

Proposal on Investigating Media Transparency

December 4, 2015

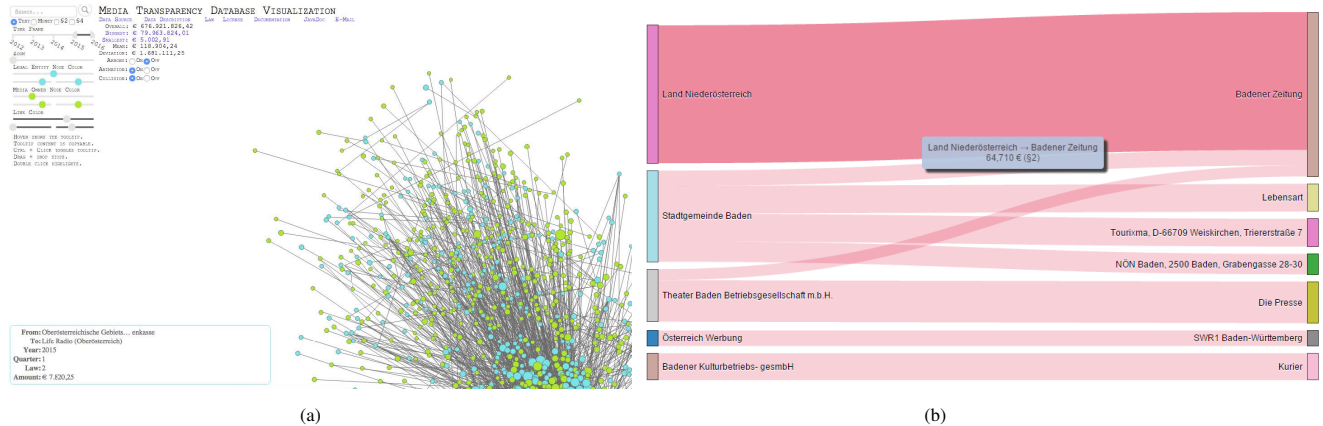
David Pfahler*
TU Wien, AustriaSupervisor: Alexander Rind†
St. Pölten University of Applied SciencesSupervisor: Wolfgang Aigner‡
TU Wien, Information Engineering Group, Austria

Figure 1: Screenshots of (1a) a force directed node link diagram. [20] and (1b) a payment flow visualization filtered by the keyword: “Baden”. [18]

Abstract

The “media transparency database” contains the accumulated amount of money spent by governmental organizations on media companies. This data can be explored as a multimodal dynamic network. Existing web projects already present solutions to visualize the dataset, but to analyse the data further a user needs more interaction methods. I am going to implement a task-tailored dashboard with multiple connected views, which implement brushing and linking to enable the user to analyse the dataset in an easy to use matter.

Keywords: Visual Analytics, Data Driven Journalism, Media Transparency, Politics, Journalism, Network, Time-Oriented-Data

1 Problem Description

Governmental advertisement in media and sponsorships are a possible way to influence press opinion. Therefore, the Austrian parliament passed a law that made it mandatory for governmental orga-

nizations to disclose their expenses for advertisements in different media (TV, radio, print, as well as online) [5].

This so-called “media transparency database” is made publicly available by the Austrian Regulatory Authority for Broadcasting and Telecommunications (RTR) via the Austrian open government data portal [1]. It contains the accumulated amount of money transferred in a certain quarter of the year for each governmental organization and media company. This database can be explored as a multimodal dynamic network.

During the project I want to implement an interactive visual analytics dashboard to explore the media transparency database efficiently.

In Section 2 previous approaches of visualizing the database are presented. In Section 3 I present the expected results of my approach and in Section 4 the used methods and the technologies are described.

1.1 Data Structure

The media transparency database is structured as a Table with each row containing a relation from one governmental organization (*Rechtssträger*) to one media company (*Medium*). This relation contains the amount of transferred money (*Euro*), the quarter of the year (*Jahr Quartal*) and the law of the reason of the payment (*Gesetz*). Table 1 shows the first four entries of this table.

The table contains over 145000 entries over 12 quarters. So that one quarter contains 13000 data entries. There are over 1000 governmental organizations and media companies.

The data quality of the database is not sufficient enough for some

*e-mail: e1126287@student.tuwien.ac.at

†e-mail: alexander.rind@fhstp.ac.at

‡e-mail: alexander.rind@fhstp.ac.at

Rechtsträger	Jahr Quartal	Gesetz	Medium	Euro
Abfallwirtschaft Tirol-Mitte GesmbH	2012 Q4	§2	Bezirksblätter Tirol	8.122,32 €
Agrarmarkt Austria Marketing GesmbH	2012 Q4	§2	Falstaff	26.418,00 €
Agrarmarkt Austria Marketing GesmbH	2012 Q4	§2	Connoisseur Circle	6.142,50 €
Agrarmarkt Austria Marketing GesmbH	2012 Q4	§2	bz-Wiener Bezirkszeitung	7.031,16 €
	⋮			

Table 1: The first 4 entries of the “media transparency database”

Media	Number of relations	Sum
Kronen Zeitung	1042	47.748.890,86 €
KRONE	15	2.524.123,07 €
Kronehit	172	2.223.498,63 €
Kronen Zeitung	1	1.387.919,29 €
Krone bunt	26	1.193.156,45 €
Kronen Zeitung	1	1.071.461,24 €
Krone	25	585.228,77 €
www.krone.at	49	524.306,82 €
KRONEHIT	42	479.324,52 €
	⋮	

Table 2: The first entries of the “media transparency database” filtered by the query string for the media entries “Krone” ordered by the sum of the amount of transfered money

data entries. These entries include spelling mistakes or are just differently formatted. Table 2 shows the data quality problems of this database on the example of the different entries of the media company “Kronen Zeitung”.

2 Related Work

The “media transparency database” is available since the third quarter of 2012. Since then several visualizations got presented:

- [15] presents a visualization that uses grouping of the media entities to reduce the screen space and complexity of the visualization. It only uses one quarter of one year of the total data. It is possible to interact with the visualization and ungroup the media entities.
- [13] presents a static visualization with bar charts and a line plot as a visualization for time oriented data by the Austrian newspaper “Der Standard”.
- One of my colleagues implemented a force directed node link diagram. The user of this visualization is able to interact with the data and filter it with different queries. But the force directed node link diagram was too slow for the huge database [20].
- [18] implemented a website to get an overview of the media dataset. It features multiple visualizations which are all interactive but not connected to one dashboard.

The first two visualizations are presentations of an analysis of the data. But the last two approaches are visualizations that support the user to analyze and investigate into the data. The force directed node link diagram has the problem that it is too slow to render a nice overview of the dynamic network. Additionally it is hard to interpret the payment flow, because it is visually encoded in the

size of the nodes of the diagram. Figure 1a shows the node link diagram with too many nodes.

The visualization of [18] is stable, easy and fast to interact. But the visualizations are distributed onto 4 different web pages, which makes it hard to combine the insight of the user from one visualization with the others. Additionally the payment flow visualization is restricted to only 800 relations. Figure 1b shows the payment flow for the filter with the keyword “Baden”.

3 Expected Results

My approach to improve the previously implemented visualizations is to combine multiple visualizations into a task-tailored visual analytics dashboard [16] and combine them with linking and brushing [10].

With this designed interface the information is more structured and in an easy to read manner [12].

3.1 Multiple Views

The main contribution of my project is the connection of multiple views into one task tailored dashboard. This dashboard should give the user the possibility of gaining an overview of the data, but also explore details of the dataset to gain insight.

The following views are going to be implemented:

Aggregation A part of the dashboard is used to aggregate the data to visualize an overview of the currently selected data. For example bar charts are used to visualize the money flow in the selected years or quarters. Additionally this view enables the user to create a high level selection, for example on a specific year or quarter.

Flow The visualization of the cash flow of a legal entity to its media entities has to be encoded with a flow visualization. (see Subsection 3.2)

Details the details of the selected entries are displayed by visualizations on demand and with a table of the entries.

3.2 Flow Visualization

In the context of this “media transparency database” a survey was created, which compared different network flow representations. [17] The results of this survey showed that all representations used combinations of existing techniques based on the node-link diagram and 5 of 6 online visualization techniques were based on sankey- or chord diagrams.

Based on this results I am going to implement a chord diagram to visualize the flow network data of the database.

3.3 Data Wrangling

In Section 1.1 the problem of the data quality is presented. To enhance this data quality it is necessary to edit the data before visualizing it. I want to include simple data wrangling techniques. [14] This includes joining a set of legal entries or media entries into one, while retaining the relations and removing a set of entries from the visualized data, including its relations.

3.4 Interactions

The interaction with the visualizations is essential for the user to explore the data and to verify or deny a created hypothesis. The user should be able to:

Filter/Sort for time and money.

Search for nodes. Which are legal and media entities.

Combine and remove a selection. (See Subsection 3.3)

Select entries to get insight into this entry with brushing and linking.

4 Methods

In this Section I present how I want to create and design the media transparency dashboard.

My ideas of the design, the used visualizations and interaction techniques were inspired by the related work (see Section 2. Additionally I held feedback with my supervisor and experimented with the data in existing visualizations in diverse technologies like Visplore [19] or Tableau [21]. The used technologies are presented in Subsection 4.1 and the software quality assurance plan is presented in Subsection 4.2.

4.1 Technologies

The following technologies are going to be used:

JavaScript a script language to create dynamic client side webpage.

Data-Driven Documents A JavaScript library for manipulating documents based on data.[11]

jQuery A fast, small, and feature-rich JavaScript library.[8]

Bootstrap A framework for developing responsive, mobile first projects on the web.[2]

Brunch A node.js [9] build tool to compile scripts and styles and to concatenate scripts and styles. [4]

clean-css Is a node.js [9] library for minifying CSS files.

uglify-js Is a node.js [9] library for minifying JavaScript files.

crossfilter Is a JavaScript library to explore multivariate datasets with coordinated views. [6]

DC Is a JavaScript library with native crossfilter support to create charts for multidimensional data exploration[7]

Git Is used as version control system.

4.2 software quality assurance plan

To ensure the quality of the software I am going to test the software in the three most common used browsers. [3]

- Chrome (68.4%): From Version C 45 (2.3%)
- Firefox (19.2%): From Version FF41 (4.2%)
- Internet Explorer (6.8%): In Edge (1.0%) and IE 11 (4.4%)

Additionally I am going to make an usability expert review with my Supervisor or Christina Niederer and to ensure that the software is also usable by non expert users I am going to make informal usability tests with users who have no experience with information visualization or computer science.

References

- [1] Austrian regulatory authority for broadcasting and telecommunications. Bekanntgegebene Daten. https://www.rtr.at/de/m/veroeffentl_medkftg_daten. Accessed: 2015-11-16.
- [2] Bootstrap. <http://getbootstrap.com/>. Accessed: 2015-03-05.
- [3] Browser statistics. http://www.w3schools.com/browsers/browsers_stats.asp. Accessed: 2015-04-12.
- [4] Brunch. <http://brunch.io/>. Accessed: 2015-11-16.
- [5] Bundesgesetz über die Transparenz von Medienkooperationen sowie von Werbeaufträgen und Förderungen an Medieninhaber eines periodischen Mediums (Medienkooperations- und -förderungs-Transparenzgesetz, MedKF-TG).
- [6] Crossfilter - fast multidimensional filtering for coordinated views. <http://square.github.io/crossfilter/>. Accessed: 2015-11-16.
- [7] dc.js - dimensional charting javascript library. <https://dc-js.github.io/dc.js/>. Accessed: 2015-11-16.
- [8] jquery. <http://jquery.com/>. Accessed: 2015-03-05.
- [9] Node.js. <https://nodejs.org>. Accessed: 2015-11-16.
- [10] C. Ahlberg and B. Shneiderman. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 313–317. ACM, 1994.
- [11] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [12] M. Elias and A. Bezerianos. Exploration views: understanding dashboard creation and customization for visualization novices. In *Human-Computer Interaction—INTERACT*, pages 274–291. Springer, 2011.

- [13] M. Hametner. Inserate: 40,6 Millionen im ersten Quartal. <http://derstandard.at/2000017464403/Regierungsinsert-406-Millionen-im-ersten-Quartal>. Web-Standard: 2015-06-15.
- [14] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, R. N. Henry, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011.
- [15] F. Lang. Medientransparenz - die zweite. <http://www.paroli-magazin.at/555/>. Paroli-Magazin: 2013-03-18.
- [16] T. Mühlbacher and H. Piringer. Task-tailored lightweight setups for interactive visual analysis of time series. In *Proceedings of IEEE Vis 2014*, 2014.
- [17] C. Niederer, W. Aigner, and A. Rind. Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism. In H.-J. Schulz, B. Urban, and U. F. von Lukas, editors, *Proceedings of the International Summer School on Visual Computing*, 2015.
- [18] S. Peter, B. Amir, S. Mercedes, and V. Stephan. Medientransparenz austria. <http://www.medien-transparenz.at/>. Accessed: 2015-11-16.
- [19] H. Piringer, C. Tominski, P. Muigg, and W. Berger. A multi-threading architecture to support interactive visual exploration. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1113–1120, 2009.
- [20] M. T. Schrempf. Media transparency database visualization. <http://web.student.tuwien.ac.at/~e0920136/Files/MTDV/src/MTDV.html>. Accessed: 2015-11-16.
- [21] T. Software. Visual analysis best practices: A guidebook. <http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook>, 2015. Last visited 2015-10-10, 2015.

A.3 Detail Visualization

To explore and browse the data it is also necessary to visualize the details of the data. This is done by detail on demand (see Figure 2c) but also with a data table (see Figure 2b)

A.4 Interaction

Different interaction techniques are needed to analyse the data. They are presented in the Section 3.4.

B Time Plan

23.12.2015 The first prototype is up and running. All features from the work packages are at least partly implemented or mocked to get an impression on how the finished application will work. It should only run on the Chrome Browser. So it is possible to get feedback from the supervisor.

22.01.2016 The project is finished and tested on the three major browsers. The usability expert review and the informal usability tests are going to be made and the results get included into the presentation.

29.01.2016 The presentation of the project.

Appendices

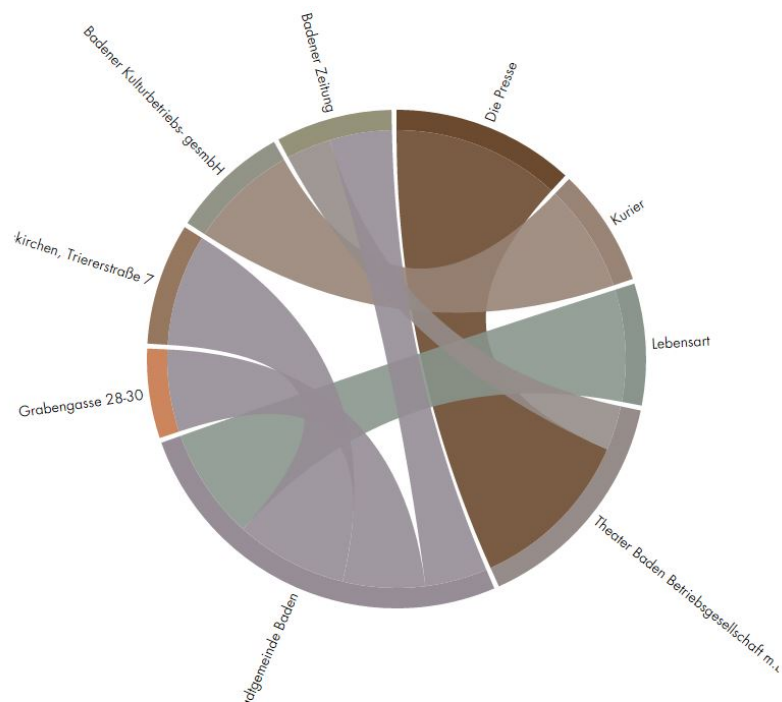
A Work Packages

A.1 Flow Visualization

This diagram is going to visualize the money flow from governmental organizations to the media companies and is therefore a key feature. Figure 2a shows a first prototype of a flow visualization and is already a expected result.

A.2 Filter Charts

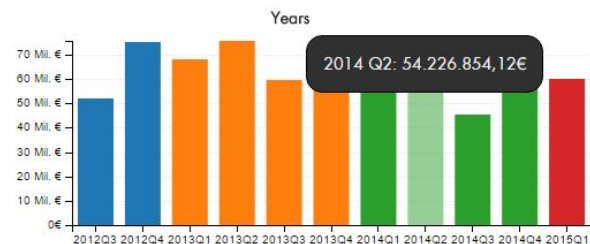
The filtering should be done partly as direct interaction with diagrams. Therefore bar charts or pie charts for ordinal data structures are a good visualization technique. Figure 2c shows a bar chart of the money flow over the last years. One bar is hovered with the mouse and details on demand are displayed as a tooltip. In Figure 2d a selection on one year and on one legal background is done by interacting directly with the visualization in the dashboard.



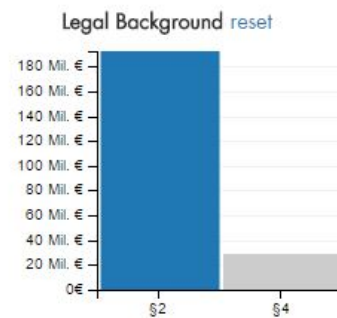
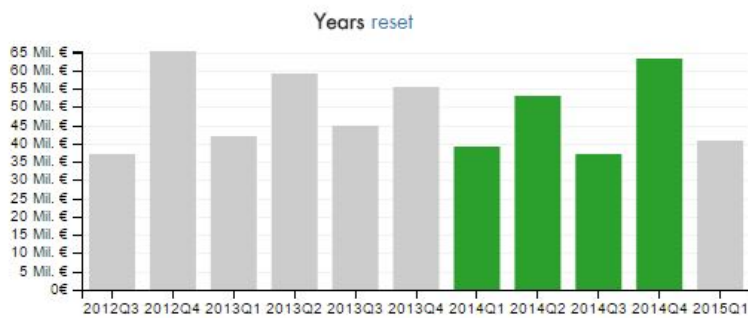
(a) A chord diagram

Rechtsträger	Anzahl an Relationen	Summe
Stadtgemeinde Baden	5	42.806,39€
Theater Baden Betriebsgesellschaft m.b.H.	2	24.670,00€
Badener Kulturbetriebs- gesmbH	1	12.600,00€

(b) A table to browse details



(c) Details on demand on a filter visualization



(d) A selection on one year and on one legal background