HYBRID CITY 2015

DATA TO THE PEOPLE

PROCEEDINGS OF THE 3RD INTERNATIONAL BIENNIAL CONFERENCE

EDITORS

I. THEONA

D. CHARITOS

UNIVERSITY RESEARCH INSTITUTE OF APPLIED COMMUNICATION

Street-level City Analytics: Mapping the Amsterdam Knowledge Mile

Sabine Niederer
Citizen Data Lab
Amsterdam University of Applied Sciences
Amsterdam, The Netherlands

Gabriele Colombo, Michele Mauri, Matteo Azzi
DensityDesign Research Lab, Design Department
Politecnico di Milano
Milano, Italy

Abstract. This paper presents digital methods for city analytics, applied to the mapping and activation of an urban area in the city of Amsterdam called the Knowledge Mile. Firstly, we map companies registered in the area and analyse their connections through on-line hyperlinking. Secondly, we use Instagram, Panoramio and Google Search data to map most-shared photos and high-ranked images of the area. Lastly, we use Foursquare data to map most-shared locations. The produced maps visualize the online presence and resonance of an urban area that is an axis cutting through the city center and crossing many district and neighborhood 'borders'. The maps have been used as navigational tools and conversation pieces during workshops and participatory design sessions with local stakeholders.

Keywords: digital methods, data visualisation, social media data, city analytics.

I. Introduction

This paper presents digital methods of city analytics, applied to the mapping and activation of an urban area in the city of Amsterdam, Weesperstraat and Wibautstraat, which is currently under development for the campuses of two major universities -the University of Amsterdam and the Amsterdam University of Applied Sciences- which will house thirty thousand students in 2018. Once declared the ugliest street of the Netherlands, the Wibautstraat and Weesperstraat still face major metropolitan challenges. The streets are a highway, cutting through the city, leading to traffic and mobility issues and poor air quality. At the same time, the area is undergoing a major transformation through the development of these campuses. Besides the universities (Amsterdam University of Applied Sciences and the University of Amsterdam), the area is increasingly populated by a diversity of users and local initiatives: shopkeepers, multinationals, underground, startups, clubs, municipal services, citizen initiatives, media companies, new media startups, museums and cafés.

This Amsterdam area, stretching from Amstelplein (the Amstel Square right next to Amstel Station) to Nieuwmarkt (a market square in the city centre) has recently been announced the Knowledge Mile, providing an ambitious local framework for bringing together citizens, research and education, and public and commercial parties based along this urban area[i]. The Knowledge Mile also functions as urban area for research into smart solutions for urban problems. For international companies and organizations, the Knowledge Mile offers an opportunity to develop, test and demonstrate such applications. In this paper we present a mapping of the Knowledge Mile, that makes use of online data to characterise the area through its 'important' places. The research follows the logic of specific web platforms and online networked content, and makes use of digital methods for social and cultural research with the web [1].

Critical views on research with web data highlight its dependency on already problematic proprietary walled gardens, and otherwise volatile ever-'innovating' commercial web platforms, such as Facebook and Twitter. Scholars particularly warn of the sheer impossibility of distinguishing between the working logic of web platforms and the exemplarity of 'platform artifacts' [2][3][1]. For example, the most 'retweeted' content on Twitter may just be the most Twitter-friendly, rather than the most relevant content. Therefore, in the example of Twitter analysis, we may only be finding out more about the logic of the platform itself, rather than the issue under study or the eventfulness of a certain tweet [2].

When dealing with web content, it has therefore been argued that researchers need to take into account the socio-technical logic of the platform as part of the analysis of its content[4]. In fact, with the explosive rise of (big) data, attention to these socio-technical logics of platforms must be further prioritized, as social research increasingly makes use of what is called 'live research' [5][3], where masses of content (with specific forms and technicities) are aggregated in real-time, copied onto other networks, and archived across the (social) web. Furthermore, data analysis and the tools that enable this are built on dynamic web services. In a critique of the famous Google Flu Trends project, David Lazer writes how Twitter, Facebook, Google, and the Internet more generally are constantly changing because of the actions of millions of engineers and consumers [6]. Understanding and studying these platforms as socio-

[i] See also: www.knowledgemile.org

technical systems for what they are, is of utmost importance, as they are “increasingly embedded in our societies” [6].

More specifically, one ‘technicity of content’ [4] of importance to our analysis is the way in which platforms organize the geo-location of their content. The mapping presented in this paper consists of three sub-projects, each with different online data sets.[ii] The overall aim as mentioned before, was to characterise the Knowledge Mile, by identifying its important places (in three ways). The maps then function as a baseline measurement for further annual research as well as a conversation piece used in participatory design sessions with stakeholders of the neighborhood.

Sociologist Tommaso Venturini, when discussing controversy maps, has described social maps as visual interface to complex issues: “To be of any use, social maps have to be less confused and convoluted than collective disputes. They cannot just mirror the complexity of controversies: they have to make such complexity legible” [7]. Similarly, visualization of data layered onto a geographic map of an area, should render legible the complexity of the area, as well as the ways in which the social media platforms it takes its data from deal with geo-location. At the same time, map-based visualizations have been criticized for their oversimplification and reductionist approach to vast and multifarious data, highlighting some information and obfuscating other data for the sake of “creating an image that can be read and judged” [8] or constructing a “top-down narrative” [9]. We would like to stress that in this project and our related research, the practice and objects of mapping are not efforts to ignore the distributed nature of today’s technologies or data, but in fact to gain a better understanding of the many threads that weave these complex patterns of competing technologies as these intertwine form and content. We try to accomplish this here by zooming in on a strictly demarcated geographical location and studying this area across multiple platforms from different viewpoints and by creating not one all-encompassing *mother map*, but a series of different maps that underline both the complexity of studying places through online content and the entanglement of content with its technicity. These maps then function as a navigational tool, rather than a reductionist narrative, and are presented to local stakeholders in workshops and collaborative design sessions.

II. Analyzing the Urban Area Through Web Data and the Question of Geo-Location

In recent years, several studies have focused on the digitally encoded metadata (such as pictures on Instagram, taken with smartphones, geo-tracks, check-ins) to provide novel views on the city [10][11]. While in this project we mainly used this kind of geographic encoding, we expanded the analysis to web entities not geo-coded with standard formats, identifying methods for their geo-location, which we will discuss later in this paper. As discussed in the introduction, the question here was to retrieve the most important places along the knowledge mile. This was then assessed by looking at the most-networked, the most iconic (visually) and the most acknowledged (or shareable). Retrieving the ‘most important places’ then was operationalized in three ways (which eventually resulted in three maps). Firstly, we defined as most important those places most referred to by other actors in the area. And here we conducted a hyperlink analysis to see which actor was most linked to by other actors in the area. (When we say actors we mean people, organisations, companies, etc.). Secondly, we defined most important places as those most shared and acknowledged by visitors and other local actors, by way of checking in on the Foursquare social app. Thirdly, we defined most important places as those most photographed and shared on Instagram and Panoramio, and those images resonating most in Google Image search results. The base layer for all the three maps is a traditional cartographic representation of the area, created here with Openstreetmap data, showing streets, cycle lanes, buildings, canals and green areas. As the resulting maps were thought to be used in a collaborative session, and watched by multiple people, we decided to design them adopting a 2.5mt long format.

In the data collection process we soon faced a first methodological issue, concerning how to locate web data in a particular geographic space. Previous studies express the need to move “beyond the geotag” when performing geoweb analysis, suggesting, among other techniques, to integrate data from sources not explicitly geographic oriented and to integrate the analysis with official sources data such as census data or news reports [12]. In this study, what has been demarcated as the Knowledge Mile is in fact a diverse area made up of different streets crossing various urban areas. In addition, the data sources identified for the analysis (platforms such as Foursquare or Instagram but also the Chamber of Commerce archive) use very different ways to embed geographical informations into their entities. Therefore different techniques to demarcate the area of interest were needed, based on data sources specificities

[ii] In previous studies we have seen how different social platforms may be used to provide different views on a city. For instance, in a comparative analysis of geo-tagged social media data, we found that Pinterest posts, or ‘pins’, provided a distinct view on ‘boutique’ Amsterdam. The photos we retrieved were mostly taken and tagged on the smaller streets and up and coming areas of Amsterdam, with small bars, art and design venues and original fashion stores. In contrast, the platform Meet-up presented a view on Amsterdam that was more tech-driven as well as athletic (depending on the respective meet-up group), highlighting the parks, sports venues and also the tech-friendly bars in the city of Amsterdam. (See also the project page of the ‘City as Interface’ project of Summer School 2014. URL: https://wiki.digitalmethods.net/Dmi/TheCityAsInterface)

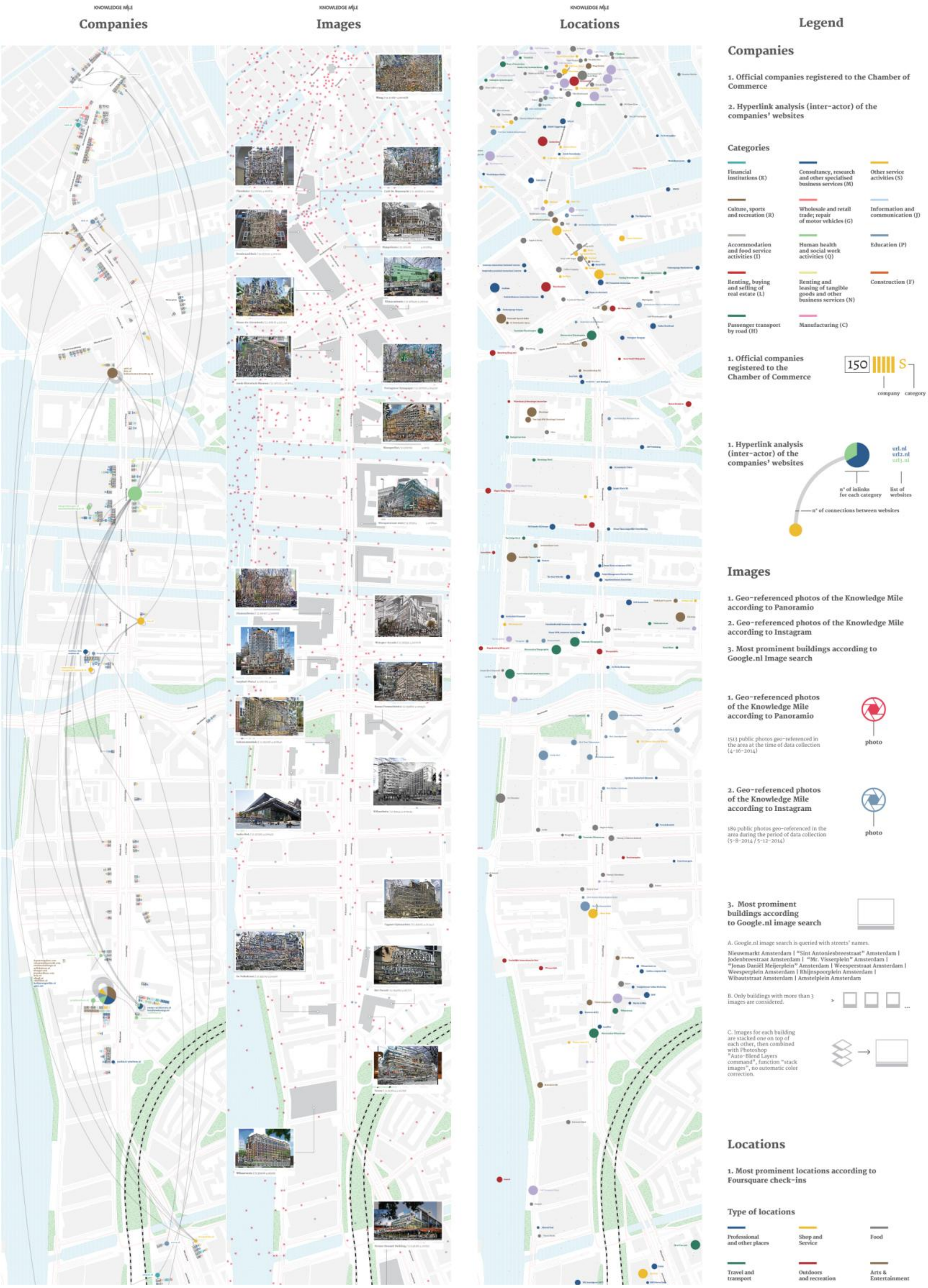

Fig. 1. The three maps and the legend. From the left: A) companies map, based on chambre of commerce data and websites. B) Images map, showing Flickr and Panoramio images, as well the first images resulting on Google.nl. C) Locations map, showing all the places found on Foursquare and the number of check-ins. D) Maps' legend.

but still allowing the collection of comparable datasets. We therefore combined three methods for geo-locating web entities on the Knowledge Mile. For the first map of the Knowledge Mile, we used the Chamber of Commerce database to retrieve a list of registered companies on the area. Here companies' data are geo-coded with street names and postal codes. We identified street names on and around the Knowledge Mile and used them to query the database and collect the names and addresses for those companies and organisations registered along the Knowledge Mile. For Instagram, Panoramio and Foursquare we collected data from their application programming interfaces (APIs), using a geo-coordinates 'bounding box', which means drawing a rectangle around the Knowledge Mile, considering all geo-coded data within that geographical area and collecting the exact longitude and latitude for each image. Lastly, we queried the same street names on the local image search engine to collect images related to the different streets of the Knowledge Mile, then manually geocoding all buildings found in the search results.

III. MAP #1: HYPERLINK NETWORK ON A GEO-MAP

The first map (Fig. 1A) provides an overview of all the companies in the Knowledge Mile area, allowing both to plot their physical presence onto the map and to analyze their online resonance. It is therefore composed by two layers. In the first layer, all companies listed in the Chamber of Commerce's Trade Register are depicted as a small square, the colour represents the commercial category as listed in the Trade Register. On the map, near each house number all the companies are listed, divided by category. With this first information layer, the user is able to see how many companies are present in the area, where there are more, if there are particular geographic clusterings of them. The second layer of information is provided by the web data, using the companies' URLs as listed in the Trade Register. As the data proved incomplete, we manually searched for each listed company on the local version of Google[iii], collecting the official websites if available. These were the starting points to be used in the IssueCrawler tool for hyperlink analysis. With the IssueCrawler we could then performed so-called 'inter-actor analysis', as the tool in the inter-actor setting "captures the starting points' outlinks and shows inter-linking between the starting points only.[iv v]

Each website was then geo-coded, according to the company's address[vi]. Using Gephi, a network analysis

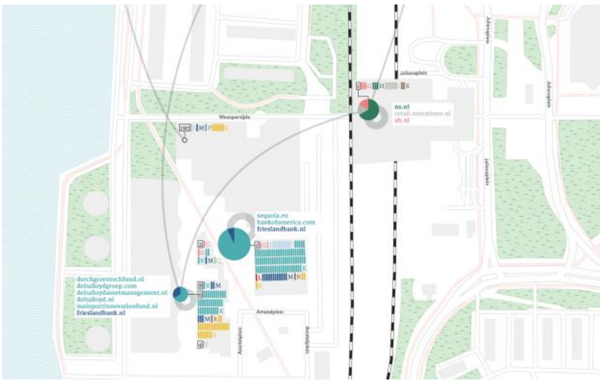

Fig. 2. Detail of the first map.

software[vii], we visualized the network, and with the "geolayout" plug-in we geographically dispersed it: each website has been placed on the geographical map, exactly where its company is physically located. With these operations, it has been possible to geolocate the web, and to perform a hyperlink analysis 'grounded' on the geographic level. For addresses with multiple companies registered there, the number of available websites is represented through a pie chart, each wedge representing a commercial category (as listed in the Trade Register).

From the Trade Register data plotted onto the map it is possible to see that commercial activities (i.e. companies) are not uniformly distributed along the Knowledge Mile. There are three main clusters within the area: small commercial activities close to the center (Nieuwmarkt area), a group of cultural and sports-related venues in the middle (Trouw and Volkskrant area), and finally a concentration of financial activities (around the Amstel Station).

The network map shows that interlinking between organisations registered along the Knowledge Mile occurs mostly within sectors, news organizations link to news, although the university does link to museums for instance.

IV. MAP #2: MOST-SHARED & HIGH RANKED IMAGES ON A GEO-MAP

The second map (Fig. 1B) returns an overview of photo sharing activity in the area, showing patterns in the amount of images shared online. Furthermore, it presents an aggregate view of the most depicted buildings, looking at which images are found online. Each photo shared on Instagram and Panoramio is plotted on the map based on its geo-coordinates. Blue glyphs represent Instagram photos while red ones represent photos shared on Panoramio. Users are then able to spot areas with intense photo sharing activity, as well as blank spaces where there are few or none image shared. On top of this quantitative overview a second layer of information is added, using data retrieved from the local search engine. Querying streets' names returns

---

[iii] URL: www.google.nl

[iv]URL: http://www.govcom.org/Issuecrawler_instructions.htm.

[v] For a discussion of the IssueCrawler software, see also Bruns, 2007.

[vi] We then fed the information in the the openstreetmap "Nominatim" database#, translating them in geo-coordinates.

[vii] The used software is an open-source and widely used platform for networks analysis and visualization (see http://gephi.github.io/)

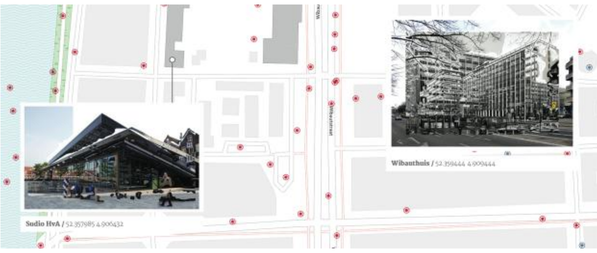

Fig. 3. Detail of the second map.

roughly 200 images per query. Images are then downloaded with a web browser plug-in[viii] and archived locally in a single collection of images. Images are manually tagged and only the ones depicting buildings in proximity of the Knowledge Mile are retained. For this analysis we only consider buildings found in the dataset more than 3 times. Images are then stacked one on top of each other, and visually combined with an 'image blending' algorithm that automatically merge images together[ix]. The generated composites are then positioned on the map in close proximity of the geographical location of the respective buildings. These algorithmically generated stacks present a merged view of the same building, showing from which angle each building is most commonly perceived and which type of image is more often used to depict it (computer generated renders, historical footage, or professional photos).

With this map it is possible to see that pictures are taken mainly in the city center, and the amount decreases when moving from it. Instagram images are mainly disposed along canals, probably due their photogenicity. Finally, the algorithmically generated images allow to observe the most significative perspectives on buildings, as well as the historical stratification of building imagery. Photos taken from the same viewpoint generate less fragmented composite versions, allowing to distinguish those buildings always observed from their facade (e.g. The Film Academy or Studio HvA, within the University premises) from those who are captured from a wide range of angles and distances (e.g. the Waag building, in the city center).

V. Map #3: Most-shared locations on a geo-map

The third and last map (Fig. 1C) represents the most shared locations on the platform of Foursquare. We collected all the places in the area listed on Foursquare and their total number of check-ins. Each venue is then visualized with a circle sized according to the number of

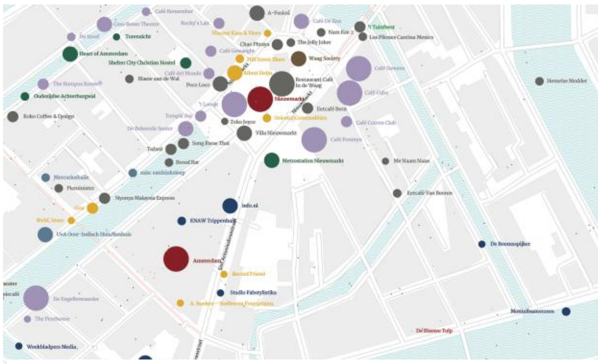

Fig. 4. Detail of the third map.

check-ins[x]. With colours we represented the categories of Foursquare venues, in order to identify possible local clusters (e.g. nightlife zone, offices blocks). The Foursquare data provide a different view on the area. The main nodes of urban transportation (train and metro stations, bus stops) become more visible, as well as supermarkets and bars. Also, as the platform allows to create multiple venues, biggest activities (such as, the HvA university) are split into their parts producing a more detailed view of the one resulting from the Chambre of Commerce data.

These Knowledge Mile maps show the online presence of an urban area under development that is as diverse as one would image from an axis that cuts through the city center and crosses many district and neighborhood 'borders'. The network map shows that interlinking between organisations registered along the Knowledge Mile occurs mostly within sectors, although the university does link to museums for instance. In the participatory design sessions, people were rather surprised to see how marginal they were on the map (a natural response, if you will, to being confronted with maps about one's own online presence), or how divided the sectors actually were. The staleness of the Chamber of Commerce data provided a problem, as nobody updates their URLs in the database, which is why we decided to organize fieldwork to collect the current URLs and social media usernames or pages from door to door, so we have a richer data set for further research.

VI. Towards participatory design: Conclusions and further research

The Knowledge Mile maps represent different online data sets of a geographic area, by using different methods of geo-demarcation, data analysis and visualisation. First, by geolocating addresses coming from administrative databases, we showed the density of and the connections between companies registered in the area. Secondly, using natively digital geo-coded

[viii] Mozilla Firefox add-on "DownThemAll! https://addons.mozilla.org/en-US/firefox/addon/downthemall/

[ix] Adobe Photoshop Auto-Blend Layers command, function "stack images", no automatic color correction. The tool is meant to combine multiple images of the same scene in order to obtain a better composite version of them. It works masking out over or underexposed areas and it is here repurposed as visual analysis tool.

[x] After a first data analysis, we clustered the amount of check-ins for each venue in six data ranges and we used a non-linear scale to visualize them (from very small to very large). This scale was adopted in order to enable the depiction of smallest elements. Venues with less than 200 check-ins are simply represented with a small dot, without citing the venue's name.

objects, such as Foursquare checkins and geo-tagged photos, we layered the social media view of the area. Finally, querying names streets in the dominant search engine, we collected the online image of each street. Each layer offered a methodological exercise in rethinking geo-location based on the specificity of each platform and the technicity of its content. What is relevant in such methods is the ability to layer the online activity on top of the map of the actual geo-location. The Knowledge Mile maps show the online presence and resonance of an urban area under development that is as diverse as one would image from an axis that cuts through the city center and crosses many district and neighborhood 'borders'. The layers on the map each represent different concepts of importance or relevance: connectedness, well-visited or often-depicted and perhaps even visually iconic.

By presenting the data that we retrieved online in a visualized form, we enabled local stakeholders (such as inhabitants, organizations, government officials) to assess and evaluate their relevance and resonance in the area, and the relationships with the other stakeholders. The network map shows that interlinking between organisations registered along the Knowledge Mile occurs mostly within sectors, with some exceptions. When presented in participatory design sessions, the participants were rather surprised to see their resonance, be it positively (in the case of the municipality) or negatively, for actors only marginally present on the map (a natural response, if you will, to being confronted with maps about one's own online presence). Another confronting view was the division between sectors, with barely any interlinking between types of organizations. Participants recognized the staleness of their records in the Chamber of Commerce Trade Register, as nobody updates their URLs in the database after registration. This is why we decided to organize fieldwork to collect the current URLs and social media usernames or pages from door to door, to have an up to date and rich data set for further research. This further result will culminate in the Knowledge Mile Atlas, an atlas offering input for participatory design sessions as well as documentation of the development of this urban area, through its online resonance in web content.

ACKNOWLEDGEMENT

The research for these mappings was conducted during a data sprint in 2014 at the Amsterdam University of Applied Sciences, in a highly motivating setting that formed the basis of the Citizen Data Lab, which was founded later that year with support of the Amsterdam Creative Industries Network (ACIN). ACIN director Matthijs ten Berge conceived of the Knowledge Mile as an area for urban development and research, and we thank him and his colleagues from the Knowledge Mile crew, including Ruurd Priester and Maarten Terpstra, for the fruitful collaboration in presenting the research results to local stakeholders in participatory design sessions. The collaboration with Density Design Lab is crucial to the data collection and visualization of this research. The tools used for the data analysis have been developed by the Digital Methods Initiative at the University of Amsterdam, Dept. of Media Studies.

REFERENCES

[1] Rogers, Richard (2013). Digital Methods. Cambridge, MA: MIT Publishing.

[2] Marres, Noortje (2015). Why Map Issues? On Controversy Analysis as a Digital Method. Science, Technology, & Human Values, online first, DOI: 10.1177/0162243915574602.

[3] Marres, Noortje and Esther Weltevrede (2013). Scraping the Social? Issues in real-time social research. Journal of Cultural Economy (6)3, 313-335.

[4] Niederer, Sabine and José Van Dijck (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. New Media & Society (12)8, 1368-1387.

[5] Lury, Celia (2012). Going live: towards an amphibious sociology. The Sociological Review 60:S1, 184-197.

[6] Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani (2014). The Parable of Google Flu: Traps in Big Data Analysis. Science 343, 1203-1205.

[7] Venturini, Tommaso (2012). Building on Faults: How to represent controversies with digital methods. Public Understanding of Science 21(7), 796-812

[8] Lovink, Geert. A critique of mapping. Forthcoming.

[9] Reichert 2015.

[10] Manovich, Lev et al. (2014). Selfiecity. URL: www.selfiecity.net.

[11] CityMurmur. URL: www.Citymurmur.org.

[12] Crampton, Jeremy W., et al. "Beyond the geotag: situating 'big data'and leveraging the potential of the geoweb." Cartography and geographic information science 40.2 (2013): 130-139.