



# Proposal on Investigating Media Transparency

David Pfahler\*

Vienna University of Technology, Austria

Supervisor: Alexander Rind†

St. Pölten University of Applied Sciences

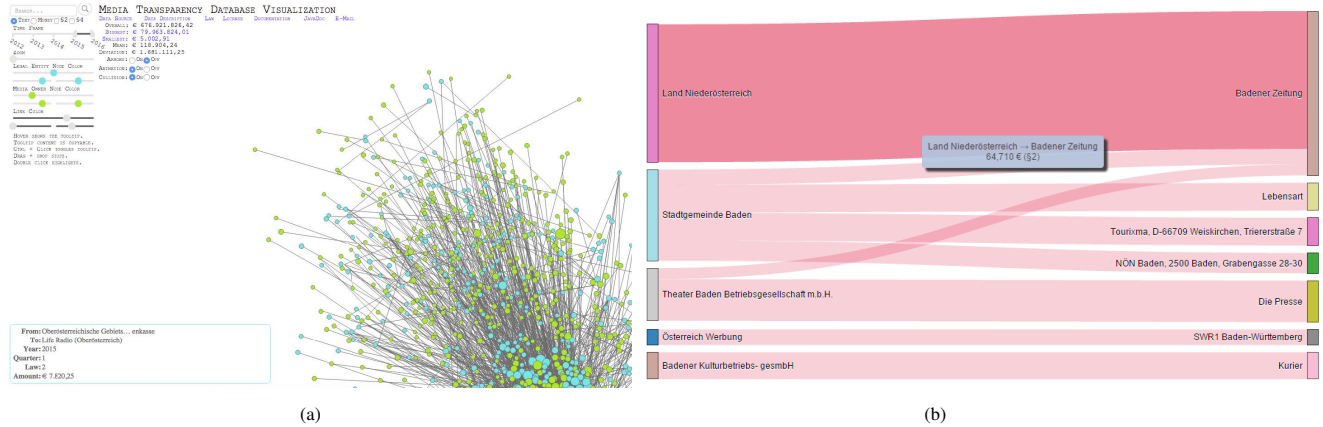


Figure 1: Screenshots of (1a) the force directed node link diagram of [Schrempf] and (1b) the payment flow visualization of [Peter Salhofer] filtered by the keyword: “Baden”

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

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

†e-mail: alexander.rind@fhstp.ac.at

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. [RTR] 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.

## 2 Related Work

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

- [Lang] 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.
- [Hametner] presents a static visualization with bar charts and a line plot as a visualization for time oriented data by the Austrian newspaper “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.

[Schrempf ]

- [Peter Salhofer ] 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 [Peter Salhofer ] is stable, easy and fast to interact. But the visualizations are distributed onto 4 different web sites, 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 implement multiple views (Subsection 3.1) and to include data wrangling techniques (Subsection 3.2)).

### 3.1 Multiple Views

My approach to improve the previously implemented visualizations is to combine multiple visualizations into a task-tailored visual analytics dashboard [Mühlbacher and Piringer 2014] and combine them with linking and brushing [Ahlberg and Shneiderman 1994]. With this designed interface the information is more structured and in an easy to read manner [Elias and Bezerianos 2011].

### 3.2 Data Wrangling

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 1: The first entries of the “media transparency database” filtered by the query string for the media entries “Krone”

The data quality of the “media transparency database” is not sufficient enough for some data entries. These entries include spelling mistakes or are just differently formatted. Table 1 shows the data quality problems of this database.

So it is necessary to edit the data before visualizing it. I want to include simple data wrangling techniques. [Kandel et al. 2011] This

includes joining a set of legal entries or media entries into one, while **remaining** the relations and removing a set of entries from the visualized data, including its relations.

## 4 Methods

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

### 4.1 Multi View Design

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. Additionally **this views are** a possibility to create a high level selection.

**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 4.2)

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

### 4.2 Flow Visualization

In the context of this “media transparency database” a survey was created, which compared different network flow representations. [Christina Niederer and Rind ] 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.

### 4.3 Interactions

The interaction with the visualizations is essential for the user to explore the data and to verify or deny **his hypothesis. He** 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 3.2)

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

### 4.4 Technologies

The following technologies are going to be used:

**Data-Driven Documents** A JavaScript library for manipulating documents based on data.[Bostock et al. 2011]

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

**Bootstrap** A Framework for developing responsive, mobile first projects on the web.[BS ]

**Brunch** A node.js[nod ] build tool to compile scripts and styles and to concatenate scripts and styles. [bru ]

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

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

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

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

MÜHLBACHER, T., AND PIRINGER, H. 2014. Task-tailored lightweight setups for interactive visual analysis of time series. In *Proceedings of IEEE Vis 2014*.

Node.js. <https://nodejs.org>. Accessed: 2015-11-16.

PETER SALHOFER, AMIR BASYOUNI, M. S. S. V. Medientransparenz austria. <http://www.medien-transparenz.at/>. Accessed: 2015-11-16.

Austrian regulatory authority for broadcasting and telecommunications. bekanntgegebene daten. [https://www.rtr.at/de/m/veroeffentl\\_medkftg\\_daten](https://www.rtr.at/de/m/veroeffentl_medkftg_daten). Accessed: 2015-11-16.

SCHREMPP, M. T. Media transparency database visualization. <http://web.student.tuwien.ac.at/~e0920136/Files/MTDV/src/MTDV.html>. Accessed: 2015-11-16.

## References

AHLBERG, C., AND SHNEIDERMAN, B. 1994. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 313–317.

BOSTOCK, M., OGIEVETSKY, V., AND HEER, J. 2011. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

Brunch. <http://brunch.io/>. Accessed: 2015-11-16.

Bootstrap. <http://getbootstrap.com/>. Accessed: 2015-03-05.

CHRISTINA NIEDERER, W. A., AND RIND, A. Survey on visualizing dynamic, weighted, and directed graphs in the context of data-driven journalism.

Crossfilter - fast multidimensional filtering for coordinated views. <http://square.github.io/crossfilter/>. Accessed: 2015-11-16.

dc.js - dimensional charting javascript library. <https://dc-js.github.io/dc.js/>. Accessed: 2015-11-16.

ELIAS, M., AND BEZERIANOS, A. 2011. Exploration views: understanding dashboard creation and customization for visualization novices. In *Human-Computer Interaction-INTERACT*. Springer, 274–291.

Bundesgesetz über die transparenz von medienkooperationen sowie von werbeaufträgen und förderungen an medieninhaber eines periodischen mediums (medienkooperations- und -förderungs-transparenzgesetz, medkf-tg).

HAMETNER, M. Inserate: 40,6 millionen im ersten quartal. <http://derstandard.at/2000017464403/Regierungsinserate-406-Millionen-im-ersten-Quartal>. Web-Standard: 2015-06-15.

jquery. <http://jquery.com/>. Accessed: 2015-03-05.

KANDEL, S., HEER, J., PLAISANT, C., KENNEDY, J., VAN HAM, F., RICHE, N. H., WEAVER, C., LEE, B., BRODBECK, D., AND BUONO, P. 2011. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization 10*, 4, 271–288.

LANG, F. Medientransparenz - die zweite. <http://www.paroli-magazin.at/555/>. Paroli-Magazin: 2013-03-18.

