

DynaMat: Open data-based dynamic comparison and summarization of financial and CSR performance of companies.

Vasiliki Gkatziaki^a, Sotiris Diplaris^a, Symeon Papadopoulos^a, Richard Mills^b and Ioannis Kompatsiaris^a

^aCERTH-ITI, Thessaloniki, Greece; ^bUniversity of Cambridge, Cambridge, UK

ARTICLE HISTORY

Compiled October 2, 2018

During the last decades, citizens and transparency initiatives put increasing pressure on governments, organizations, and companies to be more transparent and to publicize information regarding their financial, environmental, social and governance activities. Securities and Exchange Commission (SEC) is a government commission with the goal of protecting investors from deceitful and manipulative market practices. In 2008, SEC adopted a new rule (SEC 2017) that enforces public companies to disclose their financial statements using the eXtensible Business Reporting Language (XBRL¹) format in an attempt to increase financial transparency. Although several organizations have started engaging in open data practices, data quality, structure and availability is still highly inconsistent across organizations, which makes it challenging and effort-intensive to obtain and analyze such data. Financial statements and CSR reports are challenging to be understandable from the general public since no detailed explanations are provided about the metrics in use. One of open-data impediments is considered the understandability of the data (Zuiderwijk et al. 2012).

WikiRate² is an open-source platform for collecting and analyzing information about companies' Financial, Environmental, Social and Governance performance and it aims to make corporations more transparent, reactive and ethical by making data about their activities more accessible to the general public. More specifically, the platform is based on a question answer schema to make data more comprehensible by providing details about the metrics and the methodology used to derive each metric. Expert and non-expert users can define questions they want to ask to companies, and each answer to such questions is linked to a company, year of reference and a valid source of evidence. WikiRates design calls for three broad types of Metric:

- *Researched Metrics* are containers for storing answers of simple questions, raw data that come from an external source.
- *Calculated Metrics* are containers for analysis of researched metrics and perform some mathematical or logical operation on input metrics to produce output automatically.
- *Relationship Metrics* are containers for storing answers regarding existing relationships between companies.

¹<https://www.xbrl.org>

²<https://www.wikirate.org>

WikiRate’s metric system (Mills et al. 2016) is open and transparent. Users have access to the methodology of the calculated metrics, as well as to the low-level data that are leveraged to create these metrics. All metric values must have a source and a reader can easily follow links to see these sources. Everything on the platform can be discussed and edited, to facilitate discourse about how information should be interpreted and which questions are most important.

A large amount of data is available to the WikiRate platform and distilling knowledge from such data could be a challenging task. Exploratory data analysis is an approach for summarizing and gaining insights from data often by use of visual methods. A visualization approach is already used by the WikiRate platform to facilitate users to better understand the data. More specifically, a visualization of the distribution of a selected metric answer across the full set of companies is provided and highlights the value for the company of interest. WikiRate makes the data available to third parties through a RESTful API for further analysis and exploitation.

In total, the platform has more than 1,300 metrics and over 312,000 metric answers. WikiRate’s users can find a plethora of metric answers for companies but meaningful summarizations are not available to the platform making the comprehension of the data even more challenging for users. In this paper, we present DynaMat, a system that leverages WikiRate’s open data to generate dynamically comparison matrices based on users’ preferences. A comparison matrix is a table that summarizes various aspects of companies. More specifically, DynaMat is a system where users can create their own comparison matrices dynamically by selecting metrics and companies of interest to compare their performance based on the selected metrics.

WikiRate has been designed as a peer-production platform, as each user has the opportunity to engage with the construction of research projects and the analysis of the data they produce. Projects are tools to organize, gather and analyze data on companies environmental, social and governance performance (WikiRate 2018). Through Projects users can pull together a specific set of companies (by sector) and metrics, based on a specific topic, to help other WikiRate researchers contribute data. Most of Research Projects in WikiRate platform are developed mainly through partnerships with key stakeholders such as Amnesty International, GRI, PRME etc. To this end, we decided to take DynaMat one step further and make available to users summarizations of several projects since they often group metrics and companies based on a specific topic. As we mentioned, in 2008, SEC adopted a new rule that enforces public companies to disclose their financial statements using the XBRL format. A plethora of financial facts related to different aspects of financial performance have been extracted from 10-K filings in XBRL format and integrated into the WikiRate platform. DynaMat offers custom summarizations on selected financial facts to make financial performance comparison of companies easier for users. Finally, small summaries, depicted as comparison matrices, of several companies can be also found on the system based on financial metrics that they could function as indicators of tax avoidance for companies, as was demonstrated in our previous work (Gkatziaki et al. 2017).

The main goal of this work has been to develop a system that facilitates users efforts to compare the performance of companies on different aspects of financial, environmental, social and governance performance. Therefore, the evaluation of DynaMat as to its usability, capabilities and perceived usefulness constitutes an important task. Several tasks, comparing the performance of specific companies on specific aspects, were asked from a group of users to be performed by using first the WikiRate platform and then DynaMat. Then, we asked for feedback via an online questionnaire. The questionnaire was based on the USE Questionnaire (Lund 2001) and included Likert

type questions corresponding to the following dimensions: usefulness, ease of use, ease of learning and satisfaction. The evaluation task is still in progress and based on the results, we plan on improving and extending DynaMat in the future, by adding new features, more hints and help messages.

Our contributions. WikiRate platform collects and analyses a plethora of metric answers about companies in a transparent manner with the aim of making data accessible to all and using it to push companies for improvement. DynaMat is a system that leverages WikiRates open-data to offer dynamic summarizations of companies performance on specific aspects in the form of easily digestible comparison matrices. Users can choose between browsing on ready-to-use summarizations of research projects and specific financial metrics or creating their own summarizations by selecting specific metrics and companies to compare. Finally, an evaluation task was designed where a group of users were asked to compare companies performance on specific aspects by using at first WikiRate and then DynaMat to assess the usefulness, ease of use, ease of learning and satisfaction of both systems.

References

- Gkatziaki, Vasiliki, Symeon Papadopoulos, Sotiris Diplaris, and Ioannis Kompatsiaris. 2017. “Large-scale open corporate data collection and analysis as an enabler of corporate social responsibility research.” In *International Conference on Internet Science*, 266–280. Springer.
- Lund, Arnold M. 2001. “Measuring usability with the use questionnaire12.” *Usability interface* 8 (2): 3–6.
- Mills, Richard, Stefano De Paoli, Sotiris Diplaris, Vasiliki Gkatziaki, Symeon Papadopoulos, Srivigneshwar R Prasad, Ethan McCutchen, Vishal Kapadia, and Philipp Hirche. 2016. “WikiRate. org—leveraging collective awareness to understand companies environmental, social and governance performance.” In *International Conference on Internet Science*, 74–88. Springer.
- SEC. 2017. “Interactive Data to Improve Financial Reporting.” Jun. <https://www.sec.gov/rules/final/2009/33-9002.pdf>.
- WikiRate. 2018. “WikiRate Research Projects.” Sep. <https://wikirate.org/Project>.
- Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks. 2012. “Socio-technical Impediments of Open Data.” *Electronic Journal of e-Government* 10 (2).