



► D5.3 Corporate Network Mapping – Annual Development Review I

Author ► Decko Commons ► 2/28/2017

Dissemination level	Public
Contractual date of delivery	Month 12 31/12/2017
Actual date of delivery	Month 14 28/02/2017
Work package	WP5 Corporate Network Mapping - Design and Development
Deliverable number	D5.3 Corporate Network Mapping – Annual Development Review I
Type	DEM
Approval status	Submission
Version	1.2
Number of pages	22
File name	D5_3-20170228-V12_DC_Annual_Dev_Review_I.docx

Abstract. A review of progress on WP5’s objectives of creating rich data maps of corporate relationships and using those maps to build “network-aware” ratings. Includes a data strategy update, reporting on all eight WP5 tasks, and links to software results, including the code behind attainment of SMART targets 12 (ground truth dataset) and 13 (relationship metrics).

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.



History

Version	Date	Reason	Revised by
0.1	15/02/2017	Framing draft	Ethan McCutchen
0.2	16/02/2017	Introduction	Ethan McCutchen
0.3	17/02/2017	Tasks and Executive Summary	Ethan McCutchen
0.4	18/02/2017	Tasks and Results	Ethan McCutchen
1.0	19/02/2017	Strategy, Results, and Conclusion	Ethan McCutchen
1.1	27/02/2017	Integrate changes suggested by team	Ethan McCutchen

Author list

Organization	Name	Contact information
Decko Commons	Ethan McCutchen	ethan@decko.org
CERTH	Vasiliki Gkatziaki	vasgat@iti.gr

Executive Summary

WP5, *Corporate Network Mapping - Design and Development* houses software design and development supporting both:

- Building the *chain* – cultivating data maps of relationships amongst corporate entities, and
- Making it *react* – creating tools to navigate, interpret, recombine, and share that data.

This deliverable is the first of three annual reviews reporting on design and development progress and presenting the products of those efforts.

As planned, the first year of the ChainReact grant gave heavy emphasis to establishing data requirements, which drove the resolution of key questions about how best to integrate the three ChainReact “struts”: The *Whistle*, *OpenCorporates*, and *WikiRate*.

That work also drove a significant revision in our data strategy from one focused on starting with end-to-end proof of concept to one focused *first* on separately strengthening the reporting/input mechanisms (at The Whistle), the network mapping mechanisms (at OpenCorporates and WikiRate) and the rating/output mechanisms (at WikiRate); and *next* on improving the end-to-end flows between them.

The resolution of key conceptual questions combined with the updated strategy led to two Year 1 WP5 SMART Targets: a ground-truth dataset mapping WikiRate *Companies* to OpenCorporates *Entities* (ST12), and a proof-of-concept relationship metric adapting WikiRate’s metric framework to handle the representation of corporate relationships (ST13).

Both targets were achieved, and considerable progress was made towards Y2 SMART targets as well. Elaborations of the solutions and links to the software underpinning them are provided in this report.

Table of Contents

HISTORY.....	1
AUTHOR LIST.....	1
EXECUTIVE SUMMARY.....	2
TABLE OF CONTENTS.....	3
INTRODUCTION.....	4
DATA STRATEGY.....	5
PROGRESS BY TASK.....	7
T5.1 CORPORATE NETWORK MAPPING - DATA REQUIREMENTS, SCRAPING, AND PARSING.....	7
T5.2 CORPORATE NETWORK MAPPING - INTEGRATION ON WIKIRATE.ORG.....	9
T5.3 CORPORATE NETWORK MAPPING - VISUALIZATION ON WIKIRATE.ORG.....	11
T5.4 THE WHISTLE INTEGRATION.....	12
T5.5 DESIGN AND IMPLEMENTATION OF WIKIRATE.ORG FEATURES.....	12
T5.6 IMPROVEMENTS TO DECKO FRAMEWORK.....	14
T5.7 WIKIRATE SYSTEM ADMINISTRATION.....	15
T5.8 APPLICATION PROGRAMMING INTERFACE.....	16
RESULTS DELIVERED.....	17
SMART TARGET 12 – COMPLETE GROUND TRUTH DATASET.....	17
SMART TARGET 13 – PROOF-OF-CONCEPT RELATIONSHIP METRIC.....	18
LIVE CHANGES TO WIKIRATE.ORG.....	20
CONCLUSION.....	21

1. Introduction

The “Chain” in “ChainReact” can be understood as a chain of communication, of research, of impacts, or of corporate relationships. The “React” can be interpreted as stakeholder responses along the Chain or the intended positive down-stream effects of connecting the problems of the bottom of the Chain to the image of the top.

But WP5 is for techies, and we tend to interpret *Chain* more narrowly as a data chain (or network), and *React* as the way high-level outputs (ratings, reviews, etc.) dynamically change in response to changes in low-level inputs (whistleblower reports, filings, environmental measurements, disclosures, etc.). Working with this narrower data-level interpretation, WP5, *Corporate Network Mapping - Design and Development* carries the responsibility of (a) building the chain and (b) making it react.

In this report, we will first discuss our strategy for realizing this dynamic data network. We will explain why we have moved from our original data strategy, in which we planned to begin our work on the ChainReact model with an end-to-end proof of concept; to a more practical strategy, in which we first develop strong input mechanism (TheWhistle) and strong output mechanisms (network-aware WikiRate ratings) and then work to strengthen the links between them.

We then report on progress towards strategic goals through the lens of each of the eight WP5 tasks. Three WP5 tasks – T5.1, T5.2, and T5.3 – focus on building the data chain and integrating its representation on OpenCorporates and WikiRate.org. Another, T5.4, organizes the dynamic integration of inputs from TheWhistle. The final four – T5.5, T5.6, T5.7, and T5.8 – support the cultivation and dissemination of the primary mechanism of data “reaction”: network-aware metrics on WikiRate.org.

Finally we present the products of our work, including links to publicly functioning software that has been deployed live, proof-of-concept functionality that has been deployed private, and behind-the-scenes tools that contribute to the development of the ChainReact corporate data network.

2. Data Strategy

Objective 2 of the ChainReact grant is to “**Connect issues at corporate networks’ edges to network drivers**”.

The *networks’ edge* here refers *not* to an edge in the mathematical network sense (in which an edge is a link between nodes) but in the sense of *extremity*. Corporate networks can connect tiny remote factories, mines, and service providers (edges) to prominent global brands (drivers).

In ChainReact’s vision, The Whistle will one of many potential sources of information from the edge. The Whistle’s particular focus is generating reports of human rights violations; other edge sources may devise new ways to measure an edge company’s impacts on the environment, its stakeholders, or the community at large. WikiRate, meanwhile is more heavily focused on network drivers, specifically on supporting the production of rich, transparent, dynamic ratings of these drivers. The challenge of Objective 2 is thus to connect The Whistle reports to WikiRate ratings.

Our core strategy for achieving that end remains unchanged: we will map WikiRate *companies* to OpenCorporate *entities*, create rich corporate relationship maps on both, tag reports from The Whistle with specific companies, and then support network-aware metrics on WikiRate that allow those designing rating systems to make use of network relationships in their ratings.

What *has* changed from the original grant conception is the prioritization of producing an end-to-end connection. Originally, end-to-end connectivity was ranked among our first priorities. But as the consortium partners began formulating more concrete plans, it quickly became apparent that this was inappropriate, and that the better strategy would be *first* to ensure the success of (a) The Whistle as a tool for reporting campaigns and (b) WikiRate / OpenCorporates as integrated tools for representing and rating corporate networks.

There are two primary reasons for this change: (1) the need to ensure the success of The Whistle, and (2) the magnitude of the challenge of mapping supply relationships.

Unlike WikiRate and OpenCorporates, The Whistle is being developed as a brand new site. For much of the first year it has focused on carrying out user research to ensure that it does not suffer the same fate as other software intended to support human rights reporting – failure because it simply doesn’t fit the needs of people in the field. To promote its success, The Whistle team has conducted extensive research into the needs of human rights workers and has worked hard to form appropriate relationships for pilot projects. Requiring that the very first pilot projects (a) connect to companies already richly mapped in our fledgling network maps, and (b) produce outputs that are already appropriate for metric integration, would add dramatic and unnecessary constraints to a project in its infancy. ChainReact, we believe, is better served by giving The Whistle room to develop into a strong, flexible tool than by binding it to a premature proof-of-network concept.

Just as importantly, our preliminary data explorations, in particular the findings of D5.2 (Corporate Network Mapping Availability) indicate that, while there is a growing body of relationship data, supply network data remains highly limited. Thus, even if The Whistle managed to produce reports tagged to companies, it is

unlikely at this early stage that OpenCorporates and WikiRate could meaningfully connect those reports back to the network drivers.

The upgraded strategy actually fits far more naturally with the subgoals of Objective 2 as articulated in the grant:

- **Subgoal A: Seed and make navigable ChainReact’s corporate relationship mapping data by automated processing of official public documents and cross-platform integrations.** This work is well underway in year 1 and our progress towards these ends is articulated below in the task reporting, particularly T5.1 and T5.2.
- **Subgoal B: Deepen and improve corporate network maps by engaging corporations and their stakeholders via disclosure outreach.** While this work is predominantly captured under WP6, it is central to our data strategy, and is the primary mechanism through which we hope to attain end-to-end connectivity between The Whistle reports and WikiRate ratings.

It is worth noting, as well, that our achievement of Objective 2 will likely be richly entwined with our achievement of Objective 3 “**Incentivize companies to address network issues**”. While it would appear at first glance that Objective 3 depends on Objective 2, we believe the two must be advanced in parallel and can in fact support each other.

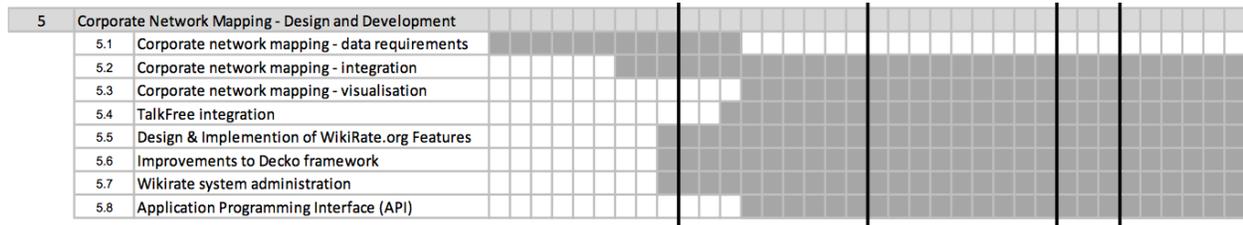
For example, Subgoal A of Objective 3 is to “Support community development of network-aware corporate metrics on WikiRate.org”. Network-aware metric answers are generated by mathematically integrating network maps (see Objective 2) with more conventional metrics answers.

Our solution to this need is to approach relationships themselves as metrics. As outlined in reporting on T5.3 (under *Progress by Task*) and SMART target 13 (under *Results Delivered*) we have already begun work on relationship metrics. This is a major step towards network-aware metrics: just as WikiRate’s calculated metrics can use researched metric answers as variables, its calculated *network-aware* metrics will be able to use researched relationship metric answers as variables.

In the explorations surrounding the relationship metrics proof-of-concept (ST13), we began engaging with the rich diversity of options available to metric designers working in this area. We anticipate that design decisions will in many cases have a significant impact on how reliably and accurately companies can report. By working with companies on metric design, we expect that we can design metrics better positioned to incentivise positively addressing problems (Objective 3A), and that companies will be more helpful with data population (Objective 2B). In this way, the work of WP5 is intersecting deeply with that of WP6, and the two work packages are coordinating extensively as both complete their exploratory first years.

3. Progress by Task

As with many Work Packages, WP5 was designed to accelerate and broaden through the grant. The Gantt chart of WPs and Tasks makes this plan clear.



The grid formed by these first twelve months and eight tasks had fewer than a third of the possible squares darkened for the first year and just one task active for the first sixth months. On the other hand, seven of eight tasks are active throughout year 2 and year 3.

As anticipated, T5.1 was indeed our most consistently active task this year, but some other tasks began receiving attention sooner than expected. For example, we anticipated beginning T5.3 (Corporate Network Mapping – visualization) in M12, but because visualization regularly entered into discussions of data requirements via scenario development, we began exploring designs months earlier.

T5.4, on the other hand, will likely not receive full attention until several months later than originally intended because, as explained in the “Data Strategy” section above, the consortium endorsed The Whistle’s decision to emphasize successful user engagement over data integration with other ChainReact struts in the first phases of its development. For that reason T5.1 may remain active longer than anticipated so that it can respond to the evolution of The Whistle.

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

The core data design challenges of T5.1 have involved creating and populating a data model complex enough to reflect the bulk of the complexity of the corporate world but simple enough to support providing new clarity to the arena.

OpenCorporates provided admirable leadership in this realm, both in teaching what they had already learned about related data and in gathering more. Having already gathered extensive data on control networks and brand networks, they led an effort to research data availability in the realm of supply networks, reporting of which, generally speaking, is neither required nor standardized by governing authorities.

Their findings from Deliverable 5.2, Corporate Network Mapping Availability, provided vital context both for designing an appropriate data model and for setting forward the data strategy outlined above. That deliverable reviewed:

- Corporate structure information freely available from the company registers in all EU member states
- Sources for brand information
- Sources for supply network information

The report showed that information about corporate structures is not at all evenly available, and where it is available it is often behind a paywall, despite it being public information, and rarely available as open data. However, this is changing, and it seems highly likely that more sources will become available during the course of the project. In addition, several new sources for corporate structure information were identified, and are being added to OpenCorporates as part of this project.

For corporate supplier information, the report confirmed the necessity of this project, highlighting the lack of publicly available supply chain information. While there are encouraging trends towards increasing disclosure of suppliers, this has yet to translate into data on supply chains, and this affords the partners in this project an opportunity to change this situation.

At Huddle III in Cambridge, the consortium was able to use the partners' preliminary research and explorations to inform a much more concrete implementation plan, both in terms of functionality and content areas. The major target data arenas for initial focus included:

1. Industries. Much progress has been made on harmonisation of industry classifications over the past 20 or so years. However, for end users, as opposed to statisticians or data modelers, dealing with cross-border industry codes remains a challenge. For example, a typical query for companies is filtering by industry. However, doing this across jurisdictions is far from trivial, even though the principle is straightforward. In the EU, for example, all industry codes map to NACE. Things get more complicated when dealing with US companies, with NAICS or SIC the usual systems, neither of them mapping directly to NACE. Even within the EU, reporting will usually be at local industry classification level, requiring mapping to the relevant NACE code in order to do such filtering. We are currently working on the practical steps for mapping to shared industry code systems, such as NACE, NAICS and ISIC, allowing multi-level filtering between each, building on the work already done by OpenCorporates. These data will be valuable for their own merit and are also a necessary building block for WikiRate's implementation of metric applicability and ultimately transparency ratings (see T5.2). ST16 ("Open Source Industry Metric") is scheduled for completion in Q6.
2. Subsidiary Indicators. Several different characteristics of a corporation can play into subsidiary relationships: board appointment rights, ownership percentage, independent governance, etc. Such indicators will be represented as relationship metrics, and WikiRate users will be able to create nuanced definitions of such concepts as "subsidiary" by creating calculated metrics based on these data. Control networks are of great interest in their own rights and are critical to understanding supply networks. While some simple subsidiary indicators were used in the proof-of-concept implementation of relationship metrics (ST13) and several others will likely be in the coming production deployment of that functionality, we expect the population and use of this data to be an ongoing concern of both automated scraping and community research efforts.

Throughout the year, as a background to the headline strategic work on resolving key conceptual issues, OpenCorporates has been enhancing its corporate relationship data on the basis of a shared consortium understanding that this data will be of use in our representation of corporate networks in the near future.

Some of this work includes extending its data model. Specifically, OpenCorporates has been adding to corporate structure information, including further refining its extraction of data from SEC filings.

However, the most significant step has been the arrival of beneficial ownership data from UK Companies House. This is the first major release of public beneficial ownership data in the world, and is important for this project for two reasons: first, it contains corporate structure information (in most cases, listing the parent companies of companies controlled by other companies); second, as it contains both data and statements of null data (i.e. why there is no controlling entity), it has required some subtle data modeling in order to represent this accurately, in a form that is useful for end-users, and in a way that can be combined with other corporate structure relationship data. OpenCorporates have created the first data model for this, and are now importing this data daily as it comes online (UK companies now have to file this information at incorporation, or at time of their annual return).

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

The most challenging *design* aspect of establishing data requirements for T5.1 has been how best to connect the legal-filing-oriented “entities” on OpenCorporates (OC) with the brand-reporting-oriented “Companies” on WikiRate.org. While much of the work designing an integration solution fell under T5.1, we will report on the solution here, because that integration is the core *development* challenge of T5.2

After Huddle I we devised a visualization of the planned integration between all three ChainReact struts (see below). As illustrated, the “Chain” of information from citizen reporters to the broader public is, in its initial implementation, primarily a flow from The Whistle to WikiRate facilitated by deep integration with OpenCorporates, which provides vast resources for grounding the world of corporate perception in the world of legal filing. It is worth noting that this primary flow will not exclude many other related flows, including citizen reporting from sources other than The Whistle (which is particularly tailored to human rights reporting) and flows to the public through varied uses of OpenCorporates data.

As the Figure below indicates, the initial integration plan centered around mapping OpenCorporates “company groups” to WikiRate Companies. The notion that a *reporting* entity – eg the corporate brand behind a CSR report, and the primary representation of a company on WikiRate.org – is often comprised of many legal entities, the primary representation of a company on OpenCorporates.

The problem with this solution was that it hinged upon a precise resolution of the intentionally –and usefully – imprecise concept of the “corporate group”. For example, different metrics aimed at measuring the performance of the same corporate brand can be designed to investigate that company’s structure at different depths. Mapping a WikiRate Company to a fixed OpenCorporates company group removes that flexibility and requires extensive research and nuanced decision-making for every single mapping.

In the scenario discussions following Huddle II, we arrived at the new solution of mapping each WikiRate company to a single OC “headquarter” entity. WikiRate users would not typically play a direct role in defining company groups, and in fact would only need to engage directly with subsidiaries if that subsidiary itself were a significant enough “reporting” entity to comprise a separate WikiRate company.

That said, WikiRate users could engage lightly with OpenCorporate entities not mapped directly to a WikiRate Company an OC widget displayed prominently on WikiRate.org Company pages. Those pages would provide a brief overview of OpenCorporates’s data about the headquarter entity and its connections to related entities. Curious users will be able to explore the data more deeply by linking to OpenCorporates.com – and pull relevant reports/filings into WikiRate as sources. As noted in reporting on T5.5, while this widget has not yet been implemented on WikiRate.org, a new implementation of the Company page designed around the widget has. (See also SMART Target 15 “OC entities in use on WR.org”, scheduled for Q5.)

During this time, we also resolved a second major representational challenge: subsidiaries. Just as WikiRate would not require a single canonical corporate grouping, it would not require a single canonical definition of subsidiaries, either. Many different considerations might be weighed in deeming a company a subsidiary or not. And, in fact, “subsidiary” itself could be treated as a gradation. Therefore, WikiRate would represent relationships among corporations as metrics, and as with typical calculated metrics, indicators could be dynamically combined via algorithm.

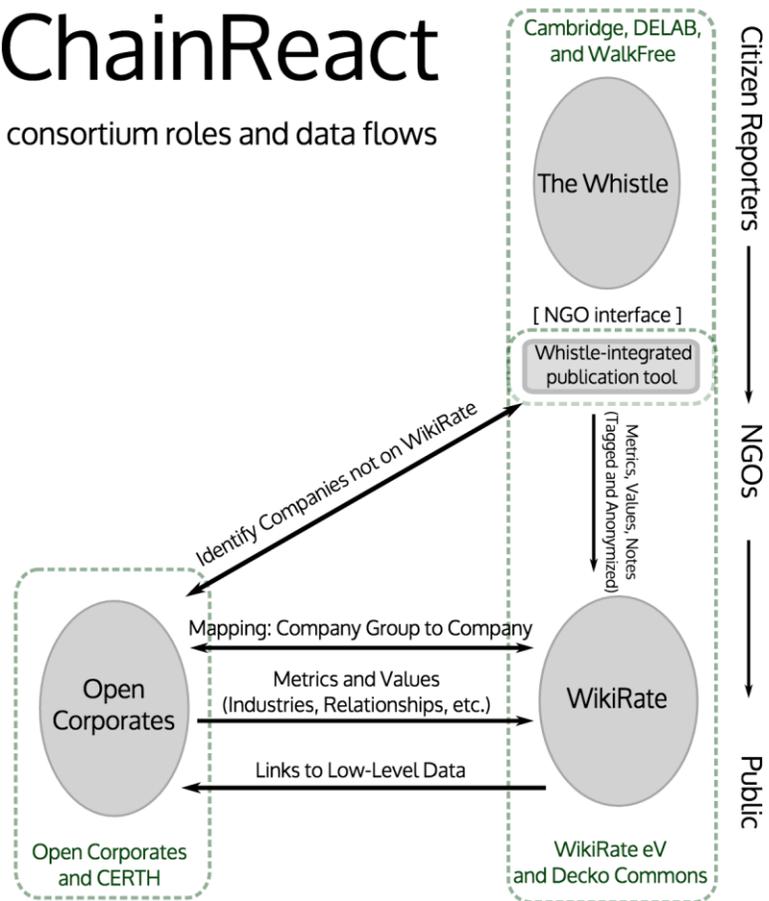
With this conceptual map, we were able to set forth appropriate SMART targets at Huddle III. Two targets centered primarily on T5.2:

- ST-12: By Q4 CERTH and WikiRate eV were to complete a ground truth dataset manually connecting 1000 WikiRate companies to OpenCorporates entities
- ST-14. Then, in Q5, the mappings will be done automatically with at least 80% accuracy

ST-12 was delivered in Q4 as planned and its results are provided in the *Results Delivered* section.

ChainReact

consortium roles and data flows



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

At the start of the grant, we did not anticipate beginning work on T5.3 until Month 13. However, the centrality of visualization to the “React” component of ChainReact (which, in its narrow data interpretation, suggests that ratings outputs should react to network inputs) led us to consider ways to visualize network relationships earlier than expected.

At Huddle III in Cambridge, it was decided that the first SMART Target directly involving changes to WikiRate.org interface would be a proof-of-concept relationship metric (ST13). A “relationship metric” is a mechanism for extending WikiRate.org’s Metric pattern in a way that allows it to represent relationships between companies. The target was to produce a small set of metric data based on actual viable data sources and to implement, on a demo site, an interface that makes this data navigable in a form compatible with other WikiRate metrics.

The advantages of representing this relationship include:

1. Integration of network data with other Metrics via Calculated metrics. **Objective 3, Subgoal A** is to “Support community development of network-aware corporate metrics on WikiRate.org”. The concept of “network-aware metrics” is built upon the notion that a company’s social and environmental impacts should be measured not solely by the individual company’s acts but also by those of its network. WikiRate’s calculated metrics will be able to use relationship metrics as part of a way of formulaically combining metric answers to generate ratings and other valuable performance measures.
2. Enabling the same capabilities for framing and scrutinizing definitions of corporate relationships that every other Metric includes. For example, there are many ways to define a “subsidiary”, and these different definitions may be of use in different situations. By treating “subsidiary” as a Metric, we support conveyance and debate of strict methodologies, distribution visualization, etc.
3. Visibility of timeline. The Metric data representation embraces the Year as a standard time increment to ease integration of data. By representing relationships as metrics, we also make it easy for users to follow our standard patterns (both current and future) of visualizing data over time

Using this data, Decko Commons implemented the proof-of-concept implementation and deployed it, as planned, to a demo copy of WikiRate.org. The effort was quite successful both in showing the utility of the data model and in flushing out potential challenges in interface design (by the WikiRate team) and metric design (by WikiRate contributors). While the eventual live implementation will involve some significant tweaks from the proof of concept implementation, we expect the core data model to stand and are boosted by the support that all we have learned from this exploration will provide to the next phase of implementation.

Given the success of the proof-of-concept implementation, it will be proposed at the upcoming Huddle IV (online) WP5 meeting that a new SMART target be introduced to plan for the full deployment of relationship metrics in Q7.

Annotated screenshots and links to software code for ST-13 are available below in the *Results Delivered* section.

T5.4 The Whistle Integration

As outlined in the *Data Strategy* section above, our original plan for ChainReact was to try to build, very early on in the project, an end-to-end proof-of-concept representation of a supply network. The alpha version, as planned, would have been capable of feeding preliminary reports from The Whistle into WikiRate, using preliminary data from OpenCorporates. WikiRate users could then produce preliminary ratings.

The conclusion reached after Huddles II and III was that this was a Garbage In – Garbage Out proposal. Worse it would divert resources from both The Whistle and the corporate networking effort to force a strained connection between endpoints before the endpoints themselves had been proven. Instead, with our new “two-path” approach, we will be working in parallel to insure that both the Whistle and the corporate network maps are robust and useful in their own rights. Then, in the latter half of the project, we will seek to connect the two to magnify their utility.

That said, we did make progress this year on a general integration model on which to build as The Whistle is developed. As early as our discussions at Huddle I in Cambridge, the consortium questioned the idea articulated in the grant’s original text that the open community processing tasks should to take place on The Whistle. Such community digestion makes more sense on WikiRate, because The Whistle is a closed system that optimizes security, while WikiRate.org is an open system that optimizes transparency. As important if not more, this new approach relieves the massive challenge of having to build a community around this narrow set of tasks, which already fall within the broader purview of the WikiRate community.

In the resultant model, therefore, the expectation is that reports that move from The Whistle to WikiRate will be fully sanitized for public consumption, and that the primary way in which they are differentiated from reporting that The Whistle might facilitate generating for other forms of publication is that the reports will be directly tied to a specific company. We expect that this publication mechanism will be comprised of a tool that brings data from all three ChainReact struts to one place, enabling NGO’s managing The Whistle campaigns to publish reports to WikiRate and tag them with specific companies with the support of contextual source material from OpenCorporates.

T5.5 Design and Implementation of WikiRate.org Features

In our first year, design of WikiRate features fell predominantly into three categories:

- Relationship metric design, most of which was captured under task T5.3 (see above)
- Updates to Company page layouts to ready the site for ST15 (“OC Entities in use on WR.org”)
- Technical design of scaling improvements to support the accommodation of relationship metric data

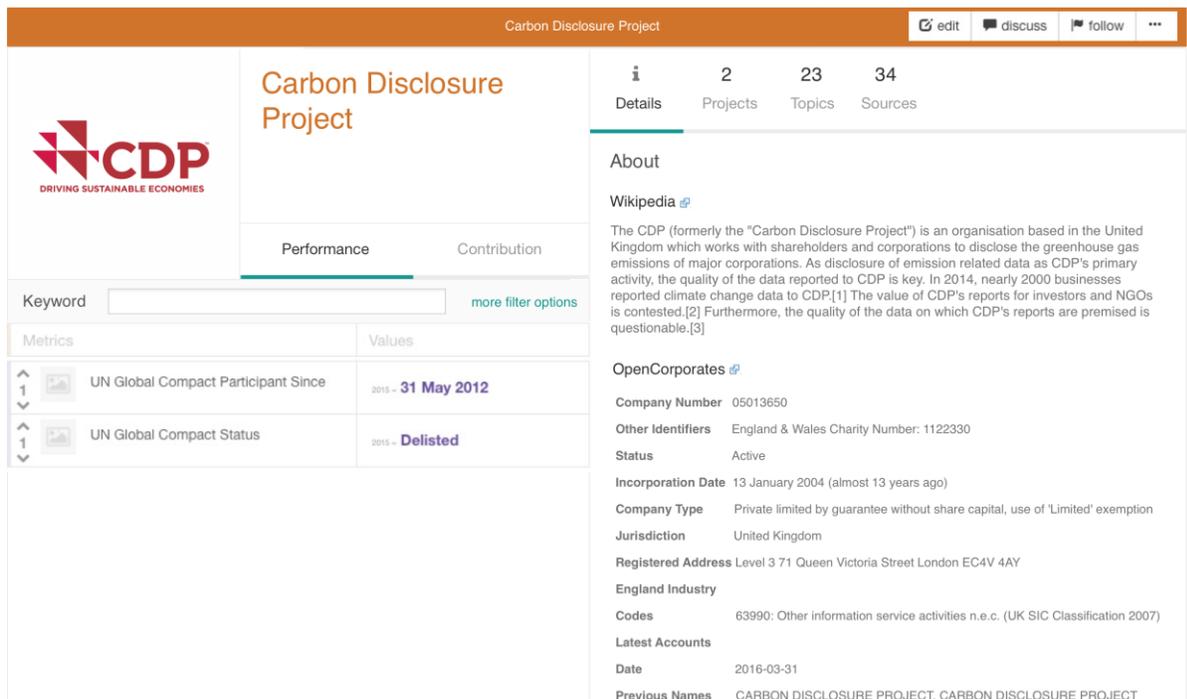
The second and third bullets are discussed below:

Company page layout

SMART Target 15 calls for the integration of OpenCorporates data directly into WikiRate.org pages. To that, Company pages were extensively redesigned to make it possible to feature OpenCorporates data prominently and thus promote its integration.

The new design was rapidly implemented, with the exception of the OpenCorporates integration, which will be introduced in Q5 (the target timeframe of ST15). The Wikipedia integration, which is now live, allowed a simple proof of concept for the use of a remote website integration in this primary WikiRate.org context.

In addition to creating space for the OpenCorporates integration, the new design also resolves the need for better visual integration of pages covering company performance (in the world) and company contributions (on the site), making clear that WikiRate can also serve to “rate the raters”. CDP, the organization used in the mockup shown below, is an example of a company that is both a metric designer and a metric subject.



Metrics	Values
UN Global Compact Participant Since	2015 - 31 May 2012
UN Global Compact Status	2015 - Delisted

OpenCorporates

- Company Number: 05013650
- Other Identifiers: England & Wales Charity Number: 1122330
- Status: Active
- Incorporation Date: 13 January 2004 (almost 13 years ago)
- Company Type: Private limited by guarantee without share capital, use of 'limited' exemption
- Jurisdiction: United Kingdom
- Registered Address: Level 3 71 Queen Victoria Street London EC4V 4AY
- England Industry
- Codes: 63990: Other information service activities n.e.c. (UK SIC Classification 2007)
- Latest Accounts
- Date: 2016-03-31
- Previous Names: CARBON DISCLOSURE PROJECT, CARBON DISCLOSURE PROJECT

In order to ensure website consistency, the design’s simplified tab structure was also implemented on Topics and Metrics; WikiRate.org’s two other core content areas.

Scaling and Performance

ChainReact expands the potential scale of WikiRate in three primary ways:

1. The investigation of supply networks will likely drive the scope of what companies are considered appropriate for representation on WikiRate to include much smaller entities and thus many more companies.

2. The addition of relationship metrics will increase quite considerably both the over-all quantity of data and the complexity of processing.
3. The planned outreach campaigns promise to increase site traffic considerably.

For those reasons, ensuring scalability and performance will remain a strong priority for WikiRate. Many of these considerations will fall to the Decko platform (see T5.6), but others will require the optimization of WikiRate-specific structures.

Decko Commons accelerated its efforts to improve WikiRate.org's website performance and scalability with two major WikiRate-specific optimizations:

- Answer lookup table (implemented and deployed), which dramatically increase the efficiency of queries pertaining to metric answers
- Cached count table (designed, not implemented), which will simplify data representation and speed up processing of count optimizations currently stored in cards

T5.6 Improvements to Decko Framework

Among the many improvements to the Wagn platform this year, two stand out as primary: view caching and layout API.

View caching.

To minimize re-rendering of repeated content, we began implementing in Q4 a new nuanced view caching mechanism that will provide foundational support to WikiRate and other Decko-based sites in scaling to large amounts of data and traffic. The mechanism embraces Decko's fractal patterns in nesting views and cards and is capable of caching that which is static, generating live that which is dynamic, and clearing the cache surgically in response to system changes. The new view caching system is currently in testing and expected to be deployed live on WikiRate.org in Q5

Layout API

Ever since the integration of the Bootstrap framework (<http://getbootstrap.com/>) into Wagn, Wagneers have been tempted to create custom layouts and views to make use of its capabilities. This customization, however powerful, pulls creators away from using shared patterns and thus complicates development and maintenance.

Therefore, the Wagn team decided to implement a new layout API that makes it easier to build complex cards that take advantage of Bootstrap features without generating an overabundance of custom code. Not only has the new API reduced code complexity, it has also reduced interface complexity by encouraging interface designers to reuse patterns. For example, Topic pages were a tertiary priority this quarter, but because tab patterns had changed on Company pages (see T5.5), and the layout API was minimal, it was straightforward to adapt Topics quickly to follow the same design patterns as Companies.

Other

Additional Decko improvements include:

- A new more object-oriented approach to card view nesting
- Generalized support for “solid cache” – search results stored as pointers
- Delayed job handing
- Updated documentation and reorganization of major Ruby Classes
- Harmonized link / path syntax
- Support for new wysiwyg editors (ProseMirror)
- Configurable remote file storage

T5.7 WikiRate System Administration

While WikiRate’s System Administration patterns are relatively well established at this point, we continue to gradually refine our deployment, testing, code scoring, style standardization, and code review mechanics.

Particular focus this year was given to deeper use of NewRelic, our application performance monitoring tool. For example, we personalized our NewRelic instrumentation of both standard requests and errors; the generic implementation uses only RubyonRails patterns, while our new instrumentation embraces Decko patterns. With the insights gained from these improvements we were able to dramatically increase the speed of many of our slowest pages.

We have also embraced NewRelic notifications more deeply, integrating those notifications with Slack, our primary team notification tool. Because relatively simple “performance bugs” like runaway queries or recursion problems can have a drastic negative impact on performance, we have increased the priority of our performance notifications and are now responding to such issues right away.

Decko Commons also streamlined its process for creating test databases. This was necessitated by the growth of the production database from which the test database is derived.

While performance improvements have made the shift less urgent, moving to a multi-server architecture remains a long-term goal. This year we began addressing the most critical issue standing in the path of this move, that of decoupling file service from application service. The low-level development platform requirements for this change have been completed, and all that remains is to make use of the platform’s capabilities by migrating files to a remote (cloud) server.

We also continued steady improvement on testing, deployment, documentation, and other infrastructural practices, all of which got broad attention thanks to the broad code changes involved in view caching and layout API upgrades.

T5.8 Application Programming Interface

Decko Commons' grant reporting convention has been to interpret "API" as referring to either of two application programming interfaces:

- Decko's "mod" API, the primary mechanism for engaging with Decko as a development framework. Mod developers use the API to build *mods*, short for *modules* or *modifications*. Improvements to this API empower mod developers with more, more powerful, or more easily understood tools for interacting with Decko cards
- The card REST API, which offers a standard pattern for creating, reading, updating, and deleting cards that can be customized for specific Decko sites (like WikiRate) with specific views and parameters

View definitions, one of the core functions of Decko's mod API and a central structure within the MoFoS (Model-Format-Set) architecture. The new view caching described above in T5.6 drove a major overhaul of our handling of view arguments. The new API is more approachable for developers and generates cleaner, more efficient code.

The card REST API also received design attention, particularly with regards to its application to retrieving company-metric answers in JSON and CSV format.

In the previous quarters (see note regarding previous quarters in T5.6), we expanded the application of the REST API for importing and exporting card data, reorganized mods extensively to make best practices easier to learn and embrace, and began generalizing wikirate metric functionality in a form better suited to eventually expanding its use to rating subjects other than companies.

Results Delivered

SMART Target 12 – Complete ground truth dataset

We approached the ST12 goal of developing a ground truth dataset by having CERTH developing a Web Interface to facilitate annotation by WikiRate. It provided information about the WikiRate company in an iframe element and a number of OC company candidates as possible matches to the query company. Additionally, a suggested match was provided based on a number of criteria. Users should provide the OC company number matched to the query company. In total, 1,000 companies were annotated from WikiRate and an additional 535 companies were annotated automatically based on their available CIK (Central Index Key) number in our database and OC database.

Next, CERTH worked with OpenCorporates API and they developed a basic methodology to query OC API and map WikiRate companies to OC. More specifically, the developed methodology was comprised of the following steps:

1. Query OC API based on company name and country of origin (if available) and order the results based on relevance (as defined by OC).
2. In most cases OC API returns more than one company entities. Thus, the similarity between WikiRate company name and OC company name is calculated (tf-idf similarity was used) and we select as a match the company with the highest similarity.
3. In case of equal similarities, the company with “branch_status”: null is selected as a match.
4. In case of multiple OC company entities having equal similarities and null values in “branch_status” field, the company entity that appears first in the result set is selected as match.

Then, the performance of the proposed methodology was evaluated by comparing the produced results over the ground truth dataset. The methodology did not perform as expected. More specifically, the achieved prediction accuracy was 49.05%. It is noteworthy that the achieved prediction accuracy was significantly higher in companies located outside of the United States over those located in the U.S.: 75.33% over 40.50%. The prediction accuracy was significantly increased when the (U.S.) state that the company is located was available. Overall, the achieved prediction accuracy was increased to 69%. Nevertheless, the lack of information on WikiRate companies makes the task of company mapping more challenging. To overcome this obstacle, CERTH is planning on extracting structured information about WR companies from already available CSR reports in order to enrich company info and facilitate the company mapping task.

As next steps, state-of-the-art methodologies on Relation Extraction will be studied and exploited to enrich companies with additional information such as headquarters location, subsidiaries etc. in order to achieve higher precision and recall.

Code behind ST12 is available at <https://github.com/MKLab-ITI/company-data-integration>.

SMART Target 13 – Proof-of-concept relationship metric

ST13’s development followed three phases:

1. Decko Commons proposed a data representation
2. A collaborative team with representatives of four partners prepared sample data
3. Decko Commons implemented pilot functionality and imported the sample data.

The conceptual space occupied by an Answer on WikiRate is the intersection of a Metric, Company, and Year. That is to say, a given metric, company, and year can have only one answer.

To represent relationships, we needed to add a second company to this list. The first data proposition involved allowing two kinds of answers – standard and relationship. When this approach ran up against abundant implementation and interface stumbling blocks, a more elegant solution was proposed: that of retaining the simple metric answer representation, but allowing each answer to a Relationship metric to be composed of multiple relationship answers.

The new proposal, including a specific card representation, was agreed upon, and a spreadsheet was structured for gathering pilot data and populating the necessary fields.

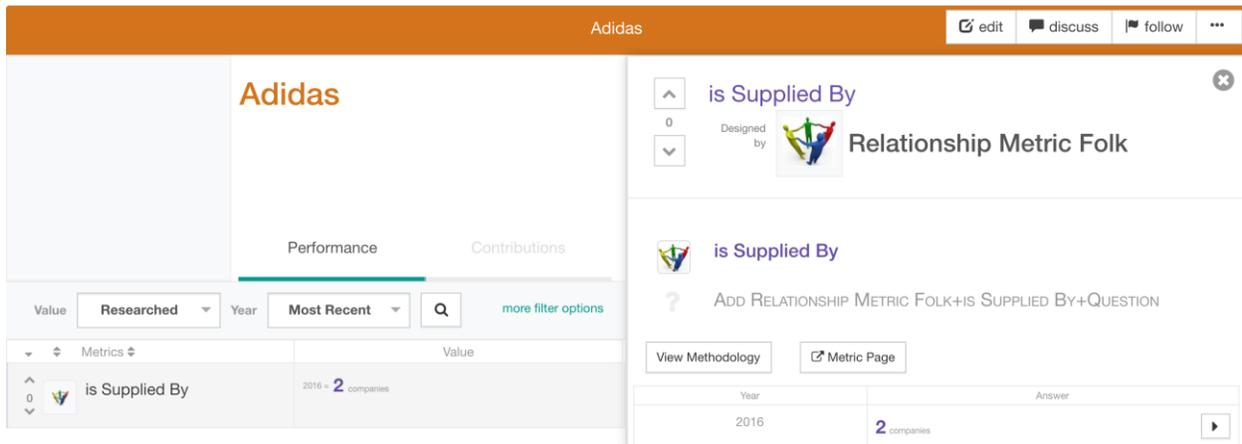
Pilot relationship metric data for ST13 were prepared by a team comprised of collaborators from WikiRate eV, CERTH, Cambridge, and Decko Commons on the basis of data availability and breadth of application. Metrics and their answers were designed and refined on Google Spreadsheets. The final sheets contained 8 relationship metrics (each with an inverse) and 63 relationship answers.

Here are the eight metrics and their inverses

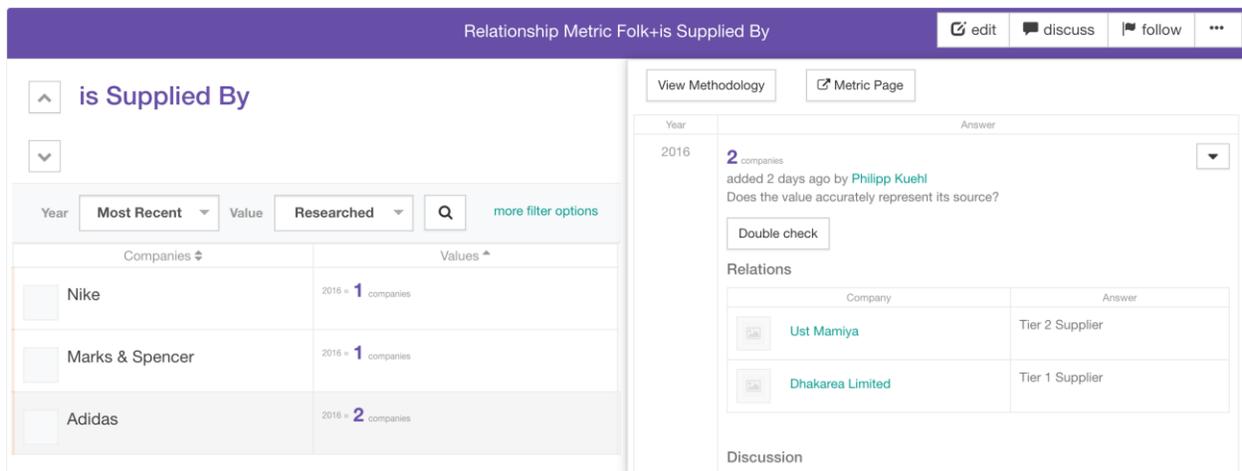
designer	title	inverse
Relationship Metric Folk	is Supplied By	is Supplier of
OpenCorporates	is Controller Of	is Controlled by
OpenCorporates	is Subsidiary Of	has Subsidiary
OpenCorporates	has Branch	is a Branch of
Proxy Voting Research Group	has voting authority over	receives shareholder votes from
Proxy Voting Research Group	has voting authority over (weighted)	receives shareholder votes from (weighted)
Proxy Voting Research Group	voted with management of	management supported by votes of
OpenCorporates	is Shareholder Of	has Shareholder

Note that the current proof-of-concept relationship metrics are not limited to supply relationships but also control and brand relationships.

Relationship Metric Folk is an illustrative name and was implemented as a Research Group. The following screenshot shows a metric designed by that Research Group as it appeared on a Company page on a demo copy of WikiRate.org. You will note that the metric, like any metric appears just once in the record list on the company’s left.



However, when an Answer of a Relationship metric is expanded, as in the following screenshot showing the same Metric Answer as it appeared on a Metric page, then multiple Relationship Answers are displayed. In this case, the user can view two different companies that supply Adidas: Ust Mamiya and Universal Menswear Ltd.



While full interface was not built for this proof of concept, each relationship already supports independent sourcing and discussion.

Code behind ST13 is available at <https://github.com/wagn/wikirate/pull/602>.

Live changes to WikiRate.org

In the *Progress By Task* section, extensive code changes to WikiRate.org are described, including

- T5.5 Company page layouts, altered in preparation for OpenCorporates widget integration (ST15)
- T5.5 WikiRate-specific performance and scaling improvements, including *answers* lookup table
- T5.6 Decko framework-level performance and scaling improvements, including new view caching API
- T5.6 New Decko layout API (improves efficiency of configuring layouts for Decko sites like WikiRate)
- T5.8 Cleaner view definitions

All of these changes have been deployed live at <http://wikirate.org>.

WikiRate-specific code (T5.5) is available at <https://github.com/wagn/wikirate>.

Shared Decko code (T5.6 and T5.8) is available at <https://github.com/wagn/wagn>.

4. Conclusion

In the original grant, we wrote:

There is also a chicken-and-egg quality to the challenge of making corporate networks sustainable. On the one hand, there is often little incentive to gather data about the smaller companies at supplier networks' edges, because the data are seen as worthless unless connected to brands familiar to western markets, and that requires network maps. But those who begin to work on network maps find that they are generally making connections between more and more companies about which there is very little data.

Having researched corporate network data more deeply, we certainly feel confirmed in our assessment of the nature of data challenges. We also feel chastened in having set forth a strategy that did not sufficiently respect this complexity and placed the goal of end-to-end data mapping too early in the process.

That said, the first year of work on WP5 has produced several significant successes, including:

1. A deeper exploration of the current state of corporate network data in D5.2
2. The adoption of a new, more practical strategy for achieving end-to-end corporate network maps
3. The generation of SMART targets that will drive success in that strategy, including two SMART targets scheduled for completion in year 1.
4. The completion of a ground truth dataset mapping WikiRate companies to headquarter OpenCorporate entities (ST12). This dataset is now supporting the development of automated mapping (ST14, due in Q5).
5. A successful proof-of-concept implementation of relationship metrics (ST13).
6. The implementation of new company pages on WikiRate.org in preparation for widget integration with OpenCorporates (ST15, due in Q5).
7. Multiple performance and scaling improvements to Decko and WikiRate.org.

Another success not explicitly noted elsewhere in this report has been the rich and energetic collaboration of all partners involved. In the original grant, immediately following the passage quoted above is the following:

Here is the deep value of the ChainReact proposal: by bringing together a savvy consortium to focus on three platforms at once, we can much more efficiently tackle a multi-faceted challenge that would defy a more piecemeal approach.

After a year, we feel confirmed in this assessment as well. These corporate network maps comprise a significant technical, social, and conceptual challenge, and we are happy both to have the opportunity to address them and to report that our progress toward that end is substantial.