a sample narrative slice data and the corresponding dropdown menu, and Figure 6.3 shows one of the generated visualisations for a narrative slice. Participant feedback was gathered through a questionnaire consisting of four statements, which focused on each visualisation's comprehensibility, accuracy, appropriateness and the assigned confidence score. The same questionnaire was completed for each of the five narrative slices.

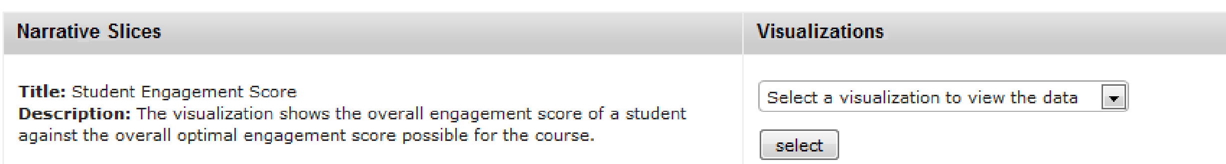

Figure 6.2 Narrative slices and visualisation lists

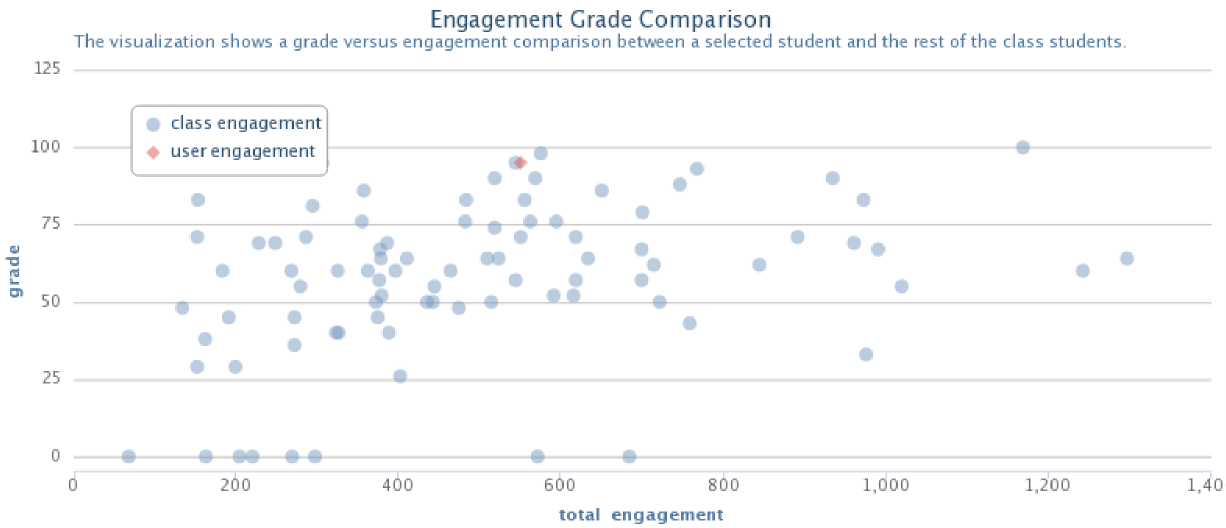

Figure 6.3 A visualisation selected from a narrative slice dropdown

6.4.3 Experimental Results: Part One

Participants were asked to read the narrative slice descriptions, select and analyse all the visualisations for each slice and respond to the following four statements using a four point scale and provide further comments:

1. the data was comprehensible when viewing it through the selected visualisation (S1),
2. the data was accurately rendered through the selected visualisation (S2),
3. the visualisation was appropriate for rendering the data (S3),
4. the score given to the visualisation was a fair reflection of how well it rendered the data (S4).

The fourth statement was displayed for cases where narrative slice data was mapped to more than one visualisation technique. The visualisation techniques for all the of the narrative slices included bar charts, time series charts, pie charts, area charts, stacked bar charts, parallel coordinates, scatterplots, gauges, Treemaps and bubble charts. Table 6-4 shows the participant feedback for the four statements by narrative slice.

Narrative Slice	Visualisation	Statement	Strongly Agree count	Agree count	Disagree count	Strongly Disagree count
Narrative Slice One	Gauge	S1[49]	7	2	0	0
		S2[50]	7	2	0	0
		S3[51]	6	2	1	0
Narrative Slice Two	Line Chart	S1	4	4	1	0
		S2	5	3	1	0
		S3	2	3	3	1
		S4[52]	2	5	2	0
	Pie Chart	S1	6	3	0	0
		S2	5	4	0	0
		S3	4	4	1	0
		S4	4	2	3	0
	Bar Chart	S1	3	4	0	2
		S2	3	3	2	1
		S3	2	4	2	1
		S4	1	4	4	0
Narrative Slice Three	Scatter Plot	S1	3	5	0	1
		S2	6	3	0	0
		S3	5	3	0	1
Narrative Slice Four	Bubble Chart	S1	1	6	1	1
		S2	3	4	2	0
		S3	2	5	1	1

[49] The data was comprehensible when viewing it through the selected visualisation.
[50] The data was accurately rendered through the selected visualisation
[51] The visualisation was appropriate for rendering the data
[52] The score given to the visualisation was a fair reflection of how well it rendered the data

Narrative Slice	Visualisation	Statement	Strongly Agree count	Agree count	Disagree count	Strongly Disagree count
Narrative Slice Five	Line Chart	S1	4	4	1	0
		S2	4	5	0	0
		S3	1	7	1	0
		S4	1	3	4	1
	Bar Chart	S1	3	4	1	1
		S2	5	3	1	0
		S3	2	6	0	1
		S4	1	4	3	1
	Stacked Area Chart	S1	4	2	1	2
		S2	4	3	1	1
		S3	3	3	1	2
		S4	1	2	5	1

Table 6-4 Questionnaire responses per narrative slice[53]

6.4.3.1 Analysis of narrative slice one responses

Narrative slice one consisted of the following description: "Overall engagement score of a student against the overall optimal engagement score possible for the course". From the responses, it can be seen that all the participants agreed and strongly agreed that the data was comprehensible and accurately rendered through the visualisation provided. A similar trend was evident for the appropriateness of the visualisation for rendering the data, with the exception of one participant, who disagreed, commenting that a linear representation could be more suitable. The responses showed that the generated visualisation (gauge) was an appropriate match for the data presented in the narrative slice.

6.4.3.2 Analysis of narrative slice two responses

Narrative slice two consisted of the following description: "A breakdown for a student's engagement score per task". The responses for S1 and S2, addressing data comprehensibility

[53] S4 was not applicable to narrative slice 1, 3 and 4 as only one visualisation was deemed appropriate and generated for the data

through the visualisation and accuracy through the visualisation respectively, were largely positive for the pie chart and line chart. The responses to S3 (addressing the suitability of the visualisation for the data) were more varied across the three visualisations, but the pie chart still had the most encouraging responses with only one participant disagreeing with this statement.

The View Generation Model assigned a low confidence score to the line chart and the bar chart and a high confidence score to the pie chart. Overall, many participants disagreed in their responses to S4, which addressed whether the confidence score was a fair reflection. Participant comments specified that the pie chart should have received a lower confidence score. Four participants commented that the bar chart should have received a medium confidence score and two participants had the same comment for the line chart. From the responses, it can be seen that the three visualisations were appropriate for the data, however, the participant believed that the confidence score rankings were inaccurate, especially for the bar chart.

6.4.3.3 Analysis of narrative slice three responses

Narrative slice three consisted of the following description: "Grade versus course engagement comparison between a selected student and the rest of the class students". The participant responses were very encouraging with the vast majority agreeing to the three statements. One participant disagreed with S3 and commented that the visualisation would not be suitable if the dataset was larger, with a greater number of nodes. However, he acknowledged that it was suitable for the data presented. The responses showed that the scatter plot was an appropriate representation of the data and no alternative visualisations were suggested by the participants.

6.4.3.4 Analysis of narrative slice four responses

The fourth narrative slice consisted of the following description: "Grade versus course engagement for the top 5 performing students". The responses for the generated visualisation (bubble chart) were positive. However, two participants believed that the data was not accurately rendered by the visualisation. Their comments highlighted that the three-dimensional data presented by the bubble chart was not easy to interpret. The responses showed that, although the data was comprehensible through the visualisation, the choice of visualisation was not the most appropriate for data analysis purposes.

6.4.3.5 Analysis of narrative slice five responses

Narrative slice five consisted of the following description: “Breakdown of task engagement score for top 5 performing students in the class”. The participant responses to S1, S2 and S3 followed a similar trend to the previous narrative slices. Some participant comments regarding the stacked area chart (which received mixed responses for S1, S2 and S3) highlighted that it was a more difficult visualisation to interpret because of its layout and how the particular data values mapped to areas and position on the chart.

The line chart and bar chart visualisations had received a low confidence score and the stacked area chart had received a medium confidence score. The responses to the statement addressing the confidence score rating (S4) were largely negative for both the line chart and the stacked area chart. Some of the participant comments suggested that both the line chart and bar chart should have received a medium confidence score and that the stacked area should have received a low confidence score. From the responses, it can be seen that the line chart and bar chart visualisations were appropriate for the data, however, the confidence score was inaccurate.

6.4.4 Experimental Setup: Part two

The second part of Experiment 2 involved the professor of the Information Management and Data Engineering course delivered annually in Trinity College Dublin (whose students worked through the AMAS PLE activities) participating in a user trial and a structured interview. The aim of this part of the experiment was to present a visual narrative with automatically generated visualisations and derived data explorations to determine the suitability of such visual narratives for his students using the AMAS PLE. This was an important step in the design of the narrative consumption part of the *approach*. The visual narrative presented a message using the AMAS student data from the 2012-2013 academic year. The professor’s task was to assess how effective the visual narrative and the derived data explorations were in informing him about his students’ activities and performance levels for the previous academic year.

The visual narrative consisted of five narrative slices: three with derived data explorations and two without. The aim of the visual narrative was to highlight students’ course engagements and included learners’ overall engagement scores and a breakdown of their engagement per activity. The visual narrative also included the time learners spent on activities and the level of interactions per activity. Finally, it included continuous assessment grades achieved by students and compared these results to the level of course engagement.

During the analysis of the visual narrative, the professor was asked to answer eleven questions about the data by interacting with the visualisations. These included identifying the time that students from various categories of engagement spent on different course activities. In some cases answering questions required the professor to analyse some of the derived data explorations. The professor spent seven minutes analysing the visual narrative and its accompanying derived data explorations and he was able to answer all the questions correctly.

Following this analysis, the professor was asked to complete a questionnaire (Appendix B) consisting of six statements which related to the usefulness of view manipulations within the narrative, and the usefulness and usability of the derived data explorations. This was followed by a structured interview which consisted of six questions focusing on whether such visual narratives with derived data explorations would be useful for his students, as they worked through their course content. In addition, the professor was asked to suggest any changes needed to the visual narrative for his students.

6.4.5 Experimental Results: Part two

The professor's responses to the statements in the questionnaire were very positive. He found the derived data explorations very useful for gaining insights into student course interactions, and he was able to follow and interact with the visual narrative effectively. He stated that the narrative slices and their corresponding visualisations represented his needs quite well. In the interview, the professor observed that the visual narratives worked very well and he was very keen on having these made available to his students in the following academic year. The professor stated that he would like to have seen more derived data explorations and felt that they worked well in the domain, and he also mentioned the usefulness of these for his students. He commented that there was a lack of comparisons, such as students who had scored similar to the one under analysis, and that these could benefit his students analysing visual narratives while working through their activities. Having consumed a visual narrative and explored its corresponding derived data explorations, the professor did not feel that the consumption and exploration of visual narratives would be in any way challenging for his students. This feedback highlighted that the visual narratives required some additions, such as more derived data and peer comparison visualisations, but in general the findings revealed that the visual narratives worked well in providing a beneficial message that could be explored and analysed through user friendly navigations and interactions.

6.4.6 Summary & Analysis

In Part One of Experiment 2, participants were presented with a total of eight visualisations. From the results of this part of the experiment, it can be seen that the responses to the statements regarding the comprehensibility, accuracy and appropriateness of the visualisations (S1, S2 and S3) were largely positive with the exception of two visualisations. One of the visualisation that did not receive largely positive feedback was the bar chart in narrative slice two. The responses were varied with suggestions that indicated that this visualisation technique did not accurately or appropriately render the particular data. The other visualisation that did not receive largely positive feedback (response was varied) was the stacked area visualisation in narrative slice five, as participants felt it was difficult to interpret the specific narrative slice data through it.

Overall, the largely positive responses for the majority of the visualisations suggested that the View Generation Model was accurately choosing appropriate visualisation techniques and rendering data in a manner that was comprehensible. However, it also showed that some modifications were needed for a minority of visualisations. The mixed responses and sometimes negative responses to statement regarding the confidence score (S4) and the participant comments showed that the ranking was not accurate. The feedback and comments were used to enhance View Generation Model for both ranking and visualisation techniques selection for the second iteration of the framework.

The aim of Part Two of this experiment was to present a visual narrative to a professor showing his students course engagement and performance data for the previous year and gather feedback about usability and usefulness of such narratives for his future students. It also aimed to include any missing requirements from suggestions made by the professor. Overall, the professor's feedback from the user trial and interview was positive. Experiment 2 addressed evaluation features EF1 (view manipulations) and EF2 (visual encodings), and the feedback from both parts of the experiment were used to finalise the design and subsequently the implementation of the *approach*.

6.5 Experiment 3: Visual Narrative Evaluations through Real World Settings

The first iteration of the VisEN framework consisted of the Visualisation Engine, which was evaluated during Experiment 2. The second iteration updated the View Generation Model following the incorporation of suggestions made in by the participants of Experiment 2, and included an implementation of the Visual Narrative Explorer component. Experiment 3 focused on evaluating the consumption of visual narratives with authentic users, through realistic use cases, and thus assessed the visual narrative consumption aspect of the *derived data approach*. The second iteration of VisEN was deployed to the AMAS PLE, which was used by 233 students (two groups of learners across two academic years: 2013-2014 and 2014-2015), studying the Information Management and Data Engineering course at Trinity College Dublin. The 233 learners consisted of third year Computer Science and fourth year Computer Engineering students. Research objective four of the thesis includes the production of individualised visual narratives for students to support the learning process in online learning environments; hence it was important to evaluate the framework with a real world use case. Experiment 3 consisted of two parts: the first part assessed the impact that the visual narratives had on the engagement and motivation of the 233 students using the AMAS PLE. It also assessed the role of the visual narratives in enhancing their performance. The second part presented a visual narrative using live student-logged data to the course professor at the mid-point of the course, which enabled him to view individual learner engagement, completed activities, time spent on activities and student comparisons. Research objective four of the thesis also includes allowing educators to view and analyse visual narratives showing student learning behaviour, so it was important to evaluate the output of the framework with the professor of the Information Management and Data Engineering course. The evaluation of this part of the experiment analysed the professor's feedback after viewing his students' study patterns during the course.

6.5.1 Experimental Goals

Students were given access to individualised visual narratives, which provided them with a message of how they were engaging with the course content to date, the resources they had used, and the time they had spent on activities. The students could view the derived data explorations that were linked to their visual narratives and which included resource usage, activity completion timings, and peer comparisons. The goal of the first part of Experiment 3 aimed to evaluate the impact that these visual narratives had on learners' course engagement and their results through five studies. The first three studies used these students' logged data containing usage of the AMAS PLE over the two academic years (2013-2014 and 2014-2015) to monitor the role played by the visual narratives in enhancing learners' course engagement. The fourth study analysed the students' feedback through a questionnaire focusing on how useful and usable the students found the visual narratives. The fifth study analysed the grades that the students received for the course to determine if there were any correlations between visual narrative usage and the results achieved.

The goal of the second part of this experiment was to present student learning behaviours to the course professor and gauge how useful it was to him for determining resource usage by his students, their engagement levels and the time they were spending on activities. The student learning behaviours were presented to the professor in a visual narrative which used his students' logged data. The evaluation was conducted by observing the professor's usage of the visual narrative (navigation, interactions) and through a structured interview.

6.5.2 Experimental Setup: Part One

The second iteration of the VisEN framework was deployed to the AMAS PLE during both the 2013-2014 and the 2014-2015 academic years. The course was run for a period of three months (October - December), during the first semester in each year, with over 120,000 student interactions logged in the course of the two academic years. The difference in setup between the two years included the addition of an extra narrative slice to the visual narratives in the 2014-2015 academic year and an update to the servers in the same year to improve the speed pages were loaded[54].

Individualised visual narratives were produced for learners using personal logged data consisting of four narrative slices, each with derived data explorations. Figure 6.4 shows some of the visualisations which form the visual narratives including the derived data explorations. The visual narratives consisted of the following narrative slices (with visualisations and descriptions):

- the learners course engagement,
- engagement breakdown by activity,
- resource Usage statistics,
- time spent on activities.

The visual narratives included the following derived data explorations, which were accessible through links in the narrative:

[54] The impact of the addition of the extra narrative slice in the 2014-2015 academic year and the server upgrade will be discussed in study 4 (analysis of students' responses to the post-course questionnaire).

- maximum and minimum time spent and resource usage per activity,
- completed timings of activities,
- peer comparisons of course engagement,
- engagement and completion time comparisons with peers,
- similar student performance.

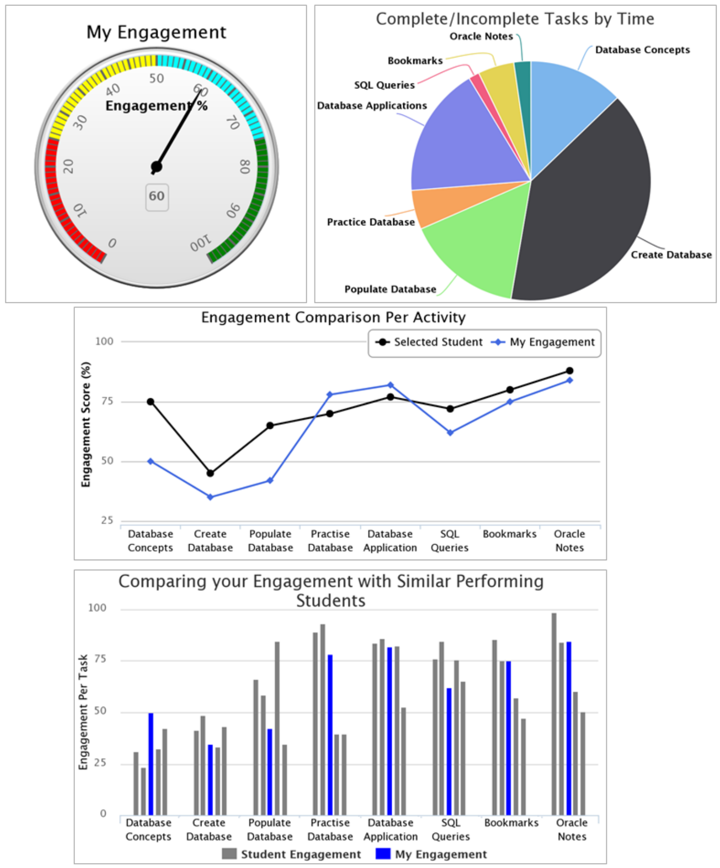

Figure 6.4 Visualisations corresponding to the four narrative slices

During the course, AMAS sent personal engagement notifications to students through email at regular intervals (one every two to three weeks) informing them of their level of engagement with the course. These notifications described the learner's current engagement with the course (bad, poor, above average, good, or excellent) and provided related advice. Course content interactions (page views) and submissions were used by AMAS to calculate engagement. Course activity and grades from the preceding years' were used to determine thresholds for calculating the level of engagement.

The motivation behind this thesis was to address the lack of learner engagement, the lack of learner motivation and the high dropout rates of students using TEL environments. Through the AMAS notifications and course content interactions, the students who had received bad or poor engagement notifications and who subsequently improved their engagement to above average or higher could be clearly identified. By separating these learners (hereafter referred to as 'improving students') from the rest, the role if any, that the visual narratives had on the enhanced engagement could be determined. From amongst the rest of the students (those who did not show this level of improved course engagement), there were strong students who did not engage well as they did not need to; students that were engaging at the top levels during the entire course; and students with poor engagement throughout the course. These students (rest of the class) did not show the same level of improvement in course engagement. Hence studies 1-3 separated the 'improving students' from the rest of the learners. Each study then analysed the impact that the visual narratives had on the improvement of these learners and compared it with the rest of the students.

To identify students improving their level of engagement, sets of engagement notifications sent in the 2013-2014 academic year were considered. In the 2013-2014 academic year, students who received a bad or poor engagement notification and subsequently received an above average, good, or excellent notification were considered as improving students. The number of notifications sent in the 2014-2015 academic year was reduced (reminder notifications were not sent) and the engagement calculation metric was updated. This meant that fewer students received an above average or higher notification (even though many of them were engaging with the course content at the same level as the previous year's students) as reminder notifications were not sent. Hence, to determine similar (to 2013-2014) improving students in the 2014-2015 academic year, a baseline using the 2013-2014 data was required. The baseline was calculated using the two variables of the students that improved their engagement from poor or bad to above average or higher. These variables included the percentage shift in course content interactions between the two periods (poor or bad to above average or higher) and the number of course content interactions during the improved engagement period. This baseline was then applied to the same course tasks as the previous year for students who had received a poor or bad engagement in the 2014-2015 academic year to identify similar improving students for this (2014-2015) academic year.

The engagement notifications did not explicitly make any references to the visual narratives and neither did they provide links to them. In addition, it is important to note that interacting or viewing these narratives did not impact engagement scores.

6.5.3 Experimental Results: Part One (Student Visual Narratives)

In total 233 students participated in the course over the two academic years. From this number, a total of 97 learners were identified as improving students. As mentioned in the experimental goals section (6.5.1), this part of the experiment consisted of five studies.

Study 1 analysed all the students that received a bad or poor engagement notification. It separated the improving students (those that improved their engagement to above average or higher - baseline used in 2014-2015 to determine this) from those that did not improve to the same level or did not improve at all. It then analysed the log data to compare the visual narrative usage of the improving students against the rest. **Study 2** analysed the student-logged data to determine the correlation between visual narrative interactions and course engagement. **Study 3** analysed the impact that the visual narratives had on the course engagement levels of historically weak students (those with an average grade of 55% or lower in previous years). It first analysed the logged data of weaker learners from amongst the improving students and then examined the logged data of the rest of the weaker students and compared both groups.

Study 4 assessed the responses to the AMAS post-course questionnaire completed by the students once the course was over. The questionnaire consisted of over fifty statements, six of which were focused on the visual narratives. The study examined the responses to the six visual narrative statements, which related to motivational factors, usefulness of the visual narratives with derived data explorations, and peer comparisons. . The analysis of the logs across both academic years found that the improving students had on average twice as much visual interactions than the rest of the class. Hence it was important to separate the responses of these students from the rest of the learner responses during the analysis. **Study 5** analysed the results achieved by the students (course continuous assessment and final exam mark) and assessed the impact that the visual narratives had on their performance.

6.5.3.1 Study 1: Visual Narrative views following a bad or poor engagement notification

In the 2013-2014 academic year, 46 learners were from amongst the category identified as improving students, and in 2014-2015, 51 learners showed a similar improvement. From the rest of the students that received a bad or poor engagement notification and did not show the same level of improvement, there were 52 learners in 2013-2014 and 46 learners in 2014-2015. Study 1 examined the usage of the visual narratives following a bad or poor engagement notification for both sets of students.

This study first analysed the number of times the students revisited their visual narratives (as opposed to interactions with individual narrative slices and derived data explorations) before and after receiving bad or poor engagement notifications. The boxplot shown in Figure 6.5 (left) presents the spread in visual narrative revisits between a period of poor or bad engagement with the course and the subsequent period of above average or higher engagement for the improving 97 students. In 2013-2014, A and B in the figure show the increase in visual narrative visits, A highlighting the number of visits during bad or poor engagement and B showing the number of visits during improved engagement. The same holds true for 2014-2015, shown through A' (poor or bad engagement) and B' (improved engagement). The shift from A to B and from A' to B' clearly indicates how these students increased their usage of their visual narratives as their course engagement improved.

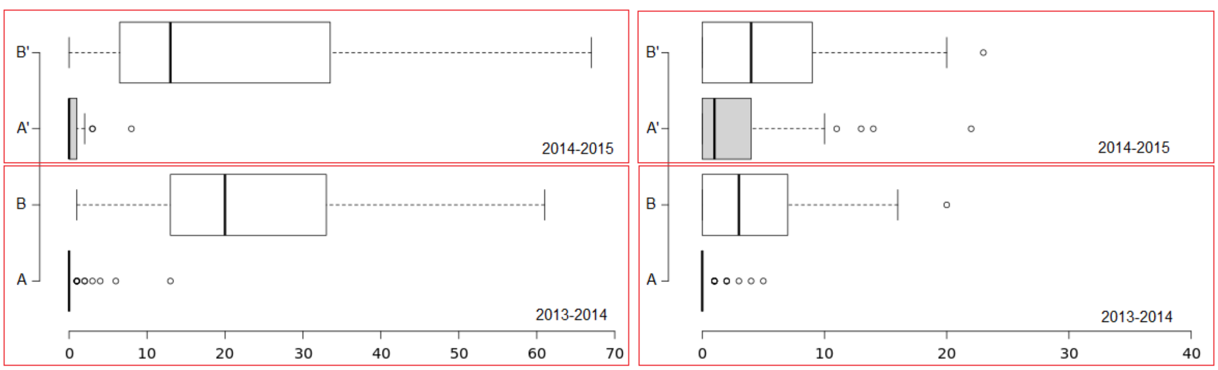

Fig. 6.5. Visual Narrative Revisits (left: improving students; right: other students)

The right of Figure 6.5 presents the spread in visual narrative revisits before and after a poor or bad engagement notification was issued to the rest of the students. These students did not improve their engagement to above average or higher (as was the case for the improving students). This boxplot shows a much smaller shift in the number visual narrative revisits in the engagement period following a poor or bad engagement notification for this group of students.

This study then analysed both visual narrative visits and visual narrative interactions following a poor or bad engagement notification. All of the 97 identified improving students (improved their engagement from bad or poor to above average or higher) showed increased interactions with their visual narratives during the period in which the engagement improved. The study found that the overwhelming majority of these students (85% in 2013-2014 and 76% in 2014-2015) showed a minimum of a fourfold increase in their visual narrative interactions and had revisited their visual narratives at least seven times during the improved engagement period.

From the rest of the students (98) that received a poor or bad engagement notification but did not improve to the same level as the former group, 21% in 2013-2014 showed a similar level of increase in their visual narrative interactions and revisits. In the 2014-2015 academic year, this figure was at 8%. All of the students from this group who showed a minimum of a fourfold increase in their visual narrative interactions (21% and 8%) had shown improvement in their engagement but it was not at the same level as the 97 identified improving students.

The findings from this study showed that the visual narratives were used to a much higher degree by the 97 improving students when their engagement increased to above average or higher. It also showed that the rest of the students who had received a bad or poor engagement notification and subsequently did not improve to a similar level, had considerably fewer interactions with their visual narratives than the improving students. This highlighted a relationship between visual narrative usage and enhanced engagement.

6.5.3.2 Study 2: Correlation between Visual Narrative Views and Engagement

Following the findings from **Study 1**, this study (i.e. **Study 2**) analysed the relationship between course content interactions and visual narrative usage. The Pearson correlation coefficient was used to determine the relationship. From the 233 students across both academic years, the baseline described in section 6.5.2 was used to categorise the learners by level of engagement:

1. the first category consisted of the 97 learners that were identified as improving students,
2. the second category of learners (59 students) consisted of those who engaged at or above the baseline figure consistently and AMAS did not identify these students as needing to show improvement. Amongst these students were some who engaged at above average or higher levels from the start of the course,
3. the third category (consisting of the rest of the students) were those with fewer course interactions than the baseline level.

The calculated Pearson correlation coefficients are presented in Table 6-5 for all three categories of students.

Category	Pearson Correlation coefficient
First Category	0.7 (strong positive correlation)
Second Category	0.52 (moderate positive correlation)
Third Category	0.43 (weak positive correlation)

Table 6-5 Correlation between course engagement and visual narrative usage

Analysis of the logs showed that the first category of learners (the 97 improving students) had on average more visual narrative interactions than the second category of students, who, in turn, had on average more visual narrative interactions than the third category of learners. The results shown in Table 6-5 highlights a clear correlation between visual narrative interactions and course engagement. This was especially the case for the improving students where it showed that an increase in visual narrative usage was strongly correlated to an increase in learner engagement.

6.5.3.3 Study 3: Assessing the impact of visual narratives on Engagement of weaker students

One of the primary aims of AMAS is to support weaker students in completing their course assignments. This study focused on analysing the impact the visual narratives had on supporting weaker learners in improving course engagement. In the 2013-2014 academic year, 108 students participated in the AMAS SQL course; 22 of these were identified as weak students as they had an average grade of or below 55% coming into the course for each of the previous two or three years of their course (two for the third year Computer Science students and three for the fourth year Computer Engineering students). In the 2014-2015 academic year, 125 students participated in the course and 23 were identified as weak students using the same criteria. From these 45 students across the two academic years, 23 were deemed to be in the category of improving students identified above (section 6.5.2).

This study first analysed the visual interactions of the 23 weaker and improving students after receiving a bad or poor engagement score. The analysis of the logged data of these 23 students was analysed and it was found that all of them immediately viewed their visual narratives and

22 of them showed a minimum of a 90% increase in interactions with their visual narratives during the period in which their engagement improved. The study analysed if there was a correlation between these students' interactions with their visual narratives and an improvement in engagement. The Pearson correlation was used to determine if an increase in visual narrative visits correlated to an improvement in course engagement for these 23 students. The Pearson correlation coefficient was found to be 0.703, which showed that an increase in visual narrative interactions amongst the improving weaker students was strongly correlated to an increase in their engagement.

For the remaining 22 weaker students who did not show the same level of engagement improvement, the analysis of the logged data found that twelve of them had not interacted with their visual narratives in the engagement period following the notification. On average the visual interactions amongst this group of learners was four times less than the visual interactions of the improving weaker students. Only one of the students from this group showed a similar level on visual interactions as the improving weaker students. For those students who did interact with their visual narratives (ten learners) following the notification, a moderate positive correlation (the Pearson correlation coefficient was 0.59) was found between engagement and visual narrative interactions.

From this study, a clear difference can be seen in the usage of visual narratives between the two groups of weaker students. This highlights the role that the visual narratives had in enhancing the engagement of the improving weaker students.

6.5.3.4 Study 4: Assessing the usefulness and usability of visual narratives through student feedback

The fourth study analysed all of the students' responses to the six of the statements (using a five-point Likert scale) related to the visual narrative from the AMAS questionnaire completed by students at the end of the course. The analysis of the logs across both academic years found that the improving students had on average a minimum of two times more visual interactions than the rest of the class. This is an important consideration when analysing the feedback and hence the responses were categorised by improving students and the rest of the class. From the 97 identified improving students, 63 completed the end of course questionnaire; 39 in 2013-2014 and 24 in 2014-2015. From the rest of the class, 98 students completed the questionnaire; 54 in 2013-2014 and 44 in 2014-2015.

The primary aim of the statements was to determine if the visual narratives had a role in motivating learners to engage with the course content after interacting with their narratives. The secondary aim of the statements were to determine how useful and usable the students found the visual narratives. The study analysed the student responses to the statements and any additional comments they provided. Chi square tests for independence were conducted using the responses to each statement, to determine if there was a relationship between the responses of the improving students and the rest of the students' responses. A chi square test for independence is used to compare frequency counts associated with categorical variables between independent groups. The results of a chi square test for independence would indicate whether there was a relationship between the student groups and the statement. Student responses were tallied across the two academic years by the options (Strongly agree, agree, undecided, disagree and strongly disagree) for the two groupings (improving students and the rest of the class). Student responses and comments were analysed for the six statements including the results of the chi square test.

Statement 1: The engagement and task duration visualisations motivated me to engage with the course.

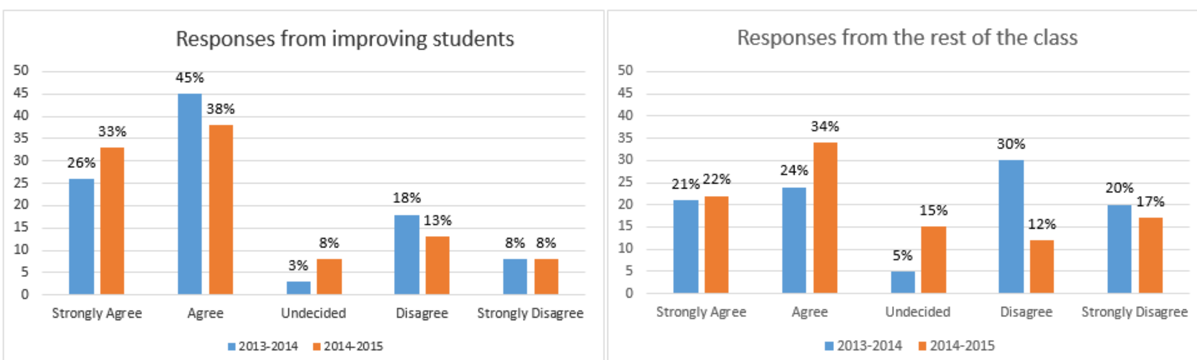

Figure 6.6 Responses from improving students (left) and the rest of the class (right)

Statement 1 incorporates all the visualisations from the visual narrative presented to the individual students. The responses showed that 71% of the improving students from the 2013-2014 and again 71% of similar students from the 2014-2015 academic year, strongly agreed or agreed that the visualisations helped motivate them to engage with the course. Such results were expected following the conclusions drawn from Studies 1 and 2 above as they confirmed that the visual narratives had an important role in motivating and supporting engagement of students who improved their engagement. The response from the rest of the class (right of Figure 6.6) was quite spread across the five options, where 45% and 56% of these students from the 2013-

2014 and 2014-2015 academic years respectively, strongly agreed or agreed. It was expected that the response from this group of students would be varied as the usage of the visual narratives was much lower amongst these students when compared to the improving learners.

A chi square test for independence was calculated using the responses of the two student groupings. The chi square statistic was 7.78, and since this value was smaller than the critical value (9.488) for the calculated degrees of freedom, this indicated that there was no relationship between the student groupings and their responses to the statement. From the analysis of the responses, it can be seen that the students with higher interactions with their visual narratives (improving students) had more positive responses. Moreover, the responses from the rest of the class students consisted of more students that agreed or strongly agreed with this statement than disagreed or strongly disagreed with it and this is reflected in the chi square statistic result.

In total, eleven comments were submitted across the two academic years from both categories of students for Statement 1. These were predominantly provided by students that disagreed, strongly disagreed or were undecided about this statement. It was found that the majority of comments discussed how the engagement metric was not a fair reflection of their engagement and thus did not motivate them. This is interesting because these students had issues with the data rather than with the visual narrative and hence did not feel motivated.

Statement 2: It was useful to explore related data through other visualisations.

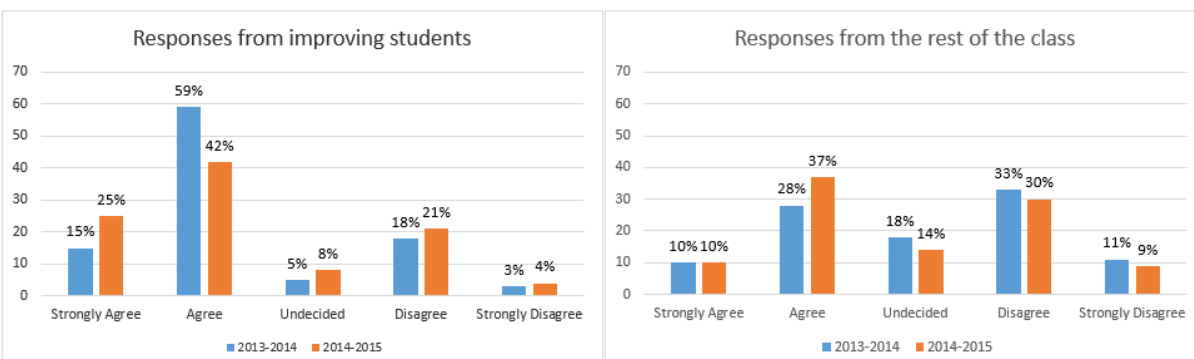

Figure 6.7 Responses from improving students (left) and the rest of the class (right)

A core and novel component of the VisEN framework is the derived data explorations presenting data related to that shown in the visualisations of the visual narrative, to enable the consumer to gain deeper insights into the message presented. The responses displayed in Figure 6.7 show that the majority of the improving students found this feature beneficial for their study, with

74% in the 2013-2014 academic year and 67% in the 2014-2015 academic year agreeing or strongly agreeing with the statement. This indicated that visual narratives consisting of derived data explorations were useful for these students. The response from the rest of the students was mixed with 38% and 48% agreeing or strongly agreeing in 2013-2014 and 2014-2015 respectively. As with all the statement responses from this group of students (rest of the class), a mixed response was expected as the visual narrative interactions were lower amongst these learners. It is also important to note that not all of the students from the rest of the class grouping viewed their derived data visualisations.

A chi square test for independence was calculated using the responses from the two student groupings. The chi square statistic was found to be 10.95, which is larger than the critical value (9.488) for the calculated degrees of freedom, showing that there was a relationship between the student grouping and the response type. Analysing the responses from both student groupings together with the chi square statistic result strongly indicated that the students with a higher level of interactions with their visual narratives found the derived data explorations more useful.

A total of nine comments were submitted with the responses to Statement 2 across the two academic years. These were submitted from both categories of students (five from students that disagreed or strongly disagreed with the statement, one from a learner that was undecided, and three from those that agreed with the statement). The comments from the students that disagreed or strongly disagreed with the statement focused on issues with how the metrics were calculated and how the engagement levels changed. One of the students (from the rest of the class category) did not interact with the visualisations or view any of the derived data exploration links as she was not aware of them. Similar to the comments from Statement 1, some the comments from students that disagreed or strongly disagreed focused on the data rather than the visual narrative.

Statement 3: I found it beneficial to be able to track my progress and compare it against other class students through personalised visualisations.

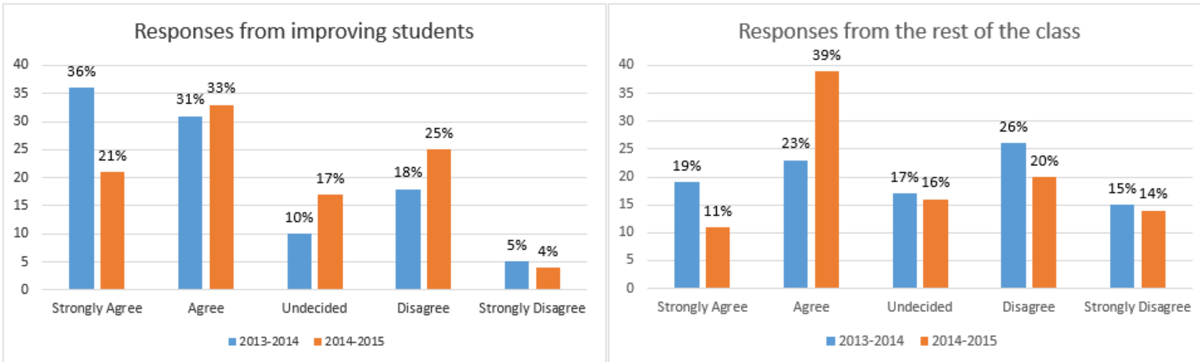

Figure 6.8 Responses from improving students (left) and the rest of the class (right)

Research (Barolli et al., 2006, Méndez et al., 2006) has shown that peer comparisons can promote self-reflection and it has also shown that students can be averse to such comparisons (Burleson et al., 2005). The responses to this statement, shown in Figure 6.8, were quite spread, with 17% (in 2014-2015) and 10% (in 2013-2014) of the students, who improved their engagement, undecided regarding the benefit of peer comparisons. Although the students that agreed or strongly agreed with the statement were in majority in this group, the responses suggested that some students were unsure about the benefit of such comparisons.

A chi square test for independence was calculated using the responses from the two student groupings. The chi square statistic was found to be 7.81, which is lower than the critical value (9.488) for the calculated degrees of freedom, and this showed that there was no relationship between the student groupings and the response type. This could be explained by the fact that responses were quite spread across the five response categories (strongly agree to strongly disagree), especially for the rest of the class students (right of Figure 6.8).

A total of thirteen comments were submitted from both categories of students for Statement 3 across the two academic years. Eight of these responses were from students that disagreed or strongly disagreed with the statement, two from students that were undecided and three from students that agreed. The comments focused on two issues: the first was that these students were unconcerned about others progress, and issues with how the metrics were calculated. The results and comments from Statement 3 showed that peer comparisons benefited the majority of students but many were averse to these comparisons.

Statement 4: I was able to follow the story provided by the visual narrative and I was able to obtain a good understanding about my own and fellow students' course related activities from these visualisations.

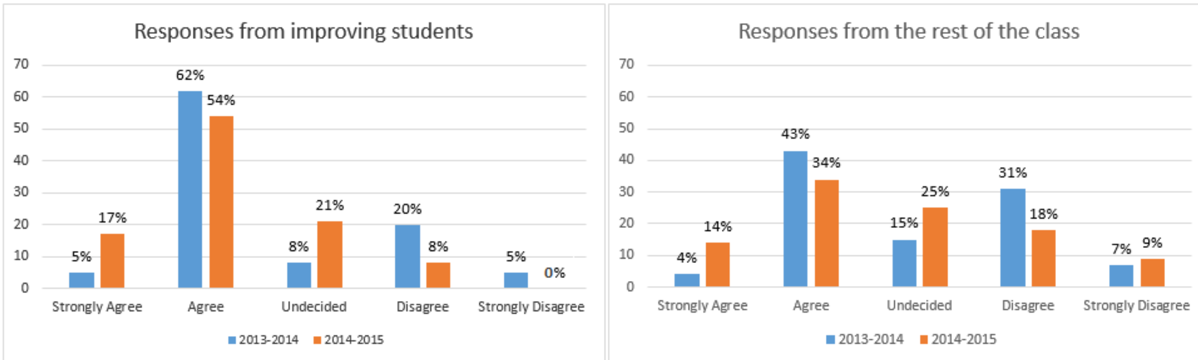

Figure 6.9 Responses from improving students (left) and the rest of the class (right)

Statement 4 addressed learners' understanding of the message communicated by their personal visual narratives. The responses, shown in Figure 6.9 (left), found that 67% and 71% of the students (who improved their engagement) agreed or strongly agreed with this statement in the 2013-2014 and 2014-2015 academic year respectively. An extra narrative slice was added to the students individualised visual narratives in 2014-2015, which may explain the slightly more positive responses to this statement in that particular academic year. The extra narrative slice presented the breakdown of the students completed tasks and showed students who had similar engagement levels. Overall, the responses to this statement show that the visual narratives were effective in communicating the intended message to students who had high visual narrative interactions. Again the responses were mixed from amongst the rest of the class students, where 47% and 48% agreed or strongly agreed with the statement in the 2013-2014 and the 2014-2015 academic years respectively.

The chi square statistic for the responses from the two student groupings for Statement 4 was 7.55, which is lower than the critical value (9.488) for the calculated degrees of freedom. This showed that there was no relationship between the student groupings and the response type. From the analysis of the responses, it can be seen that the students with a higher number of interactions with their visual narratives (improving students) were better able to follow the narrative. Nevertheless, the responses from the rest of the class students consisted of more students that agreed or strongly agreed with this statement than disagreed or strongly disagreed with it. Hence there was no relationship between the student groupings and their responses.

There were a total of five comments submitted from both categories of students for this statement across the two academic years; four of these were from students that disagreed or strongly disagreed with the statement and one from an undecided student. The comments expressed concerns about how the metric was calculated. These comments are similar to those raised with Statements 1, 2 and 3, but do not reflect Statement 4, which addresses whether the students were able to understand the message in the visual narrative.

Statement 5: I did not find it useful to be able to see how much time others spent on activities.

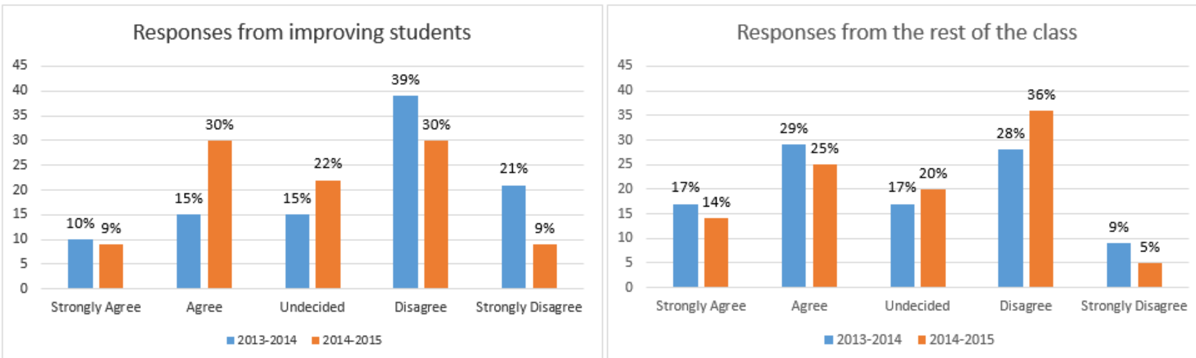

Figure 6.10 Responses from improving students (left) and the rest of the class (right)

Some of the derived data explorations of the visual narrative provided students with the time peers spent on activities that had been completed. This allowed students to compare the time they were spending against peers, but more importantly, allowed learners to estimate how much time it might take them to complete an activity yet to be started or an activity in progress. Statement 5 addressed how useful learners found this within their visual narratives. The responses to this statement presented in Figure 6.10 (left), showed that 60% and 39% found it useful from amongst the improving learners in the 2013-2014 and 2014-2015 academic years respectively. The percentage of undecided responses was also relatively high with 15% and 22% of these students undecided in the 2013-2014 and 2014-2015 academic years respectively. In addition 25% and 39% of these students across both academic years did not find it useful to view peer comparisons. The response from the rest of the class, Figure 6.10 (right), was also varied with a large percentage of these students undecided or who did not find it useful across both academic years. The mixed responses across both academic years coincides with the view that many students do not find peer comparisons beneficial.

The chi square statistic for the responses from the two student groupings for Statement 5 was 2.68, which is lower than the critical value (9.488) for the calculated degrees of freedom. This showed that there was no relationship between the student grouping and the response type. This can be explained by the fact that the feedback were varied across the five response options (Strongly Agree, Agree, Undecided, Disagree and Strongly Disagree) from both sets of students.

A total of eight comments from all the students (both categories) across the two academic years were submitted with the responses to this statement. The comments predominantly discussed how some of these students were not concerned with peer comparisons and one comment (from the rest of the class category) mentioned that the student was unaware of the comparison views.

Statement 6: Navigating from visualisation to visualisation was not user friendly.

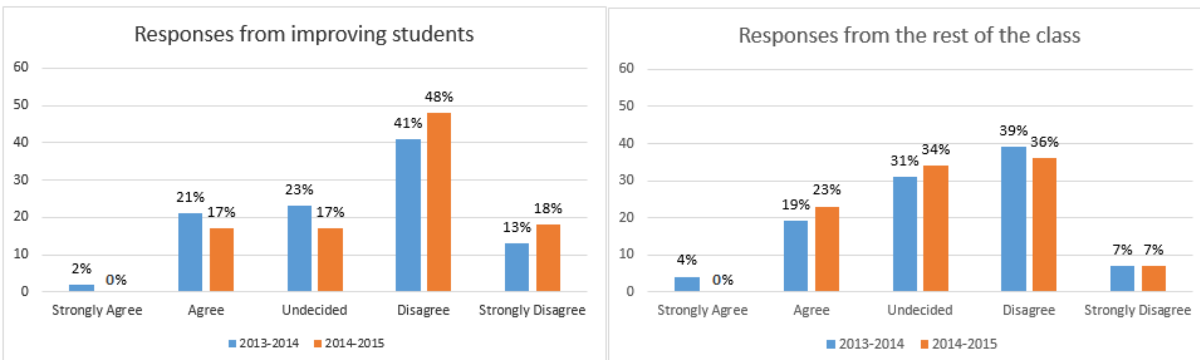

Figure 6.11 Responses from improving students (left) and the rest of the class (right)

The final statement addressed the usability of the visual narrative and in particular the navigation process. 54% and 66% of the improving students found it user friendly to navigate between the visualisations in the visual narrative in the 2013-2014 and 2014-2015 academic years respectively, as shown in Figure 6.11 (left). Several comments made by users in the 2013-2014 academic year that mentioned that the visualisations were slow to load and this comment only appeared twice in the 2014-2015 feedback from these learners. Hence the difference of 12% between this group of students in 2013-2014 and 2014-2015 may be attributed to the server upgrade that took place before the 2014-2015 academic year commenced, which meant that visualisations loaded faster. The responses from the rest of the class were varied with 31% and 34% undecided in 2013-20-14 and 2014-2015 respectively.

The chi square statistic for the responses from the two student groupings for Statement 6 was 3.107, which is lower than the critical value (9.488) for the calculated degrees of freedom. This showed that there was no relationship between the student grouping and the response type. It can be seen from the responses that the two-thirds of the improving students found the navigation user friendly when the server was upgraded. It can also be seen that although one-third of the rest of the class students were undecided, more of them found the navigation user friendly; hence there was no relationship between the student groupings and the responses given.

In total six comments were submitted with the responses to Statement 6 across the two academic years. Three were from students who did not find the navigation users friendly and the other three were from undecided learners. All six comments addressed the time it took to load the visualisations. Although the server upgrade improved the time to generated and load a visualisations, two of the students in 2014-2015 were still concerned about this and believed the navigation was not user friendly.

Summary of Study 4

Study 4 highlighted that the improving students (on average had two times more visual interactions than the rest of the class), by in large had very positive responses to the statements regarding the motivational factors, usefulness, and usability of the visual narratives. Nevertheless, the feedback from this group regarding peer comparisons were more varied across the five response options (five-point Likert scale). The responses from the rest the class was mixed, but this was expected as their visual narrative usage was much lower than the former group. The Chi Square test for independence result for Statement 2 was significant between the two groups of students. However, the Chi Square result was not significant for the other five statements. This is not necessarily a negative result, as although the students from the rest of the class provided mixed feedback, the majority of their responses fell into positive options (five-point Likert scale). Hence, the Chi Square test for independence results were not significant for these statements.

6.5.3.5 Study 5: Analysing Student Results

Studies 1-3 focused on the role of the visual narratives in enhancing the engagement levels of students. Study 4 also focused on engagement, usefulness and usability of the visual narratives and highlighted that the improving students had on average more than double the interactions with their visual narratives than the rest of the learners. Hence those studies separated students that were improving their engagement from the rest of the class. The final study focused on the role, if any, that the visual narratives had on enhancing student performance (grades). It used the students’ grades to determine if there was a correlation between their visual narrative usage and the marks they achieved in the course. Students were given two grades from the course; one related to their continuous assessments (derived from course submission markings and project work), and the second grade was based on their examination results. It should be noted that the examination consisted of questions related to databases that the class had studied through the AMAS PLE, and other questions related to topics covered by the professor during the course. This study analysed the grades students achieved from their continuous assessments from the AMAS PLE and the grades achieved from their answers to the examination questions, which related to the online course, to determine if there was any relationship between students’ grades and visual narrative interactions.

Correlations between visual narrative interactions and continuous assessment performance for all students. The continuous assessment grade included the mark a student achieved for his/her SQL course submissions that he/she worked on using the AMAS PLE. The grade also included the marks that a student achieved for his/her project submission at the end of the course. In the 2013-2014 academic year, both these marks were available and the Pearson correlation was used to determine if there was a relationship between visual narrative usage and the marks achieved by the students for the AMAS PLE submissions. The Pearson correlation was 0.53, which indicated a moderate positive relationship between performance and visual narrative usage. In the 2014-2015 academic year, only a combined mark for AMAS PLE submissions and the end of course project was available, and the Pearson correlation between this combined mark and visual narrative usage was 0.36, which indicated a weak positive correlation. Project work did not require students to use the AMAS PLE to the same degree as the required submissions (which were due prior to commencing work on the project), hence it was expected that a combined mark would result in a weaker correlation.

No correlations between visual narrative interaction and final year examination. This study also analysed the relationship between visual narrative interactions and the examination grade achieved by the students in both academic years. Pearson correlation coefficients for visual narrative interactions versus examination grade (answers related to the AMAS PLE) were calculated for these students. The Pearson correlation coefficients for both academic years was close to 0 and hence it was concluded that there was no relationship between visual narrative interactions and the examination grade. It should be noted that this is not necessarily a negative result as any usage of the AMAS PLE prior to the examination was not in the attainment of course work but rather as a revision tool. At that point the visual narratives are of less value as they focused on course work and the students at this point were revising for their examination.

6.5.4 Experimental Setup: Part Two

The second part of Experiment 3 used live student data at the mid-point of the 2014-2015 course, when students had worked through some of their course activities. A visual narrative consisting of narrative slices including course engagement per student, resource usage, time spent on activities and completed activities, was constructed and presented to the professor of the course. The data source also included student activity logs from the previous academic year, and these were used in the visual narrative to present student comparisons between the two academic years. The derived data explorations presented engagement breakdown per activity for each student, comparisons between students across the two academic years, resource usage and time spent on activities by top and bottom engaging students. The evaluation for the second part of the experiment focused on how useful the professor found the visual narrative for monitoring student learning behaviour.

6.5.5 Experimental Results: Part Two (Monitoring Student Learning Behaviour)

The professor was asked to analyse and explore the visual narrative. The professor spent fifteen minutes analysing the visual narrative and the derived data explorations. He analysed students with high course engagement and identified patterns regarding resources used and time spent by these students on the course. The professor also analysed and identified patterns in the resource usage and time spent by the students that were engaging poorly with the course. The previous year's source data included the students' continuous assessment results (as the course was completed the previous year) and thus he could analyse correlations between grade, engagement, resource usage and time spent on activities. The comparison visualisations between the two years enabled the professor to estimate student performance in the 2014-2015 academic year, using the grades, engagement levels, resource usage and time spent on activities in the previous year. Following the analysis, the professor was interviewed to determine:

- whether he found it useful to view and explore a visual narrative presenting his students' learning behaviour, and whether he could follow and understand the message,
- if he could find out everything he wanted to know or was he ever left frustrated, and
- his views on the derived data explorations.

In the interview, the professor observed that he had found the visual narrative very useful for analysing his students' learning behaviours, including when they started working on activities, the resources they were using, and the time they were spending on them. He stated that the visualisations representing the data were appropriate, and found that the derived data explorations were useful for gaining insights into the data and uncovering data not presented in the visual narrative. The professor had no difficulties in following and understanding the message communicated in the visual narrative. In addition, he was very keen on having this approach available at all times during the course in the following academic year, as this would enable him to monitor learning behaviour from the outset and allow him to tailor messages to individual students. Monitoring student learning behaviour is commonly found in LA and it has been shown to be very useful for educators (Govaerts et al., 2012, Jovanovic et al., 2008). The feedback from the professor has shown similar positive results after using the visual narratives.

6.5.6 Analysis and Summary

This experiment made some key findings through two real world use cases. Study 1 from the first part of Experiment 3 found that the majority of improving students, using the AMAS PLE over two academic years, showed a fourfold increase in visual narrative interactions during the engagement improvement. This was not the case for the rest of the students, who had far fewer visual narrative interactions than the improving students following a bad or poor engagement notification. Study 2 found a statistically strong correlation between the visual narrative interactions and course engagement amongst the improving students. Study 3 analysed the logged data of weaker students and found that all 45 of these students received a bad or poor engagement notification at some point during the course and 23 were from amongst the category of improving students. This study found that these 23 students had a much higher usage of their visual narratives during the period of improved engagement than the other 22 weaker students (that did not improve their course engagement to the same degree). This study also found a statistically a strong correlation between visual narrative usage and course content interactions amongst the 23 improving weaker students. These findings showed that the visual narratives helped weaker students in understanding their engagement score and supported them in engaging better with the course. Studies 1, 2 and 3 analysed student-logged data with their individualised visual narratives which included of derived data explorations. The analysis from these three studies found that the visual narrative interactions of the improving students was well spread across both the narrative slices and the derived data explorations.

Study 4 analysed the students' responses to the visual narrative section of the end-of-course questionnaire across both academic years. The responses showed that 71% of the improving students felt motivated to engage with the course content following their visual narratives interactions. Over 71% of the same students found the derived data explorations useful and were able to follow and understand their visual narrative. The responses from the rest of the class were varied, but importantly their visual narrative interactions on average were less than half of the number of visual interactions of the improving students. The responses to the statements related to the peer comparisons and usability were mixed amongst both groups of students. Study 4 also indicated that statistically students with higher visual narrative interactions found the derived data explorations more useful. Studies 1-4 clearly showed that the visual narratives had a very important role in enhancing the course engagement of students, especially for the improving students. Study 5 found that the visual narratives also supported the students in improving their performance in their continuous assessments grades but this enhancement was not at the same level as the course engagement improvement.

The second part of Experiment 3 showed that the professor of the Information Management and Data Engineering course was able to analyse the learning behaviour of his students, through a visual narrative, as they worked through their course activities. He was effectively able to identify trends regarding resource usage and activity times and was able to compare these trends with the data regarding student resource usage and activity times from previous academic year. Overall, he found the visual narratives very useful and stated that he would like to have access to such visual narratives throughout the duration of the course.

Experiment 3 focused on evaluating consumption of visual narratives through two real world use cases and addressed evaluation features EF1 (view manipulations) and EF2 (visual encodings) and EF3 (supporting consumers explore visual narratives through derived data explorations). The feedback, logged data analysis and the examination of the breakdown of grades achieved by students that were conducted in this experiment provided compelling evidence regarding the success of the narrative consumption aspect of the *derived data approach*.

6.6 Experiment 4: Authoring and Consuming Visual Narratives

The first iteration of the VisEN framework included the implementation of the Visualisation Engine component, which was upgraded during the second iteration. The second iteration also included the implementation of the Visual Narrative Explorer component. Experiment 2 focused on evaluating features in the first iteration and Experiment 3 focused on evaluating features in the second iteration. The third iteration was a full manifestation of the implementation discussed in Chapter 5, which also included the Narrative Builder component and supported the end-to-end process of visual narrative construction and consumption. The aim of VisEN is to support authors in constructing visual narratives in the TEL domain which can be navigated and explored by consumers. Hence it was important to evaluate the end-to-end process, including visual narrative construction and the consumption of these constructed narratives. Experiment 4 focused on evaluating both the construction and consumption of visual narratives and thus assessed both aspects of the *derived data approach*, (the *approach* was designed to address the Research Question of this thesis). It consisted of a user trial, where participants assumed the role of authors to construct a visual narrative, and then as consumers, to consume a different visual narrative.

6.6.1 Experimental Goals

Experiment 4 had four goals: the first goal was to evaluate the support offered to authors in constructing visual narratives using the VisEN Narrative Builder Component through a questionnaire, task completion times and structured interviews. This included evaluating the usefulness of presenting derived data slices to authors to support the narrative construction process, as they created narrative slices. Part of research objective four (section 1.3) was to support the construction of visual narratives in the TEL domain. Experiment 4 evaluated the support offered by the VisEN framework to users by asking them to produce visual narratives and assessing their feedback.

The second goal of Experiment 4 was to evaluate the usefulness and usability of visual narratives including the derived data explorations that were available through the narratives. Experiment 3 evaluated visual narrative consumption through real world use cases; however, the students' and professor's visual narratives were produced by the VisEN developer. In addition, the students' visual narratives were approved by the professor (Part Two of Experiment 2) prior to the deployment. The aim of the second goal of this experiment was to have users produce visual narratives that were consumed by other users, and to assess the latter through a questionnaire and structured interviews. Part of research objective four (section 1.3) was to produce visual narratives for educators of online learning environments. To this end, Part Two of Experiment

3 evaluated the visual narrative produced for the professor of the Information Management and Data Engineering course, showing student learning behaviour as they worked through their course. However, it was important to further evaluate this objective with a larger user base in order to derive sound conclusions.

The third goal of Experiment 4 was to determine whether the framework could be used to produce visual narratives using tabular and numeric data from a domain other than TEL. It also involved comparing the time required to create narrative slices (without descriptions) with VisEN against the time to create visualisations using a commonly used tool to create charts. The comparison test was carried out by a control group. Irish import and export data was chosen to achieve this goal. MS Excel was used by the control group, which consisted of ten users with expertise in creating charts through Excel.

The final goal of this experiment was a technical assessment to determine whether the full implementation of VisEN complied with the design descriptions to realise the *approach*.

6.6.2 Experimental Setup

Participant recruitment for Experiment 4 (first and second goals) involved emailing two university-based research groups. The ADAPT centre research group, spread across four universities in Ireland (Trinity College Dublin, Dublin City University, University College Dublin and Dublin Institute of Technology), and the Knowledge and Data Engineering group, in Trinity College Dublin were sent emails requesting researchers to participate in a user trial. Both research groups had a combined number of over 150 members consisting of post-graduate students, post-doctoral researchers, professors and administrative staff. A total of 40 members responded, most of whom were researchers and professors. Similar to the previous experiments, age and gender were not considered important for the evaluation.

Data from two sources, including the 2013-2014 academic year student-logged data from the AMAS PLE and Irish economic data showing top Irish imports and exports from 2008-2012, was used to evaluate the construction and consumption of visual narratives. The Irish import/export data was used to determine whether the framework could be used to produce explorable visual narratives through a domain other than TEL. In addition the Irish import/export data was freely available and conformed to the structure required by the VisEN framework (tabular and numeric). Participants were first asked to analyse and explore a previously constructed visual narrative, created using one of the two data sources and then to construct a visual narrative using the other data source. The visual narrative constructed by a participant was analysed and explored by the next participant. This construction and exploration process alternated between participants and data sources and can be explained as follows: participant N analysed and explored a visual narrative produced by participant N-1 and then produced a visual narrative to be viewed and analysed by participant N+1. This allowed participants to analyse other participants' visual narratives and helped in determining whether the description provided to participant N-1 to produce a visual narrative was similar to that offered by participant N (asked via the questionnaire).

Prior to the commencement of the user trial, two tutorial videos[55] (one for each data source) were produced which presented the construction of a visual narrative. The aim of the videos was to support participants in getting familiar with the process of constructing narrative slices and to get acquainted with the data source. The two videos were comparable with training videos provided with product releases. At the start of the user trial, each participant was shown one of the tutorial videos. As the two data sources were alternated between the participants, individuals were shown the tutorial video corresponding to the data source that they were using to construct the visual narrative.

To address the third goal of Experiment 4, an email to recruit experienced MS Excel users who regularly used the tool to create visualisations was sent to the ADAPT centre members and ten participants responded. Their task was to create three visualisations with titles and legends that were the same as those created by the participants from the first goal of Experiment 4. In general, the technical abilities of both groups of participants were deemed to be similar. The time taken to create the three visualisations by the control group was compared with the time taken by the VisEN users to determine the difference.

The rest of this section (6.6.2.1-6.6.2.3) describes the experiment setup to address the first and second goals of Experiment 4.

6.6.2.1 First Task

The participants' first task after watching the tutorial video was to view, interact with and analyse the visual narrative created by the previous participant. This visual narrative was not

[55] https://www.youtube.com/watch?v=nJR_63m7fMs, https://www.youtube.com/watch?v=ekeOLx3GWnY

from the same data source that was presented in the video to the user. As discussed in Chapters 4 and 5, visual narratives consisted of automatically generated derived data explorations that were customisable and were tailored to consumers' preferences. Before viewing the visual narratives, participants specified the level and type of data exploration through checkboxes and a slider on the Visual Narrative Interaction and Explorer interface, as shown in Figure 6.12. Participants were required to select some of the checkboxes before clicking the "view narrative" button. Participants were asked to view and interact with each narrative slice, read the corresponding descriptions and explore as many of the derived data explorations as they pleased.

Figure 6.12 Visual Narrative Interaction and Explorer interface

As discussed in Chapters 4 and 5, derived data explorations were generated using the data set used by narrative slices and by applying various mappings to it. For example, a visual narrative presenting student grades versus course engagement could consist of data related to two data sets (grades and engagement, including a combination of both). As participants navigated through the derived data explorations, the framework adjusted the ordering of the derived data exploration links based on the data that the user was spending most viewing or revisiting.

Once participants had completed the visual narrative analysis, they were asked to complete a questionnaire and a System Usability Score (SUS) for viewing and exploring visual narratives. The questionnaire consisted of one question and four statements focusing on the how well the message in the narrative was communicated, the usefulness of derived data explorations and whether customising and tailoring these explorations was beneficial.

6.6.2.2 Second Task

The second task consisted of six steps: the first four of these involved participants' constructing a visual narrative consisting of four narrative slices. The data set used to construct the visual narrative was different from the one used in the previous task (analysing a previously produced visual narrative). However, it was the same data set that had been used in the tutorial video shown to the participant, which allowed the participant to get familiar with it. Participants were given a description of what was required in each narrative slice, for example: *produce a narrative slice presenting the task durations of the top 5 performing students*. Participants were also asked to analyse the derived data slices that were automatically generated for each narrative slice. Steps 1-4 required participants to construct four narrative slices and analyse the automatically generated derived data slices. The narrative slices that participants were required to produce had several similarities to those produced in the tutorial video shown to them. Step 5 required the participants to assemble the narrative by connecting the four narrative slices.

Step 6 required users to author three more narrative slices from a different data source. These narrative slices were more challenging to create and they did not have any similarities with those produced in the tutorial video. The aim of this step was to determine whether participants having used the framework in steps 1-4 were able to produce more sophisticated narrative slices using a different data source. The data source used in step 6 was either the Irish import/export data or the 2013-2014 student data depending on what was used by the participant in steps 1-4. Students had some familiarity with the data source used in step 6 as the first task involved them analysing a visual narrative that was produced using this data. However, the three narrative slices required by step 6 were different from the slices analysed in task 1.

Once participants had finished the second task, they were asked to complete a questionnaire and a System Usability Score (SUS) for authoring visual narratives. The questionnaire consisted of eight statements focusing on the support offered in authoring visual narratives, usefulness of derived data slices, appropriateness of the visualisations generated for the narratives slices, and how difficult they found authoring visual narratives from a different data source (step 6).

6.6.2.3 Structured Interview

The final part of the user trial was a structured interview with the participants to discuss their experience with the framework. The interview focused on the:

- level of confidence the participant felt during the construction of the visual narrative,
- level of support offered by the framework in constructing visual narratives,
- support offered by the derived data slices/explorations (construction of visual narratives, exploration and gaining insight into produced visual narratives),
- tailoring and customisation of explorations within a produced visual narrative,
- how confident the participant would be using the tool with their own data.

6.6.3 Experimental Results

Experiment 4 consisted of two user tasks followed by a questionnaire for both tasks. This section examines the results of the feedback from both questionnaires and discusses findings from the interviews. This section also analyses the suitability of VisEN for producing visual narratives with non-TEL domain data by analysing narrative slice creation times.

6.6.3.1 First Task Questionnaire Results

Having completed the analysis of the visual narrative, participants were asked to complete a questionnaire consisting of four statements and one question. The question asked the participants to provide a description of what he/she understood from the data presented in the visual narrative. This was important as it allowed a comparison to be drawn between what was constructed by one participant and how the next person understood it. The four statements used a four-point Likert scale and focused on whether it was difficult to understand the visual narrative, the usefulness of the derived data checkboxes and slider, the usefulness of the derived data explorations and the usefulness of the tailoring of the derived data exploration links. The participants were encouraged to provide comments with their answers. Participants were also asked to complete a SUS test for the usability of the visual narrative consumption. The results are discussed for both the questionnaire and SUS below. From a total of 40 participants, half of them analysed visual narratives presenting student learning data (AMAS PLE data) and the other half analysed visual narratives presenting Irish import/export data (domain other than TEL).

Question 1: How would you describe the narrative viewed?

The primary aim of asking each participant to describe the visual narrative viewed was to determine whether the intended message was understood. A secondary aim was to determine if the description provided by the participant matched the specifications given to the previous person to construct the visual narrative.

The results showed that 26 of the 40 participants were able to fully and accurately describe the visual narrative (primary aim), six participants partially described and the remaining eight misunderstood the questioned (established in the interview). The eight participants that misunderstood the question described the layout and support offered by Visual Narrative Interface. The description from the 26 participants that accurately outlined the visual narrative matched the specification given to the users when asked to construct the narrative (secondary aim). The six participants that partially described the visual narrative understood the message (established in the interview) but decided to describe it with one bullet point. This showed that the framework allowed authors to construct visual narrative that communicated an intended message that could be fully understood by consumers.

Statements 1-4: First Task Questionnaire

Table 6-6 specifies the four statements and quantifies the participant responses.

Statements	Strongly agree count	Agree count	Disagree count	Strongly disagree count
S1: The message communicated by the visual narrative was difficult to understand.	0	3	25	12
S2: The data exploration check boxes and slider are useful in allowing one to select the data they would like to explore in the narrative.	14	22	3	1
S3: The explorations were not useful in enabling me to gain insight into the visual narrative.	0	1	25	14
S4: The re-ordering of explorations links based on my interactions with the explorations is useful in enabling me to quickly find related data that I am interested in.	4	16	19	1

Table 6-6 First task questionnaire statements responses

From Table 6-6, it can be seen that the participants' responses to statements 1-3 were largely positive, which showed that the majority found the message communicated easy to understand, gained insight from the derived data explorations, and found their customisation useful. The responses from statement 4 were quite varied, with only half of the participants finding the tailoring of derived data exploration links useful.

Participant Comments. Comments were provided by several participants with answers to each of the statements. Five comments were submitted with the responses to Statement 1, two of which were from the participants that found the message communicated by the visual narrative difficult to understand. The comments stated that the visualisations were confusing and they had issues understanding the metric used for the axes. The remaining comments suggested that the message was easy to follow and the combination of visualisations together with descriptions was beneficial.

Eight participants provided comments with their responses to Statement 2. From amongst the four participants that specified that customising derived data exploration links via checkboxes and a slider was not useful, two of them provided comments. One of them outlined that he had not tried various alternatives provided by the interface (combination of various checkboxes) and hence it was difficult to make a judgement. The other outlined that it was difficult to judge when he was not looking for something specific in the derived data explorations. The remaining comments highlighted the usefulness of customising explorations and mentioned that it was intuitive and facilitated specific insight.

Six participants provided comments with their responses to Statement 3, including the one user that did not find the derived data explorations useful for gaining insight into the visual narrative. His comment mentioned an indexing issue in one of the bar charts in the explorations which he found confusing. The other comments reinforced the outcome from the responses to this statement as they outlined that the derived data explorations were useful and interesting.

Sixteen participants provided comments with their responses to statement 4. Fourteen of these were from participants that did not find the tailoring of the derived data exploration links useful. These comments generally outlined that the participants were unaware that the framework was inferring their preferences, or that the tailoring would have suited data with a larger number of derived data exploration links. It was later established through the interview that the majority of participants did not like the links being re-ordered when they went back to revisit a narrative slice as the links were sometimes in different positions from where they remembered that they had been. They preferred the tailoring to only occur on narrative slices they were yet to visit.

System Usability Scale. The final part of first task involved the participants completing a SUS and the score was an average of 77 with a standard deviation of 12.7. The SUS score indicated that the visual narrative consumption and exploration interface usability was high.

Hypotheses Testing. The analysis of this questionnaire involved conducting hypotheses tests for proportion, where the population size was over ten times larger than the sample size used in the experiment. One-tailed z-tests were conducted for each statement, to determine whether the null hypotheses could be rejected. Table 6-7 outlines the null hypothesis for each statement, the Z score, p-value and whether the null hypothesis was rejected or not.

Null Hypotheses	**P-value**	**Significance level**	**Reject Null hypothesis**
S1 H0: At least 90% of the population would disagree or strongly disagree that the visual narrative was difficult to understand (P>=90).	0.702	0.05	No
S2 H0: At least 90% of the population would agree or strongly agree that the check boxes and slider are useful (P>=90).	0.05	0.05	No
S3 H0: At least 90% of the population would agree or strongly agree that the derived data explorations were useful for gaining insight (P>=90).	0.94	0.05	No
S4 H0: At least 90% of the population would agree or strongly agree that the re-ordering of derived data exploration links based on interactions were useful (P>=90).	0	0.05	Yes

Table 6-7 Hypotheses Tests results

As shown in Table 6-7, the null hypotheses corresponding to first three statements was not rejected. This can be regarded as indicative that VisEN can be used to produce visual narratives that are not difficult to understand, and can provide derived data explorations that can be useful for gaining insights into the message communicated by the narrative.

6.6.3.2 Second Task Questionnaire Results

Once the visual narrative was constructed, participants were asked to complete a questionnaire consisting of eight statements. These addressed whether participants found it difficult to produce visual narratives, the appropriateness of the visualisations for the data, and the level of support offered by the framework.

Statements 1-8: Second Task Questionnaire

Table 6-8 shows the eight statements and quantifies the responses from the questionnaire.

Statements	Strongly agree count	Agree count	Disagree count	Strongly disagree count
S1: Constructing Narrative Slices was difficult.	0	1	22	17
S2: I was not able to construct the different narrative slices required for the narrative.	2	2	18	18
S3: The visualisations used in the narrative were appropriate for rendering the data.	14	25	1	0
S4: I was frustrated with the limitations of the user interface.	0	4	20	15
S5: The framework supported me in telling the story required	11	26	2	0
S6: I do not think the explorations[56] would be useful for narrative authors	0	3	19	18
S7: I did not find step 6 difficult even though it involved constructing narrative slices from a different data source, which was not introduced to me in the tutorial video	17	20	3	0
S8: Although I was not familiar with the data source used in step 6, the interface was quite intuitive and my experiences executing the previous tasks enabled me to construct the narrative slices required	19	20	1	0

Table 6-8 Second Task Questionnaire responses

[56] Derived data transformations were referred to as explorations in the third iteration of VisEN

From Table 6-8, it can be seen that the participants' responses to all eight statements were largely positive, which showed that the majority did not find it difficult to construct the required data slices. It also showed that the majority found that the framework supported them in constructing the required narrative slices, and did not cause them to become frustrated. The responses highlighted that the visualisations were appropriate for the data and the derived data slices were deemed useful to the general narrative authoring process. Finally, the majority of participants said they did not find step 6 difficult, highlighting that the framework provided an intuitive interface that supported users with unfamiliar data sources.

Participant comments. Comments were provided by several participants with answers to each of the statements. Three participants provided comments regarding Statement 1 (constructing narrative slices was difficult), including the one user who agreed with the statement, outlining that the process of dragging and dropping objects and specifying filter was confusing. The remaining comments discussed their increase in confidence as they progressed and the framework offering guidance.

There was only one comment provided regarding Statement 2 (I was not able to construct the different narrative slices required for the narrative), which came from a participant who agreed with the statement. She outlined that she found one of the four narrative slices challenging as she was unsure of how to specify the filter.

Six participants provided comments regarding statement 3 (the visualisations used in the narrative were appropriate for rendering the data), which all centred on the choices of the visualisations offered. As discussed in Chapters 4 and 5, the View Generation Model created a set of visualisations that could appropriately render the narrative slice data and all of the visualisations in the set were made available to authors to choose from. The comments mentioned that there was always one visualisation that could appropriately render the data to the participants' liking, and that the choices were very useful. One participant stated that a line chart would have been more appropriate for one of the narrative slices, which was not offered for that particular slice by the framework.

Four participants provided comments regarding statement 4 (I was frustrated with the limitations of the user interface), three from those that agreed. These comments stated that the interface should have a reset button to undo the last operation in the event of mistakes being made and not having to restart; and also that the mouse cursor should change when hovering over draggable items.

Three participants provided comments regarding statement 5 (The framework supported me in telling the story required), and one outlined that the interface should provide more tooltip support to help users get started with their visual narratives. The other two comments mentioned that they would be interested in using the framework to create visual narratives with other data.

Four participants provided comments regarding statement 6 (I do not think the explorations would be useful for narrative authors). The comments outlined that the presentation of the derived data slices were sequential rather than being able see all of them at once. One of the comments mentioned that such views would be useful for some people but not to everyone.

Four participants provided comments regarding statement 7 (I did not find step 6 difficult even though it involved constructing narrative slices from a different data source, which was not introduced to me in the tutorial video). There were no comments from the participants who disagreed with the statement. The comments that were provided were quite positive mentioning that it was not difficult to produce narrative slices once it was done through one data source.

Four participants provided comments regarding statement 8 (Although I was not familiar with the data source used in step 6, the interface was quite intuitive and my experience executing the previous tasks enabled me to construct the narrative slices required). These comments suggested providing tooltip support for setting filters, as the items in the filter list were not obvious.

System Usability Scale. The final part of task 2 involved the participants completing a SUS and the score was an average of 82 with a standard deviation of 11. The score frequency distribution showed that 36 of the 40 participants scored the system usability at 70 or higher. The SUS score showed that the usability of the framework for authoring visual narratives was high.

Hypotheses Testing. As with the first questionnaire, the analysis of this questionnaire involved conducting hypotheses tests for proportion, where the population size was over ten times larger than the sample size used in the experiment. One-tailed z-tests were conducted for each statement, to test whether the null hypotheses could be rejected. Table 6-9 outlines the results.

Null Hypotheses	P-value	Significance level	Reject Null hypothesis
S1 H0: At least 90% of the population would disagree or strongly disagree that constructing narrative slices was difficult.	0.94	0.05	No
S2 H0: At least 90% of the population would agree or strongly agree that they were able to construct the required narrative slices.	0.05	0.05	No
S3 H0: At least 90% of the population would agree or strongly agree that the narrative visualisations were appropriate.	0.94	0.05	No
S4 H0: At least 90% of the population would disagree or strongly disagree that they were frustrated with the limitations of the user interface.	0.68	0.05	No
S5 H0: At least 90% of the population would agree or strongly agree that the framework supported them in telling the story required.	0.94	0.05	No
S6 H0: At least 90% of the population would agree or strongly agree that derived data slices were useful for authors.	0.7	0.05	No
S7 H0: At least 90% of the population would agree or strongly agree that step 6 was not difficult.	0.7	0.05	No
S8 H0: At least 90% of the population would agree or strongly agree that it was not difficult to construct the narrative slices specified in step 6.	0.94	0.05	No

Table 6-9 Hypotheses Tests results

As shown in Table 6-9, the null hypotheses corresponding to statements 1-8 could not be rejected. This can be regarded as indicative that it would not be difficult for the majority of researchers to produce visual narratives using VisEN (with appropriate data). In addition it highlighted that the level of support offered by the framework can be deemed sufficient.

6.6.3.3 Determining the suitability of VisEN to construct narrative slices with non-TEL data

Part of research objective four was to determine whether the VisEN framework could be used to produce visual narratives using tabular and numeric data from a domain other than TEL. This study analysed the time it took participants to complete step 6, which required half them to construct three narrative slices using the student data and the other half to complete the three narrative slices using Irish import/export data. Step 6 was chosen for this study as it involved producing narrative slices that had no similarities with those produced in the tutorial video. In addition, there were only slight differences in the steps due to the data but the level of difficulty in completing the tasks across the two data sources was the same, as it was originally decided to have comparable datasets and narrative slices. Participants had not used the data source in previous steps and it was chosen to avoid users' familiarity with the data affecting the results. Table 6-10 shows the description provided to participants for each of the three narrative slices that were required.

Data Source	Student Data	Import/Export Data
Step 6a Narrative Slice Description	Present the total task duration for individual students	Present the total imports by individual products
Step 6b Narrative Slice Description	Present the task durations of the top 5 performing students	Present the top 5 imported products for all years
Step 6c Narrative Slice Description	Present the task durations for any student of your choice	Present the 2008 and 2012 values for imported products

Table 6-10 Narrative slice descriptions by data source for step 6

Formally, the evaluation was to attempt to detect if there was a statistically significant difference in the time taken for the tasks by the users in each data group. The null hypothesis was: "*There is no significant difference in the amount of time it takes to complete step 6 using either the student data or the Irish import/export data*" (H0). A z-test was conducted with 40 data points. The p-value for the calculated Z score (0.0078) was 0.9938, which showed that the result was not significant at $p < 0.05$ and hence the null hypothesis was not be rejected.

The nature of statistical testing means that it is not possible to assert conclusively from this test that VisEN is cross-domain[57]. However, the evidence from the test is that no difference in performance was found between the user groups in the two data sets. From this evidence, it cannot be concluded that the VisEN framework can only be used in the TEL domain. This can be regarded as indicative that VisEN is not limited to TEL.

[57] Failing to reject the null hypothesis does not confirm it.

6.6.3.4 Participant Interview

Having completed tasks 1 and 2 and the accompanying questionnaires, all the participants were interviewed and they were each asked the same set of eight questions. The aim of the interview was to get feedback from participants on their experience using the framework and also to solicit a deeper understanding to answers to key statements in the questionnaires of the two user tasks. Hence the majority of questions in the interview corresponded to statements in the two questionnaires. This section discusses the responses and suggestions made by the participants.

Question 1. What did you find easy/difficult about constructing the visual narrative?

This question corresponds to Statement 1 of the second task questionnaire: *constructing narrative slices was difficult*. The 36 participants that disagreed or strongly disagreed with Statement 1 were asked what they found easy about constructing visual narratives. There were a number of common responses, with 29 mentioning that the drag and drop operations and immediate response through a visualisation allowed them to progress through the narrative slice rather quickly. Other comments included an intuitive user interface (fourteen participants) and consistent operations (ten participants), where the narrative author needed to only know about two operations (dragging/dropping and setting filters). Many participants (fifteen participants) liked that they were shown the best visualisations for their data without the need to analyse a number of views before deciding themselves. The four participants that agreed with Statement 1 highlighted that setting filters was challenging for them, but also mentioned that it was very quick to redo the visual narrative to correct any mistakes made.

There were a number of suggestions made to improve the interface, which included, changing the mouse cursor when hovering over selectable items, providing tooltips for filtering and having undo functionality to delete the last action of the narrative author. Overall, the participant responses highlighted that the drag-and-drop operations and the immediate response to actions through visualisations were features that worked very well and made the process easy.

Question 2. How confident did you feel constructing the narrative slices?

All 40 of the participants said they felt confident by the time they reached steps 6 of the second task (constructing three narrative slices from a different data source) and eight mentioned that they were confident using VisEN after watching the tutorial video. Six participants highlighted that they had initially some problems with the terminology but once they got used to it by working through the first two narrative slices, they felt more confident. The responses showed that after the initial learning curve, the participants were confident using the framework, which in general is comparable with other tools and systems that users are new to.

Question 3. How well did you feel the framework supported you in constructing the narrative slices?

This question corresponds to statement five of the second task questionnaire: *the framework supported me in telling the story required*. The majority of the participants said they were adequately supported by the framework, highlighting the data presentation, and the visualisations that were displayed upon dragging the data or specifying filters as the key contributors. One of the participants highlighted the need for tooltip support to help one get started with building a visual narrative. One of the participants suggested that the narrative descriptions should be automatically generated and entered into the description field, which could be reviewed and edited by authors. Overall, the feedback showed that the data presentation and visualisations adequately supported the participants' authoring visual narratives.

Question 4. If you had data to visualise and a story to tell, would you consider using VisEN and how confident would you be with it?

38 of the 40 participants said they would be happy to use VisEN with their own data and two highlighted that they had specific use cases in mind that they would like to use it to examine. One participant raised concerns regarding the lack of customisation offered with the visualisations (such as automatically generated legends and labels that could not be edited) and another participant was concerned regarding his level of competence with the framework itself as a hindrance factor. Overall, the responses showed that after using the framework to author a visual narrative and three further narrative slices (step 6), the majority of participants were confident, and in two cases, eager to use it with their own data, highlighting the positive impact the framework had on participants.

Question 5. When constructing visual narratives with your own data, do you think explorations (derived data slices) would show you views of data that you not have considered and like to include?

This question corresponded to Statement 6 of the second task questionnaire: *I do not think the derived data visualisations would be useful for narrative authors*. Three participants agreed and the rest disagreed or strongly disagreed with the statement. In the interview, the participants who had disagreed with this statement said that they would most likely find interesting relationships in the data they may not have considered. Three of these participants said that when analysing your own data, one can be focused on one aspect of the data and miss out on others, which can be seen through the derived data slices. Both the responses to statement 6 and the participant responses highlighted how derived data slices have the ability to support the authoring process by presenting data that may not have been considered by the author.

Question 6. What kind of insight do you think a consumer would get from viewing the explorations (derived data explorations)? Should the author have the ability to control which of these explorations are presented to the consumer?

This question corresponded to statement 3 of the first task questionnaire: *The explorations were not useful in enabling me to gain insight into the visual narrative*. Only one participant agreed this statement. In the interview, this participant felt he did not find them useful and would have preferred the narrative on its own as there was information overload when he viewed the derived data explorations. The rest of the participants disagreed or strongly disagreed with this statement, and said that consumers could see the full picture of the data, find more information about the visual narrative that is not in the main body, see the context of the narrative and allow them to be more confident in the message communicated. The responses showed that the derived data explorations could allow consumers to make key insights and that they could be more confident in the message communicated by the narrative author. All except three of the participants felt that the derived data explorations presented to the consumers should be controlled by the authors. Their reasons included concerns that some of the data may be sensitive and in such cases needs to be controlled. Another cited reason was that the number of explorations may be too many and hence should be controlled to an overload of data for consumers.

Question 7. How useful was the customisation of the explorations (derived data explorations) to narrative consumption?

This question corresponded to Statement 2 of the first task questionnaire: *The data exploration check boxes and slider are useful to allow one to select the data they would like to explore in the narrative*. Four participants disagreed or strongly disagreed with this statement. In the interview they said it was confusing that they should need to specify this at the start when they had not seen the visual narrative; they suggested that it would be better to do this once consumers had viewed the narrative. The remaining participants agreed and strongly agreed with the statement in the questionnaire; and in the interview they mentioned various uses of such functionality including consumers expressing their data interests, providing a more personalised experience, and preventing consumers getting overwhelmed with data. The responses highlighted several benefits of customising the derived data explorations. The four participants that marked these as not being useful in the questionnaire acknowledged their importance in the interview but mentioned that the process should be reversed (customising after viewing the visual narrative).

Question 8. What were your opinions on the tailoring of the explorations (derived data explorations)?

This question corresponded to statement 4 of the first task questionnaire: *The re-ordering of derived data links based on my interactions with the explorations is useful to enable me to quickly find other explorations that I am interested in*. In the interview, twenty participants said they did not find them useful in the visual narrative that they had examined, but that they could be useful when using the framework continuously, or with more explorations, where only the ones deemed useful to the consumer are shown. The derived data exploration links were continuously re-ordered across all narratives slices, including those already visited by the participants. This feature (re-ordering links in previously viewed narrative slices) was disliked by all the participants, as they could see that a link was no longer in the same place. When asked whether the links should only be re-ordered in narrative slices not visited, 38 participants agreed. Four participants suggested colouring the links to highlight that they have been repositioned based on usage patterns. The responses highlighted the issues that participants had with the tailoring of the links and suggested improvements that may enhance the consumer experience.

6.6.3.5 Control Group Test

In addition to determining whether VisEN could be used for numeric and tabular data from a domain other than TEL, goal 4 of this experiment also included a test to analyse if the time taken to produce narrative slices (without descriptions) was similar to creating the same visualisations with MS Excel. MS Excel was chosen as it is a commonly used tool to create visualisations that are similar to those produced by VisEN. In addition, experienced MS Excel users were available at the ADAPT centre. Ten experienced MS Excel users were recruited by emailing all of the ADAPT centre members. Recruited participants were given access through MS Excel to the Irish Import and Export data used in the first and second tasks above. They were required to create three visualisations that were exactly the same as those produced in the VisEN visual narrative with the same data. The results showed that it took participants more time to create the three visualisations with MS Excel than it did to complete the same tasks with VisEN. This is highlighted by Table 6.10.

VisEN			**MS Excel**		
Sample size	**Mean completion time (seconds)**	**Standard deviation**	**Sample size**	**Mean completion time (sec)**	**Standard deviation**
20	136.5	32.31	10	283.1	37.13

Table 6-11 Mean time and standard deviation for creating three visualisations

Formally, the evaluation was to attempt to detect if there was a statistically significant difference in the time taken for the tasks to be completed with VisEN and MS Excel. The null hypothesis was: "*There is no significant difference in the amount of time it takes to create visualisations for Irish import/export data using VisEN or MS Excel*" (H0). A two-tailed t-test was conducted and the calculated T value was |11.15|. As 11.15 > 2.048 (critical value) for alpha = 0.05, the null hypothesis was rejected. It was also hypothesised: "*It is not significantly faster to create visualisations for Irish import/export data using VisEN than MS Excel*" (H0). A one-tailed t-test was conducted. As the calculated T value (-11.15) < 1.701 (critical value) for alpha = 0.05, the null hypothesis was rejected.

6.6.4 Technical Feasibility Study

Finally a technical feasibility study was conducted testing the framework with several data sources, including MySQL databases and uploading of data from MS Excel and CSV files to determine that VisEN appropriately handled tabular and numeric data from several supported sources. The study involved connecting to MySQL databases and uploading MS Excel and CSV files using the VisEN Data Connector Component and observing the metadata displayed through the Narrative builder Interface, ensuring it was all present and that the table and column names from the MySQL databases were formatted. The study also ensured that the framework was able to produce visual narratives with dynamically generated derived data transformations for these data sources.

Three CSV files with US open data[58] and consisting of education data were downloaded and imported into the framework. Two MS Excel files containing Irish import and exports (2008 - 2012) and Google and Bing search engine search result counts were imported into the framework. The student data from three academic years (2012-2013, 2013-2014 and 2014-2015) each saved in a separate MySQL database was used. The results of this study showed that the data from the three CSV files, the data from two MS Excel files, and the data from the three academic years saved in the MySQL databases were correctly formatted, and were used to produce sample visual narratives. From a technical assessment point of view, the outputs of Visual Narrative Builder Interface, the Visualisation Engine and the Visual Narrative Explorer component corresponded to what was expected from the design requirements.

6.6.5 Analysis & Summary

The primary aim of Experiment 4 was to evaluate the end-to-end process of visual narrative construction and consumption in the TEL domain. The evaluation consisted of 40 participants constructing visual narratives and consuming visual narratives produced by fellow participants. The participant were then asked to provide feedback about their experience with the process through two questionnaires and structured interviews. The secondary aim of Experiment 4 was to determine whether the framework could also be used for domains other than TEL. This was evaluated using the same cohort of users producing narrative slices applying Irish import/export data and comparing it to completion times when using the TEL data.

The responses to the two questionnaires and the interview discussions showed that the participants had favourable experiences with both the visual narrative construction and consumption processes. The authoring process was evaluated through a questionnaire consisting of eight statements and a SUS score. The responses to the statement showed that the vast majority of participants felt they did not have difficulty in constructing narrative slices and believed that the visualisations were appropriate for the data used in the slices. A large majority of participants believed that the framework supported them in telling the story required and that the derived data slices could assist authors during visual narrative construction. Narrative consumption was also evaluated through a questionnaire and the responses showed that the majority of participants did not find it difficult to understand the message and found the derived data explorations and their customisations useful. The results from both questionnaires were

[58] http://www.data.gov/

further reinforced through the interview discussions. However, the tailoring of derived data exploration links received a mixed response from the participants.

The average SUS score for the narrative construction process was 82 with a standard deviation of 10.96 and the average SUS score for narrative consumption was 77 with a standard deviation of 12.7, which are scores above the average[59], indicating a high level of system usability.

The study to determine whether VisEN could be used for a domain other than TEL analysed the completion time for three similar narrative slices across two data sources. The first data source was from the TEL domain (2013-2014 AMAS PLE student-logged data) and the other was from a non-TEL domain (Irish 2008-2012 import/export data). A two-sample z-test found no significant difference in completion times between the two data sources, and this was regarded as indicative that VisEN is not limited to TEL. The control group test using MS Excel and the Irish 2008-2012 import/export data to produce three visualisations, showed that it took less time to produce the same visualisations with VisEN than MS Excel.

Experiment 4 focused on evaluating both the narrative construction and consumption of the *approach* and thereby addressed all five of the evaluation features, which were a decomposition of it. These evaluation features included EF1 (view manipulations), EF2 (visual encoding), EF3 (supporting consumers explore visual narratives through derived data explorations), EF4 (supporting authors construct visual narratives) and EF5 (tailoring derived data explorations to consumers). The technical assessment of the implementation of the end-to-end process corresponded to the description in the design to realise the *approach*.

6.7 Summary

This chapter discussed four experiments, the first of which evaluated a prototype of the *derived data approach* and the second assessed its Visualisation Engine. Experiments 3 and 4 focused on evaluating the *approach* by authentic users through realities use cases.

The first experiment, conducted through a proof of concept prototype helped define the *approach* which was used to address the Research Question. The remaining three experiments highlighted:

[59] http://www.measuringu.com/sus.php

1. that the View Generation Model generated appropriate visualisations for narrative slices and derived data transformations (Experiments 2-4),
2. that students could navigate, analyse, and explore individualised visual narratives showing their course engagement, completed tasks and durations, and supported peer comparisons in real world use cases. In addition, it showed that the visual narratives could help motivate students to engage with course content (Experiment 3). This showed that VisEN could be used to produce explorable visual narratives that are useful to consumers in real world contexts,
3. that authors with some knowledge of their data could be supported in producing visual narratives that could be understood and explored by consumers (Experiment 4),
4. from a technical point of view, the implementation of the VisEN framework produced the expected outputs, which adhered to the design requirements outlined in Chapter 4.

In conclusion, Experiments 2, 3 and 4 provided evidence to show the effectiveness and usability of the *approach* and highlighted that authors with knowledge of their data could be supported in producing visual narratives that could in turn be understood and explored by consumers.

7 Conclusions

This chapter presents the conclusions of this thesis by describing the degree to which the Research Question and corresponding research objectives outlined in Chapter 1 have been fulfilled and discusses the contributions made by this research. The work undertaken in this thesis was focused on TEL and specifically on LA but has the potential to be extended beyond this domain. This chapter ends with a Future Work section that discusses possible routes for further investigation.

7.1 Objectives and Achievements

The Research Question of this thesis examined whether *explorable visual narratives can be used to support consumers in having a better understanding of, benefiting from and gaining insights into data from the TEL domain; and to what extent can authors be appropriately supported in producing these narratives?* The Research Question consists of two parts (supporting consumers in understanding, benefiting from and gaining insights into student-logged data and supporting authors to create visual narratives using this data). Four objectives stemming from this Research Question were defined in Chapter 1, which included:

1. Research objective one was to analyse the state of the art to identify how well authors were supported in communicating student-logged data in a visual format that is easily consumable. It also included analysing the state of the art to determine the extent to which consumers were supported in understanding student-logged data and gain deeper insights into it through explorations within the same dataset. These analyses included:
   i. The identification of best practices in Information Visualisation in accessing and communicating data and in interacting and exploring it
   ii. Using relevant features from these best practices to analyse the TEL state of the art systems that present visualisations using student-logged data, to support users in making sense of data.
2. Research objective two was to use the analysis of the state of the art (best practices and limitations) to define an *approach* to address both parts of the Research Question.
3. Research objective three was to design and implement a framework (VisEN) to realise the *approach* defined by objective two, using tabular and numeric data.

4. Research objective four was to evaluate this *approach* using authentic users and realistic use cases to determine the extent to which VisEN supports the construction and consumption of visual narratives (both parts of the Research Question).

Research objective one was addressed through:

- the review of the state of the art in Information Visualisation in supporting users in making sense of data, and how it has supported the communication of this data (Chapter 2),
- The analysis of the state of the art in visualisations in TEL (Chapter 3) that generate visualisations using student-logged data (LA).

Chapter 2 identified best practices that supported users in making sense of complex datasets, and also identified gaps in the state of the art. The best practices included:

- supporting data transformations to enable users to explore datasets through statistical measures and related data,
- supporting popular interaction techniques including select, filter, explore, elaborate and navigate,
- guiding users through datasets using visual narratives that support interactive visualisations that enable data explorations.

Chapter 3 highlighted best practices in visualisations that use student-logged data in TEL and identified a number of limitations including:

- the lack of support offered to authors in constructing visual narratives,
- data transformations only focused on statistical measures,
- the lack of support for visual narratives in TEL for guiding students and educators.

The analysis from Chapters 2 and 3, (the identification of the gaps or limitations and the best practices) greatly influenced the definition of *the derived data approach* and the design to realise this *approach* (research objectives two and three). The limitations (points 1-3 below) and best practices (points 4-7 below) that influenced the definition and design the *approach* included:

1. supporting authors with access to and an understanding of data, to produce visual narratives, through suggested sequences (derived data slices),

2. supporting consumers explore visual narratives through derived data explorations and presenting these as visualisations,
3. tailoring the ordering of the derived data explorations, by inferring end user preferences based on usage of these visualisations,
4. providing mechanisms to support appropriate access to data sources,
5. providing a process to facilitate visual encoding by mapping narrative slice data to an appropriate set of visualisation techniques,
6. providing operations to enable the authoring of narrative slices, and providing mechanisms to piece together these narrative slices to produce a visual narrative,
7. supporting consumers manipulate narrative slice views through appropriate interactions techniques.

This *derived data approach* has been used to address the Research Question. The *approach* was realised through the implementation of the VisEN framework (research objective three), which was tightly coupled with the design requirements of the *approach*. This meant that the successful evaluations of the VisEN framework validated the usefulness of the *approach* and thus answered the Research Question.

Research objective four addressed the evaluation of the *approach* through realistic use cases. A proof of concept prototype was first evaluated (Experiment 1) by users experienced in the development of TEL systems. It was a partial implementation of the *approach* and consisted of inter-connected and explorable visualisations that supported drilldown, zooming, details-on-demand and filtering. The feedback from the participants of this evaluation highlighted the usefulness of a framework implementing this *approach* in the TEL domain. The second experiment evaluated the first iteration of the VisEN framework, which included one of the key components of the *approach* (the Visualisation Engine), which generates a set of appropriate and ranked visualisations for specified data and filters. The Visualisation Engine output (set of ranked visualisations) was evaluated by users experienced in using visualisations to analyse data and their feedback was used to improve the ranking of the generated visualisations.

The second iteration of the VisEN framework included functionality to support the consumption of visual narratives that consisted of derived data explorations and addressed the first part of the Research Question. It was evaluated through Experiment 3, which involved the framework's deployment to the AMAS PLE (which sent regular engagement notifications), used by undergraduate students learning SQL programming. It was deployed for a period of two academic years and used by a total of 233 students. VisEN provided individualised visual narratives to each learner using their logged data (generated by the PLE). The evaluation analysed the impact the visual narratives had on the course engagement level of the learners, their responses to the post-course questionnaire (visualisation statements), and whether the visual narratives had any influence on their continuous assessment and final examination grades.

Experiment 3 identified students whose engagement improved during the course ('improving students'). It then compared these improving students' visual narrative usage and post-course questionnaire responses with the rest of the learners, to determine the role, if any, that the visual narratives had on their (improving students) course engagement enhancement. The improving students on average had two times more visual narrative interactions than the rest of the learners. The findings from Experiment 3 showed a strong positive correlation between course engagement and visual narrative usage for improving students. Analysis of the post-course questionnaire responses found that the majority of improving students (over 70%) believed that the visual narratives motivated them to engage with the course content. The majority of these improving students (67%-74%) also indicated that they benefited from the derived data explorations during the consumption of their visual narratives. In addition, the majority of these improving students (67%-71%) believed that they were able to gain a good understanding of their own data through the visual narratives. The responses from the rest of the learners to the questionnaire were quite mixed. These findings clearly showed that the visual narratives were very beneficial for the improving students. The findings also showed a moderate positive correlation between continuous assessment grade (AMAS PLE submissions) and visual narrative usage for all of the students.

The final experiment evaluated the third version of the framework, which included functionality supporting both the construction and consumption of visual narratives. This version was evaluated by 40 participants who were asked to construct a visual narrative and consume one created from the previous participant. Two data sources were used, one from the TEL domain and the other with Irish import and export data. The findings showed that the participants were able to follow and understand the visual narratives they were consuming and were able to gain insights into it through the derived data explorations (established via the post-consumption questionnaire). In the interview, the participants indicated that the derived data explorations enabled them to explore data not present in the main body of the visual narrative, view the data on a larger scale and be more confident in the communicated message. The findings also showed that the participants were able to successfully construct the required narrative slices using both data sources. Although the participants were constructing narrative slices for a given data source, they were asked to view the derived data slices and comment on any insights gained into the data through these slices. In the interview the participants stated that the derived data slices would support them in constructing the visual narratives (having viewed them during the visual narrative construction task), as they presented visualisations of data one might not initially have considered including. The findings from this experiment also showed that VisEN could also be used to produce visual narratives for Irish import/export data and that one could produce visualisations faster with VisEN than with MS Excel when using Irish import/export data.

This thesis introduced the following Research Question:

Can explorable visual narratives support consumers in understanding, benefiting from and gaining insights into data from the TEL domain; and to what extent can authors be appropriately supported in producing these narratives?"

This Research Question consists of two parts: the first part involves supporting consumers in understanding, benefiting from and gaining insight into TEL data that they are interested in through explorable visual narratives. The second part involves supporting authors in constructing explorable visual narratives.

This thesis defined *the derived data approach to the construction and consumption of explorable visual narratives in TEL (derived data approach)* to address both parts of the Research Question. It designed and implemented the VisEN framework to realise this *approach*. The findings from Experiment 3 showed that the majority of improving students highly benefited from their visual narratives containing derived data explorations and that they enabled them to understand their logged data, especially when they were aiming to improve their engagement with their course content. The findings from Experiment 4 showed participants were able to gain insights into the visual narratives through the derived data explorations. These findings from both Experiments 3 and 4 showed that the visual narratives containing derived data explorations supported users in understanding, benefiting from and gaining insights into TEL data, and thus answered the first part of the Research Question. Other findings from Experiment 4 highlighted that the derived data slices could support the construction of visual narratives by presenting data to the author that might not have been originally considered to be included in the visual narrative. In addition, the participants' comments gathered through Experiment 4 highlighted that VisEN provided powerful mechanisms to support authors in constructing visual narratives. The review of the state of the art showed that authors are supported in the construction of visual narratives through data access functionality, ability to specify data and filters and the automatic generation of visualisations for the data. This thesis showed that in addition to supporting these functionalities, the derived data slices can also be used to improve the level of support offered to authors and thus has addressed the second part of the Research Question.

7.2 Contributions to the State of the Art

This thesis used an innovative *approach* using derived data transformations to support the construction and consumption of visual narratives. The major contribution of this thesis is *the derived data approach to the construction and consumption of explorable visual narratives in TEL* (*derived data approach*) and its accompanying technologies and models that supported it. Specifically this *approach* supported authors in constructing visual narratives, by generating sets of appropriate visualisations for specified data, and suggested narrative sequences through derived data slices by showing data related to the narrative slices. The *approach* also supported the consumption of visual narrative by supporting navigation, visual interactions and explorations of related data visualisations generated through the derived data explorations. The *approach* contributed and progressed the state of the art by:

1. supporting authors in constructing visual narratives by presenting dynamically generated derived data slices,
2. supporting the consumption of visual narratives through derived data explorations that did not exclusively focus on statistical measures,
3. introducing tailored visual narratives for both students and educators.

The minor contribution of this thesis is the implementation of the VisEN framework to support this *approach*. Chapter 5 discussed the implementation details of the components of the framework including its models and engines. VisEN was deployed to the Science Foundation of Ireland funded AMAS project to provide individualised visual narratives to undergraduate degree students studying SQL programming via the AMAS PLE. Chapter 6 described several evaluations highlighting the benefits and insights gained by authentic users through realistic use cases, specifically the enhancement of the engagement and motivation levels of improving undergraduate students using the AMAS PLE. From the evaluation of the VisEN framework deployed to the AMAS PLE and as a standalone application, it can be concluded that the *approach* can be successfully used to support the construction and consumption of visual narratives.

Evidence to the overall contribution to the state of the art comes in the form of eight peer-reviewed publications, the first four of which are directly related to this thesis.

1. Yousuf, B., & Conlan, O. (2015). VisEN: Motivating Learner Engagement through Explorable Visual Narratives. In Design for Teaching and Learning in a Networked World (pp. 367-380). Springer International Publishing.

This publication detailed the design of the VisEN framework and focused on the motivational impact the individualised visual narratives had on students working through the AMAS PLE over two academic years through qualitative and quantitative analysis.

2. Yousuf, B., & Conlan, O. (2014). Enhancing Learner Engagement through Personalised Visual Narratives. In Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on (pp. 89-93). IEEE.

This publication focused on the role that visual narratives had on student engagement during the inaugural deployment of VisEN to the AMAS PLE.

3. Yousuf, B., and Conlan O. (2014). Constructing Narrative Visualisations as a means of Increasing Learner Engagement. Hypertext 2014 Extended Proceedings, Santiago, Chile, September 1-4, 2014.

This publication discussed the approach defined by this thesis, and described the analysis of the usage of the VisEN framework by the Information Management and Data Engineering course professor to generate visual narratives for his learners, and the impact the visual narratives had on historically weaker students.

4. Yousuf, B., and Conlan O. (2012). Supporting Users to Analyse Digitised Data from Heterogeneous Sources through Tailored Visual Exploration. Seventh European

Conference on Technology Enhanced Learning, Saarbrücken (Germany), 18 - 21 September 2012

This publication introduced a preliminary version of the approach underpinning this thesis.

5. Staikopoulos, A., OKeeffe, I., Yousuf, B., Conlan, O., Walsh, E., & Wade, V. (2015). Enhancing Student Engagement through Personalised Motivations. In Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on (pp. 340-344). IEEE.

This publication compared two iterations of the AMAS PLE: the 2012-2013 version (without personalised notifications and without visual narratives) versus 2013-2014 version (with personalised notifications and with visual narratives).

6. Staikopoulos, A., O'Keeffe, I., Rafter, R., Walsh, E., Yousuf, B., Conlan, O., & Wade, V. (2014). *AMASE:* A framework for supporting personalised activity-based learning on the web. Computer Science and Information Systems, 11(1), 343-367.

This publication discussed the AMAS Engine (AMASE) and the inclusion of visualisations.

7. O'Keeffe, I., Staikopoulos, A., Rafter, R., Walsh, E., Yousuf, B., Conlan, O., & Wade, V. (2012). Personalised activity based eLearning. In Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (p. 2). ACM.

This publication introduced the AMAS Architecture.

8. Staikopoulos, A., O’Keeffe, I., Rafter, R., Walsh, E., Yousuf, B., Conlan, O., & Wade, V. (2012). AMASE: A framework for composing adaptive and Personalised Learning Activities on the Web. In Advances in Web-Based Learning-ICWL 2012 (pp. 190-199). Springer Berlin Heidelberg.

This publication discussed the AMAS framework and described a case study focusing on a personalised SQL Course.

These are the publications to date and further publications stemming from this thesis are intended in the near future.

7.3 Future Work

This thesis has shown that the VisEN framework has had a significant impact on the engagement and motivation levels of undergraduate students across two academic years through the consumption of visual narratives. This framework will continue to be used by the AMAS PLE to produce individualised visual narratives for students taking this module during the 2015-2016 and future academic years. In addition, this thesis has shown that the *approach* also supports the construction of visual narratives through derived data transformations. However this thesis did not explore the levels of support that the derived data transformations could offer authors with various degrees of understanding of their data when constructing visual narratives. It is envisaged that this concept will be explored in future work.

It is also envisaged that the underpinning *approach* will be used in domains outside of LA, such as in Open Learner Models and Educational Data Mining and outside of TEL using VisEN as a standalone application for Visual Analytics.

7.3.1 Extending VisEN

Although the implementation of the VisEN framework closely followed the design requirements, four of these requirements (which were not required to address the Research Question) were not implemented, including the following:

1. to provide a list of databases the author had previously connected to,
2. to enable authors to decide which of the derived data explorations should be made available to consumers of the visual narrative,
3. to enable the addition of new mappings (used for the generation of derived data objects) at run time,
4. to support the addition of new visualisation techniques at run time.

These requirements support the usability of the framework and will be included to enhance future iterations of the framework and thus making it fully compliant with all the design requirements.

Although the VisEN framework has been deployed to the TEL domain, it has also been evaluated for the construction and consumption of visual narratives using government data (also numeric). Supporting non-numeric data would enable the framework to be used to produce visual narratives for a larger set of domains. This could include the construction of visual narratives using semantic data from an individual's usage history to present what he/she has been prioritising over a specified time. This extension has already been planned and will require updates to the Derived Data Visualisation Model and the View Generation Model. In addition, this extension will require support for a larger set of visualisation techniques; however, the design of the framework already handles this through visualisation schemas.

Use cases have been identified where VisEN will be used as a Visual Analytics tool for the LeanBack Learning project which is one of the Science Foundation Ireland funded ADAPT Centre[60] projects. LeanBack Learning provides personalised audio summaries of multilingual text content in response to requested user topics. VisEN will be used to present sequenced visualisations of content consumed by users. In addition, the ADAPT Centre has an active process of engaging with industry and government and it is envisaged that several opportunities to use VisEN as a Visual Analytics tool will arise, such as presenting visual narratives of the Irish general election polls.

7.3.2 Open Learner Models (OLM) and Educational Data Mining (EDM)

VisEN has been used by the AMAS PLE to present individualised visual narratives to undergraduate students which have received very positive responses. The state of the art chapter (Chapter 3) briefly discussed OLM (which presents students' current understandings and beliefs) and EDM (which uses algorithms to detect patterns and predict outcomes), which have both been applied extensively in TEL. An attractive area to explore is the incorporation of either OLM or EDM into the VisEN framework to present open learner models or performance predictions to students in addition to individualised visual narratives. As the open learner models and performance predictions present the students' current knowledge and future predictions respectively, they would be presented separately to the visual narratives. The visual narrative would be used as a prelude to the student model or performance predictions to help students understand the rationale behind the model or prediction. The performance predictions would also be beneficial to educators enabling them to identify students experiencing difficulties understanding course material.

[60] http://adaptcentre.ie/

Bibliography

Ahuja, N. J., & Sille, R. (2013). A critical review of development of intelligent tutoring systems: Retrospect, present and prospect. International Journal of Computer Science Issues, 10(4), 39-48.

Albert, D., Nussbaumer, A., & Steiner, C. M. (2010). Towards generic visualisation tools and techniques for adaptive e-learning. In Proceedings of the 18th International Conference on Computers in Education ICCE (Vol. 2010).

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: using learning analytics to increase student success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267-270). ACM.

Arnold, K. E. (2010). Signals: Applying Academic Analytics. Educause Quarterly, 33(1), n1.
Bakharia, A., & Dawson, S. (2011, February). SNAPP: a bird's-eye view of temporal participant interaction. In Proceedings of the 1st international conference on learning analytics and knowledge (pp. 168-173). ACM.

Austin, M. (2011). *Useful Fictions: Evolution, Anxiety, and the Origins of Literature*. U of Nebraska Press.

Barolli, L., Koyama, A., Durresi, A., & De Marco, G. (2006). A web-based e-learning system for increasing study efficiency by stimulating learner's motivation. Information Systems Frontiers, 8(4), 297-306.

Battista, V., & Cheng, E. (2011). Motion charts: Telling stories with statistics. In American Statistical Association Joint Statistical Meetings (Vol. 4473).

Bederson, B. B., Clamage, A., Czerwinski, M. P., & Robertson, G. G. (2004). DateLens: A fisheye calendar interface for PDAs. ACM Transactions on Computer-Human Interaction (TOCHI), 11(1), 90-119.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps.

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

Bull, S., & Kay, J. (2007). Student models that invite the learner in: The SMILI. Open learner modelling framework. International Journal of Artificial Intelligence in Education, 17(2), 89-120.

Bull, S., & Kay, J. (2013). Open learner models as drivers for metacognitive processes. In International Handbook of Metacognition and Learning Technologies (pp. 349-365). Springer New York.

Bull, S., Gardner, P., Ahmad, N., Ting, J., & Clarke, B. (2009). Use and trust of simple independent open learner models to support learning within and across courses. In User Modeling, Adaptation, and Personalization (pp. 42-53). Springer Berlin Heidelberg.

Bull, S. (2004). Supporting Learning with Open Learner Models. 4th Hellenic Conference with International Participation: Information and Communication Technologies in Education, Athens, 2004.

Burleson, K., Leach, C. W., & Harrington, D. M. (2005). Upward social comparison and self-concept: Inspiration and inferiority among art students in an advanced programme. British Journal of Social Psychology, 44(1), 109-123.

Burns, R., Schwartz, S. E., & Carberry, S. (2013). Towards Adapting Information Graphics to Individual Users to Support Recognizing Intended Messages. In UMAP Workshops.

Burns, R., Carberry, S., Elzer, S., & Chester, D. (2012). Automatically recognizing intended messages in grouped bar charts. In Diagrammatic Representation and Inference (pp. 8-22). Springer Berlin Heidelberg.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: using vision to think. Morgan Kaufmann.

Carenini, G., Conati, C., Hoque, E., & Steichen, B. (2013). User Task Adaptation in Multimedia Presentations. In UMAP Workshops.

Carver, C. S., & Scheier, M. F. (2001). On the self-regulation of behavior. Cambridge University Press.

Carr, D. B., Littlefield, R. J., Nicholson, W. L., & Littlefield, J. S. (1987). Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, *82*(398), 424-436.

Cernea, D., Ebert, A., & Kerren, A. (2013). A Study of Emotion-triggered Adaptation Methods for Interactive Visualization. In UMAP Workshops.

Charleer, S., Klerkx, J., Odriozola, S., Luis, J., & Duval, E. (2013, December). Improving awareness and reflection through collaborative, interactive visualizations of badges. In ARTEL13: Proceedings of the 3rd Workshop on Awareness and Reflection in Technology-Enhanced Learning (Vol. 1103, pp. 69-81).

Charleer, S., Odriozola, S., Luis, J., Klerkx, J., & Duval, E. (2014, September). LARAe: Learning analytics reflection & awareness environment. In ARTEL14: Proceedings of the 4th Workshop on Awareness and Reflection in Technology-Enhanced Learning.

Cheema, S., Gulwani, S., & LaViola, J. (2012, May). QuickDraw: improving drawing experience for geometric diagrams. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1037-1064). ACM.

Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. In Information Visualization, 2000. InfoVis 2000. IEEE Symposium on (pp. 69-75). IEEE.

Conati, C., Carenini, G., Harati, M., Toker, D., Fitzgerald, N., Flagg, A., ... & Ng, R. (2011, August). User-Adaptive Visualizations: Can Gaze Data Tell Us When a User Needs Them?. In Scalable Integration of Analytics and Visualization.

Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics (No. PNNL-SA-45230). Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

Derntl, M., Erdtmann, S., & Klamma, R. (2012, September). An embeddable dashboard for widget-based visual analytics on scientific communities. In Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies (p. 23). ACM.

Dias, M. M., Franco, C., Rabelo, E., & Yamaguchi, J. K. (2012). *Visualization Techniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base?*. INTECH Open Access Publisher.

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. Review of Educational Research, 78(4), 828-879.

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. Computers & Education, 63, 380-392.

Duval, E. (2011, February). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 9-17). ACM.

Eccles, R., Kapler, T., Harper, R., & Wright, W. (2008). Stories in geotime. Information Visualization, 7(1), 3-17.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making.

Essa, A., & Ayad, H. (2012, April). Student success system: risk analytics and data visualization using ensembles of predictive models. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 158-161). ACM.

Essa, A., & Ayad, H. (2012). Improving student success using predictive models and data visualisations. Research in Learning Technology, 20.

Falakmasir, M. H., Hsiao, I. H., Mazzola, L., Grant, N., & Brusilovsky, P. (2012, July). The Impact of Social Performance Visualization on Students. In Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on (pp. 565-569). IEEE.

Festinger, L. (1954). A theory of social comparison processes. Human relations, 7(2), 117-140.

France, L., Heraud, J. M., Marty, J. C., Carron, T., & Heili, J. (2006, July). Monitoring virtual classroom: Visualization techniques to observe student activities in an e-learning system. (pp. 716-720). IEEE.

Gao, T., Hullman, J. R., Adar, E., Hecht, B., & Diakopoulos, N. (2014, April). NewsViews: an automated pipeline for creating custom geovisualizations for news. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems (pp. 3005-3014). ACM.

García-Solórzano, D., Cobo, G., Santamaría, E., Morán, J. A., Monzo, C., & Melenchón, J. (2012, April). Educational monitoring tool based on faceted browsing and data portraits. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 170-178). ACM.

Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User profiles for personalized information access. In The adaptive web (pp. 54-89). Springer Berlin Heidelberg.

Golberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addion wesley, 1989.

Goligoski, E. (2012). Motivating the learner: Mozilla's open badges program. Access to Knowledge: A Course Journal, 4(1).

Gómez, D. A., Suárez, C., Therón, R., & García F. (2010). Visual analytics to support e-learning, in Advances in learning processes. M. B. Rosson, Ed. INTECH, pp. 207–228

Gonzalez, H., Halevy, A., Jensen, C. S., Langen, A., Madhavan, J., Shapley, R., & Shen, W. (2010, June). Google fusion tables: data management, integration and collaboration in the cloud. In Proceedings of the 1st ACM symposium on Cloud computing (pp. 175-180). ACM.

Gotz, D., & Wen, Z. (2009, February). Behavior-driven visualization recommendation. In Proceedings of the 14th international conference on Intelligent user interfaces (pp. 315-324). ACM.

Govaerts, S., Verbert, K., Klerkx, J., & Duval, E. (2010, January). Visualizing Activities for Self-reflection and Awareness. In ICWL (pp. 91-100).

Govaerts, S., Verbert, K., & Duval, E. (2011). Evaluating the student activity meter: two case studies. In Advances in Web-Based Learning-ICWL 2011 (pp. 188-197). Springer Berlin Heidelberg.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012, May). The student activity meter for awareness and self-reflection. In CHI'12 Extended Abstracts on Human Factors in Computing Systems (pp. 869-884). ACM.

Graham, M., & Kennedy, J. (2010). A survey of multiple tree visualisation. *Information Visualization*, *9*(4), 235-252.

Grann, J., & Bushway, D. (2014, March). Competency map: visualizing student learning to promote student success. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 168-172). ACM.

Hanrahan, P. (2006, June). Vizql: a language for query, analysis and visualization. In Proceedings of the 2006 ACM SIGMOD international conference on Management of data (pp. 721-721). ACM.

Harrer, A. (2015, July). A Design Proposal for Learner-Centered Visualisations of Learning Analytics in Collaborative Scenarios. In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on* (pp. 208-210). IEEE.

Hearst, M. A. (1995, May). TileBars: visualization of term distribution information in full text information access. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 59-66). ACM Press/Addison-Wesley Publishing Co.

Heer, J., & Boyd, D. (2005, October). Vizster: Visualizing online social networks. In Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on (pp. 32-39). IEEE.

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. Queue, 10(2), 30.

Heer, J., Viégas, F. B., & Wattenberg, M. (2007, April). Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 1029-1038). ACM.

Heer, J., Agrawala, M., & Willett, W. (2008, April). Generalized selection via interactive query relaxation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 959-968). ACM.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Commun. Acm*, *53*(6), 59-67.

Hsiao, I. H., & Brusilovsky, P. (2012). Motivational social visualizations for personalized E-learning. In 21st Century Learning for 21st Century Skills (pp. 153-165). Springer Berlin Heidelberg.

Hsiao, I. H., Bakalov, F., Brusilovsky, P., & König-Ries, B. (2013). Progressor: social navigation support through open social student modeling. New Review of Hypermedia and Multimedia, 19(2), 112-131.

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. Visualization and Computer Graphics, IEEE Transactions on, 17(12), 2231-2240.

Hullman, J., Drucker, S., Riche, N. H., Lee, B., Fisher, D., & Adar, E. (2013). A deeper understanding of sequence in narrative visualization. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2406-2415.

Hullman, J., Diakopoulos, N., & Adar, E. (2013, April). Contextifier: automatic generation of annotated stock visualizations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2707-2716). ACM.

Jerding, D. F., & Stasko, J. T. (1998). The information mural: A technique for displaying and navigating large information spaces. Visualization and Computer Graphics, IEEE Transactions on, 4(3), 257-271.

Jovanovic, J., Gaševic, D., Brooks, C., Devedžic, V., & Hatala, M. (2007). LOCO-analyst: A tool for raising teachers' awareness in online learning environments. In Creating New Learning Experiences on a Global Scale (pp. 112-126). Springer Berlin Heidelberg.

Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., & Richards, G. (2008). LOCO-Analyst: semantic web technologies in learning content usage analysis. International journal of continuing engineering education and life long learning, 18(1), 54-76.

Kay, J., Li, L., & Fekete, A. (2007, January). Learner reflection in student self-assessment. In Proceedings of the ninth Australasian conference on Computing education-Volume 66 (pp. 89-95). Australian Computer Society, Inc..

Kerly, A., Ellis, R., & Bull, S. (2008). CALMsystem: a conversational agent for learner modelling. Knowledge-Based Systems, 21(3), 238-246.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. Computer, (5), 44-50.

Kosara, R., Hauser, H., & Gresh, D. L. (2003). An interaction view on information visualization. State-of-the-Art Report. Proceedings of EUROGRAPHICS.

Koutropoulos, A., Gallagher, M. S., Abajian, S. C., de Waard, I., Hogue, R. J., Keskin, N. O., & Rodriguez, C. O. (2012). Emotive Vocabulary in MOOCs: Context & Participant Retention. European Journal of Open, Distance and E-Learning.

Krishnamoorthy, S., & North, C. (2005, July). Learnability of interactive coordinated-view visualizations. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on* (pp. 306-311). IEEE.

Kump, B., Seifert, C., Beham, G., Lindstaedt, S. N., & Ley, T. (2012, April). Seeing what the system thinks you know: visualizing evidence in an open learner model. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 153-157). ACM.

Le, D. T. (2013). Bringing data to life into an introductory statistics course with Gapminder. Teaching Statistics, 35(3), 114-122.

Lee, J. P., & Grinstein, G. (1995, October). An architecture for retaining and analyzing visual explorations of databases. In Proceedings of the 6th conference on Visualization'95 (p. 101). IEEE Computer Society.

Lee, B., Kazi, R. H., & Smith, G. (2013). SketchStory: Telling more engaging stories with data through freeform sketching. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2416-2425.

Li, I., Dey, A., & Forlizzi, J. (2010, April). A stage-based model of personal informatics systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 557-566). ACM.

Linton, F., & Schaefer, H. P. (2000). Recommender systems for learning: building user and expert models through long-term observation of application use. User Modeling and User-Adapted Interaction, 10(2-3), 181-208.

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. Computers in Human Behavior, 47, 90-97.

Liu, S., Wu, Y., Wei, E., Liu, M., & Liu, Y. (2013). Storyflow: Tracking the evolution of stories. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2436-2445.
Ma, K. L., Liao, I., Frazier, J., Hauser, H., & Kostis, H. N. (2012). Scientific storytelling using visualization. Computer Graphics and Applications, IEEE, 32(1), 12-19.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education, 54(2), 588-599.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. Acm Transactions On Graphics (Tog), 5(2), 110-141.

Mahoney, J. (2013). Data visualisation and statistical analysis within the decision making process.

Matkovic, K., Gan, H., Ammer, A., Bennett, D., Purgathofer, W., & Terblanche, M. (2012). Interactive Visual Analysis of Intensive Care Unit Data-Relationship between Serum Sodium Concentration, its Rate of Change and Survival Outcome. In GRAPP/IVAPP (pp. 648-659).

May, M., George, S., & Prévôt, P. (2011). TrAVis to Enhance Students' Self-monitoring in Online Learning Supported by Computer-Mediated Communication Tools. Computer Information Systems and Industrial Management Applications, 3, 623-634.

Mazza, R., & Botturi, L. (2007). Monitoring an online course with the GISMO tool: A case study. Journal of Interactive Learning Research, 18(2), 251-265.

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. International Journal of Human-Computer Studies, 65(2), 125-139.

Mazza, R., & Milani, C. (2004, November). Gismo: a graphical interactive student monitoring tool for course management systems. In International Conference on Technology Enhanced Learning, Milan (pp. 1-8).

Mazza, R. (2006, May). Evaluating information visualization applications with focus groups: the CourseVis experience. In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization (pp. 1-6). ACM.

Mazzola, L., & Mazza, R. (2010). GVIS: a facility for adaptively mashing up and representing open learner models. In Sustaining TEL: From Innovation to Learning and Practice (pp. 554-559). Springer Berlin Heidelberg.

Méndez, J., Lorenzo, C., Acosta, L., Torres, S., & González, E. (2006). A web-based tool for control engineering teaching. Computer Applications in Engineering Education, 14(3), 178-187.

Merceron, A., & Yacef, K. (2005a). Educational Data Mining: a Case Study. In AIED (pp. 467-474).

Merceron, A., & Yacef, K. (2005b). Tada-ed for educational data mining. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 7(1), 267-287.

Mitrovic, A., & Martin, B. (2007). Evaluating the effect of open student models on self-assessment.

Monroy, C., Rangel, V. S., & Whitaker, R. (2013, April). STEMscopes: contextualizing learning analytics in a K-12 science curriculum. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 210-219). ACM.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. The Internet and Higher Education, 8(3), 221-231.

Morton, K., Bunker, R., Mackinlay, J., Morton, R., & Stolte, C. (2012, May). Dynamic workload driven data integration in tableau. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 807-816). ACM.

Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., & Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell. In Proceedings of the Workshop on Educational Data Mining, National Conference on Artificial Intelligence (pp. 15-22). AAAI Press, Pittsburgh, PA.

Muñoz-Merino, P. J., Valiente, J. A. R., & Kloos, C. D. (2013, April). Inferring higher level learning information from low level data for the Khan Academy platform. In Proceedings of the third international conference on learning analytics and knowledge (pp. 112-116). ACM.

Nazemi, K., Stab, C., & Kuijper, A. (2011). A reference model for adaptive visualization systems. In Human-Computer Interaction. Design and Development Approaches (pp. 480-489). Springer Berlin Heidelberg.

Nistor, N., Derntl, M., & Klamma, R. (2015). Learning Analytics: Trends and Issues of the Empirical Research of the Years 2011-2014. In Design for Teaching and Learning in a Networked World (pp. 453-459). Springer International Publishing.

North, C.: A User Interface for Coordinating Visualization Based On Relational Schemata: Snap-Together Visualization, PhD thesis, University of Maryland, 2000.

Odriozola, S., Luis, J., Verbert, K., & Duval, E. (2012, October). Empowering students to reflect on their activity with StepUp!: two case studies with engineering students. In Proceedings of ARETL'12 2nd workshop on Awareness and Reflection.

Ogawa, M., & Ma, K. L. (2010, October). Software evolution storylines. In Proceedings of the 5th international symposium on Software visualization (pp. 35-42). ACM.

Olson, J. F., Howison, J., & Carley, K. M. (2010, August). Paying attention to each other in visible work communities: Modeling bursty systems of multiple activity streams. In Social Computing (SocialCom), 2010 IEEE Second International Conference on (pp. 276-281). IEEE.

Perez, R. P., Romero, C., & Ventura, S. (2010, June). A Java desktop tool for mining Moodle data. In Educational Data Mining 2011.

Popescu, E., & Cioiu, D. (2011). eMUSE-integrating Web 2.0 tools in a social learning environment. In Advances in Web-Based Learning-ICWL 2011 (pp. 41-50). Springer Berlin Heidelberg.

Popescu, E. (2012). Project-Based Learning with eMUSE. In Advances in Web-Based Learning-ICWL 2012 (pp. 41-50). Springer Berlin Heidelberg.

Rao, R., & Card, S. K. (1994, April). The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 318-322). ACM.

Riche, N. H., Lee, B., & Plaisant, C. (2010, June). Understanding interactive legends: a comparative evaluation with standard widgets. In Computer graphics forum (Vol. 29, No. 3, pp. 1193-1202). Blackwell Publishing Ltd.

Romero, C., Gutiérrez, S., Freire, M., & Ventura, S. (2008, June). Mining and visualizing visited trails in web-based educational systems. In Educational Data Mining 2008.

Romero, C., Castro, C., & Ventura, S. (2013, July). A Moodle Block for Selecting, Visualizing and Mining Students' Usage Data. In Educational Data Mining 2013.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368-384.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Kloos, C. D. (2014). A demonstration of ALAS-KA: A learning analytics tool for the Khan Academy platform. In Open Learning and Teaching in Educational Communities (pp. 518-521). Springer International Publishing.

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2011). Visualizing PLE usage. In Proceedings of EFEPLE11 1st Workshop on Exploring the Fitness and Evolvability of Personal Learning Environments (pp. 34-38).

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012, April). Goal-oriented visualizations of activity tracking: a case study with engineering students. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 143-152). ACM.

Santos, J. L., Verbert, K., Govaerts, S., & Duval, E. (2013, April). Addressing learner issues with StepUp!: an Evaluation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 14-22). ACM.

Santos, J. L., Charleer, S., Parra, G., Klerkx, J., Duval, E., & Verbert, K. (2013). Evaluating the use of open badges in an open learning environment. In Scaling up Learning for Sustained Impact (pp. 314-327). Springer Berlin Heidelberg.

Satyanarayan, A., & Heer, J. (2014, June). Authoring narrative visualizations with ellipsis. In Computer Graphics Forum (Vol. 33, No. 3, pp. 361-370).

Scheuer, O., & Zinn, C. (2007). How did the e-learning session go? The Student Inspector. Frontiers in Artificial Intelligence and Applications, 158, 487.

Schmitz, H. C., Scheffel, M., Friedrich, M., Jahn, M., Niemann, K., & Wolpers, M. (2009). CAMera for PLE. In Learning in the Synergy of Multiple Disciplines (pp. 507-520). Springer Berlin Heidelberg.

Segel, E., Heer, J. 2010. Narrative visualization: telling stories with data. IEEE Transactions on Visualization and Computer Graphics 16(6): 1139-1148.

Sharma, S., Ghorpade, S., Sahni, A., & Saluja, N. (2014, March). Survey of Intelligent Tutoring Systems: A Review on the Development of Expert/Intelligent Tutoring Systems, various Teaching Strategies and Expert Tutoring System Design Suggestions. In International Journal of Engineering Research and Technology (Vol. 3, No. 11 (November-2014)). ESRSA Publications.

Shepherd, M. M., Briggs, R. O., Reinig, B. A., Yen, J., & Nunamaker Jr, J. F. (1995). Invoking social comparison to improve electronic brainstorming: Beyond anonymity. Journal of Management Information Systems, 155-170.

Shi, L., Liao, Q., Tong, H., Hu, Y., Zhao, Y., & Lin, C. (2014, March). Hierarchical focus+ context heterogeneous network visualization. In Pacific Visualization Symposium (PacificVis), 2014 IEEE (pp. 89-96). IEEE.

Shneiderman, B. (1996, September). The eyes have it: A task by data type taxonomy for information visualizations. In Visual Languages, 1996. Proceedings., IEEE Symposium on (pp. 336-343). IEEE.

Spenke, M., & Beilken, C. (2000). InfoZoom-Analysing Formula One racing results with an interactive data mining and visualisation tool. Proc. Data Mining 2000, 455-464.

Spenke, M., & Beilken, C. (2003, October). Visualization of trees as highly compressed tables with InfoZoom. In Proceedings of the IEEE Symposium on Information Visualization (pp. 122-123).

Spenke, M. (2001). Visualization and interactive analysis of blood parameters with InfoZoom. Artificial Intelligence in Medicine, 22(2), 159-172.

Staikopoulos, A., OKeeffe, I., Yousuf, B., Conlan, O., Walsh, E., & Wade, V. (2015, July). Enhancing Student Engagement through Personalized Motivations. In Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference on (pp. 340-344). IEEE.

Stefaner, M., Dalla Vecchia, E., Condotta, M., Wolpers, M., Specht, M., Apelt, S., & Duval, E. (2007). MACE–enriching architectural learning objects for experience multiplication. In Creating New Learning Experiences on a Global Scale (pp. 322-336). Springer Berlin Heidelberg.

Steichen, B., Carenini, G., & Conati, C. (2012). Adaptive Information Visualization-Predicting user characteristics and task context from eye gaze. In UMAP Workshops (Vol. 872).

Steichen, B., Carenini, G., & Conati, C. (2013, March). User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In Proceedings of the 2013 international conference on Intelligent user interfaces (pp. 317-328). ACM.

Stolte, C., Tang, D., & Hanrahan, P. (2002). Polaris: A system for query, analysis, and visualization of multidimensional relational databases. Visualization and Computer Graphics, IEEE Transactions on, 8(1), 52-65.

Talavera, L., & Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In Workshop on artificial intelligence in CSCL. 16th European conference on artificial intelligence (pp. 17-23).

Tanahashi, Y., & Ma, K. L. (2012). Design considerations for optimizing storyline visualizations. Visualization and Computer Graphics, IEEE Transactions on, 18(12), 2679-2688.

Tang, D., Stolte, C., & Bosch, R. (2004). Design choices when architecting visualizations. Information Visualization, 3(2), 65-79.

Thille, C., & Smith, J. (2011). Cold Rolled Steel and Knowledge: What Can Higher Education Learn About Productivity?. Change: The Magazine of Higher Learning, 43(2), 21-27.

Tufte, E. R., & Graves-Morris, P. R. (1983). The visual display of quantitative information (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Tufte, E. R. (2006). Beautiful evidence. New York.

Underwood, J., & Luckin, R. (2011). 1. What is AIED and why does Education need it?

Upton, K., & Kay, J. (2009). Narcissus: group and individual models to support small group work. In User Modeling, Adaptation, and Personalization (pp. 54-65). Springer Berlin Heidelberg.

Van den Elzen, S., & van Wijk, J. J. (2013, June). Small multiples, large singles: A new approach for visual data exploration. In Computer Graphics Forum (Vol. 32, No. 3pt2, pp. 191-200). Blackwell Publishing Ltd.

Van Ham, F., & Perer, A. (2009). “Search, Show Context, Expand on Demand”: Supporting Large Graph Exploration with Degree-of-Interest. Visualization and Computer Graphics, IEEE Transactions on, 15(6), 953-960.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. (1994). The think aloud method: A practical guide to modelling cognitive processes (Vol. 2). London: Academic Press.

Vassileva, J., & Sun, L. (2007). Using community visualization to stimulate participation in online communities. E-Service Journal, 6(1), 3-39.

Vassileva, J. (2008). Toward social learning environments. Learning Technologies, IEEE Transactions on, 1(4), 199-214.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. American Behavioral Scientist, 0002764213479363.

Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., & McKeon, M. (2007). Manyeyes: a site for visualization at internet scale. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1121-1128.

Wang, A. Y., Newlin, M. H., & Tucker, T. L. (2001). A discourse analysis of online classroom chats: Predictors of cyber-student performance. Teaching of Psychology, 28(3), 222-226.

Wattenberg, M., & Kriss, J. (2006). Designing for social data analysis. Visualization and Computer Graphics, IEEE Transactions on, 12(4), 549-557.

Weaver, C. (2004, October). Building highly-coordinated visualizations in improvise. In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on (pp. 159-166). IEEE.

Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking actual usage: the attention metadata approach. Journal of Educational Technology & Society, 10(3), 106-121.

Woodruff, A., Olston, C., Aiken, A., Chu, M., Ercegovac, V., Lin, M., ... & Stonebraker, M. (2001). DataSplash: A direct manipulation environment for programming semantic zoom visualizations of tabular data. Journal of Visual Languages & Computing, 12(5), 551-571.

Xiao, L., Gerth, J., & Hanrahan, P. (2006, October). Enhancing visual analysis of network traffic using a knowledge representation. In Visual Analytics Science And Technology, 2006 IEEE Symposium On (pp. 107-114). IEEE.

Yee, K. P., Swearingen, K., Li, K., & Hearst, M. (2003, April). Faceted metadata for image search and browsing. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 401-408). ACM.

Yi, J. S., Melton, R., Stasko, J., & Jacko, J. A. (2005). Dust & magnet: multivariate information visualization using a magnet metaphor. Information Visualization, 4(4), 239-256.

Yi, J. S., ah Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1224-1231.

Zafra, A., Romero, C., & Ventura, S. (2013). DRAL: a tool for discovering relevant e-activities for learners. Knowledge and information systems, 36(1), 211-250.

Zhang, H., & Almeroth, K. (2010). Moodog: Tracking student activity in online course management systems. Journal of Interactive Learning Research, 21(3), 407-429.

Zhang, H., Almeroth, K., Knight, A., Bulger, M., & Mayer, R. (2007, June). Moodog: Tracking students' online learning activities. In World Conference on Educational Multimedia, Hypermedia and Telecommunications (Vol. 2007, No. 1, pp. 4415-4422).

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into practice, 41(2), 64-70.

Appendices

Appendix A

This appendix discusses a number of systems that that aim to support educators in monitoring student learning patterns (two proposed systems and a system used in schools). The appendix also discusses storyline visualisations, which consist a narratives presenting entity interactions and lifespan.

A.1 Proposed tools to monitor student learning patterns

Two proposed tools are described below; the first proposed tool supports educators in monitoring student learning using faceted browsing and the second, Classroom View, uses Chernoff faces to enable educators to analyse how students are progressing with assigned tasks.

Monitoring based on faceted browsing

Garcia-solorzano et al. propose a monitoring tool that uses visualisations and faceted browsing to allow educators to query and view student performance and forum participation (García-Solórzano et al., 2012). Faceted browsing consists of three stages (Yee et. al., 2003): the opening stage, which shows the entire collection; the middle stage, which allows users to define constraints and narrow down the items; and the end game, where a user selects one item and ends the search. Such browsing can be useful in educational settings as educators can run searches iteratively and gain insights into student behaviour and monitor progress.

The paper proposes a browser with data collected from an LMS and a menu with facets related to gender, status (new student, repeat student), performance (studying pace, grade) and forum participation. The browser allows educators to narrow down searches and apply filters and thereby chose which students to include in the visualisation. The facets are shown on the left hand side of the browser and those that facilitate gradation are represented by sliders. The visualisation shows photos of students (if available), with a background colour to indicate the studying pace. The cards may have a red boarder to indicate a student who is performing poorly and the photos may have a dashed lines to indicate if the learner is repeating the course, as shown in Figure 1.

The proposal discusses a way the educator can view a more detailed visualisation, called a data portrait, of an individual student's forum participation. The educators can interact with the card visualisation to see a bar divided into five squares representing the number of initial posts, replies, read messages, highlighted messages and the average grade of the messages.

Figure 1 Monitoring based on faceted browsing

Classroom View

Classroom View (France et. al., 2006) proposes to support educators through visualisations to monitor the current activities students are working on, the time it takes to complete activities and how long students take to perform an activity in a virtual classroom. Chernoff faces are used to represent students where one Chernoff face is used to represent a student. The Chernoff faces evolve and change colour as the student performs activities, as shown in Figure 2. Each Chernoff face is initially represented with points for organs that grow as the system is used. The faces change colour, starting off as white and get darker representing the time the student has worked. Finally the eyes on the faces are animated and can be staring, closed, looking down, falling eyelids and looking to the sides to represent how a student is working on his/her current activity. Students are grouped by the activity they are currently working on using a bubble which also uses colour to communicate the relative delay compared to the average completion time for the activity. The bubble also shows the quality of the student depending on how much the Chernoff face in it has evolved.

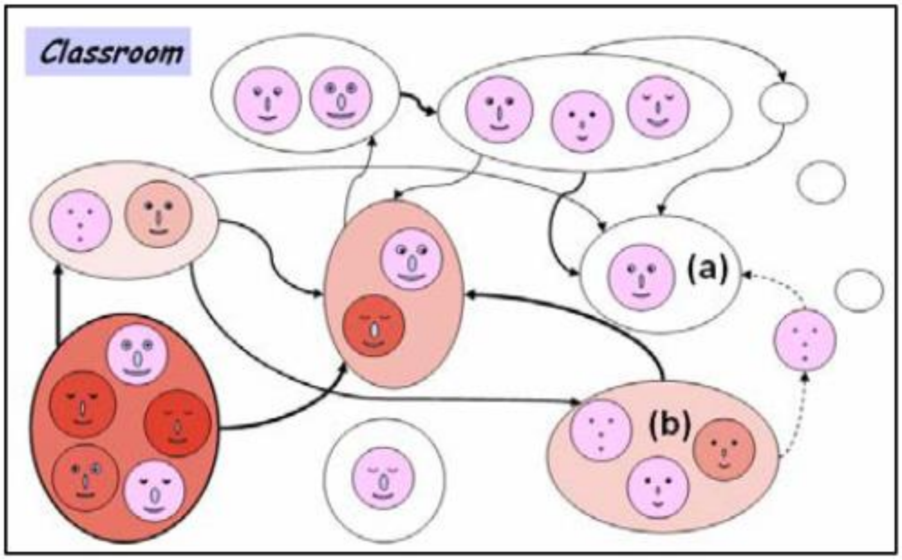

Figure 2 Classroom Views

Teachers can click on a face to analyse a visualisation showing the time the learner spent on previous and current tasks together with cumulated logs for all previous and current tasks. A bar chart is used to represent the time and the logged data. Educators can also interact with the bubbles to analyse an activity view through a heat map visualisation showing the times spent by group of students working on the activity compared to the average activity time.

Both proposals consist of interactive visualisations that support select, filter and drilldown interaction techniques. The visualisations enable educators to monitor student learning and progress and the visual interactions allow the view to be manipulated. The proposal discussing monitoring based on faceted browsing supports drill-down interactions using student cards to view forum interactions for an individual learner. The Classroom Views proposal supports drilldown to view student and group activity times and accumulated logs. These drilldown views enable educators to explore and understand individual student behaviour and support analysis of learner interactions over time. In addition, the faceted browsing based proposal supports filtering using checkboxes to narrow searchers for students. From the analysis of these two proposals, it can be seen that the visualisations present real-time student data and support visual interactions to enable educators to look back and make comparisons with other students. However, similar to the systems presenting visualisation to support educators in monitoring learners discussed in Chapter 3, both of these proposals do not support explorations of data related to that presented in the visualisations and do not support guidance through visual narratives.

A.2 Visually monitoring Student performance in schools

STEMscopes (Monroy et al., 2013) is an online science curriculum that offers a number of resources to teachers of K-12, including interventions and inquiry activities. STEMscopes offers teachers timeline and heat map visualisations to help them understand the part of the curriculum with the highest and lowest usage. The visualisations support teachers in analysing when course content is accessed, the access sequence and the time spent by learners on activities, which can allow the teachers to understand learning outcomes. The heat map shows the course content taught by schools per district. The colour intensity indicates high and low usage and allows teachers to view the levels of teaching time dedicated to the various course topics.

Figure 3 STEMscopes Dashboard

STEMscopes also visualises student performance through a dashboard, displayed in Figure 3 showing student names and green and red arrows under assessments. The green arrows indicate grade improvements since the last assessment and red arrows indicate a fall in grade since the last assessment. The dashboard groups students using coloured cells by the level of intervention (moderate and intense) and acceleration material required. Highlighting students' grade improvements and grade drops and grouping them by degree of intervention and acceleration helps teachers understand the level of help required by learners and which of these learners can be challenged with advanced material (acceleration group).

As part of the evaluation of the system, focus groups were conducted with fifteen participants in 2012 to try to understand STEMscopes teachers' usage of the dashboard. A number of suggestions were raised during this session, namely, mechanisms to share experiences with other teachers and an ability to view student progress metrics.

From the above description, it can be seen that STEMscopes provides teachers with a number of visualisations presenting student learning behaviour and progress; however, these visualisations do not allow teachers to visually explore data related to what is shown. The system gathers student-logged data which includes when course content is accessed, the access sequence and the time spent by learners on activities; however, it does not present transformations of this data including derived data transformations that could allow teachers to understand a drop or improvement in grade or support teachers in identifying patterns amongst learners.

A.3 Storyline Visualisation

Storyline visualisation systems produce temporal patterns of entity relationships in narratives. The visualisations consist of lines along a time axis, which represent entities, which are usually characters. The lines start from the left and go to the right, with the start and end points denoting an entity's lifespan and bundled lines indicate interactions between entities. In addition, filled contours are used to present location and hierarchy of entities in the story. Interactions between entities (for example, characters in movies) are shown as bundled lines. The data used for storyline visualisations consist of interactions between entities detailing the time, duration of interactions per entity. Although storyline visualisations present messages in a narrative format of interactions between and lifespan of entities, they are different from visual narratives described in this thesis which consist of an ordered sequence of linked visualisations with descriptions.

Ogawa and Ma (2010) present a prototype implementation of a storyline visualisation using a technique specifying the layout and visual encoding of entities. The layout of entities or lines follows four rules which specify how clusters need to be in adjacent lines with spacing between them. The lines should have minimal change from their original y-axis position and the crossing on lines should be minimised. The technique was evaluated using repository data from open source projects, exporting commit history including the files modified, time of commit and the developer performing the commit. A hash of the username of developer performing the commit was used to colour the lines, which were labelled using the username. The display consists of a stacked barchart below the storyline visualisation highlighting the number of files committed at each timestep per person. Each bar is a position on a timestep and the stacks are colour coded to represent the number of source code files, modules and document committed. The prototype supports the highlighting of a line and the corresponding developers commit activity in the barchart when the user hovers over a line to enable it to be easily traced and analysed.

Tanahashi & Ma (2012) present a technique for building storyline visualisations that uses genetic algorithm (Goldberg, 1989) to compute the layout of entities through three steps. The first step identifies interaction session and assigns slots and positions on the layout. The second step calculates positions and rearranges lines with respect to other entities, which minimises the number of wiggles and crossovers of lines. The final step removes white spaces, which can involve bringing lines closer to each other and at the same time maintaining the order. Once the steps of the algorithm are completed, the aesthetics and legibility of the visualisations are addressed. Aesthetics and legibility address issues of the flow of the storyline to ensure it is smooth and addresses the clarity of lines, to allow users trace a line of interest. Tanahashi & Ma discuss two techniques; one uses Line Relaxation, which smoothens jagged lines and the other de-emphasises lines by dynamically adjusting the width of lines to support users in following an entity.

StoryFlow (Lui et al., 2013) is a storyline visualisation system that produces temporal patterns of entity relationships in stories and supporting user interactions. Figure 4 shows a portion of a storyline visualisation generated by StoryFlow for the movie The Lord of the Rings. The StoryFlow pipeline requires XML file input consisting of entities and their relationships. The XML file also contains the location hierarchy, known as a location tree. The pipeline first constructs relationship trees, which presents the hierarchical relationship of entities as time frames. When presenting a hierarchy, StoryFlow tries to reduce line crossing by sorting location and ordering entity and session nodes. In order to increase straight lines, the lines between frames are aligned and optimisation is performed to minimise white space, line crossings and line wiggles.

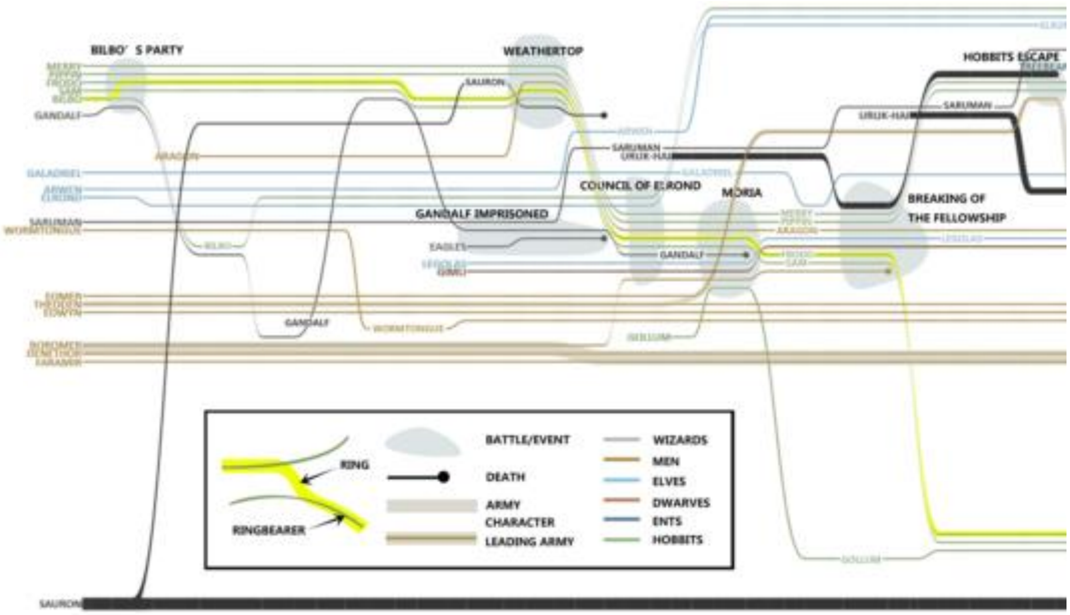

Figure 4 StroyFlow

StoryFlow provides users with interaction mechanisms for narrative exploration. The first type of interaction allows users to add and delete entities into and out of the story, which alters the relationship trees and the layout of the storyline. The second type of interaction is called ‘Line Straightening’ and allows users to analyse one entity without line wiggles, which can make it difficult to follow. Users can interact with a chosen line and straighten it. The third type is ‘interactive ordering’, which allows users to drag lines and reorder the storyline. The final interaction mechanism is called ‘bundling’. It is quite common due to large numbers of entities that not all of these can be shown, especially when entities are overlapping. StoryFlow bundles these entities into a single line and blends the line colour during the timeframe where they overlap. Users can select these blended lines to examine details to reveal the individual blended entities.

Appendix B - Experiments 1 – 4 Questionnaires and Interview Questions

Experiment 1 data analysis and narrative slice construction Questionnaire

1. I found it easy to query the source data through the Data Analysis Interface.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

2. Once it was demonstrated how to build a data table[61], I had little trouble in building the one required.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. The Data Analysis Interface used to build the data table was user friendly.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

[61] In the prototype and the first version of VisEN, narrative slices were referred to as data tables.

Experiment 1 Questionnaire 2

1. It was useful to analyse the progress of user activities through visualisations.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

2. It was useful to explore the data behind[62] a completed task bar.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. Loading the time-series chart to show the data behind a completed task bar was user friendly.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

4. The visualisations used to show the recommended titles are well connected and easy to explore and interact with

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

5. Loading the bar chart to show the average ratings for the recommended titles was user friendly.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

6. Viewing the breakdown of the title ratings was useful.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

[62] In the prototype, data behind a visualisation referred to derived data explorations

Experiment 2 Part One Questionnaire

Part One of Experiment 2 involved participants answering the following set of questions for each narrative slice (which all had a further comments section):

1. The data was comprehensible when viewing it through the selected visualisation.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

2. The data was accurately rendered through the selected visualisation.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. The visualisation was appropriate for rendering the data.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

4. The score given to the visualisation was a fair reflection of how well it rendered the data.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

Experiment 2 Part Two Questionnaire

Part Two of Experiment 2 involved the professor of the Information Management and Data Engineering module answering the following questions (each had a further comments section):

1. The visual narrative was easy to follow and it was useful to interact with the visualisation

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

2. When looking at the grade versus duration scatterplot, I found it useful to explore related data showing students with similar grades through a visualisation which showed their interactions.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. When looking at the grade versus interactions scatterplot, I found it useful to view related data showing the interaction per task for the selected student.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

4. The data slices represented my needs quite well.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

5. I was satisfied with the corresponding visualisations for these data slices.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

6. Navigating from one data slice to another was user friendly.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

Experiment 2 Part Two Interview

1. The visual narrative consisted of interactive visualisations. Did this work well? What interactions worked best?
2. The visual narrative consisted of related data explorations[63]. Did the concept of viewing related data work well? Did it work well in this instance and would it be suitable for students?
3. Would it have been useful to have more related data explorations? Which types of explorations would you like to have seen more of (to support your students)?
4. What would you see as a major challenge that your students would face when exploring a visual narrative with related data explorations?
5. Are there any additional visualisations you would have liked to have seen in the visual narrative (to support your students)?
6. If you were provided with an interface to construct narrative slices, with training, do you think you could create them?

[63] Derived data explorations were referred to as related data explorations in the first iteration of VisEN

Experiment 3 Part One Questionnaire

1. The engagement and task duration visualisations motivated me to engage with the course.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

2. It was useful to explore related data through other visualisations.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

3. I found it beneficial to be able to track my progress and compare it against other class students through personalised visualisations.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

4. I was able to follow the story provided by the visual narrative and I was able to obtain a good understanding about my own and fellow students' course related activities from these visualisations.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

5. I did not find it useful to be able to see how much time others spent on activities.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

6. Navigating from visualisation to visualisation was not user friendly.

   ☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

Comments ____

Experiment 3 Part Two Interview

1. Did you find the visual narrative useful?
2. Could you follow and understand the message?
3. Where you ever frustrated with the limitations of the user interface?
4. What did you like and dislike about the interface?
5. How appropriate were the visualisations for rendering the data?
6. Did you think that the related data explorations were useful and comprehensive?
7. In your opinion, what uses could consumers make of related data explorations?
8. Do you feel the related data explorations could uncover data not shown in the narrative?
9. Do you think related data explorations could uncover data hidden in visualisations?
10. Do you think the related data explorations showed visualisations that would be useful for consumers in order to gain insights into each narrative slice?

Experiment 4 Visual Narrative Construction Questionnaire

1. Constructing Narrative Slices was difficult.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

2. I was not able to construct the different narrative slices required for the narrative.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. The visualisations used in the narrative were appropriate for rendering the data.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

4. I was frustrated with the limitations of the user interface.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

5. The framework supported me in telling the story required.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

6. I do not think the explorations[64] would be useful for narrative authors.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

7. I did not find step 6 difficult even though it involved constructing narrative slices from a different data source, which was not introduced to me in the tutorial video
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

8. Although I was not familiar with the data source used in step 6, the interface was quite intuitive and my experiences executing the previous tasks enabled me to construct the narrative slices required.
   ☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

[64] Derived data transformations were referred to as explorations in the third iteration of VisEN

Experiment 4 Visual Narrative Consumption Questionnaire

1. How would you describe this narrative?
2. The message communicated by the visual narrative was difficult to understand.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

3. The data exploration check boxes and slider are useful in allowing one to select the data they would like to explore in the narrative.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

4. The explorations were not useful in enabling me to gain insight into the visual narrative.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

5. The re-ordering of explorations links based on my interactions with the explorations is useful in enabling me to quickly find related data that I am interested in.

☐ Strongly Agree ☐ Agree ☐ Disagree ☐ Strongly Disagree

Comments ____

Experiment 4 Participant Interview

1. What did you find easy/difficult about constructing the visual narrative?
2. How confident did you feel constructing the narrative slices?
3. How well did you feel the framework supported you in constructing the narrative slices?
4. If you had data to visualise and a story to tell, would you consider using VisEN and how confident would you be with it?
5. When constructing visual narratives with your own data, do you think explorations (derived data slices) would show you views of data that you not have considered and like to include?
6. What kind of insight do you think a consumer would get from viewing the explorations (derived data explorations)? Should the author have the ability to control which of these explorations are presented to the consumer?
7. How useful was the customisation of the explorations (derived data explorations) to narrative consumption?
8. What were your opinions on the tailoring of the explorations (derived data explorations)?

Appendix C

System Usability Scale

	Strongly Disagree				Strongly Agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5