



► Corporate Network Mapping – Annual Development Review II

**Author** ► Decko Commons ► 2/16/2018

Dissemination level	Public
Contractual date of delivery	Month 24   31/12/2017
Actual date of delivery	Month 26   16/02/2018
Work package	WP5   Corporate Network Mapping - Design and Development
Deliverable number	D5.3   Corporate Network Mapping – Annual Development Review I
Type	DEM
Approval status	Approved
Version	1.0
Number of pages	37
File name	D5_4-20180202-V1_0_DC_Annual_Dev_Review_II-ethan.docx

Demonstrates strong WP5 progress in Year 2, including the integration of OpenCorporates and WikiRate, the introduction of relationship metrics, and the availing of extensive datasets on corporate networks. Presents progress by task, by SMART target, by public product, and by objective.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.



Co-funded by the European Union

## History

Version	Date	Reason	Revised by
<b>0.1</b>	29/01/2017	Framing draft	Ethan McCutchen
<b>0.2</b>	02/02/2018	Details about CERTH's achieved Smart Targets	Vasiliki Gkatziaki
<b>0.3</b>	02/02/2018	Intro and Task drafting	Ethan McCutchen
<b>0.4</b>	05/02/2018	Details about CERTH's work in T5.2	Vasiliki Gkatziaki
<b>0.5</b>	05/02/2018	Edits to T5.4 reporting	Richard Mills
<b>0.6</b>	08/02/2018	Complete draft of task reporting	Ethan McCutchen
<b>1.0</b>	12/02/2018	Complete draft of full report	Ethan McCutchen

## Author list

Organization	Name	Contact information
<b>Decko Commons</b>	Ethan McCutchen	ethan@decko.org
<b>CERTH</b>	Vasiliki Gkatziaki	vasgat@iti.gr
<b>CERTH</b>	Sotiris Diplaris	diplaris@iti.gr
<b>CERTH</b>	Yiannis Kompatsiaris	ikom@iti.gr
<b>Cambridge</b>	Richard Mills	rm747@cam.ac.uk

---

## Executive Summary

WP5's central achievements of the first year of the grant were (1) resolving the conceptual problems of how best to integrate WikiRate and OpenCorporates and, (2) after surveyed the corporate networks data domain, designing and prototyping an elegant representation for these data on WikiRate.org.

Having navigated these challenges successfully, progress accelerated considerably in year 2, with an emphasis on completing and deploying the proof-of-concept work begun in year 1.

WikiRate and OpenCorporates are now integrated (T5.2). To this end, CERTH automated mapping of WikiRate companies to headquarter entities on OpenCorporates (ST14), which was then integrated directly into the WikiRate.org platform (ST58). Decko Commons implemented live integration of the mapped entities via OpenCorporates API calls on WikiRate.org company pages (ST15).

WikiRate.org has also introduced relationship metrics (ST49), the central mechanism for mapping corporate networks (T5.3). CERTH drove the scraping and import of relationship metrics from PDF data (ST52) and worked with all WP5 partners to design and populate an initial set of relationship metrics (ST58). Calculated metrics, which will combine with relationship metrics to form *network-aware metrics* were improved from early beta to readiness for wide adoption.

While the full integration of TheWhistle (T5.4) still hinges upon finding Whistle project partners open to aggregating and publishing report results as metric answers, solid progress was made in WikiRate.org's API (T5.8), the integration mechanism for remote integrations.

The abundant new data on WikiRate.org drove many improvements to the site, including sorting and filtering, new navigation and editing of answers, and project upgrades. It also drove performance and scaling enhancements to Decko (T5.7), now formally released under that name (ST50). Thanks largely to these improvements, WikiRate.org performed strongly and reliably throughout the year (T5.6).

These advances constitute strong WP5 contributions to progress towards reaching grant Objective 2 (subgoals A and B) and Objective 3 (subgoal A). High ambitions (and important SMART targets) remain for year 3.

## Table of Contents

HISTORY.....	1
AUTHOR LIST.....	1
EXECUTIVE SUMMARY.....	2
TABLE OF CONTENTS.....	3
INTRODUCTION.....	5
PROGRESS BY TASK.....	5
T5.1 CORPORATE NETWORK MAPPING - DATA REQUIREMENTS, SCRAPING, AND PARSING.....	5
T5.2 CORPORATE NETWORK MAPPING - INTEGRATION ON WIKIRATE.ORG.....	6
T5.3 CORPORATE NETWORK MAPPING - VISUALIZATION ON WIKIRATE.ORG.....	11
T5.4 THE WHISTLE INTEGRATION.....	18
T5.5 DESIGN AND IMPLEMENTATION OF WIKIRATE.ORG FEATURES.....	20
PROJECTS.....	20
ANSWERS.....	21
SORTING AND FILTERING.....	22
TRANSPARENCY SCORE.....	22
T5.6 IMPROVEMENTS TO DECKO FRAMEWORK.....	23
FOR WEB USERS.....	23
FOR ADMINISTRATORS AND DEVELOPERS.....	24
TOOL ADOPTION.....	25
T5.7 WIKIRATE SYSTEM ADMINISTRATION.....	25
PERFORMANCE.....	25
MAINTENANCE AND DEPLOYMENT.....	26
T5.8 APPLICATION PROGRAMMING INTERFACE.....	26
PROGRESS BY SMART TARGET.....	27
SMART TARGET 14: AUTOMATED WIKIRATE-TO-OPENCORPORATES MAPPING.....	27
SMART TARGET 15: OPENCORPORATES ENTITIES IN USE ON WIKIRATE.ORG.....	29
SMART TARGET 49: RELATIONSHIP METRICS.....	29
SMART TARGET 50: WAGN BECOMES DECKO.....	29
SMART TARGET 52: PDF EXTRACTOR TOOL.....	30
SMART TARGET 53: INTEGRATE COMPANY MAPPING METHODOLOGY ON WIKIRATE.ORG.....	31
SMART TARGET 54: TRANSPARENCY SCORE SPECS (WRIT).....	32
SMART TARGET 58: IMPORT RELATIONSHIP METRICS ON WIKIRATE.ORG.....	33
PROGRESS IN PUBLIC.....	34
PROGRESS TOWARDS GRANT OBJECTIVES.....	35
CONCLUSION.....	36



## Introduction

WP5, *Corporate Network Mapping - Design and Development* organizes software design and development in order to both:

- **Build the *Chain*** – cultivate data maps of relationships amongst corporate entities, and
- **Make it *React*** – allow users to navigate, interpret, recombine, and share the data.

This deliverable is the second of three annual reviews reporting on design and development progress and presenting the products of those efforts. This document presents progress to date from four main perspectives:

1. ***Progress by Task*** considers progress from the original grant framing, going through each of the eight tasks associated with WP5. Here we focus heavily on the *product* of WP5 work.
2. ***Progress by SMART Target*** evaluates our advancement as organized by SMART Targets conceived and refined over the course of the grant’s execution. This section adds additional *process* reporting.
3. ***Progress in Public*** is a short section listing points of access to publicly available work products for further exploration.
4. ***Progress towards Grant Objectives*** casts WP5 progress in the broader context of ChainReact’s objectives, particularly Objective 2: “Connect issues at corporate networks’ edges to network drivers”

## Progress by Task

### T5.1 Corporate network mapping - data requirements, scraping, and parsing

This task is considered completed by our Y1 efforts on data requirements. Ongoing efforts in scraping and parsing are now tracked within T5.2 as part of continuing data integration.

**T5.2 Corporate network mapping - integration on WikiRate.org**

The main focus of CERTH during Y2 of the project concerned the automated WikiRate to OpenCorporates integration and more specifically on mapping WikiRate company entities to OpenCorporates. In Figure 1, the integration plan between WikiRate and OpenCorporates is presented. In detail, the *Company Mapper* component (part of Company Mapping Module) is responsible for mapping a WikiRate company entity to an OpenCorporates company entity. Except from company number, we plan in Y3 to perform additional queries in the API of OpenCorporates to get additional information about the mapped companies. Consequently, except from the company number, the network of the company, granulated industry codes and trademarks are going to be integrated to WikiRate platform.

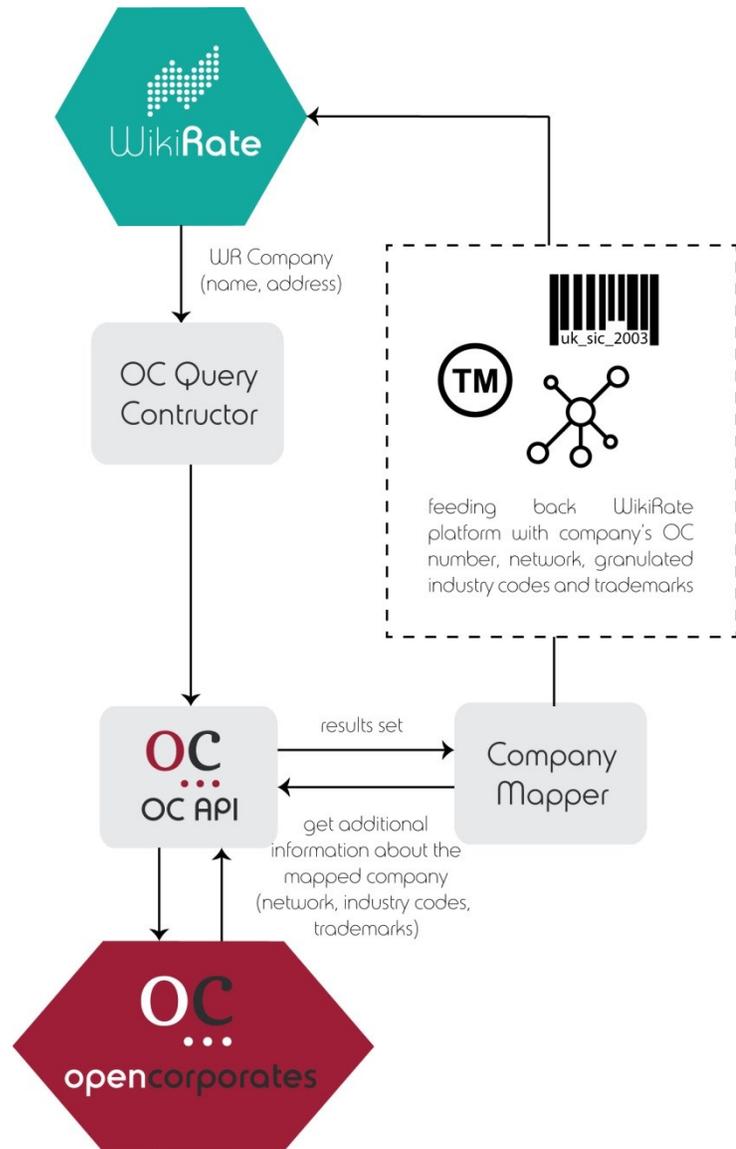


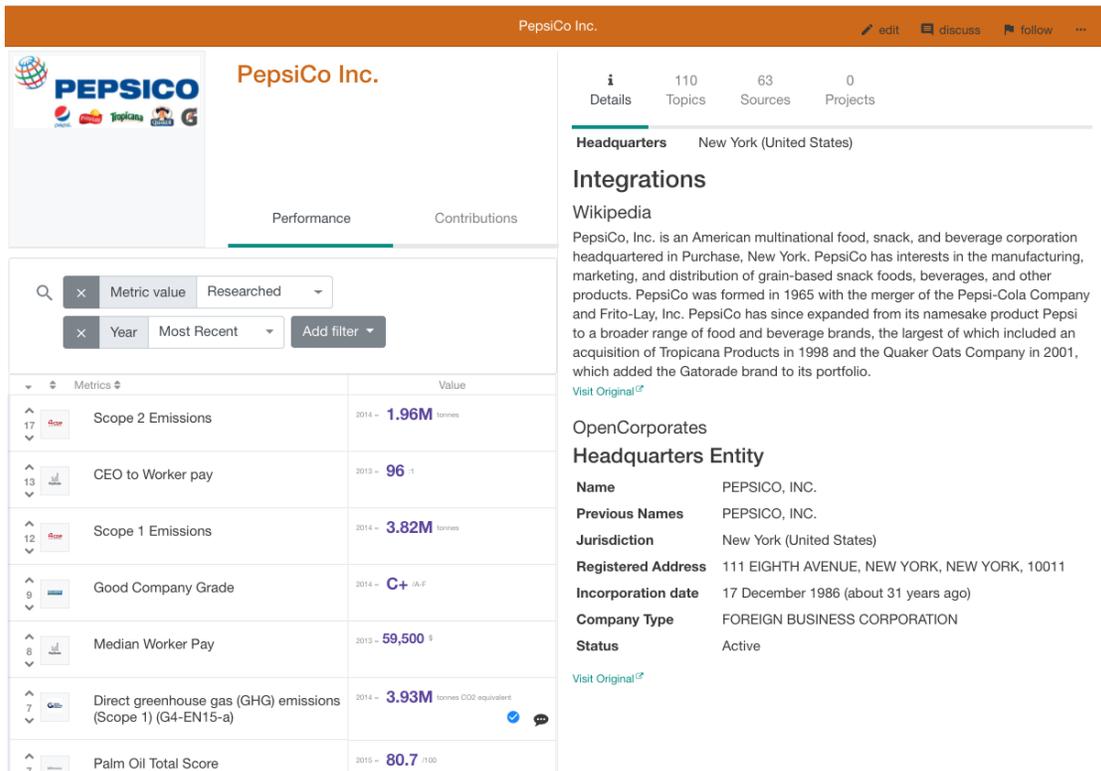
Figure 1: WikiRate – OpenCorporates Integration Plan

The first and more important step achieving the integration between WikiRate and OpenCorporates is automating the procedure of mapping WikiRate companies to OpenCorporates. One of the main challenges discussed is the different definition of companies in the two platforms. WikiRate defines a company entity more loosely than OpenCorporates. To this end, we decided to map a WikiRate company with the corresponded legal entity of the headquarters in OpenCorporates. After long discussions with the consortium several Smart Targets were set for Y2 regarding the integrations between the two struts (WikiRate-OpenCorporates):

- *ST14: Automated WikiRate-to-OpenCorporates Mapping.*

- In Y2, CERTH developed a methodology for mapping WikiRate-to-OpenCorporates companies. The main goal was to achieve accuracy ~80%. During Y1, CERTH in collaboration with WikiRate e.V. created a ground truth dataset in order to evaluate the developed methodology. The achieved accuracy was ~85% during our last experiments. In Progress by Smart Target section more details about the methodology and the results can be found.
- *ST15: OpenCorporates Entities in use on WikiRate.org*
  - This work is discussed at length as part of T5.3
- *ST53: Integrate Company Mapping methodology on WikiRate*
  - After developing and testing the mapping methodology between WikiRate-to-OpenCorporates, it was important to integrate the mappings smoothly to the WikiRate platform. Consequently, CERTH developed a publicly available RESTful Web Service to provide WikiRate-to-OpenCorporates company mappings on demand. More details can be found in section Progress by SMART Target.

A visitor to WikiRate.org can now easily view the product of our corporate network mapping efforts on any Company page. The following screenshot of [http://wikirate.org/PepsiCo\\_Inc](http://wikirate.org/PepsiCo_Inc), taken directly from the live site, shows the two primary “Integrations”, with a descriptive paragraph pulled dynamically from Wikipedia and seven key data points pulled dynamically from OpenCorporates.



**Performance**      Contributions

Metric	Value
Scope 2 Emissions	2014 - 1.96M tonnes
CEO to Worker pay	2013 - 96 <sup>1</sup>
Scope 1 Emissions	2014 - 3.82M tonnes
Good Company Grade	2014 - C+ <sup>ALF</sup>
Median Worker Pay	2013 - 59,500 <sup>3</sup>
Direct greenhouse gas (GHG) emissions (Scope 1) (G4-EN15-a)	2014 - 3.93M tonnes CO2 equivalent
Palm Oil Total Score	2015 - 80.7 <sup>100</sup>

**Integrations**

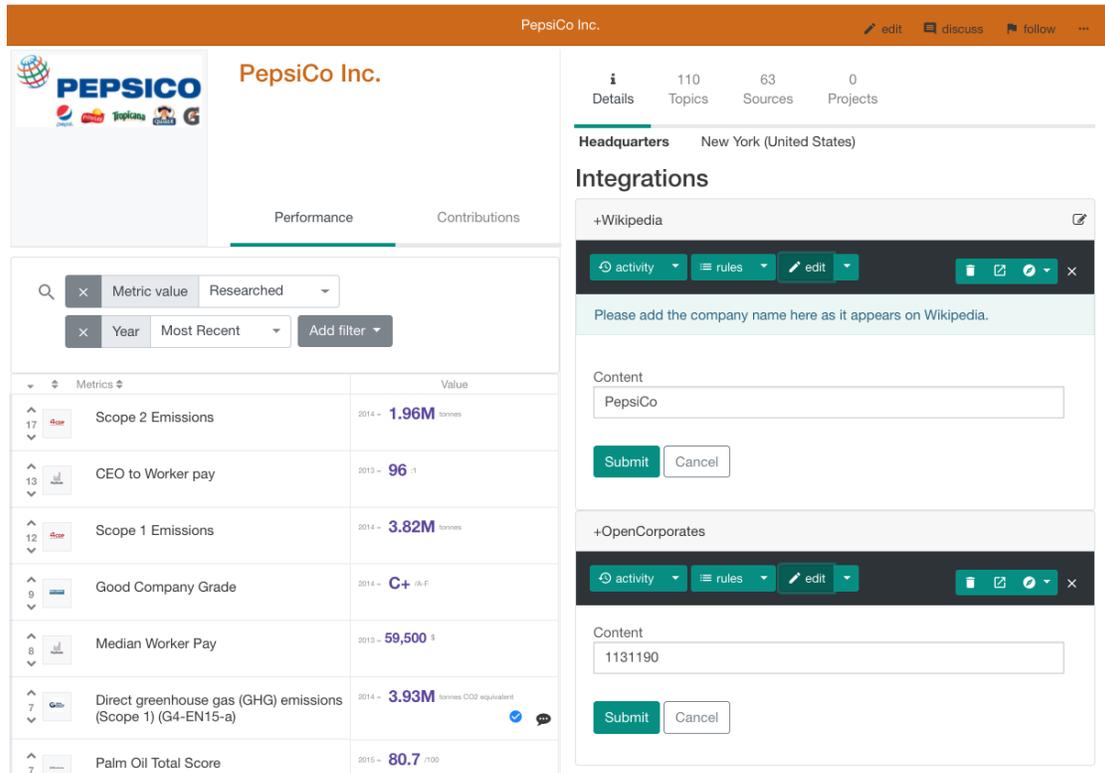
**Wikipedia**  
 PepsiCo, Inc. is an American multinational food, snack, and beverage corporation headquartered in Purchase, New York. PepsiCo has interests in the manufacturing, marketing, and distribution of grain-based snack foods, beverages, and other products. PepsiCo was formed in 1965 with the merger of the Pepsi-Cola Company and Frito-Lay, Inc. PepsiCo has since expanded from its namesake product Pepsi to a broader range of food and beverage brands, the largest of which included an acquisition of Tropicana Products in 1998 and the Quaker Oats Company in 2001, which added the Gatorade brand to its portfolio.

**OpenCorporates**

**Headquarters Entity**

<b>Name</b>	PEPSICO, INC.
<b>Previous Names</b>	PEPSICO, INC.
<b>Jurisdiction</b>	New York (United States)
<b>Registered Address</b>	111 EIGHTH AVENUE, NEW YORK, NEW YORK, 10011
<b>Incorporation date</b>	17 December 1986 (about 31 years ago)
<b>Company Type</b>	FOREIGN BUSINESS CORPORATION
<b>Status</b>	Active

As with most content on WikiRate.org, hovering over the integrated content provides an edit icon that allows users to edit data, examine its history, discuss concerns, and so forth (the icon is always visible on mobile devices, on which there is no hover behavior). In the case of the integrations, the content of the “card” is just an identifier. The Wikipedia identifier is simply the page’s name on en.wikipedia.org; the OpenCorporates identifier is its OC entity id. Incorrect mappings can thus easily be discussed and corrected.



The screenshot shows the WikiRate.org profile for PepsiCo Inc. The top navigation bar includes 'edit', 'discuss', and 'follow' options. The main content area is divided into 'Performance' and 'Contributions' tabs. A search filter is set to 'Metric value' and 'Researched'. A table of metrics is displayed, and two integration cards are visible on the right side.

Metric	Value
Scope 2 Emissions	2014 - 1.96M tonnes
CEO to Worker pay	2013 - 96
Scope 1 Emissions	2014 - 3.82M tonnes
Good Company Grade	2014 - C+ (A-F)
Median Worker Pay	2013 - 59,500
Direct greenhouse gas (GHG) emissions (Scope 1) (G4-EN15-a)	2014 - 3.93M tonnes CO2 equivalent
Palm Oil Total Score	2015 - 80.7 /100

**Integrations**

- +Wikipedia**: Content field contains 'PepsiCo'. Buttons: Submit, Cancel.
- +OpenCorporates**: Content field contains '1131190'. Buttons: Submit, Cancel.

We also drastically upgraded our import tool this year. While it is still predominantly used by team members, it is becoming both more powerful and more user friendly, and we plan to engage a wider circle of users in directly importing data during Y3.

Select:  all  exact matches  alias matches  partial matches  no matches

Conflicts with existing entries:  skip  override

Select	Row	Metric	Company	in file	on WikiRate	Year	Value	Source	Comment
<input checked="" type="checkbox"/>	9	Clean_Clothes_Campaign+Address	Crystal Group	Crystal D	Crystal Group	2016	505 Atwater Circle, St Paul, Minnesota, 55103, USA	Page_000052781	
<input checked="" type="checkbox"/>	353	Clean_Clothes_Campaign+Address	Pacific Impressions	Impressions	Pacific Impressions	2016	200 High Castle Ln, Longwood, Florida, 34747, USA	Page_000052781	
<input checked="" type="checkbox"/>	497	Clean_Clothes_Campaign+Address	AZX Sports	AZX	AZX Sports	2016	8545 Arjons Dr Unit M, San Diego, California, 92126, USA	Page_000052781	
<input checked="" type="checkbox"/>	560	Clean_Clothes_Campaign+Address	Callaway Golf Company	Callaway Golf Inc	Callaway Golf Company	2016	2180 Rutherford Rd, Carlsbad, California, 92008, USA	Page_000052781	
<input checked="" type="checkbox"/>	30	Clean_Clothes_Campaign+Address	Delta Apparel, Inc.	Delta Apparel	Delta Apparel, Inc.	2016	3355 Breckinridge Blvd, Ste 100, USA	Page_000052781	

Companies are complex ecosystems whose actions have complex ramifications. Companies are not independent entities but they are related with each other. One of the most important challenges of ChainReact project is to enrich WikiRate platform with relationship and research metric answers regarding the network of one/more companies. The actions of a subsidiary/supplier of a company should have an effect in the total image of the company. Nevertheless, collecting data about the network of a company is quite challenging. Numerous relationship metric answers are available in OpenCorporates and we plan to integrate such metric answers by the end of Y3.

During Y2, we focused on extracting relationship metric answers from Conflict Minerals Reports (available by Securities and Exchange Commission) and from suppliers list collected by Clean Clothes Campaign. The following SMART Targets were defined regarding that task:

- *ST52: PDF Extractor Tool.*
  - The need of extracting data from PDF reports was raised early on. Clean Clothes Campaign provided a list of sources where the supply chains of several corporations were available in PDF files. In Y2, CERTH developed an extractor tool for extracting tables from PDF files. Additional details can be found in section Progress by SMART Target.
- *ST58: Import Relationship Metric on WikiRate.org*
  - A plethora of relationship metric answers were imported into the WikiRate platform regarding the relationship “supplied by”, obtained from the list sources made available by the Clean Clothes Campaign.

---

In Q8, CERTH introduced the PDF Table Extractor tool, the Excel to CSV tool and the easIE framework to WikiRate e.V. during a two-day pre-huddle WP5-WP8 workshop that aimed at testing several integration mechanisms between the project's three struts. During the workshop, an introduction to Data Extraction, HTML and CSS Selectors was performed. Demos were performed of the PDF Table Extractor tool, Excel to CSV tool and to the easIE framework and two data extraction exercises were performed.

Except from PDF reports, several lists of suppliers were available on the Web. CERTH built custom wrappers to extract supply chains related to these specific companies. In total, the supply chains of 8 corporations were extracted resulting 4,222 relationship metric answers (available through zenodo<sup>1</sup>). Finally, an additional data extraction task was also performed with the aim of extracting smelters and refiners from conflict minerals reports collected from Securities and Exchange Commission. More specifically, we managed to extract 140,063 relationship metric answers involving 5,668 smelters/refiners and about 250 companies. These metric answers are going to be imported gradually to the WikiRate platform.

Through the years, CERTH has collected a vast amount of data and performed a research based on data extracted from 10-K reports in XBRL format available in Securities and Exchange Commission EDGAR platform. More specifically, CERTH investigated how open financial data can be leveraged to extract useful inferences regarding the social performance of companies. In total, 39,029 10-K reports in XBRL format from SEC were collected and after processing these reports about 529,000 financial facts (the dataset is available through zenodo<sup>2</sup>) were extracted related to different aspects of the financial performance of companies. In addition, 3,442 Conflict Minerals Reports were collected and then a data analysis task was performed on two datasets to study research hypotheses regarding specific aspects of Corporate Social Responsibility. The first dataset we studied comprised of 25,500 observations regarding 7,700 companies, while the second comprised of 465 observations regarding 465 companies and contained financial facts extracted from XBRL reports as well as green scores defined by Newsweek in 2016.

The main contributions of our work can be summarized as follows:

- *Collecting a large amount of data from the Web in relation to the financial performance of companies:* A large database of over 500,000 financial facts about more than 50,000 companies was extracted from 10-K filings (available by SEC). Additionally, a REST API was developed to make the collected data available to third parties for further research.

---

<sup>1</sup> <https://doi.org/10.5281/zenodo.824126>

<sup>2</sup> <https://doi.org/10.5281/zenodo.824138>

- 
- *Demonstrating the value of open data for CSR research:* We argue that financial open data can lead to useful inferences regarding the social performance of companies. We support our claim by testing four hypotheses on two datasets that integrate different types of data regarding the social and environmental performance of companies.

This work<sup>3</sup> was presented in the 4<sup>th</sup> International Conference on Internet Science in Thessaloniki Greece in November of 2017.

Also see *Progress by SMART Target* below.

### T5.3 Corporate network mapping - visualization on WikiRate.org

In the grant's first year, the consortium conceived of "relationship metrics" as the primary building block for representing corporate networks on WikiRate.org and built a successful proof-of-concept implementation. In its second year, we refined and advanced the relationship metric functionality, deployed it to the live site, and began populating the data in earnest by importing structured datasets. In the third year, we will further expand and improve these metrics and their use within "network-aware metrics," dynamically calculated metrics that use corporate network data as inputs.

Relationship metrics are an extension of WikiRate.org's standard metric data pattern, in which at most one answer is possible for a given company, metric, and year. Accordingly, answers to Relationship Metrics appear alongside other answers, as in the screenshot below, showing research records on the "Benetton Group Spa" Company page. In this example, the *Supplied By* metric (designed by the *Clean Clothes Campaign*) is a relationship metric, and it appears alongside several standard Research metrics/answers.

---

<sup>3</sup> Gkatziki, Vasiliki, Symeon Papadopoulos, Sotiris Diplaris, and Ioannis Kompatsiaris. "Large-Scale Open Corporate Data Collection and Analysis as an Enabler of Corporate Social Responsibility Research." In *International Conference on Internet Science*, pp. 266-280. Springer, Cham, 2017.

UNITED COLORS OF BENETTON.

Benetton Group Spa

Performance
Contributions

🔍

x Metric value
Researched
x Year
Most Recent
Add filter

Metrics	Value
<span style="font-size: 0.8em;">^ 1</span> <span style="font-size: 0.8em;">v</span> <span style="margin-left: 10px;">🔍 UNGC Status</span>	2015 = <b>Active</b> /Active,Deleted,Non-Communicating
<span style="font-size: 0.8em;">^ 1</span> <span style="font-size: 0.8em;">v</span> <span style="margin-left: 10px;">🔍 UNGC Participant Since</span>	2015 = <b>04 November 2015</b>
<span style="font-size: 0.8em;">^ 0</span> <span style="font-size: 0.8em;">v</span> <span style="margin-left: 10px;">📊 Timeframe Pledge Realization</span>	2017 = <b>2,017</b> Year <span style="float: right;">⬇</span>
<span style="font-size: 0.8em;">^ 0</span> <span style="font-size: 0.8em;">v</span> <span style="margin-left: 10px;">📊 Supply Chain Transparency Pledge Alignment</span>	2017 = <b>Partial alignment</b> /Full alignment,Partial alignment,No <span style="float: right;">⬇</span>
<span style="font-size: 0.8em;">^ 0</span> <span style="font-size: 0.8em;">v</span> <span style="margin-left: 10px;">📊 Supplied By</span>	2016 = <b>329</b> companies <span style="float: right;">⬇</span>

The short “answer value” shown in here is a simple count of the number of companies (329) related to *Benetton Group Spa* via this metric.

As with any answer in the list, clicking on a relationship metric record will expand it to provide information about the metric and the answer(s) for that company. However, unlike standard metrics, relationship metrics have compound answers comprised of a series of *relationship answers*, each outlining, in this case, a distinct supplier relationship.

Benetton Group Spa
edit discuss follow

UNITED COLORS OF BENETTON.

Benetton Group Spa

Performance
Contributions

Metric value
Researched
Year
Most Recent
Add filter

	Metrics	Value
1	UNGC Status	2015 - <b>Active</b> <small>(Active, Deceased, Non-Communicating)</small>
1	UNGC Participant Since	2015 - <b>04 November 2015</b>
0	Timeframe Pledge Realization	2017 - <b>2,017</b> <small>Year</small>
0	Supply Chain Transparency Pledge Alignment	2017 - <b>Partial alignment</b> <small>(Full alignment, Partial alignment, No)</small>
0	Supplied By	2016 - <b>329</b> <small>companies</small>
0	Frequency of Supply Chain Disclosure	2017 - <b>1</b> <small>per year</small>
0	Factory Disclosure	2017 - <b>Names, Country, Address, Product Type</b> <small>(Name, Country, Address, Ownership, Product Type, Number of Workers)</small>
2	Elimination of PFCs	2015 - <b>Yes</b> <small>Yes/No</small>
1	Elimination of APEOs and Phthalates	2015 - <b>Yes</b> <small>Yes/No</small>
1	Detox Transparency	2015 - <b>Yes</b> <small>Yes/No</small>
1	Detox Leader	2015 - <b>Yes</b> <small>Yes/No</small>
0	Bangladesh Accord Signatory	2017 - <b>Yes</b> <small>Yes/No, Yes, not in good standing due to non-payment of membership fees, Yes, ...</small>

Export: [csv](#) / [json](#)

Supplied By

0

Designed by

Clean Clothes Campaign

Clean Clothes Campaign

? What companies supply this company?

Research answer
View Methodology
View About
Metric Page

	Year	Answer																																	
0	2016	<div style="display: flex; align-items: center;"> <div style="background-color: #2e8b57; color: white; padding: 2px 5px; margin-right: 5px;">329</div> <div style="font-size: x-small; margin-right: 5px;">companies</div> <div style="font-size: x-small; margin-right: 5px;">updated 23 days ago by Vasiliki Gkatziki</div> </div> <div style="margin-top: 5px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;"></th> <th style="width: 60%;">Company</th> <th style="width: 30%;">Answer</th> </tr> </thead> <tbody> <tr><td>0</td><td>UGS GIYIM SAN.VE TIC.A.S</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>IDEA HOME TEKSTIL SANAYI VE TICARET A.S.</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>TEKKAN TEKSTIL SAN.VE TIC.LTD STI</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>BROSS TEKSTIL SANAYI VE TICARET A.S.</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>DEBATEKS TEKSTIL SAN.VE TIC.LTD STI KIRKLARELİ BRANCH</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>STYLE FASHION SARL</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>VENUS CONFECTION SARL</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>STE BSB EXPORT</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>STE MB SARL</td><td><b>Tier 1 Supplier</b></td></tr> <tr><td>0</td><td>2F SARL</td><td><b>Tier 1 Supplier</b></td></tr> </tbody> </table> </div>		Company	Answer	0	UGS GIYIM SAN.VE TIC.A.S	<b>Tier 1 Supplier</b>	0	IDEA HOME TEKSTIL SANAYI VE TICARET A.S.	<b>Tier 1 Supplier</b>	0	TEKKAN TEKSTIL SAN.VE TIC.LTD STI	<b>Tier 1 Supplier</b>	0	BROSS TEKSTIL SANAYI VE TICARET A.S.	<b>Tier 1 Supplier</b>	0	DEBATEKS TEKSTIL SAN.VE TIC.LTD STI KIRKLARELİ BRANCH	<b>Tier 1 Supplier</b>	0	STYLE FASHION SARL	<b>Tier 1 Supplier</b>	0	VENUS CONFECTION SARL	<b>Tier 1 Supplier</b>	0	STE BSB EXPORT	<b>Tier 1 Supplier</b>	0	STE MB SARL	<b>Tier 1 Supplier</b>	0	2F SARL	<b>Tier 1 Supplier</b>
	Company	Answer																																	
0	UGS GIYIM SAN.VE TIC.A.S	<b>Tier 1 Supplier</b>																																	
0	IDEA HOME TEKSTIL SANAYI VE TICARET A.S.	<b>Tier 1 Supplier</b>																																	
0	TEKKAN TEKSTIL SAN.VE TIC.LTD STI	<b>Tier 1 Supplier</b>																																	
0	BROSS TEKSTIL SANAYI VE TICARET A.S.	<b>Tier 1 Supplier</b>																																	
0	DEBATEKS TEKSTIL SAN.VE TIC.LTD STI KIRKLARELİ BRANCH	<b>Tier 1 Supplier</b>																																	
0	STYLE FASHION SARL	<b>Tier 1 Supplier</b>																																	
0	VENUS CONFECTION SARL	<b>Tier 1 Supplier</b>																																	
0	STE BSB EXPORT	<b>Tier 1 Supplier</b>																																	
0	STE MB SARL	<b>Tier 1 Supplier</b>																																	
0	2F SARL	<b>Tier 1 Supplier</b>																																	

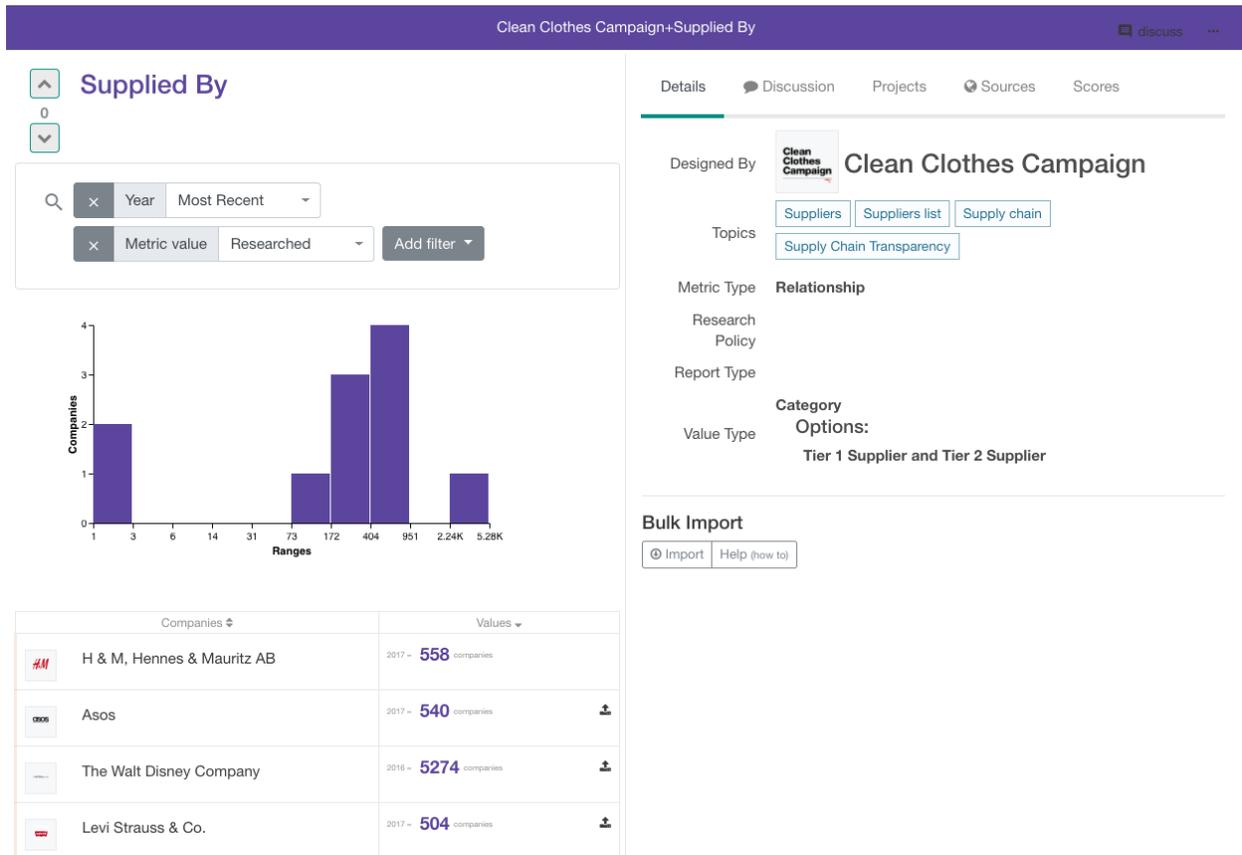
1
2
3
...
41

Discussion

Within the above expanded view, each relationship can be further expanded to reveal verification, sourcing, and discussion. This illustrates a key benefit of the *relationship answer* approach: it retains WikiRate.org’s organizing principle that each discreet value can be independently sourced, verified, and discussed.

Year	Answer										
2016	<div data-bbox="326 489 690 552"> <p><b>329</b> companies             updated 23 days ago by <a href="#">Vasiliki Gkatziaki</a></p> </div> <div data-bbox="326 579 462 611"> <p><b>Relations</b></p> </div> <table border="1" data-bbox="326 625 1404 1728"> <thead> <tr> <th data-bbox="326 625 852 657">Company</th> <th data-bbox="852 625 1404 657">Answer</th> </tr> </thead> <tbody> <tr> <td data-bbox="326 657 852 737">  <a href="#">UGS GIYIM SAN.VE TIC.A.S</a> </td> <td data-bbox="852 657 1404 737"> <b>Tier 1 Supplier</b>  </td> </tr> <tr> <td data-bbox="326 737 852 837">  <a href="#">IDEA HOME TEKSTIL SANAYI VE TICARET A.S.</a> </td> <td data-bbox="852 737 1404 837"> <b>Tier 1 Supplier</b>  </td> </tr> <tr> <td data-bbox="326 837 852 1656">  <a href="#">TEKKAN TEKSTIL SAN.VE TIC LTD STI</a> </td> <td data-bbox="852 837 1404 1656"> <b>Tier 1 Supplier</b>            updated 23 days ago by <a href="#">Vasiliki Gkatziaki</a>   <b>Checks (0)</b>            Nobody has checked this value since it was created            Does the value accurately represent its source?   <input type="button" value="Yes, I checked"/> <input type="button" value="No, I'll fix it"/>   <b>Citations (1)</b>  <div data-bbox="865 1178 1398 1346" style="border: 1px solid #ccc; padding: 5px; background-color: #fff9c4;"> <p>wikirate.org → BENETTON SUPPLIER LIST 2016              added 23 days ago by <a href="#">Vasiliki Gkatziaki</a>   2016  1 Metrics  0 Notes   Download</p> </div>  <b>Comments</b>  <div data-bbox="878 1436 1386 1566" style="border: 1px solid #ccc; height: 60px; margin-bottom: 5px;"></div> <input type="button" value="Comment"/> </td> </tr> <tr> <td data-bbox="326 1656 852 1728">  <a href="#">BROSS TEKSTIL SANAYI VE TICARET A.S.</a> </td> <td data-bbox="852 1656 1404 1728"> <b>Tier 1 Supplier</b>  </td> </tr> </tbody> </table>	Company	Answer	 <a href="#">UGS GIYIM SAN.VE TIC.A.S</a>	<b>Tier 1 Supplier</b> 	 <a href="#">IDEA HOME TEKSTIL SANAYI VE TICARET A.S.</a>	<b>Tier 1 Supplier</b> 	 <a href="#">TEKKAN TEKSTIL SAN.VE TIC LTD STI</a>	<b>Tier 1 Supplier</b> updated 23 days ago by <a href="#">Vasiliki Gkatziaki</a>  <b>Checks (0)</b> Nobody has checked this value since it was created Does the value accurately represent its source?  <input type="button" value="Yes, I checked"/> <input type="button" value="No, I'll fix it"/>  <b>Citations (1)</b> <div data-bbox="865 1178 1398 1346" style="border: 1px solid #ccc; padding: 5px; background-color: #fff9c4;"> <p>wikirate.org → BENETTON SUPPLIER LIST 2016              added 23 days ago by <a href="#">Vasiliki Gkatziaki</a>   2016  1 Metrics  0 Notes   Download</p> </div> <b>Comments</b> <div data-bbox="878 1436 1386 1566" style="border: 1px solid #ccc; height: 60px; margin-bottom: 5px;"></div> <input type="button" value="Comment"/>	 <a href="#">BROSS TEKSTIL SANAYI VE TICARET A.S.</a>	<b>Tier 1 Supplier</b> 
Company	Answer										
 <a href="#">UGS GIYIM SAN.VE TIC.A.S</a>	<b>Tier 1 Supplier</b> 										
 <a href="#">IDEA HOME TEKSTIL SANAYI VE TICARET A.S.</a>	<b>Tier 1 Supplier</b> 										
 <a href="#">TEKKAN TEKSTIL SAN.VE TIC LTD STI</a>	<b>Tier 1 Supplier</b> updated 23 days ago by <a href="#">Vasiliki Gkatziaki</a>  <b>Checks (0)</b> Nobody has checked this value since it was created Does the value accurately represent its source?  <input type="button" value="Yes, I checked"/> <input type="button" value="No, I'll fix it"/>  <b>Citations (1)</b> <div data-bbox="865 1178 1398 1346" style="border: 1px solid #ccc; padding: 5px; background-color: #fff9c4;"> <p>wikirate.org → BENETTON SUPPLIER LIST 2016              added 23 days ago by <a href="#">Vasiliki Gkatziaki</a>   2016  1 Metrics  0 Notes   Download</p> </div> <b>Comments</b> <div data-bbox="878 1436 1386 1566" style="border: 1px solid #ccc; height: 60px; margin-bottom: 5px;"></div> <input type="button" value="Comment"/>										
 <a href="#">BROSS TEKSTIL SANAYI VE TICARET A.S.</a>	<b>Tier 1 Supplier</b> 										

Once more richly populated, relationship metrics will lend themselves to a host of rich graphic visualizations. Already, we are able to show distribution graphs based on the number of suppliers a company is known to have:



Companies	Values
H & M, Hennes & Mauritz AB	2017 - 558 companies
Asos	2017 - 540 companies
The Walt Disney Company	2016 - 5274 companies
Levi Strauss & Co.	2017 - 504 companies

As noted, relationship metrics are the building blocks of corporate network mapping on WikiRate.org, but our aim is not merely to make network data available, but to make it functional in measuring corporate performance. Or – as articulated in Objective 3, Subgoal B – to “support community development of network-aware corporate metrics on WikiRate.org.” Our progress towards that grant objective is examined in more detail below in the “Progress towards Grant Objectives” section. Its relevance here is that “network-aware metrics” has now been translated into concrete functionality: calculated metrics that use relationship metrics as inputs.

Put simply, Calculated Metrics + Relationship Metrics = Network-Aware Metrics.

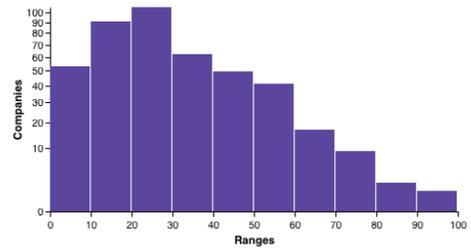
*Calculated Metrics*, in WikiRate terms, are metrics whose answers are dynamically calculated from other metrics. As a simple example, the metric *Female workers %* (designed by the research group *WikiRate SDG Metric design*) is derived by dividing the number of female employees by the number of total employees (both of which are *Research Metrics* designed by the *Global Reporting Initiative*).

edit discuss follow
WikiRate SDG Metric design+Female Workers %

### Female Workers %

1 What percentage of the company's workers are female?

Year: Most Recent
Metric value: Researched
Add filter



Companies	Values
Comcel S.A	2015 - <b>90.4</b>
Estée Lauder Companies Inc.	2016 - <b>85</b>
Bancolombia	2016 - <b>81</b>
Mango Group	2014 - <b>79.6</b>

Details Discussion Projects

---

Designed By **WikiRate SDG Metric design**

Topics SDG5: Gender Equality

Metric Type **Formula**

---

**Formula**

$$= \left( \frac{\text{Female employees (G4-10-a)}}{\text{Employees (G4-10-a)}} \right) * 100$$

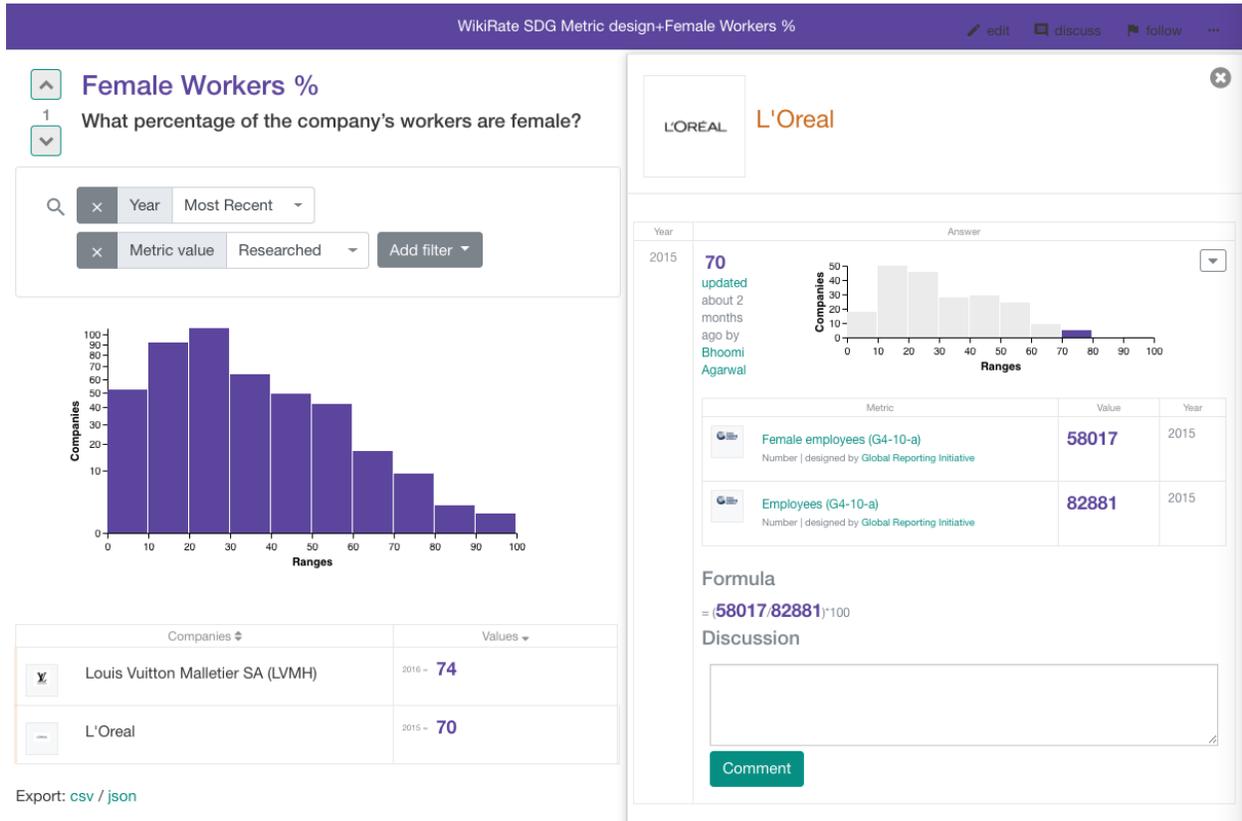
Number | designed by Global Reporting Initiative

---

**About**

36% in 2016 as written in the report used to answer the same question with real number and no %

Expanding an answer on a calculated metric will reveal data for each year available for the given company and metric, including the values for metric answers used as inputs in the calculation. The following shows an expansion of the record for the metric *Female Workers %* and the Company *L'Oreal*.



WikiRate SDG Metric design+Female Workers %

**Female Workers %**  
1 What percentage of the company's workers are female?

Year: Most Recent  
Metric value: Researched

Companies

Company	Year	Value
Louis Vuitton Malletier SA (LVMH)	2016	74
L'Oreal	2015	70

Export: [csv](#) / [json](#)

Year: 2015  
Answer: 70  
updated about 2 months ago by Bhoomi Agarwal

Metric	Value	Year
Female employees (G4-10-a) <small>Number   designed by Global Reporting Initiative</small>	58017	2015
Employees (G4-10-a) <small>Number   designed by Global Reporting Initiative</small>	82881	2015

Formula  
= (58017 / 82881) \* 100

Discussion

Comment

At the beginning of year 2, calculated metrics were in early beta and not ready for broad use. Improving calculated metrics for readiness in their role in network-aware metrics was a major priority in year 2. Having optimized answer data storage, resolved numerous bugs, and drastically upgraded the interface, calculated metrics are now ready for much wider usage, including the soon-to-come addition of functionality supporting the use of relationship metric answers as formula variables.

We are also developing two key enhancements to calculated metrics that we believe will support much broader adoption in network-aware and other formulaic metrics: hybrids and ancestry.

Hybrid metrics are metrics whose answers that either be *calculated* or *researched* (WikiRate's terms for dynamic and static data respectively). These were conceived to address the issue that a company sometimes makes public the value of a calculated metric (eg ratio of CEO to Worker pay) but not the variable values (CEO pay and Worker pay).

---

Metric ancestry allows one metric to inherit answers from one or more other metrics. This solution was conceived to handle the common problem of derivative metrics conceived to expand upon the methodology or research policy of existing metrics. Ancestry avoids rampant duplication, so that a given answer may be used in many contexts but its sourcing, discussion, verification, and (where applicable) correction is all managed in one place.

By combining ancestry and hybrids, researchers can more easily make open derivatives of previously restrictive metrics. This concept protects both the original metric designers, whose ancestral metrics can remain “designer assessed” – meaning that the designer tightly controls answers to their metrics – and derivative designers, who can clearly credit the contributions of the original metric designers while building upon their methodology.

We expect such research patterns will be particularly critical to the success of network-aware metrics, because developing excellent data pertaining to supply networks will entail integrating and expanding upon many interrelated datasets.

Also see *Progress by SMART Target* below. Relevant targets include *ST49: Relationship Metrics* and *ST58 Import Relationship Metrics on WikiRate.org*

#### T5.4 The Whistle Integration

As described in detail in our first Annual Development Review, the strategy for Whistle Integration evolved considerably in our first year from the original plan. Initially, we expected to build an end-to-end proof-of-concept representation of a supply network. The planned alpha version would have been capable of feeding preliminary reports from The Whistle (or “TalkFree” as it was called) into WikiRate, using preliminary data from OpenCorporates.

The first flaw in that plan was underestimating of the difficulty of generating the system’s primary inputs, Whistle Reports. In year one, the Whistle undertook a series of rigorous design tests in an effort to ensure that the team spent its energy developing something that would actually be used in the field by human rights groups. Given the sensitivity of reporting material, the value of (and limits upon) human rights’ workers time, and the breadth of potential reporting needs, it would have been a grave mistake to skip this step and rush in to build a prototype *not* based on real use cases just to prove a network model. Thus the team settled on a “two-path” approach, in which The Whistle development (WP3) and corporate network mapping (WP5) proceeded in parallel to insure that both the Whistle and the corporate network maps are robust and useful in their own rights.

The second flaw was that the initial plan conceived of generalized automated publication of reporting details from The Whistle to WikiRate.org. As The Whistle has learned more about the needs of their key initial users (human rights groups) and WikiRate has learned more about theirs

---

(nonprofit and academic researchers), both have reached the conclusion that there is as of yet little perceived value in the kinds of *generic* WikiRate metrics that could be automatically generated from Whistle Reports (number of reports, number of verified reports, etc). By contrast, both teams see considerable value in *project-specific* metrics that could be designed to fit the needs and capacities of the human rights teams processing the raw Whistle Reports.

Our integration strategy, then, is for the various ChainReact development teams to collaborate closely around the first Whistle pilot projects whose reporting (a) involves WikiRate.org companies and (b) has reached a point of readiness for metric design. Finding an NGO partner for such a campaign has proven difficult, for reasons outlined in D6.2 – in short, reporting outputs that relate to companies, even when aggregated, are perceived as presenting risks both to individual reporters and the NGO that reviews and aggregates these reports. D6.2 presents various ways in which the proposition can be adjusted to mitigate these risks and barriers to adoption by NGOs.

However, the basic outline for the integration remains in place: The Whistle users will be able to proceed smoothly from their project interface to publishing metric answers on WikiRate.org. It is expected that The Whistle project organizers will be able to affiliate a project with one or more WikiRate metrics and report validators will be able to connect a report to one or more WikiRate companies. Both will use WikiRate.org’s API to identify the correct record. Having done so, permissioned project contributors will be able to assess and submit a WikiRate answer connecting a project’s metric to a report’s company.

The search for suitable NGO partners has reinforced the idea that the Whistle is not the only source of reports relevant to companies’ performance which could be integrated with WikiRate. Labor Voices for example sidesteps many of the risks associated with verifiable reports of rights violations by instead collecting survey-type data from anonymous reporters. The collection of this kind of data presents less of a concern to companies and Labor Voices have obtained the cooperation of many companies. We are presently exploring collaboration with Labor Voices and Unicef’s Ureport, both of which could collect relevant data from grassroots users to be published on WikiRate. WikiRate’s API is therefore being developed with an eye to facilitating integrations with a number of sources of bottom-up data.

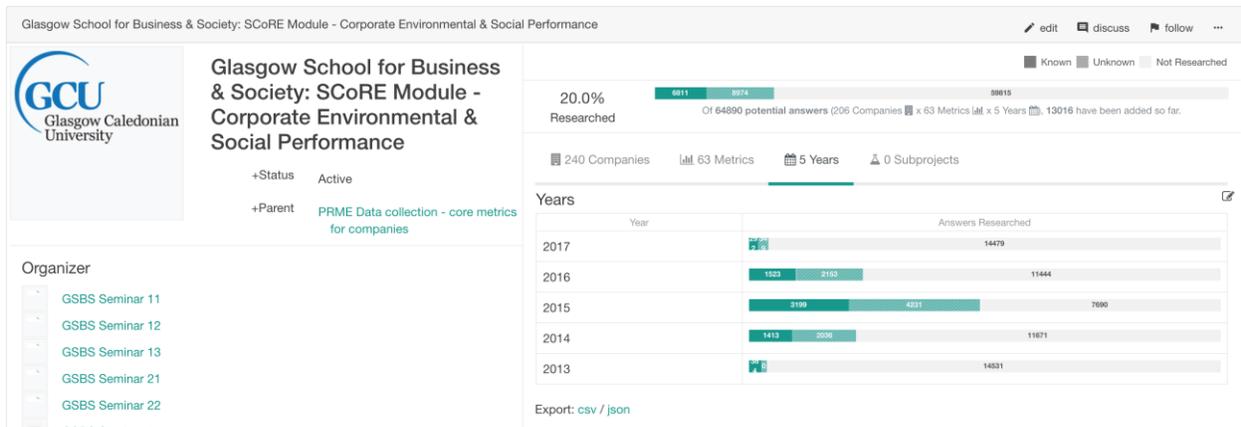
Progress on WikiRate.org’s API is reported below in T5.8.

**T5.5 Design and Implementation of WikiRate.org Features**

Ongoing development of WikiRate.org is primarily driven by the needs of the research program that is the primary long-term driver for creating new open corporate network mapping data.

**Projects**

One major need for Research Projects is the support for constricting projects to specific years. We designed and deployed a solution that allows project organizers to configure a set of years and see progress bars for each year.



If a project has years specified, all progress (company, metric, and overall) are measured against completion of all possible answers for all years. If not, progress is measured towards having an answer for *any* year for each given record.

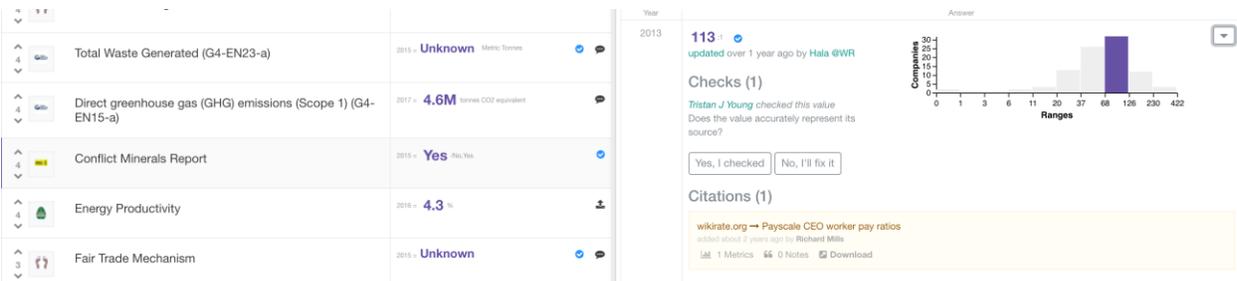
We also added support for subprojects, which involves ancestral linkage between related projects. A subproject is functionally independent but derives many defaults from its parent. Subprojects have already proven helpful tools in organizing larger research collaborations and making participant relationships visible.

Further project improvements include browsing status (active/inactive) and the application of the new pointer interface described in T5.6 to projects.

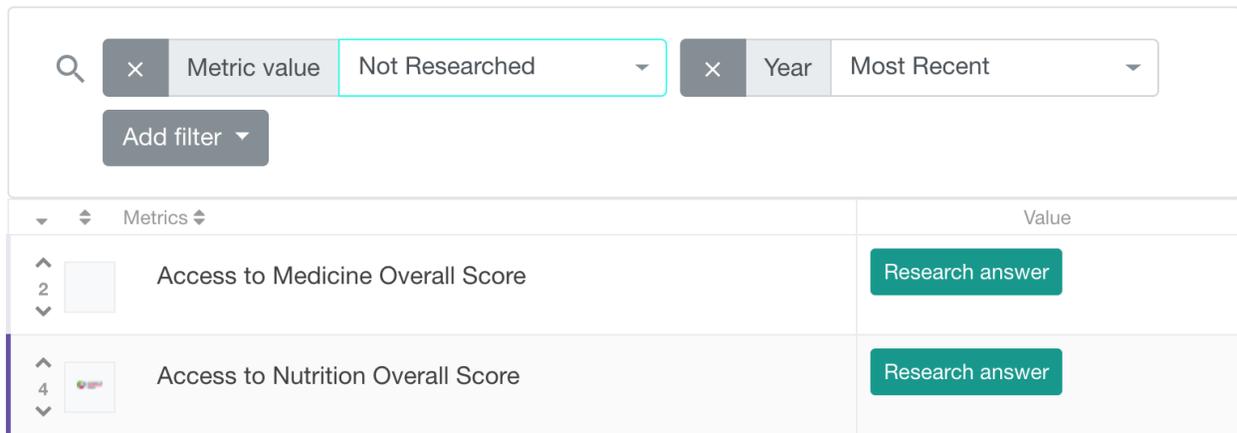
Answers

Projects link directly to the answer research interface, the primary workspace for WikiRate researchers. Improvements to Answers include:

- Improvements to verification (double-checking) workflow, including simpler interface, clearer guidance, and support for multiple checkers
- Support for multi-category answers
- Answer flags to make it easy to see which answers have been imported, checked, discussed, etc.
- An upgrade to the visualization library (Vega)



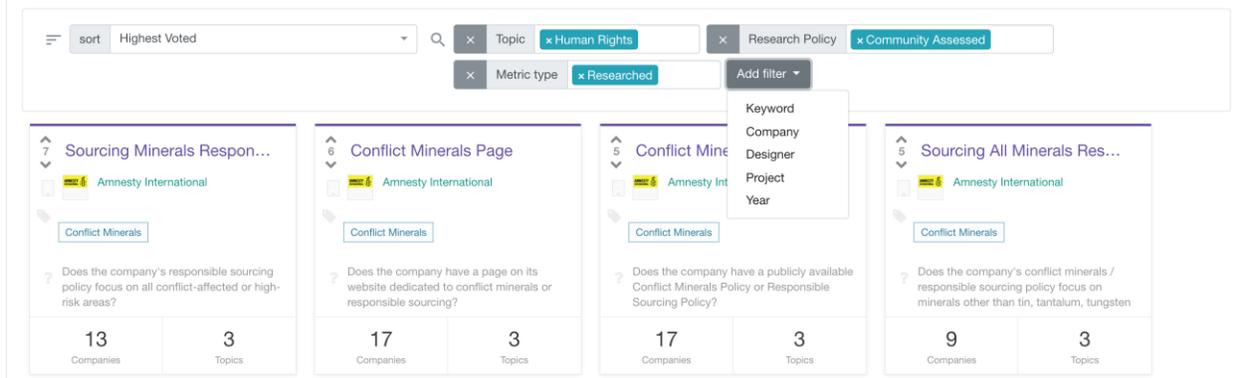
We also embraced a new interface path to answer research based on returning “not researched” answers on company and metric pages. This allows convenient prepopulation of company and metric on research pages.



This path is made far more compelling by the new sorting and filtering interface.

Sorting and Filtering

The rapid growth of data on WikiRate.org has driven the need for more and better means of sorting and filtering. The new interface implemented in Q8 is more accessible, more powerful, and more compact than its predecessor.



Filter pages are typically configured to have few or no filters on by default, but users can quickly add and use additional filters. The above screenshot shows a search already filtering by topic, research policy, and metric type and a dropdown preparing to add one of many other available filters.

The new filter interface appears on browse pages (browse metrics, companies, topics, etc), on the answer listings on company and metric pages, and on the new filtered pointer interface described in T5.6.

Transparency Score

Though we have not yet introduced the Transparency Score, it is anticipated to become a central component of company engagement, so a detailed five-page specification was prepared. See further details below under *Progress by SMART Target*. The relevant target is *ST54: Transparency Score Specs (WRIT)*.

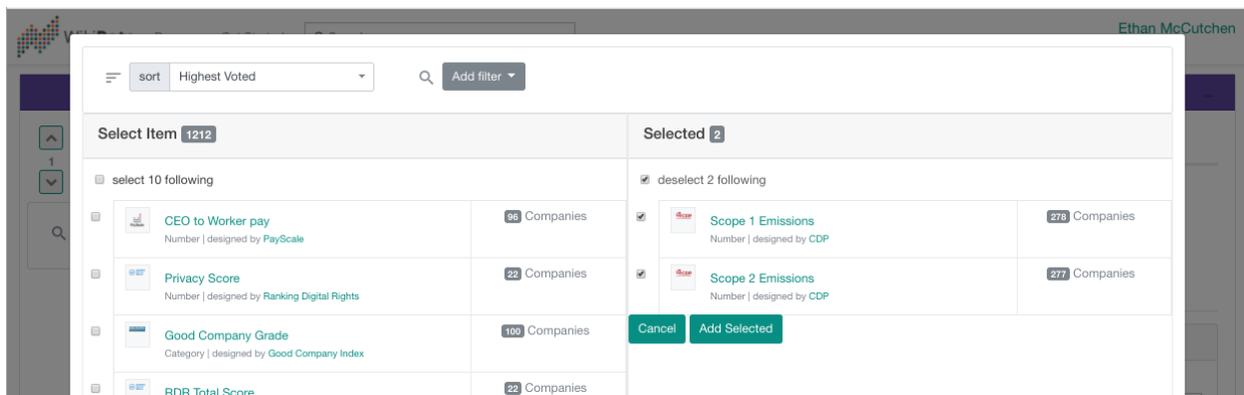
**T5.6 Improvements to Decko Framework**

The most significant Decko improvements this year were directly driven by immediate ChainReact / WikiRate needs, including the new pointer interface (by Project needs) and the new WYSIWYG. Others were driven by maintenance concerns (library upgrades), administration needs (view caching, tmp file handling), and broader tool adoption (edit interface designs, and seeding).

**For Web Users**

While the new filtering interface described in T5.5 was conceived and designed for handling growing data needs on WikiRate.org, it was generalized in implementation to support a general pattern of filtering and sorting that could be used on any deck. The best-defined use case involves adding “Pointer” cards (explicit lists), especially from a large data set that cannot easily be navigated without the use of filters.

The screenshot below shows the filtered list interface being used to create a list of metrics.



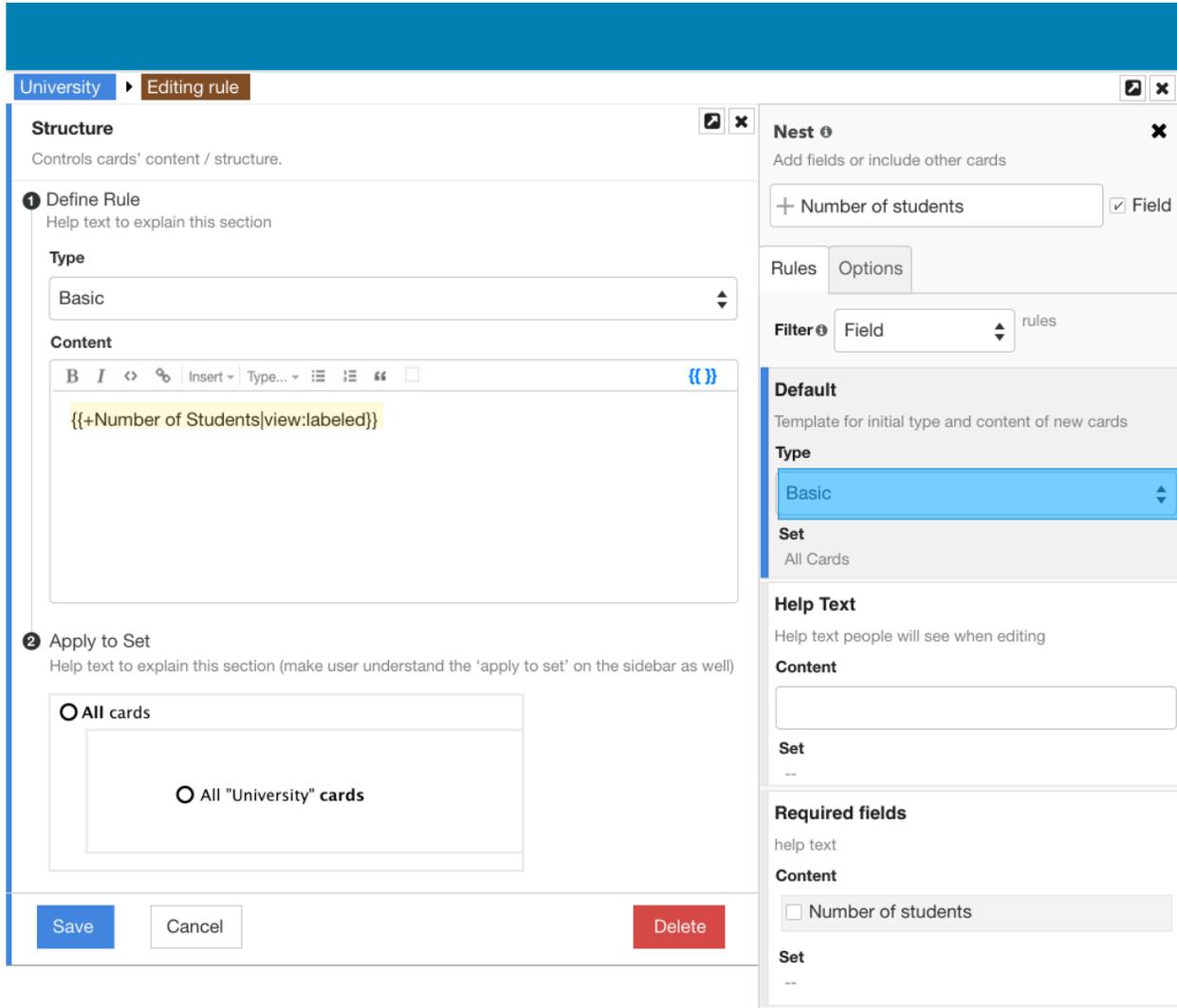
On WikiRate, this interface can now be used within Projects to create lists of associated metrics and companies. It is also used within Calculated Metrics to create lists of metrics to be used as variables.

This year we also developed designs for a new core edit interface pattern for Decko. The current inline-editing pattern is grows increasingly untenable as sites like WikiRate grow increasingly complex, because sometimes cards, like images, can be displayed in very small spaces, and yet editing them takes a lot of space. Simply put, we cannot fit all of Decko’s configuration power into the space of an avatar.

Therefore we have designed a new “modal window” pattern that keeps the current system’s strength of supporting in-context editing but relieves its core problem of space constraints.

The mockup below shows a modal window supporting a user in the complex task of building a pattern structure of nested cards. This kind of nesting is Decko’s primary mechanism for creating patterned structures, but as of yet this kind of structuring is not readily discoverable. The new

interface seeks to address that need by upgrading the rules interface and adding a new “nest” GUI. The mockup shows a nest (at right) being added to a structure rule (at left).



### For Administrators and Developers

View caching optimizes decko system performance by caching rendered card views. View definitions can configure caching rules, and card changes automatically clear relevant views. Development on the system began in 2016, but it was completed in 2017 and is now fully deployed on WikiRate.org.

Decko’s module loading system was also upgraded to support multiple means of interpreting set module files, the primary tool for expanding card functionality, eg. by defining views and events. The new loading system supports the old system of generating temporary files but can also interpret

code dynamically. The different systems allow flexibility in debugging, performance, test coverage, and other development and administrative processes.

### Tool Adoption

We made a concerted effort this year to prepare Decko for a coming push for wider adoption.

The most significant change was rebranding the wagn as *decko*. The Ruby gem is now provided under that name at [rubygems.org](http://rubygems.org), and the code is distributed under that name at Github.

As part of that change, we upgraded most of the major libraries that Decko depends on, most notable Ruby on Rails and Bootstrap. We also isolated asset libraries into separate repositories to make the codebase more approachable.

And, finally, we improved the process for generating seed data to improve the experience eating and populating new decks.

Also see *Progress by SMART Target* below. Relevant targets include *ST50: Wagn becomes Decko*.

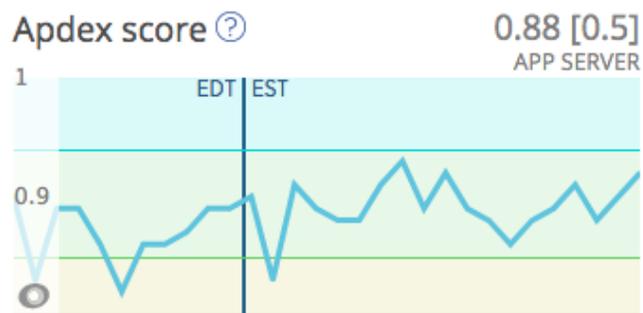
### T5.7 WikiRate System Administration

WikiRate.org performed well in Y2, and refinements to our deployment process have yielded smoother feature introductions.

### Performance

Despite considerable increases in data and site usage, WikiRate’s performance continued to improve in 2017, thanks to performance optimizations driven by regular, consistent attention to New Relic, our performance monitoring tool.

The graph at right shows the APDEX (Application Performance inDEX) score for the final quarter of 2017 (three months is the maximum time span supported for this visualization). APDEX is a measure of performance conducive to user satisfaction. The green range is considered “good”; the blue range is considered “excellent”.



The new view caching system was activated at around the quarter’s midpoint, after which performance grew more robust. However, aim to improve performance further in 2018 so as to have a consistently “excellent” performance, and we are targeting a monthly APDEX above 0.95 by the year’s second quarter.

---

There is also evidence that the site would continue to perform at its current strength even with significant increases in traffic, thanks to the view caching system. Early in the year we began using Loader (<http://loader.io>) and Flood (<http://flood.io>) for stress testing WikiRate. Because initial page renders are much slower than those using cached views, and the percentage of pages available from caching improves with higher traffic as modeled, the system load increases at a much slower rate than simulated traffic.

Other performance improvements this year include:

- The introduction of a “counts” table, which maintains various tallies (of metrics, companies, answers, etc.) used throughout the site. These counts were previously stored as card content, but this lighter weight mechanism is considerably faster and slows the growth of the cards table.
- We have similarly optimized the storage of card changes by ending the practice of storing a card’s initial type, name, and content as “changes”. This has reduced the size of the card\_changes table and improved the speed of queries using it.
- All uploaded WikiRate files and images were moved to cloud hosting. This reduced server load, simplified the generation of development/testing database subsets, and availed a host of straightforward data security tools.

#### Maintenance and Deployment

In moving from the old *wagn* gem to the new *decko* gem, we conducted a general review of our deployment tools and performed general system updates, including to Ruby, Passenger, Apache, and MySQL. We also automated cache clearing upon deployment and moved Pivotal Tracker “story” approval to a non-developer to maximize development resources and insure independent, non-technical review of submitted work. Further, we automated multiple stages of story “progression” so that a story was automatically “Finished” when a Github pull request was merged into the master branch and “Delivered” once deployed to a staging branch for review.

#### T5.8 Application Programming Interface

With interest in WikiRate’s REST API is growing, design and development is increasing focused on two primary use cases.

The first, exemplified by The CleanClothesCampaign, is essentially a read-only scenario in which remote technologies dynamically request data maintained on WikiRate.org for use in applications with topic-specific target audiences. With WikiRate’s help, the CCC has designed a widget that will make requests for lists of companies. Having selected one, users will see a view of that companies’ supply relationships.

---

The second, exemplified by TheWhistle (as discussed in T5.4) involves more dynamic interactions, including contributions from remote sources. This will involve the use of token-based authentication and, most likely, the use of WikiRate research forms within iframes.

While WikiRate continues to use the core Decko REST API as the basis for its web API interactions, it has enhanced them this year in two key ways:

1. By driving the development of a filtering and sorting API, which is central to the new querying functionality discussed in T5.5
2. With the development of new WikiRate-specific CSV and JSON views, both of which are now regularly featured within search results as a means of making the API more discoverable.

## Progress by SMART Target

### SMART Target 14: Automated WikiRate-to-OpenCorporates Mapping

During Y1, CERTH familiarized with OpenCorporates concepts, studied the problem of company mapping, theoretically and practically and conducted some preliminary work on company mapping. Finally, a ground truth dataset<sup>4</sup> comprised of 1,455 companies was developed in collaboration with WikiRate e.V. to facilitate the company mapping task. By the end of the Y1, ST14 was formulated with the goal of developing a methodology for WikiRate-to-OpenCorporates company mapping and achieving accuracy ~80%.

Early on some challenges were identified. The first and most important was the different definitions of company entities in the two platforms. WikiRate defines companies more conceptually while OpenCorporates has a strict definition and each company is a legal entity with specific characteristics. After discussions with the consortium, we decided to map WikiRate companies with the legal entity of the headquarters in OpenCorporates. That decision led to the identification of the second challenge: lack of information about companies on WikiRate.org. More specifically, a preliminary work on company mapping by using only the company name achieved 49% accuracy. To achieve higher accuracy, we decided to extend WikiRate company entity to include information about company's headquarters location in the WikiRate company page.

Given as input the company name as defined in WikiRate and the address of the headquarters, the developed *Address Features Extractor* component of the *Company Mapping Module* extracts address features (*country\_code*, *jurisdiction\_code*) and then *OC Query Contractor* component formulates appropriate requests to OpenCorporates API. Then, the *Company Mapper* component takes as input

---

<sup>4</sup> <https://doi.org/10.5281/zenodo.824192>

the result set and searches for the best candidates based on the name similarity. If the maximum similarity does not exceed a given threshold then the algorithm considers that no matching company has been found. If it exceeds a given threshold, we return as the best match the company that is higher in the list as returned by OpenCorporates and has branch status equal to null. If no company in the set of best candidates has branch status equal to null then the first best candidate is returned as the matched company. The Company Mapping Module Architecture is presented in Figure 2.

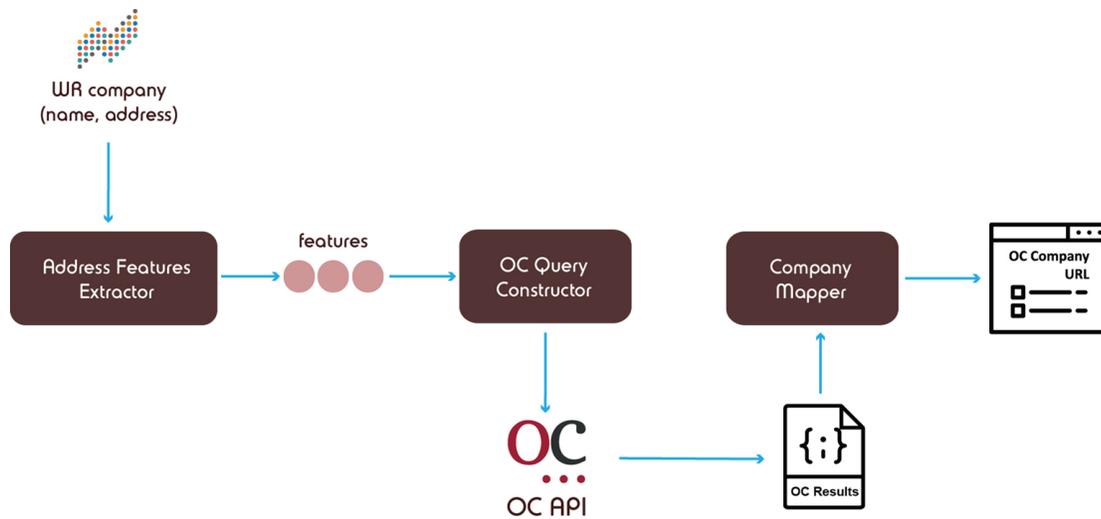


Figure 2: Company Mapping Module Architecture

The performance of the developed Company Mapping Module was evaluated by comparing the produced results over the ground truth dataset, which was created in Y1 of the project. An important increase in the achieved performance was detected compared to the preliminary work on company mapping conducted in Y1. More specifically, the achieved prediction accuracy was 49% during our first attempts while the achieved prediction accuracy was 84.9% during our last experiments. The *Company Mapping Module* is developed in Java and the code is available on GitHub<sup>5</sup>.

One of the most important components of the *Company Mapper* is the Similarity function that calculates the similarity between two company entities. CERTH developed and evaluated two similarity functions for comparing tuples:

- *Company Match Similarity function* which takes into consideration the company name and headquarters address and calculates the tf-idf similarity of the names and addresses. Appropriate weights were applied for each attribute (0.85 in case of the name 0.15 of the address). By taking into consideration only the company name by using the Company Match Similarity function, we achieved 84.9% accuracy and by taking into consideration also the address the achieved accuracy was 85.9%.

<sup>5</sup>

- *Fuzzy Match Similarity function*: which is a token based edit distance similarity function and takes into consideration both company name and address attributes. The achieved accuracy by using *Fuzzy Match Similarity function* was ~80%.

In Y3 of the project, we plan to improve *Fuzzy Match Similarity function* to achieve higher accuracy.

#### SMART Target 15: OpenCorporates Entities in use on WikiRate.org

As described in detail in T5.2, the achievement of ST15 means that two ChainReact struts – OpenCorporates and WikiRate – are now formally connected. Nearly four thousand WikiRate companies have been mapped to OpenCorporates headquarter entities. WikiRate users are able to review, discuss, and update mappings, and when new companies are created on WikiRate, a search for matching companies on OpenCorporates is automatically triggered (see ST53).

While the technical details of the target are recorded in the task reporting, it is worth emphasizing here that this work rests on the work of two major earlier phases of effort. First, there was the need to cross the significant conceptual chasm between the legal conception of a corporate entity, which underpins all OpenCorporates data, with the cultural conception of a company, which underpins both CSR reporting and WikiRate Companies.

Once this challenge was resolved with the concept of “headquarter entities”, the next challenge was automating a sufficiently substantial mapping for initial population. (ST12 and ST14).

Design for the integrative WikiRate features proceeded mostly along conventional lines, though additional input was sought from OpenCorporates due to the use of their brand and API.

#### SMART Target 49: Relationship Metrics

In Y1 we completed SMART Target 13, *Proof-of-Concept Relationship Metrics*. With the concept proven, the challenge of getting Relationship Metrics ready for deployment on WikiRate.org was quite tractable, with the greatest challenges involving making answers easily navigable. The double expansion (first of answers, then of relationship answers) as described above in T5.3 is robust and comprehensive, but further visualizations, filtering, sorting, etc will be necessary to make long and/or complex answers (eg for companies with many suppliers) easily navigable.

Relationship Metrics were deployed live to WikiRate.org at the end of Q6, and further enhancements were deployed in Q8.

#### SMART Target 50: Wagn becomes Decko

The rebranding of *Wagn* has been planned since the middle of 2014, the year in which *Decko Commons* was formed. Concerns about the name *Wagn* had already persisted for years by that point including:

1. Inconsistent metaphors. *Wagn* was a tool created by *Grass Commons* to organize *cards*. But what did wagons and grass and cards have to do with each other?
2. Wrong technical connotations. A “wagon” suggested a rickety, out-of-date, aesthetically uninteresting technology. We wanted to be crisp and current.
3. Wrong social connotations. The old logo idea had been to use the conestoga wagon to suggest “pioneering.” But the actual connotation received was often one of US manifest destiny.
4. Wrong narrative. The genesis of the name was combining “wiki” and “tagging”, because *tagging* was the expected main usage of compound names. But those names are now used primarily as either fields or for establishing sets.

A few names, most memorably *Cardicle*, were proposed to the *Wagn* community, and they were fairly roundly rejected. But *Decko* was widely and immediately embraced. Its virtues included:

1. Short and easy to spell
2. Metaphoric consistency. *Decko Cards* = *Deck of Cards*
3. Suggestion of technical artistry. The *Art Deco* movement touched all kinds of technical design.
4. Domain name availability. We were able to register *decko.org*.

The specific aim of this SMART target was to finalize all *technical* aspects of renaming, so that the tool was fully distributed as *decko* not *wagn*. This was formally achieved at the end of Q7, with *decko* replacing *wagn* as the canonical GitHub repository and RubyGems gem. However, further *communications* aspects of rebranding remain, such as moving to *decko.org*. That move is planned to coincide with the release of *Decko* 1.0. (At the time of writing, *decko* is released as version 0.3.8).

### SMART Target 52: PDF Extractor Tool

Early on the project, the need of extracting data from PDF reports was raised. More specifically, Clean Clothes Campaign provided a list of sources where the supply chains of several companies were available. In most cases, the data were organized in tables in PDF files. The need of extracting such data and importing them on WikiRate platform led to the formulation of ST52.

A first version of the PDF Table Extractor tool was available by the end of Q7 and it was using rulings as separators of rows and columns to detect and extract data from tables lying in PDF files. A second version of the tool was completed by the end of Q8 that detects also tables that do not use rulings as separators. The tool was developed based on the Java library of *Tabula*<sup>6</sup> and is available online<sup>7</sup>. A snapshot of the tool is depicted in Figure 3.

<sup>6</sup> <https://github.com/tabulapdf/tabula-java>

<sup>7</sup> <http://easie.iti.gr/pdf-tables/>

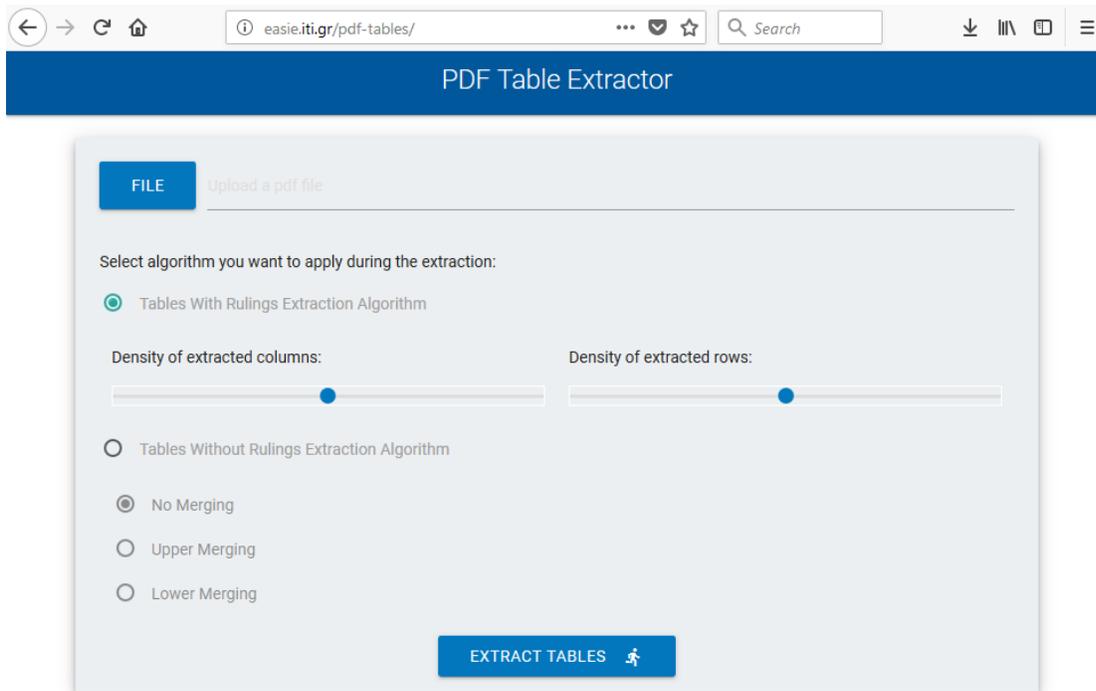


Figure 3: Screenshot of PDF Table Extractor Tool

Using the PDF Table Extractor tool, we managed to extract the suppliers’ list for 23 companies, trapped in PDF reports, resulting to the extraction of numerous of relationship metric answers. In addition, the development of a new algorithm for table data extraction has been initiated and is based on the work<sup>8</sup> “PDF-TREX: An Approach for Recognizing and Extracting Tables from PDF Documents”.

Finally, CERTH introduced the PDF Table Extractor tool during a two-day pre-huddle WP5-WP8 workshop that aimed at testing several integration mechanisms between the project’s three struts.

**SMART Target 53: Integrate Company Mapping methodology on WikiRate.org**

One of the main and most challenging goals of the project is the integration of the three struts WikiRate-OpenCorporates-Whistle. The formation of ST53 targeted on achieving partially this goal by integrating WikiRate and OpenCorporates. Decko Commons used CERTH’s company mappings and OpenCorporate’s web API to upgrade WikiRate.org company pages with live OpenCorporates widgets as presented in Figure 4.

<sup>8</sup> Oro, E. and Ruffolo, M., 2009, July. TREX: An approach for recognizing and extracting tables from PDF documents. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on* (pp. 906-910). IEEE.

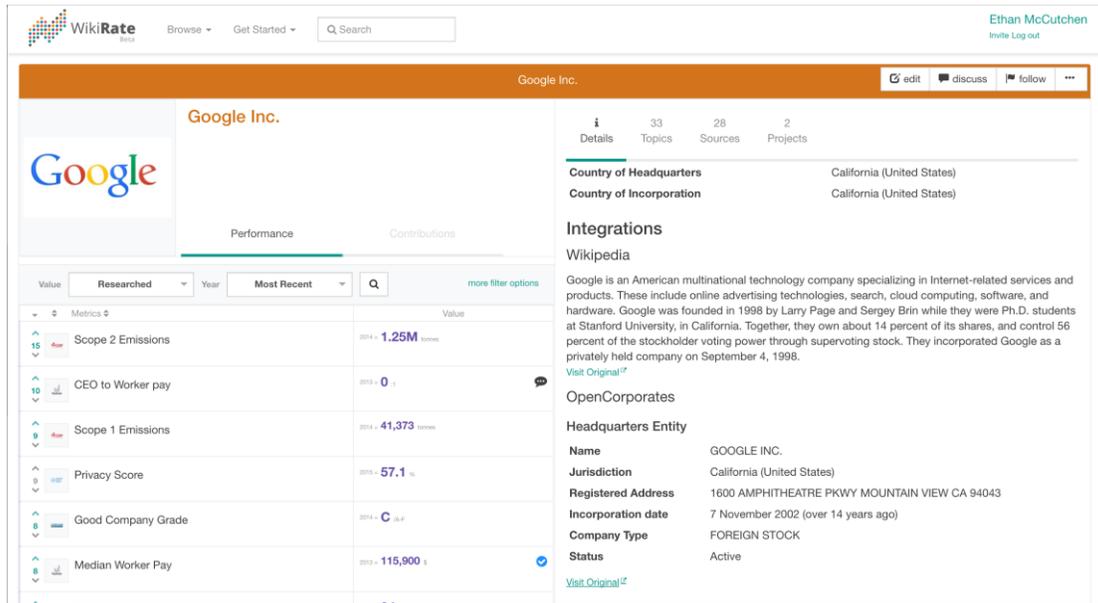


Figure 4: WikiRate company pages updated with live OpenCorporates widgets.

More specifically, CERTH developed a publicly available RESTful Web Service in order to integrate WikiRate and OpenCorporates smoothly by providing automated WikiRate-to-OpenCorporates company mappings on demand. The Web Service was developed in Java and returns the mappings in a JSON format. The endpoint is `easIE.iti.gr/oc_company_mapping` and a basic method is available:

- GET `getMappedCompany` which is responsible for returning the mapped OpenCorporates company number along with jurisdiction code of incorporation and jurisdiction code (if it is not given as input). Users can search for a company by giving either the company name and jurisdiction code (where company is located) or its Wikipedia URL.

The consortium is now working on deepening the integration between WikiRate and OpenCorporates with a focus on *Industries* and *Trademarks*. To this end, two additional SMART targets (ST56, ST16) have been set for the third year of the project focusing on integrating industry metrics and trademarks into the WikiRate platform.

### SMART Target 54: Transparency Score Specs (WRIT)

The WRIT Transparency Score was conceived in 2014 as part of the same *Ratings Proposal* that first proposed Metrics (both Researched and Calculated) as a fundamental organizing framework for quantitative data on WikiRate.org.

The WRIT (still a working name) remains a central component of a long-term strategy for driving corporate responsiveness by rewarding transparency. However, there was growing concern early this

year that the original designs did not do enough to consider concepts such as applicability and materiality, two central concepts in ESG tracking and disclosure.

Therefore in Q7 Decko Commons completed an updated and more detailed specification for the WRIT that embraced applicability as key element in calculating WRIT scores.

In short, “WRIT-ten” metrics (or WRIT input metrics) are defined as metrics are both researchable by and applicable to a given company. The formula for the company’s WRIT score:

$$\text{For all WRIT-ten Metrics:} \\ \frac{\sum (\text{Answer Transparency} * \text{Metric Importance})}{\sum (\text{Metric Importance})}$$

In other words, a company’s WRIT score is an average of its metric answers’ transparency weighted by the metric’s *importance*. Importance, in turn, is measured by adding (a) net votes on a metric to (b) a proportional allotment of importance of applicable calculated metrics of which the metric is a variable.

Applicability factors into the final WRIT score both via researched metrics and calculated metrics. What is most changed about this specification is that *metric importance is now company specific*; in previous conceptions, a metric’s importance was absolute. This does not require new inputs, just a more nuanced – and common sense – interpretation of existing inputs. For example, consider a rating of car companies’ impact on climate change (in WikiRate terms, a WikiRating that is applicable only to companies whose industry is Automotive). If that rating is considered highly important, then the variable metrics that feed it should become more important, *but only to car companies*.

In this way, the WRIT uses the simple inputs of applicability, voting, and appearance as variable to generate *importance*, which serves as a dynamic approximation of materiality. Applicability, while far simpler in concept than materiality, itself is not yet well handled on WikiRate.org, and is the last major obstacle to implementation of the WRIT as specified.

### SMART Target 58: Import Relationship Metrics on WikiRate.org

The following steps were performed in order to import relationship metrics on WikiRate platform:

1. Identify sources containing information about the relationships between companies.
2. Design an appropriate relationship metric into the WikiRate platform.
3. Export relationship metrics available on online sources (webpages), pdf reports, OpenCorporates API.
4. Format appropriate csv import files.

---

## 5. Import relationships into the WikiRate platform.

ST58 was focused on importing “supplied by” relationships into the platform obtained from the lists of sources made available by the Clean Clothes Campaign. After extracting the data, an issue was detected. The majority of companies report their suppliers in factory level. After discussions with all partners, we decided to import the relationships of companies with specific factories (importing factories as companies) and define an additional relationship named “has factory”, if the owner company of the factory is known. A detailed depiction of the relationships between companies and suppliers’ facilities and owner companies is presented in Figure 5.

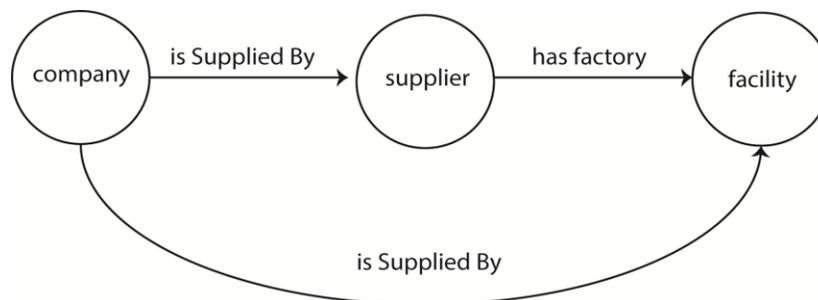


Figure 5

By the end of Q8, numerous relationships regarding the supplier network of several multinational corporations were imported into the WikiRate platform. Importing relationship metric answers into the WikiRate platform is a work in progress and we plan to import more answers by the end of the project.

## Progress in Public

[Wikirate.org](https://wikirate.org) houses OpenCorporates mappings (T5.2, ST14, ST15, ST53) and relationship metrics (T5.3, ST49, ST58). Other improvements to the site are detailed under T5.5. Source code is available at <https://github.com/wikirate/wikirate>

[OpenCorporates.com](https://opencorporates.com) houses data on over 130 million companies

The newly rebranded open-source *decko* gem (T5.6, ST50) is available for download at <https://rubygems.org/gems/decko>. Source code is at <https://github.com/decko-commons/decko>

---

## Progress towards Grant Objectives

WP5 places a key role in the achievement of both Objective 2 (subgoals A and B) and Objective 3 (subgoal A).

### **Objective 2: Connect issues at corporate networks' edges to network drivers**

#### **Subgoal A: Seed and make navigable ChainReact's corporate relationship mapping data by automated processing of official public documents and cross-platform integrations.**

In Y2, we significantly advanced progress towards this objective. The core data representation for ChainReact's corporate relationship mapping data – Relationship Metrics – was completed and deployed to WikiRate.org (T5.3). Relationship metrics have thus far been seeded with initial data, and further seeding will continue during Q9 and Q10. Answers to relationship metrics have been made navigable with Relationship Answers, and additional improvements and visualizations will soon further enhance navigability.

The planned cross-platform integration of WikiRate and OpenCorporates (T5.2) was also achieved this year. Fuller achievement of this objective will be realized with SMART target 74, “Integrated supplier relationship metrics,” at which point relationship data on OpenCorporates can automatically be translated into Relationship Metrics. CERTH's PDF extractor tool has not only enabled automated processing of public documents, but has extended this capability to nontechnical contributors.

#### **Subgoal B: Deepen and improve corporate network maps by engaging corporations and their stakeholders via disclosure outreach.**

WP5's role in reaching this subgoal is comprised of (1) making it easy to contribute to Relationship Metrics and (2) expanding WikiRate more inviting for companies.

We made great progress this year on item 1; WikiRate.org users can now contribute relationship answers both directly and via imports (see T5.3). The import interface is now far more usable, but it is still the domain of power users and needs to be made more robust and discoverable.

As for increasing the appeal to companies, the progress in year 2 was more incremental. While company representatives can use the site as normal users, the plan for year 3 is to create functionality that much more directly targets company representatives, including verified company accounts and WRIT integrations that organize metrics by *importance* (see ST54) and quantify the transparency score benefits of new disclosures.

---

**Objective 3: Incentivize companies to address network issues****Subgoal A: Support community development of network-aware corporate metrics on WikiRate.org.**

As explained in T5.3 reporting, “network-aware corporate metrics” have now been operationalized as calculated metrics that use relationship metrics as variables. Both relationship metrics and calculated metrics took great strides in year 2. What remains is their integration.

We expect to deploy network-aware corporate metrics in Q10 and to see them become focal to WikiRate projects (ie community development) in Q11.

## Conclusion

In year 2, the WP5 partners made strong progress towards grant objectives by reaching eight key SMART targets. WikiRate.org and OpenCorporates have now been integrated by linking WikiRate companies to their headquarter entities on OpenCorporates. WikiRate.org is now mapping corporate networks via relationship metrics.

These relationships, together with calculated metrics, form the building blocks of network metrics, which are on target for full delivery in year 3. Year 3 will also feature enhanced functionality for corporate engagement, greater adoption of relationship metrics, and a deepening of the integration between WikiRate and OpenCorporates.

As reported more extensively in WP3 deliverables, the integration of TheWhistle and WikiRate (T5.4) is the task most behind schedule, primarily due to challenges in finding partners of TheWhistle who are interested in and comfortable with publishing metric answers based on their work. Whether or not such answers become available, WP5 expects to develop a full API for such integrations, which were never expected to be limited to just TheWhistle. Other API use cases will be active by Q10, and we hope to expand its application throughout year 3.