# System Design Decision Support Tools

Deliverable D7.1

Version FINAL

**Authors:** W.R. Van Hage[1], T. Ploeger[1]
**Affiliation:** (1) SynerScope B.V.

Building structured event indexes of large volumes of financial and economic data for decision making

ICT 316404

| | |
|---|---|
| **Grant Agreement No.** | 316404 |
| **Project Acronym** | NEWSREADER |
| **Project Full Title** | Building structured event indexes of large volumes of financial and economic data for decision making. |
| **Funding Scheme** | FP7-ICT-2011-8 |
| **Project Website** | http://www.newsreader-project.eu/ |
| **Project Coordinator** | Prof. dr. Piek T.J.M. Vossen<br>VU University Amsterdam<br>Tel. + 31 (0) 20 5986466<br>Fax. + 31 (0) 20 5986500<br>Email: piek.vossen@vu.nl |
| **Document Number** | Deliverable D7.1 |
| **Status & Version** | FINAL |
| **Contractual Date of Delivery** | July 2013 |
| **Actual Date of Delivery** | November 4, 2013 |
| **Type** | Report |
| **Security (distribution level)** | Public |
| **Number of Pages** | 31 |
| **WP Contributing to the Deliverable** | WP7 |
| **WP Responsible** | SynerScope B.V. |
| **EC Project Officer** | Susan Fraser |

**Authors:** W.R. Van Hage[1], T. Ploeger[1]

**Affiliation:** (1) SynerScope B.V.

**Keywords:** decision support, system design

**Abstract:** In this deliverable we describe the design of the Decision Support Tool Suite (DSTS), the main point of end-user interaction with event indexes extracted from news in the NewsReader project. We describe SynerScope Marcato, a visual analytics application that delivers real time interaction with dynamic network-centric data and show how it can be adapted for use as the DSTS within the NewsReader-project.

# Table of Revisions

| Version | Date | Description and reason | By | Affected sections |
|---|---|---|---|---|
| 0.1 | 13 Sept 2013 | Deliverable skeleton | Thomas Ploeger | All |
| 0.2 | 16 Sept 2013 | Content outline | Willem R. Van Hage, Thomas Ploeger | All |
| 0.3 | 17 Sept 2013 | Revised content outline | Willem R. Van Hage, Thomas Ploeger | All |
| 0.4 | 18 Sept 2013 | Initial content for sections 1, 3, 4, and 5 | Willem R. Van Hage, Thomas Ploeger | 1, 3, 4, 5 |
| 0.5 | 19 Sept 2013 | Initial content for sections 2, 6, and 7 | Willem R. Van Hage, Thomas Ploeger | 2, 6, 7 |
| 0.6 | 8 Oct 2013 | Processed VUA feedback | Willem R. Van Hage, Thomas Ploeger | All |
| 0.7 | 1 Nov 2013 | Processed SCW feedback | Willem R. Van Hage, Thomas Ploeger | 1, 3, 4, 5, 6 |
| FINAL | 4 Nov 2013 | Finalized Document | Willem R. Van Hage, Thomas Ploeger | 2, 4, 6 |

# Executive Summary

The NewsReader project aims to support decision making by building structured event indexes of large volumes of news articles and financial data. The main point of interaction with the NewsReader event indexes will be the Decision Support Tool Suite (DSTS). The DSTS is a graphical user interface that is meant to support users when making strategic decisions. The DSTS does this by providing insight into the sequences of events that led up to a current situation so that a user can extrapolate to what might happen in the future.

The technical implemenation of the DSTS will largely consist of the SynerScope Marcato tool, a visual analytics application that delivers real time interaction with dynamic network-centric data. Within the NewsReader project, Marcato will interface with the KnowledgeStore and with external Web sources. Marcato allows the user to display several visualization methods ('views') simultaneously. Interactions with one view (e.g. highlighting, selecting) are replicated in all other open views.

The Hierarchical Edge Bundling View (HEB) is the primary network visualization in Marcato. Each Node is visualized as a point on a circle, and each Link is visualized as a curved line between its source and target Node. The Nodes are grouped hierarchically, based on an ordering defined by the user. In addition to the existing views, Marcato can be customized by developing plugin views using HTML5 and JavaScript.

Within the NewsReader project, one of the additions to Marcato will be an implementation of interactive Narrative Charts as a Marcato plugin. Another addition that will be made to Marcato is a Web Lookup plugin that performs automatic lookup of contextual information over a HTTP connection.

Marcato can be run locally on the end users own machine, if said machine meets the minimum hardware and software requirements. Alternatively, Marcato can be run remotely on a virtual machine in the cloud and streamed to the end users machine.

# Contents

# List of Figures

# 1 Introduction

The NewsReader project aims to support decision making by building structured event indexes of large volumes of news articles and financial data. In this deliverable we describe the design of the Decision Support Tool Suite (DSTS), the main point of interaction with the NewsReader event indexes for the end user.

In Section 1, we describe the purpose of the DSTS within NewsReader, the design principles that serve as its foundation, and its technical embedding within the rest of the project. In Section 2 we introduce SynerScope Marcato. An adapted version of Marcato will form the basis of the DSTS. In the next sections, we describe Marcato in more detail.

We show how data is currently imported into Marcato and how this process will be adapted for NewsReader in Section 3. Marcato offers several visualization techniques or views, each of which are explained in detail in Section 4. Section 5 describes the interaction methods Marcato makes available to the end user. In Section 6 we show how we will take advantage of Marcato's plugin system to develop new views specifically for NewsReader. Section 7 describes the options and requirements for deploying Marcato.

The overall DSTS system design as outlined in in this deliverable is subject to change during over the course of the project

## 1.1 What is the Decision Support Tool Suite?

The DSTS, is a graphical user interface that is meant to support users when making strategic decisions. In essence, we assist the decision making task by providing insight into the sequences of events that led up to a current situation so that a user can extrapolate to what might happen in the future. The DSTS will not aim to predict future events, but aid users in drawing their own conclusions. An important reason for this decision is that a predictive system can only work well under the assumptions that all relevant data is available in the system and that the mechanics of consequence are modeled in a sufficiently correct and complete manner. Both of these assumptions are likely to fail in an open domain such as the news.

The DSTS focuses on providing maximal insight into the complex changing network of events, actors, and places. The main problem with investigating such complex information structures is that there is a huge space of possible interactions and correlations that might prove to be relevant. A common way to search for relevant parts of such large feature spaces is to use feature selection, regression, and other statistical techniques.

In Work Package 7, we attempt to solve this problem in a new way, which has only recently become feasible due to the advances made in parallel and distributed systems and hardware graphics accelleration. As opposed to preselecting features we aim to allow the end user to interactively navigate through the large space of correlated features. The main motivations for this design decision are that many facts relevant to the decision making process are rare events, that by themselves will never stand out in statistics, and that the relevance of facts can only really be judged by the end users themselves.

A DSTS that aids users with interactive exploration of large feature spaces needs to perform consistently well in both speed and space scalability to prevent undesirable interruptions in the users interaction flow.

Another requirement for the DSTS to be usable by end users is that the information is presented in such a way that it is intuitive to users with knowledge about the information domain. This leads us to two decisions. One being that the DSTS needs to be a visual environment. The other, that the DSTS needs to show the actual data (in the case of NewsReader, the underlying news articles from which events were extracted), with which the user is acquainted, and not derived information that can appear unfamiliar to the user.

Not only should the DSTS be a visual environment, the interaction possibilities and visual representation of the patterns in the information structures, i.e. the storylines and involved entities, should be determined by the data, and not some predefined view. If this were the case, then new insights about future events would be limited by models made of past events, and hence inappropriate to new situations.

## 1.2 Design Principles

As discussed above, the DSTS needs to have a number of characteristics in order to be a useful tool for decision making by end users.

**No Divination** The main goal of the DSTS should be to provide insight in past events. The DSTS does not provide extrapolations based on confidence values or probabilities. However, if the events happen to be about extrapolations (such as speculations, plans, or strategies) these are the facts and they can be shown.

**No Tunnel Vision** The DSTS must be able to rapidly change the perspective of the investigation, depending on the user's changing understanding of and thus changing interest in the data.

**No Reports** The DSTS must be a graphical tool.

**No Magic** The DSTS must show actual data, which is recognizable.

**No Hiding** The DSTS must show all the data, not just aggregate statistics that hide important details.

**No Waiting** The DSTS must be fast and scalable.

**No Bias** The DSTS should not use a predefined vocabulary of concepts, but show whatever occurs in the data.

## 1.3 Technical Embedding

Within the NewsReader project, the DSTS will interface with the KnowledgeStore and with external Web sources. All the primary decision making information produced in other Work Packages from raw data will be collected in the KnowledgeStore. The DSTS will import selections of the KnowledgeStore for investigation. External knowledge sources that are not processed in other Work Packages should be accessed directly from the DSTS. The DSTS will be limited to accessing external sources over HTTP, i.e. Web pages.

The place of the Decision Support Tool Suite in the rest of the NewsReader system is illustrated in Figure 1.

# 2 SynerScope Marcato and the Decision Support Tool Suite

The technical implemenation of the DSTS will largely consist of the SynerScope Marcato tool, a visual analytics application that delivers real time interaction with dynamic network-centric data. Marcato supports simultaneous views and coordinates user interaction, enabling the user to identify causal relationships and to uncover unforeseen connections. These features are explained in detail in Sections 4 and 5. In this section we discuss the technical specifications of the existing SynerScope Marcato tool and of the to be developed extensions that allow integration with the rest of the NewsReader system and NewsReader-specific visualizations.

Marcato has all of the properties that are required for the interactive manipulation of events from the news, but has a few technical shortcomings. Most notably, it requires interfacing with the NewsReader KnowledgeStore, and a number of NewsReader specific visualization techniques, such as narrative charts (i.e. the metro line pictures of the life lines of actors) and context visualization (e.g. hypertext view of the original news articles that describe parts of the investigated story).

**SIS** Marcato is designed to work with a very basic information schema, called the SynerScope Interface Schema (SIS). SIS consists of two object types: Nodes and Links. Links connect two Nodes. Both Nodes and Links can have additional attributes of a number of data types, including integers, floating point numbers, free text, date and time, latitude and longitude. Nodes and Links need a key attribute. This attribute is used to connect Nodes and Links. It is not predefined which kinds of data objects can be Nodes, Links, or attributes of Nodes or Links. This can be decided at the time of import.

A common decision, in the case of simple events such as transactions, communication, and interpersonal relations, is that events are modeled as Links between entity Nodes. Alternatively, events and entities can both Nodes while Links are simple associations between them.

**Multiple and Coordinated Views** The central interaction paradigm of Marcato is Multiple and Coordinated Views. Marcato shows a number of different perspectives on the data, for example, relations and time, and each selection made in either of these views causes an equivalent selection to be made in all other views. This enables the user to explore correlations between different facets of stories. For instance, it can be used to

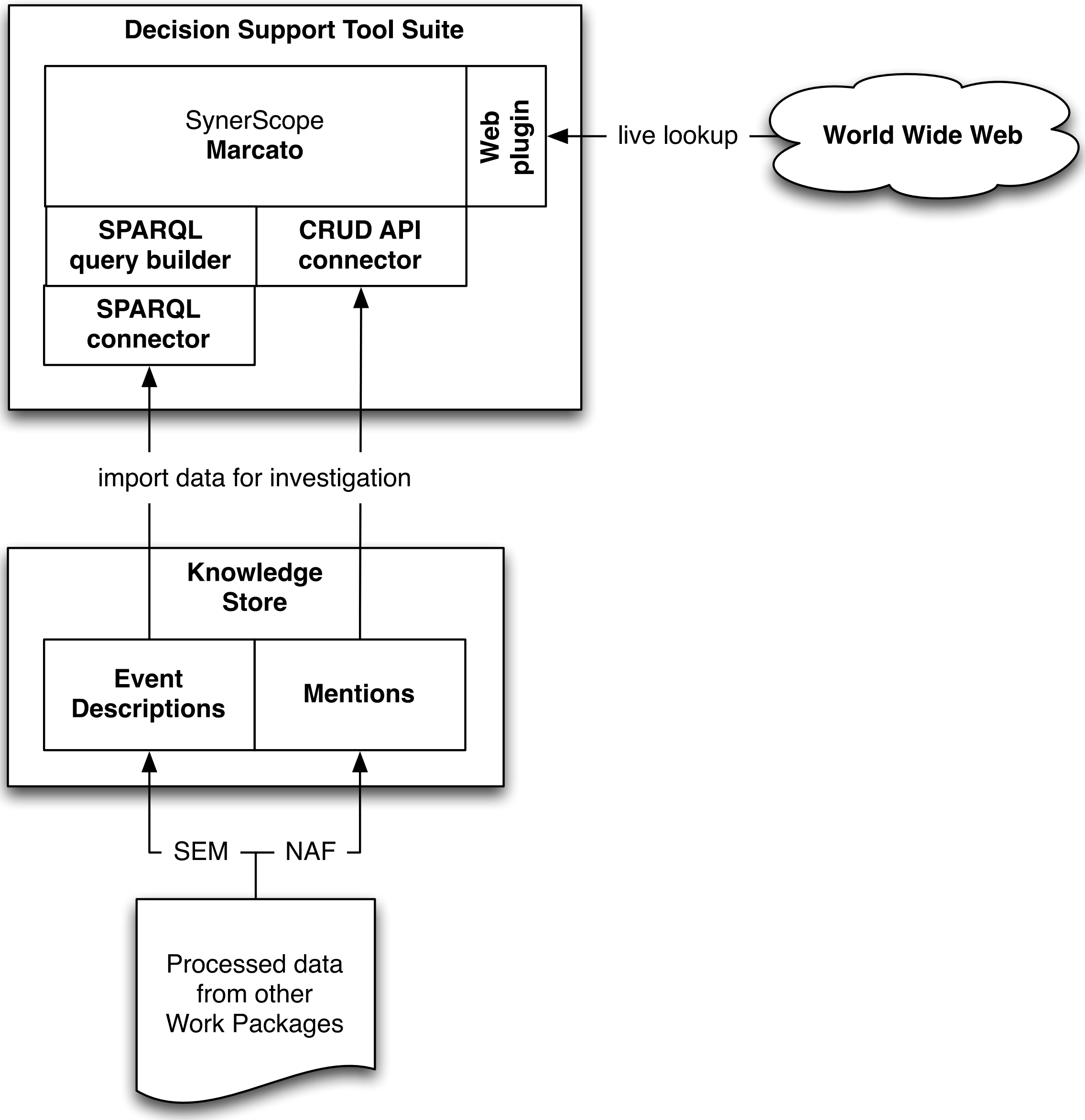



Figure 1: Technical embedding of the Decision Support Tool Suite in the NewsReader system. SynerScope Marcato is existing software. The adjacent components SPARQL query builder, SPARQL connector, CRUD API connector, and Web plugin, are software deliverables within the NewsReader project.

explore whether actors that interact are geographically collocated by selecting part of a social network in a network view, then checking which corresponding locations are selected in a geographical view and whether these are near to each other.

**Technical Specifications** Marcato can work with (sparse) tabular data from Relational Database Systems, like PostgreSQL[1], ParAccel[2], or MonetDB[3], through an ODBC connection or from Character[4] Separated Value (CSV) files stored on disk or a FUSE Mountable Hadoop Filesystem. To run, Marcato needs a recent Intel CPU and nVidia graphics card.

**Necessary Additions** In addition to SynerScope Marcato, the DSTS will consist of a number of Open Source plugin extensions for Marcato, and a specific SPARQL and CRUD API connector to import data from the KnowledgeStore.

Plugin extensions will be used to build new NewsReader-specific visualizations and interaction mechanisms. These additions are illustrated in Figure 1.

In the following sections we will discuss the various visualization techniques that exist in Marcato.

# 3 Data Importing / Exporting

Tabular data can be imported into SynerScope Marcato from databases or from files. To import graph data from the NewsReader KnowledgeStore, two new custom connectors will be developed, one to interact with SPARQL endpoints to retrieve event descriptions, and another to interact with the CRUD API to fetch the original documents from which the event descriptions were extracted.

Figure 2 shows the two-stage process that imports and transforms data for visualization.

**Import workflow** Currently, data can be imported into SynerScope Marcato directly from a database using ODBC middleware or from Character Separated (CSV) files. Any imported data is stored as a table in Marcato's internal database.

The import workflow starts with creating a new Marcato Project. In a project, the user specifies from which source data is to be imported (Database or CSV-files) and where on the file system Marcato's internal database should be stored.

Within the newly created project, a user can create Investigations. Creating an Investigation essentially means creating a data mapping from the imported data to the SIS, essentially identifying the Nodes and Links in the imported data.

[1]http://www.postgresql.org/
[2]http://www.paraccel.com/
[3]http://www.monetdb.org/
[4]Traditionally a comma, but can be an arbitrary character.

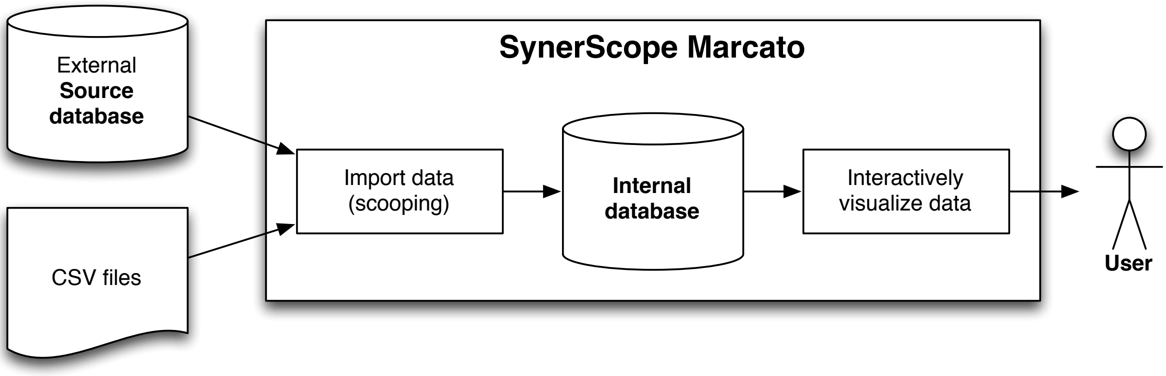



Figure 2: The two-stage process importing data from external sources into a fast internal database that is used for interactive visualization.

Suppose a user has imported two CSV-files into a project: One with Node data and one with Link data. In that case, creating an Investigation consists of the following sequential steps:

1. **Specification of CSV properties**: For both imported CSV files, the user must specify if that file starts with a header row and what the seperator and quotation characters are.

2. **Column definition**: Marcato autodetects the data type (e.g. integer, text, date, coordinate) and name of each column per imported CSV-file. If necessary, the user can specify manual overrides during this step. Both CSV files are now storable as individual tables in Marcato's internal database, so Marcato now creates a nodes and a links table.

3. **Node definition**: The user can now create a new Node type and specify that the data is contained in the nodes table that was created in the previous step. The user is able to choose which column from the table will be the unique identifier for the node. This identifier will be used to link nodes to other nodes via links.

4. **Link definition**: The user can now create a new Link type and specify that the data is contained in the links table. The user specifies which columns of the link table contain the incoming and outgoing unique node identifiers.

5. **Filtering**: Finally, the end user can set filters based on node or link attributes to reduce the amount of data visibile in Marcato.

After completing these steps, the resulting Investigation is opened and visualized. An illustration of the current SynerScope Marcato importer that allows users to graphically build Node and Link definitions is shown in Figure 3.

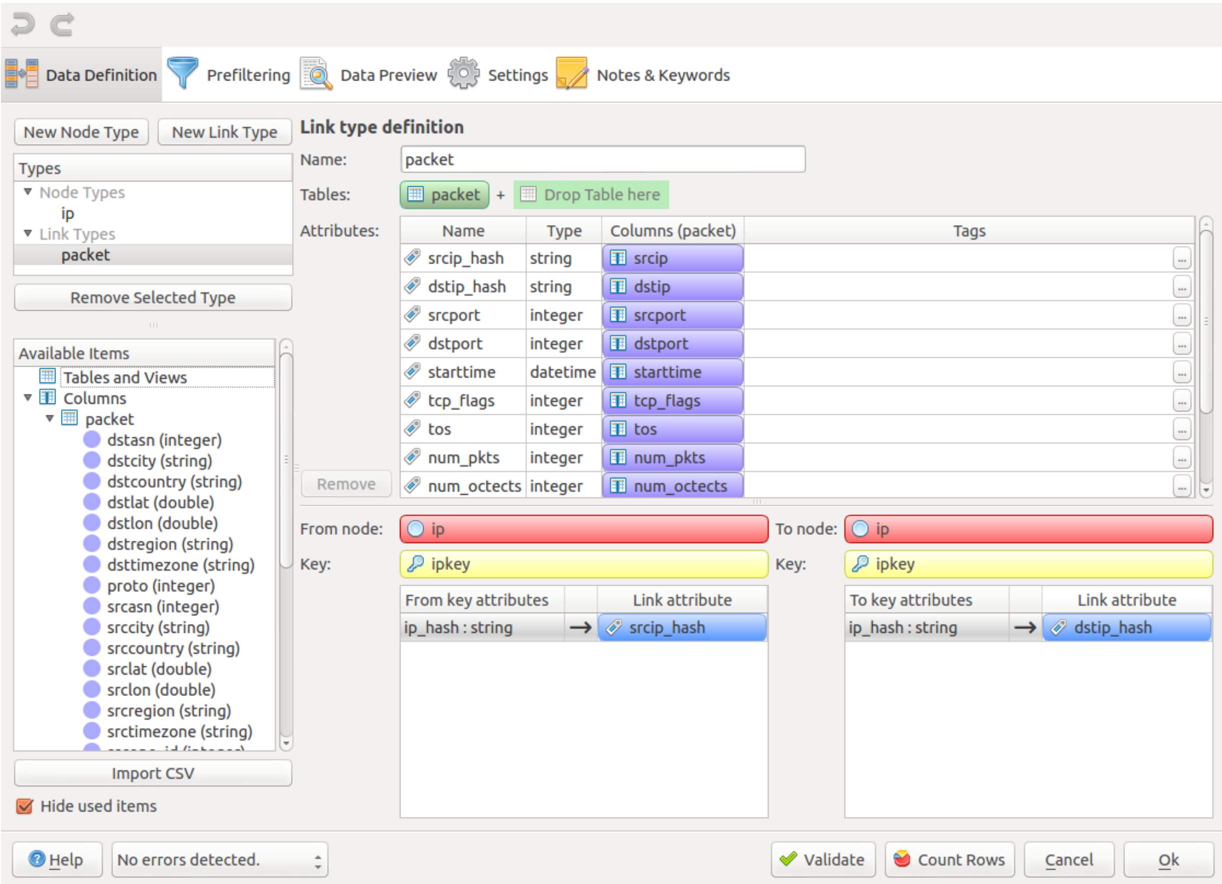



Figure 3: The Marcato importer, an interactive schema mapper that is used to define how tables and columns in the source database are mapped onto the Nodes and Links of the SIS.

**KnowledgeStore Connectors** The current Marcato importer provides interactive mapping functionality for the projection of tabular data onto the SIS. Projecting RDF data onto the SIS is a slightly more complex and currently unsupported process, that requires a new specific importer to be made.

Within the NewsReader project, SynerScope will develop an interactive, graphical SPARQL query builder that allows the user to define a mapping from event data in RDF to SIS without writing program code. This query builder will be guided by the available data and help the user with building queries. The resulting query will be used to extract tabular results from a SPARQL endpoint that will then automatically be imported into Marcato without further mapping.

Within Marcato, the user should have access to the original documents from which a certain event was derived. The KnowledgeStore's CRUD API will allow for the retrieval of the original document given a certain event retrieved through its SPARQL endpoint, so SynerScope will develop a connector for the CRUD API as well.

**Exporting Data** Some decisions will require additional processing of discovered patterns in the data. For such tasks it is necessary to be able to export (parts of) the data from Marcato to other programs, such as statistical tools, spreadsheets, and machine learning tools. It is possible to make a dump of all the data in the current Selection in Marcato. The data can be stored in CSV format on a disk accessible from the computer running Marcato. The file can then be imported in other tools to continue the analysis process.

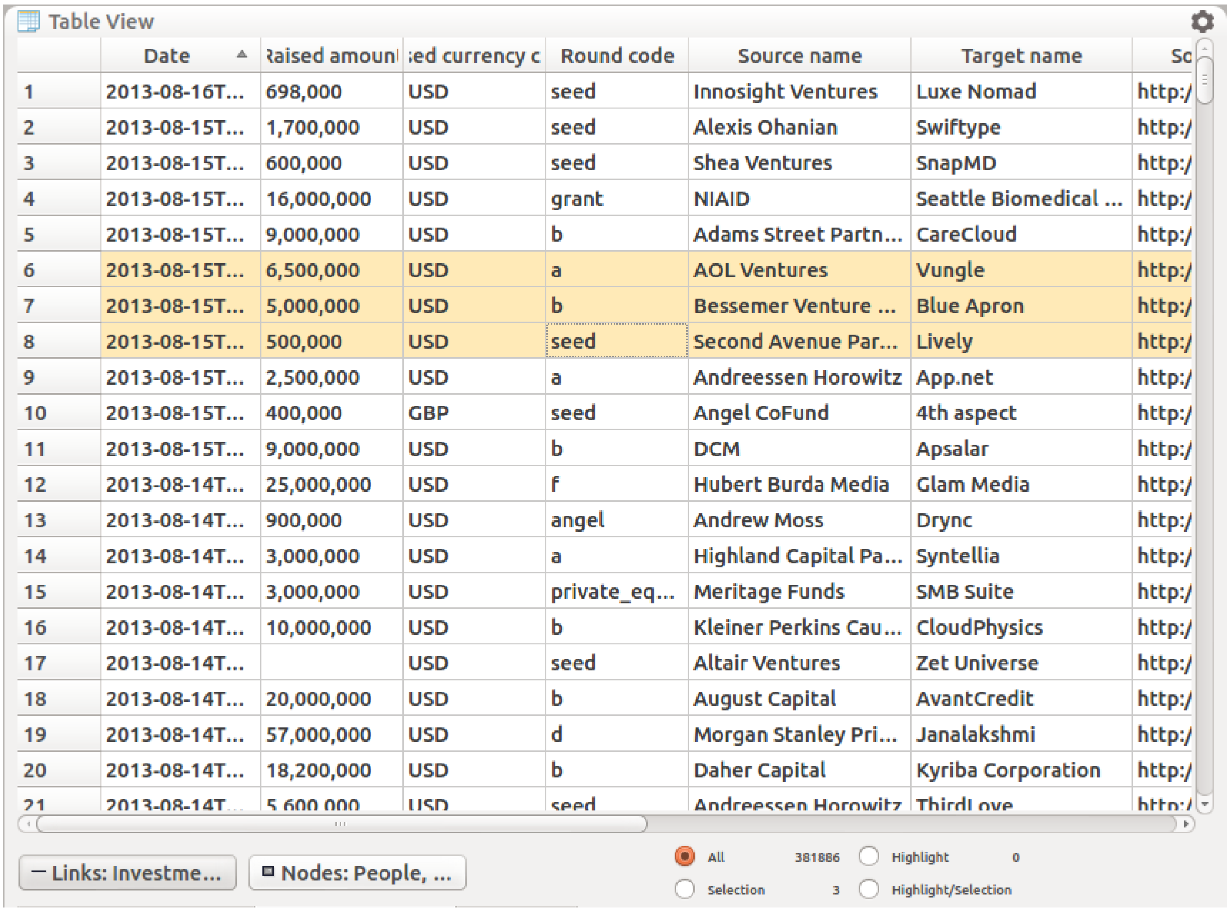
# 4 Visualization Techniques

After constructing a new or opening an existing Investigation, the user has access to all visualizations of Marcato. The visualizations are rendered using OpenGL to take advantage of hardware-accelerated rendering. Each visualization is described in detail in its own section below.

## 4.1 Table View

The Table View provides a traditional spreadsheet view on the data. For each of the SIS data source types, i.e. each type of Nodes and each type of Links, there is a separate sheet. The Table View shows all the data as a table of values. Rows in the table correspond to either Links or Nodes, depending on which sheet is selected. The columns represent the attributes of the Nodes or Links. The user can select rows, which are then also highlighted in other Views. The Table View can also be used to sort Nodes or Links by the values of a certain field by clicking the header cells of the columns in the table. Furthermore, from the Table View, the user can export the current Selection to a spreadsheet file for later processing with other analysis software such as Microsoft Excel, SAS, or R.

Figure 4 shows a typical Table View.




| | Date | Raised amoun | ed currency c | Round code | Source name | Target name | So |
|---|---|---|---|---|---|---|---|
| 1 | 2013-08-16T... | 698,000 | USD | seed | Innosight Ventures | Luxe Nomad | http:/ |
| 2 | 2013-08-15T... | 1,700,000 | USD | seed | Alexis Ohanian | Swiftype | http:/ |
| 3 | 2013-08-15T... | 600,000 | USD | seed | Shea Ventures | SnapMD | http:/ |
| 4 | 2013-08-15T... | 16,000,000 | USD | grant | NIAID | Seattle Biomedical ... | http:/ |
| 5 | 2013-08-15T... | 9,000,000 | USD | b | Adams Street Partn... | CareCloud | http:/ |
| 6 | 2013-08-15T... | 6,500,000 | USD | a | AOL Ventures | Vungle | http:/ |
| 7 | 2013-08-15T... | 5,000,000 | USD | b | Bessemer Venture ... | Blue Apron | http:/ |
| 8 | 2013-08-15T... | 500,000 | USD | seed | Second Avenue Par... | Lively | http:/ |
| 9 | 2013-08-15T... | 2,500,000 | USD | a | Andreessen Horowitz | App.net | http:/ |
| 10 | 2013-08-15T... | 400,000 | GBP | seed | Angel CoFund | 4th aspect | http:/ |
| 11 | 2013-08-15T... | 9,000,000 | USD | b | DCM | Apsalar | http:/ |
| 12 | 2013-08-14T... | 25,000,000 | USD | f | Hubert Burda Media | Glam Media | http:/ |
| 13 | 2013-08-14T... | 900,000 | USD | angel | Andrew Moss | Drync | http:/ |
| 14 | 2013-08-14T... | 3,000,000 | USD | a | Highland Capital Pa... | Syntellia | http:/ |
| 15 | 2013-08-14T... | 3,000,000 | USD | private_eq... | Meritage Funds | SMB Suite | http:/ |
| 16 | 2013-08-14T... | 10,000,000 | USD | b | Kleiner Perkins Cau... | CloudPhysics | http:/ |
| 17 | 2013-08-14T... | | USD | seed | Altair Ventures | Zet Universe | http:/ |
| 18 | 2013-08-14T... | 20,000,000 | USD | b | August Capital | AvantCredit | http:/ |
| 19 | 2013-08-14T... | 57,000,000 | USD | d | Morgan Stanley Pri... | Janalakshmi | http:/ |
| 20 | 2013-08-14T... | 18,200,000 | USD | b | Daher Capital | Kyriba Corporation | http:/ |
| 21 | 2013-08-14T... | 5,600,000 | USD | seed | Andreessen Horowitz | ThirdLove | http:/ |

Figure 4: The Table View with a Selection in orange. In this figure, each line is one investment round with properties such as a date, currency, and amount of money.

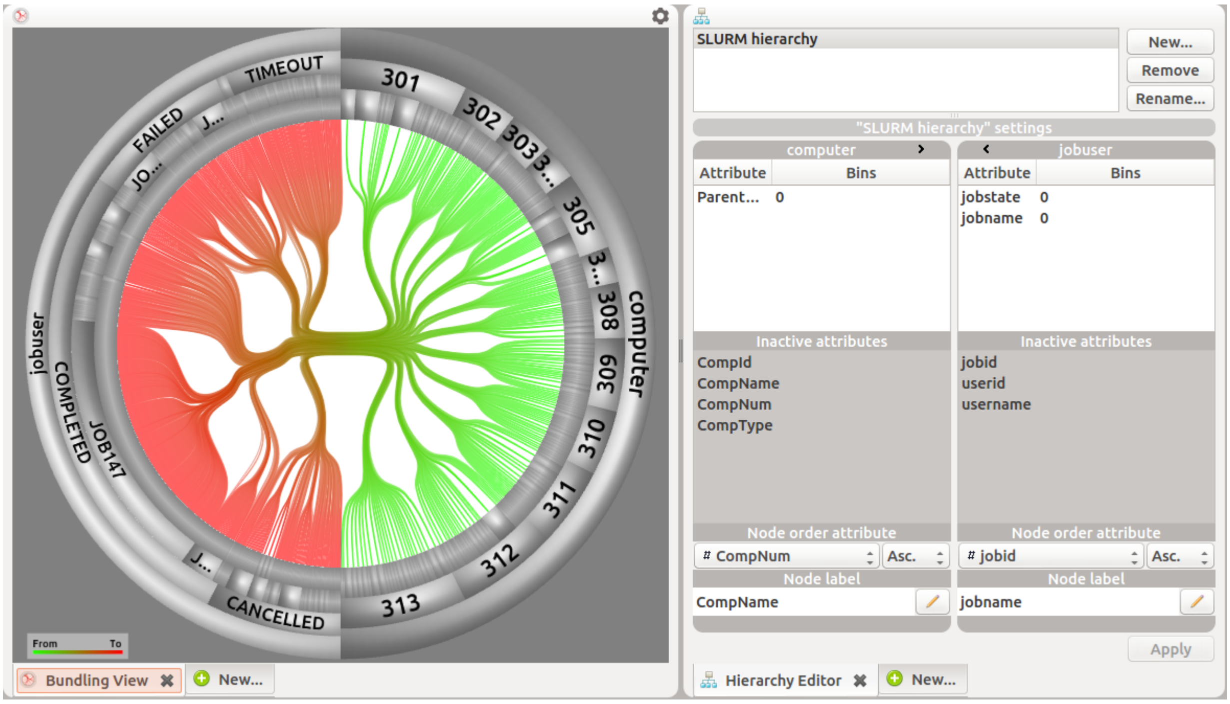



Figure 5: The Hierarchy Editor View on the right side defining a hierarchy for the two Node types in this data set. The resulting respectively one and two-level hierarchy is applied to the HEB on the left.

## 4.2 Hierarchy Editor View

The SIS allows for Nodes and Links. There can be multiple types of Nodes or Links. For example, if we consider Nodes to be people or books and links to be book reviews linking a person to a book they wrote a review about, then we have two Node types, people, and books, and one Link type, reviews.

In the Hierarchy Editor View it is possible to define a hierarchical ordering (of arbitrary depth) on each Node type. Each layer of the hierarchical ordering groups nodes together based on some common attribute value. For example, if the people Nodes have two attributes, "city of residence" and "age", we can define two hierarchical levels, one based on city, the other on age. Categorical attributes are grouped by distinct values, numerical attributes are distributed over a certain number of bins. The Node hierarchies can be changed in real time. All Nodes in a hierarchy categories can be selected together, which allows for quick exploration of common properties of the Nodes in that category in other Views.

Figure 5 shows the hierarchy editor on the right and the resulting hierarchy applied in the Hierarchical Edge Bundling (HEB) view (see Section 4.3) on the left.

## 4.3 Hierarchical Edge Bundling View

The Hierarchical Edge Bundling View (HEB) is the primary network view in Marcato. Each Node is visualized as a point on a circle, and each Link is visualized as a curved line between its source and target Node. The Nodes are sorted so that changes in the network over time have a minimal impact on the location of the Nodes on the circle. This makes for a stable view of the network.

The Nodes are grouped hierarchically, based on the ordering defined in the Hierarchy Editor View. The Links between Nodes of the same hierarchical category are bundled together (as if they were tied together with a cable tie). The thickness of the bundle is caused by overlapping lines (each line is still shown individually) and shows the amount of interaction between categories. Bundles can be selected together by making a slicing gesture by clicking and dragging the mouse over a bundle. The HEB is illustrated in Figure 6.

## 4.4 Massive Sequence View

The Massive Sequence View (MSV) is the primary temporal view in Marcato. Each Node gets a fixed position on the horizontal axis. Nodes are grouped hierarchically in the same fashion as in the HEB. Links between Nodes are represented by a horizontal line between the respective positions of the Nodes. On the vertical axis the user can select a scalar attribute, typically a time or date. This orders the Links temporally.

The MSV is typically used to inspect patterns in time. When certain categories of Nodes interact with each other in some temporal pattern, this becomes instantly recognizable in the MSV. An example of a time line, showing the same Nodes and Links as Figure 6 is shown in Figure 7.

## 4.5 Map View

The Map View is the primary spatial view in Marcato. The user can select two attributes from any Node or Link data source to interpret as WGS84 latitude and longitude coordinates. These attributes are used to plot the Nodes (not the Links) on a map as points. By default, the Open Street Map[5] tile server is used over an internet connection to show a map background behind the points. The points are plotted in such a way that the user can see the point density. An additional scalar attribute can be selected to indicate point size on the map. Figure 8 shows a typical Map View presentation.

## 4.6 Scatter Plot View

The Scatter Plot View uses Cartesian coordinates to relate the values of two attributes of either Nodes or Links. Dots are drawn on a two-dimensional chart, the positioning relative

[5] http://www.openstreetmap.org

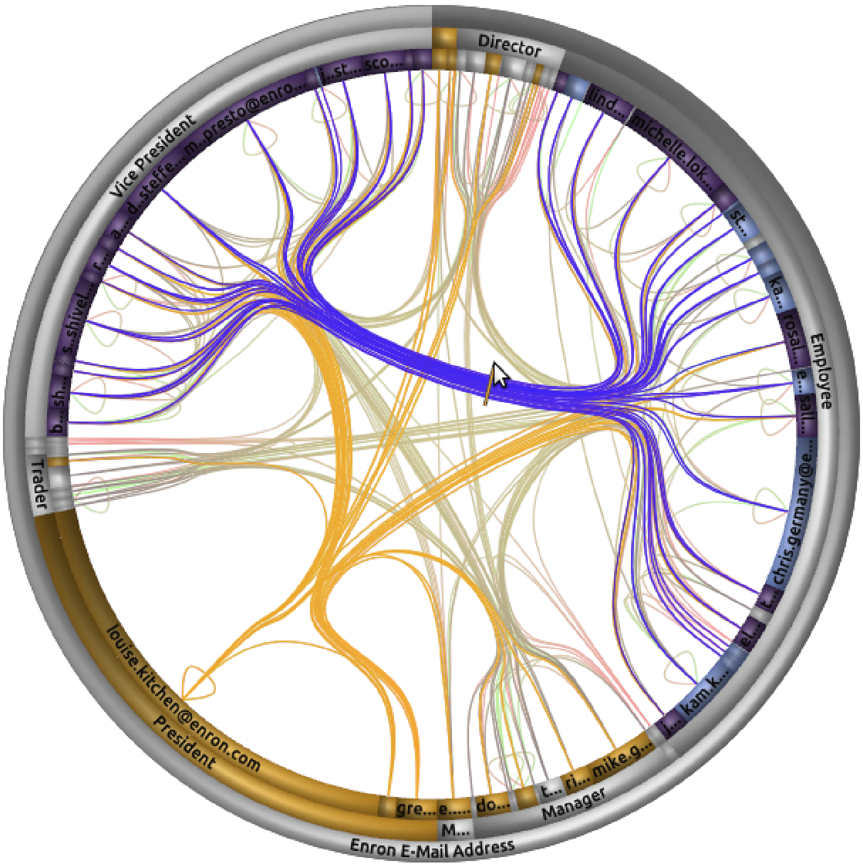



Figure 6: The Hierarchical Edge Bundling View, showing a Selection in orange, a Highlight in blue, and the overlap between the two as dark blue. The nodes on the circle represent employees (grouped categorically) while the edges between them represent e-mails. The bundle between the Vice President category and Employee category is being Highlighted with a slicing mous gesture.

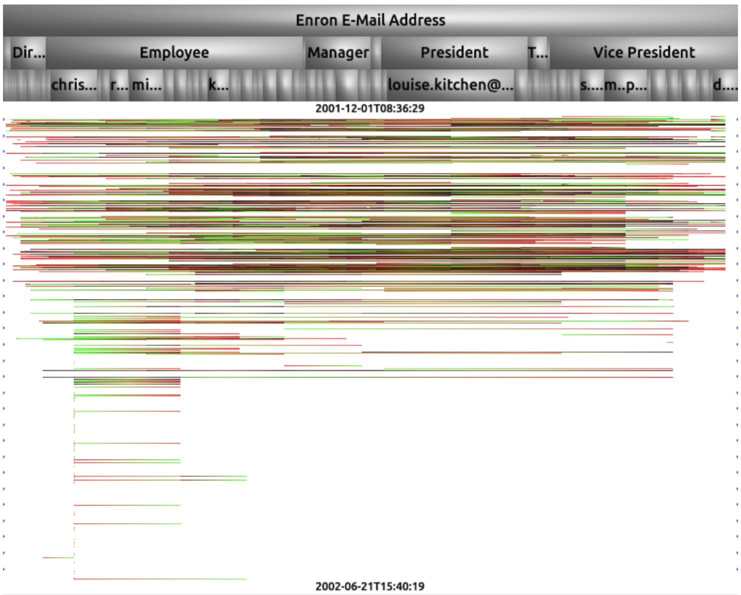



Figure 7: The Massive Sequence View, showing the same interactive hierarchy as Figure 6. Again, the nodes (now on top) represent employees (grouped categorically) while the edges between (now ordered by time from top to bottom) them represent e-mails.

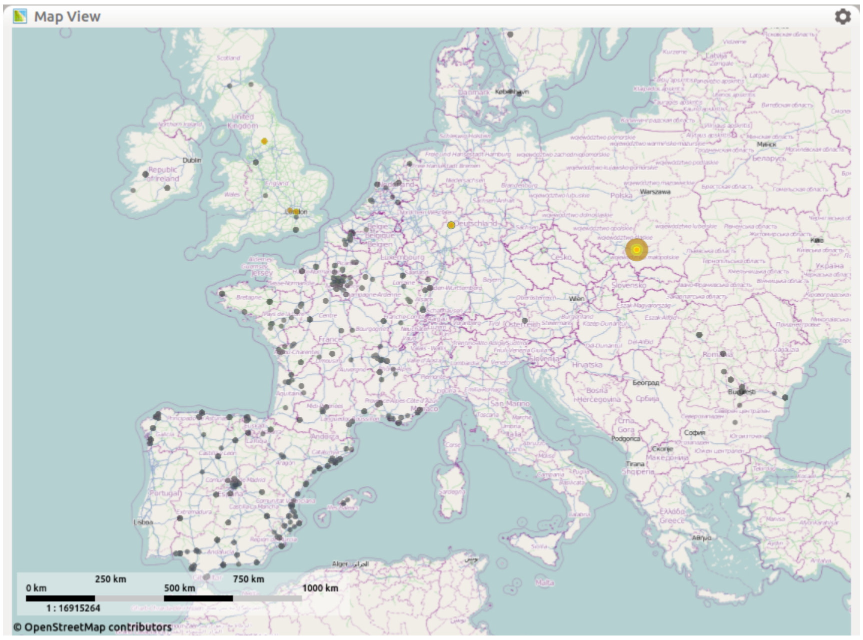



Figure 8: The Map View, showing geographic point density and size.

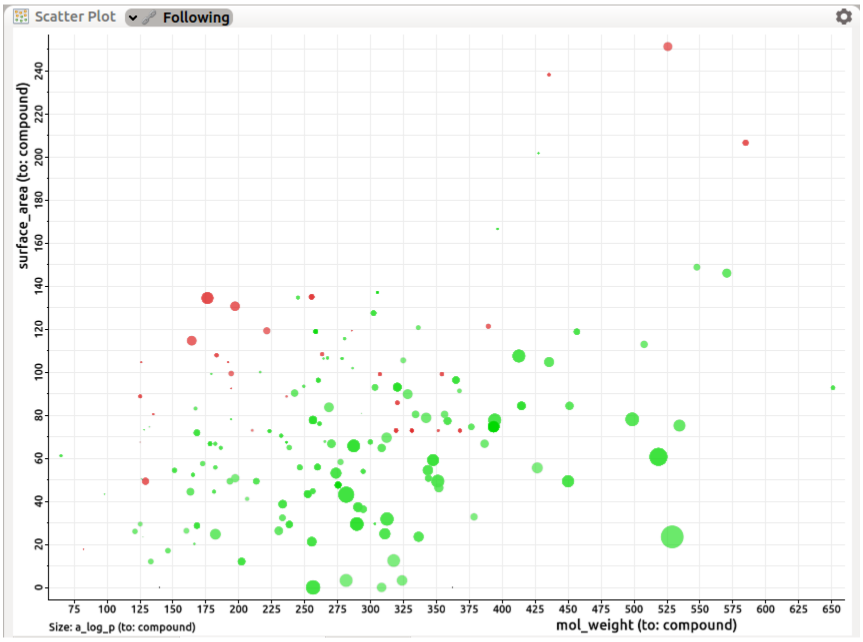



Figure 9: The Scatter Plot View, showing three variables: $x$, $y$, and magnitude, with the sign of the magnitude depicted as red (negative) or green (positive). Shown is a plot of biochemical data, specifically molecule weight ($x$) versus surface area ($y$).

to the horizontal and vertical axis being determined by the attribute's values. A third attribute can used to set the size of the dots.

### 4.7 Search and Filter View

The Search and Filter View is an interactive view that allows the user to select Nodes or Links by searching by value. The user can type in a search query that is matched against the value of a certain attribute for a given Node or Link type. The current version of the Search and Filter View allows for case sensitive or insensitive exact string matching.

## 5 Interaction Methods

The user can interact with SynerScope Marcato's views in several ways: By selecting and highlighting, drilling down or up, and expanding selections. Every interaction method is coordinated across multiple views and over the node hierarchy. This principle, together with the details of each interaction method, is described below.

### 5.1 Coordinated Views

SynerScope Marcato uses multiple coordinated or linked views to provide different but simultaneous visualizations of the same data. Selections and Highlights are applied across

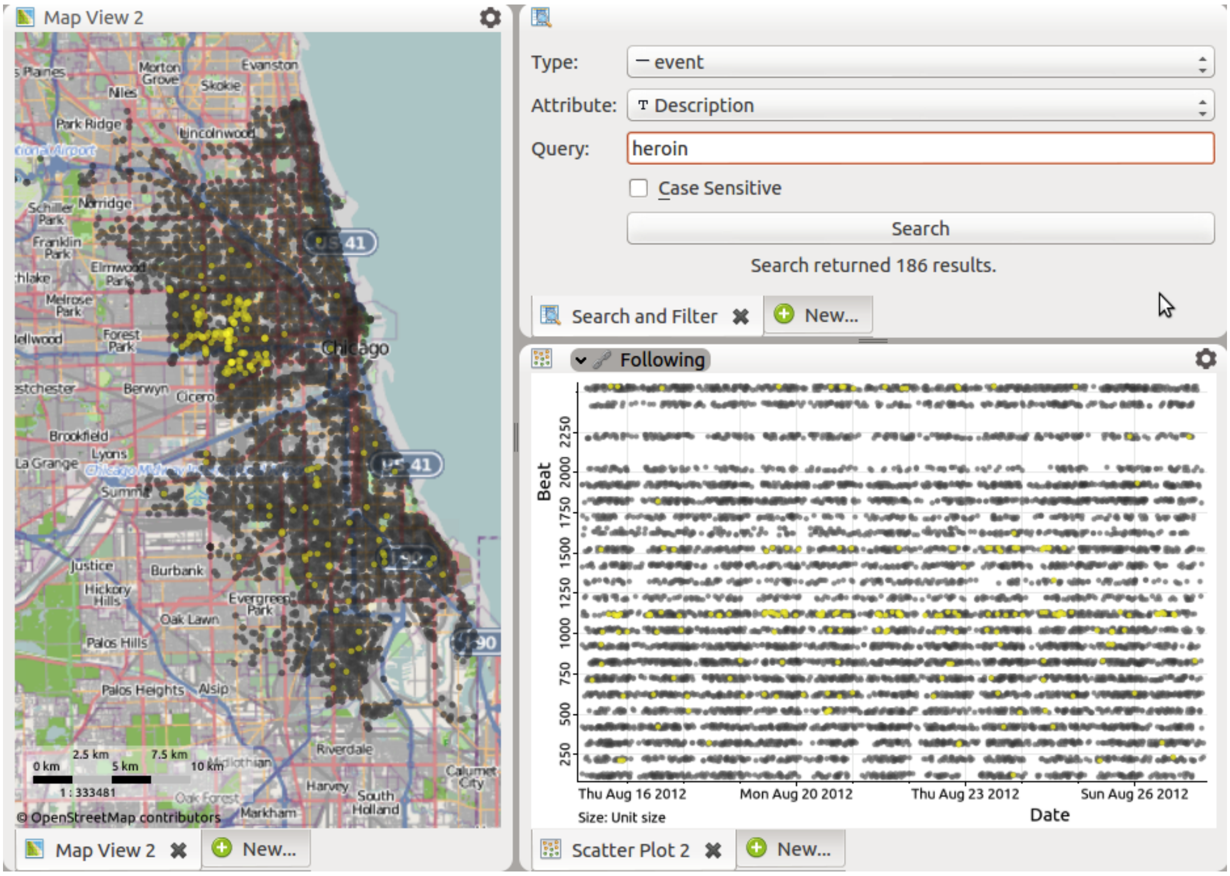



Figure 10: The Search and Filter View, showing that the search results are Selected in all other Views, in this case, the Map View and Scatter Plot View. This Figure also shows that all color coding can be customized to provide accessibility for the color blind. In this case the Selection is shown in yellow instead of orange. Shown are criminal events in the city of Chicago.

views. This means that if a node or link is highlighted in one particular view, it will also be highlighted in all other views. Similarly, if a node or link is selected in one view, it is selected in all other views.

## 5.2 Selection and Highlight

One of the main interaction paradigms of Marcato is the combined use of Selections and Highlights. When a user expresses that his or her attention is focused on a certain Node, Link, or category/bundle of Nodes or Links, Marcato Highlights these objects with a color that makes them stand out from the other objects. This Highlight is simultaneously performed in all visible Views. The Highlight is changed dynamically and designed to support fast real-time interaction. When a user confirms that the Highlight is of interest by clicking then the Highlight turns into a more static Selection. The Selection stays the same until a new Selection is made. The Selection typically gets a different color than the Highlight, which allows the user to contrast the two sets to see commonalities and differences. Commonalities between the Selection and the Highlight get their own color that makes it stand out from the other objects.

## 5.3 Drill down/up

There are two methods to navigate large data sets in Marcato. The first is controling the set of items that is imported into Marcato from the source database with filters. The second is Drilling down and Drilling up.

Drilling down rescopes the domain of investigation to the current Selection. Essentially, it allows the user to zoom in on an interesting part of the data set to get a more detailed view. Marcato always shows all the data, therfore drilling down does not add more detail, but by hiding the rest of the data it provides the Selection with more room to be visualized, which provides a clearer view of the individual data points.

Drilling up performs the converse operation. It rescopes the domain of investigation to the entire data set, preserving the current Selection. This can allow the user to select small parts of the data that are of interest in a Drilled down view, after which the user can Drill up to see the context of the Selection in the entire data set. An important point to note is that Drilling down can change the distribution of the data values, which makes different patterns in the data stand out.

## 5.4 Network expansion

Network expansion allows the user to extend the currently Selected set of Nodes or Links with respectively Links or Nodes that are directly connected. This way it is possible to let the entire Selection grow hop-by-hop over the Link network. Network connectivity, or the lack of connectivity, can be investigated in this way.

Drilling down provides more detail, but it also cuts away the context, so another common use is to gather the local network context of a Selection of Nodes before Drilling down, which can provide the user with a minimal relevant collection of contextual information to judge the patterns in the Drilled down view. A number of fine-grained expansion tools are also available, such as expansion within the domain of the selection. This expands the currently Selected set of Nodes and Links with all the Links between currently selected Nodes that are not selected.

## 5.5 Node hierarchies

The Node hierarchy that was made in the Hierarchy Editor View is always visible in a number of different Link-centric views, such as the HEB and MSV. The individual Nodes always reside at the bottom of the hierarchy as leaves on a tree. The branches of the tree are partitioned at every level, and the root of the tree corresponds to the entire current scope (either the entire data set, or the current Drill down). Each category in the hierarchy functions as a button, which when clicked selects all the Nodes it subsumes. An important point to node about the hierarchy is that it can be changed in real time. This makes it possible to rapidly change the categorization of the data, and hence the bundles of links that appear in the HEB or MSV. This means creating a specific hierarchy essentially constitutes a hypothesis that the grouping imposed by that hierarchy correlates with the actual occurrence of Links in the data set. When this hypothesis is true, then patterns will show up in the HEB or MSV, otherwise adding the hierarchy does not reveal any order.

For example, the data shown in Figure 6 is communication traffic between people. The hierarchy on the nodes is based on the role of the communicating people in their organization. The corresponding implicit hypothesis is that people's respective roles influence whether or not they talk to each other. In any company one would expect this hypothesis to hold. If this were not the case, then the internal communication does not correspond to the organizational structure of the company. The fact that there are clearly visible bundles in Figure 6 shows that the hypothesis is true.

# 6 Plugin System

In addition to the existing HEB, MSV, map, scatter plot, and search-and-filter views, SynerScope Marcato can be customized by developing plugin views using HTML5 and JavaScript with AJAX calls. These plugin views can interact with the other views as part of the Multiple and Coordinated Views set through a JavaScript object that provides bidirectional communication between the Plugin View and the rest of Marcato. This allows users to make Javascript routines that influence the current Highlight and Selection or that access the data. HTML5-based views are typically much less performant than the other, QT and OpenGL-based Views.

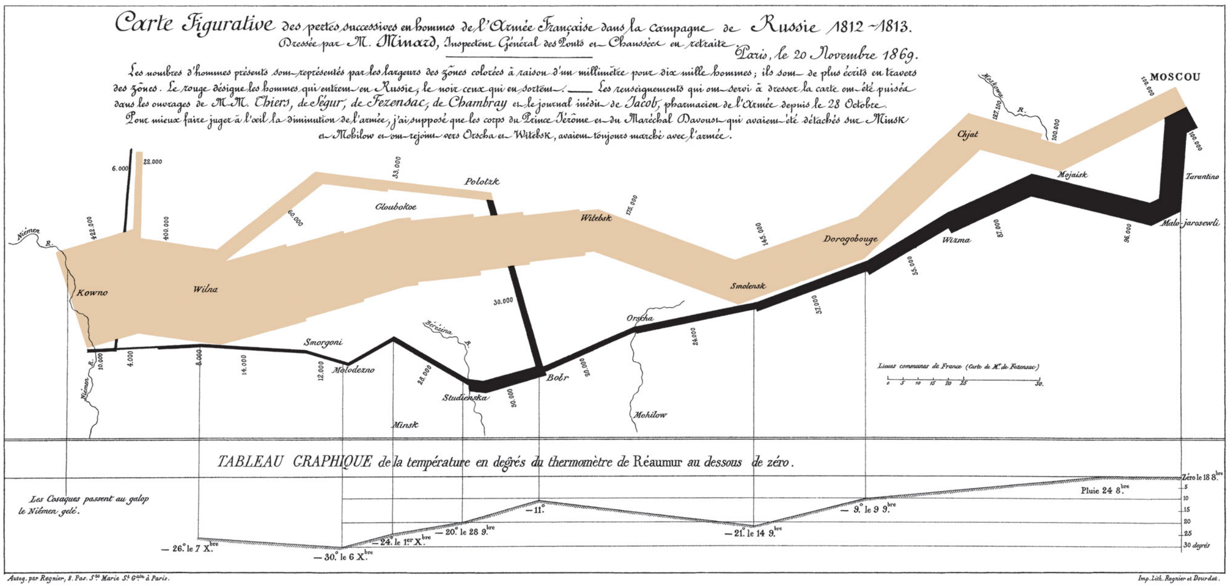



Figure 11: From left to right, the thickness of the beige line represents the size of Napoleon's army as it marches from Kaunas to Moscow. From right to left, the black line represents the army's size as it retreats. The line chart at the bottom shows the temperature during the retreat.

## 6.1 Narrative Charts

Within the NewsReader project, one of the additions to Marcato will be an implementation of interactive Narrative Charts as a Marcato plugin. Narrative charts show the movements and interactions of actors over time. An early example of a narrative chart (shown in Figure 11) is the flow map drawn by Charles Joseph Minard to visualize Napoleon's troop movements during his Russian campaign of 1812 [Minard, 1869]. A more recent example developed at the University of Waterloo [6] is a chart of character interactions in a Tintin comic, shown in Figure 12.

Suppose a user has configured Marcato to show entities as Nodes with the Links between entities being events. This will allow the user to visualize the storylines of a limited Selection of Nodes/Entities as a Narrative Chart. The Narrative Chart will be a Coordinated View in Marcato. This means, for example, that the user can Select a number of Links/Events in the HEB and the corresponding Narrative Chart will be rendered automatically when the Narrative Chart Plugin View is open. Highlighting parts of the Narrative Chart will consequently Highlight the corresponding parts of other views, including the HEB.

Narrative Charts are related to the Massive Sequence View. The main difference between the two is that in the MSV Nodes have a fixed position depending on the current hierarchy, which in Narrative Charts they have a varying location depending on which

[6] http://csclub.uwaterloo.ca/~n2iskand/?page_id=13

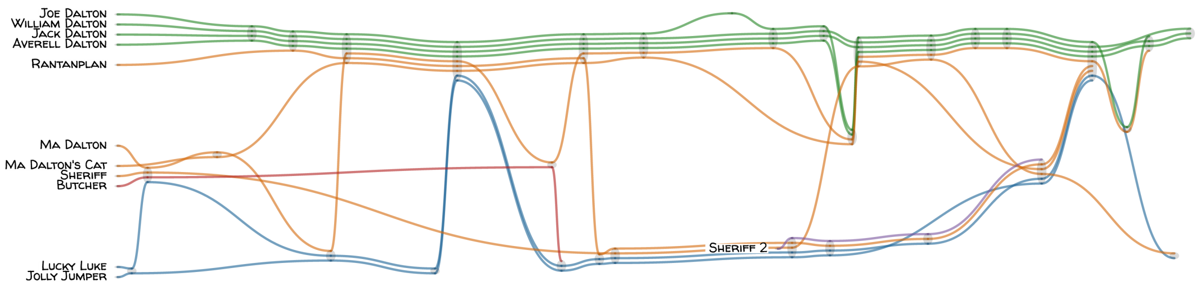



Figure 12: The point where the lines (originating at a character's name on the left) meet represents an interaction of those characters in the story.

Nodes/Entities they share Links/Events with in a certain time frame. Another difference is that in the MSV, Links/Events are shown as lines between Nodes/Entities that bridge the gap between the source and target Node/Entity, while in Narrative Charts the Nodes/Entities move towards each other so that Links/Events become a point. This makes it impossible to draw a consistent visualization when a Node/Entity can have simultaneous Links/Events to two unrelated and hence distant Nodes/Entities.

Force direction can partially overcome this problem, but will not solve it completely. Narrative Charts can be very informative for a limited number of events and participants, but become incomprehensible for large numbers. This, combined with the limited performance of Web plugins motivates the decision to limit the scope of Narrative Charts to a small number of storylines in the initial version of the DSTS.

## 6.2 Web Lookup

Another addition that will be made to Marcato is a Web Lookup plugin that performs automatic lookup of contextual information over a HTTP connection and presents the results either in the form of a ranked result list, like results from a search engine, or in the form of the actual displayed contextual information, like Web pages in a browser. The Web Lookup plugin will have two modes of interaction.

First, users will be able to find documents without typing in a query, but simply by selecting interesting patterns in the storylines shown in the rest of Marcato. The corresponding source documents will be automatically retrieved and displayed. For instance, there is a sudden burst of activity surrounding a certain Node that strikes the user in the MSV. Selecting this burst will yield the corresponding news items that allow the user to directly judge the relevance of the burst.

Second, users will be able to use the Web Lookup plugin as a search engine to find documents and automatically see the surrounding storyline context of the content of that document in the rest of Marcato. For instance, the user will be able to search for a certain topic by tying in a query in a search box which will yield a number of documents, ranked by relevance. These documents discuss parts of the storylines that touch upon this topic. The corresponding Nodes and Links will be automatically selected in the rest of Marcato.

# 7 Technical Deployment of the Decision Support Tool Suite

SynerScope Marcato can be run locally on the end users own machine, if said machine meets the minimum hardware and software requirements specified below. Alternatively, Marcato can be run remotely on a virtual machine in the cloud and streamed to the end users machine.

## 7.1 Local

For a dataset consisting of one hundred thousand Nodes and one million links, SynerScope Marcato runs smoothly on a machine that meets or exceeds the following specifications:

**Processor** Intel Core i5 or equivalent. Intel Core i7 is recommended.

**Memory** At least 2GB–4GB RAM is recommended for large datasets.

**Screen Resolution** At least 1024x768, 1440x900 or higher is highly recommended.

**Graphics** An NVIDIA Quadro 4000 or equivalent professional graphics card is recommended, although experimental support for consumer-level graphics is present.

## 7.2 Server / Cloud

SynerScope Marcato can run in a "private" cloud environment, as a graphically accellerated virtualized application. This allows users to user the Decision Support Tool Suite from their own computer regardless of the technical specification, even a tablet computer is sufficient, as long as the network access to the cloud server is sufficiently stable.

Another advantage of hosting the DSTS in a cloud setting is that a **private** cloud can be set up, so that the application can be run on the same server as the data. This way sensitive data can remain contained on the server.

The validated sufficient specifications of a cloud server to host Marcato for multiple users are the same time are the following:

**Graphics** System should be certified with NVidia VGX/GRID. One or two NVidia GRID K1 cards

**Network** At least two 10GigE connectors

**Processor** Dual-CPU Xeon E5 8-core

**Memory** At least 128GB of RAM

**Disk** At least 8*1TB of Diskspace, preferably SAS 10K RPM

### 7.3 Amazon Web Services (AWS)

It is technically possible to run the Decision Support Tool Suite, on the Amazon [7] cloud. This has been demonstrated to work. Practical issues include the cost of storage, CPU time, GPU time, and network bandwidth, and the possible limitations on accessing other Web resources from AWS. A further investigation is necessary to decide whether or not AWS or some other "public" Cloud platform can constitute a part of the NewsReader setup.

## 8 Conclusion

In this deliverable we documented the design of the Decision Support Tool Suite, a graphical user interface that is meant to support users when making strategic decisions. The DSTS provides insight into sequences of events, allowing users to extrapolate to what might happen in the future.

The DSTS will largely consist of the SynerScope Marcato tool, a visual analytics application that delivers real time interaction with dynamic network-centric data. Within the NewsReader project, Marcato will interface with the KnowledgeStore (through a to be developed query builder), as well as with external Web sources. Marcato features several simultaneous visualization methods ('views'). Interactions with one view (e.g. highlighting, selecting) are replicated in all other open views.

The Hierarchical Edge Bundling View (HEB) is the primary network visualization in Marcato. In addition to the existing views, Marcato can be customized by developing plugin views. Within NewsReader, Narrative Charts and a Contextual Web Lookup will be developed as a Marcato plugins.

## References

[Minard, 1869] Charles Joseph Minard. *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*. Regnier et Dourdet, Paris, 1869.

[7] http://aws.amazon.com/