EUROPEAN JOURNAL OF MEDIA STUDIES
www.necsus-ejms.org
Published by: Amsterdam University Press

The Third Party Diary

Tracking the trackers on Dutch governmental websites

Lonneke van der Velden

NECSUS 3 (1): 195-217
DOI: 10.5117/NECSUS2014.1.VELD

Abstract

This article discusses how the browser plugin Ghostery contributes to a particular understanding of contemporary consumer surveillance by making Web tracking transparent. The Tracker Tracker is a digital methods tool that, by following Ghostery, detects trackers on specific sets of URLs. It was used to examine all the websites of the Government of the Netherlands on a regular basis. Ghostery also invokes a particular informational genre which has an effect on how we understand the issue of Web tracking. The use of such a tool therefore raises a question: what happens when we repurpose an 'issue device' as 'research device'?

Keywords: consumer surveillance, cookies, digital methods, traces, Web tracking

Introduction: Web tracking as data and Web tracking as issue

There are a range of 'privacy enhancing' tools on the Web. In this article I will discuss how the browser plugin Ghostery transcends individual usage. By making Web tracking transparent it empirically and conceptually contributes to a particular understanding of contemporary consumer surveillance.

Ghostery detects techniques (called 'third party elements') that collect data on Internet users when they visit certain websites; Ghostery also gives the user an alert with a small visualisation in the Web page. The fact that Ghostery has specific detection principles makes the tool useful for Web researchers as well. Building upon the work of the Digital Methods Initiative (DMI) which specialises in repurposing Web devices for research I have

explored the 'Tracker Tracker'.[1] The Tracker Tracker mobilises Ghostery's capacities for the study of third party elements on specific sets of URLs. In this way it enables the comparison of the presence of third party elements in a more systematic manner.

In my case study I used this tool to look more closely into a sample of Dutch governmental websites in 2012. The reason for doing this case study was twofold. First of all, online tracking by Dutch governmental websites was controversial at the time. There was discussion about the Dutch implementation of the EU e-Privacy Directive and the extent to which the Dutch government was still tracking Internet users without their consent, hence failing to obey the law. My question was whether it was possible to measure the governmental response to this debate by using the Tracker Tracker to map the presence of third parties on governmental websites over time. The results pointed to an average of almost 60% presence of third parties and indicated that the government responded only slowly, if at all, to the affair. The results also showed clusters of websites sharing similar third parties. This raises questions about the way governmental websites perform different roles online; in addition to their expected and visible role as the main public service providers they also have an active role in contributing to the information economy by sharing (personal) data with major corporations.

A second reason for using the tool in the context of a particular Dutch local affair was that it was a way of 'situating' Digital Methods. This should be seen as a more experimental attempt to discuss how the Tracker Tracker tool performs in relation to a particular data set. Some of my results made me think about Ghostery's method of working and its capabilities, an issue that links up to wider academic debates about the increasing role of digital devices in social research.[2] The Digital Methods program mobilises digital devices explicitly for knowledge production. However, as Marres & Weltevrede argue, devices come with 'epistemology built in'.[3] This subsequently also raises questions about the politics of knowledge that these devices bring along, questions that a variety of digital methods researchers are currently examining.[4] For example, Marres has questioned the kind of methods that are remediated by Web devices and how that affects the work that comes out of the research assemblage in which these devices participate.[5]

Ghostery also lends itself to a more in-depth inquiry; it is an example of a device that brings Web tracking into view in order to make Internet users aware of the fact that their browsing behaviour is being monitored. That means that Ghostery is implicated in a particular issue and uses a specific

repertoire to explain what Web tracking is about. Therefore an important question arises about what way Ghostery brings this issue to the fore.

Digital devices in action

This article has a running concern with how Ghostery brings Web tracking into view and what that means for the way it participates in the research project. According to Gitelman & Jackson, data are often imagined as being picked up from some 'undifferentiated blur'. In many discourses data are talked of as being 'collected', 'piled', or 'mined'. However, as these authors go on to argue, data always depend on operations of knowledge production. Data, as they quote Lev Manovich, do not just 'exist' but need to be 'generated'.[6] In *Raw Data is an Oxymoron*, Gitelman & Jackson aim to pursue the question of 'how different disciplines have imagined their objects and how different data sets harbor interpretative structures of their own imagining'.[7]

When using Web devices for research a reframing of this concern would be a need to consider how these devices imagine data and how this feeds back into our data sets. The specific use of the term 'device' by Ruppert & Law & Savage is useful here. They state: '[w]ithin these cascades [of applications and software] a device can make, compile and transmit digital data and/or remake, analyse and translate data into information and interventions.'[8] They stress the organisational activity of devices in which both knowledge and social action get distributed. By doing so devices are constitutive of emergent social relations. Similar to the performativity of devices of the social sciences and economics,[9] say Ruppert et al., digital devices 'enact' the social. They 'inscribe' something into the very thing they attempt to analyse. This is a reason for them to say that key to what we as digital researchers ought to do with regard to digital devices is to get close. That is, to

> get our hands dirty and explore their affordances: how it is that they collect, store and transmit numerical, textual, aural or visual signals; how they work with respect to standard social science techniques such as sampling and comprehensiveness; and how they relate to social and political institutions.[10]

As I hope to illustrate, Ghostery proves to be a good opportunity for such an exploration. I will look at the context in which it operates, its method, assumptions, affiliations, and suggestions for actions, and how that is constitutive for the issue of online tracking. In line with other work in science

and technology studies (STS) I will look at the 'situated, material conditions of knowledge production'.[11] In other words I will first approach the device as an 'object' of study before repurposing it as a 'method' for research, a distinction made in the work by Marres & Weltevrede.[12] Another way of putting it would be that this is an investigation into a 'device in action'. By setting the study up in this way there will be several instances in which the generation of data is made explicit. I discuss how Ghostery imagines data, how the output of the Tracker Tracker tool shows in what ways third parties get their data, and how I treated the data set myself. In all these moments I try to show how data is organised through different formats and how these formats, in the context of the case study, interact.

Getting close to Ghostery

Ghostery operates in the context of a data market in which website optimisation coincides with behavioural advertising. Webmasters make use of corporate tools to keep track of their visitors and often share the data with third parties, for example advertising networks. As McStay explains: '[b]ehavioral advertising tracks users' browsing activities between websites over a period of time for the purposes of serving advertising tailored to what advertisers assume are users' interests.'[13] These assumed interests are extracted from the type of websites and other indicators of browsing behaviour (such as location, time, type of device, etc.). After the data are collected, stored, and aggregated, profiles are sold at real-time biddings. Advertisers can bid for advertising space delivered to specific users – the more detailed the profile the higher its value.[14] Just as in the 'regular' financial sector this market comes complete with 'data brokers' and 'data speculation'.[15] To characterise the culture of data trade metaphors such as 'Data Wild West' circulate among marketers themselves as well as among their critics.[16] For individual users it is not easy to know what happens with data that are collected because the privacy policies of companies are not very transparent.[17]

In this context a range of tools are developed that tell users that their online behaviour data is being monitored.[18] To give a few examples: Lightbeam (previously called 'Collusion') is a Firefox browser plugin developed by Mozilla that will display your online traces through a real-time network graph; another tool is Disconnect, a Chrome extension that will visualise third party trackers per site you visit and provide you with a bar chart estimating the time that you saved yourself if you decided to block the

trackers. Ghostery, which is the central actor in this study, delves deep into the trackers. Whereas privacy policies that are supposed to clear up what happens to user data remain opaque Ghostery brings the instruments that are crucial in this process into view. As stated on the website, it 'shows you the invisible web – cookies, tags, web bugs, pixels and beacons – and gives you a roll-call of over 1,800 ad networks, behavioral data providers, web publishers and other companies interested in your activity'.[19]

Ghostery is above all a visualisation tool that focuses on the *collectors* of data; it makes a translation from pieces of code in the page source to the specific type of tool it recognises this code to be a trace of. For example, 'http://b.scorecardresearch.com/beacon.js?_=1391171393485' is recognised as 'ScoreCardResearch Beacon'. Ghostery proceeds to bring this finding to the screen by displaying a pop-up. In the screenshot below you can see that when one visits this particular website (of the police) there are also two third parties present: Google Analytics and ShareThis. In this particular example Ghostery shows that this computer is not only communicating to the server of the website but also to the servers of other third party companies.

Fig. 1: Pop-up Third Parties, http://kombijdepolitie.nl, January 2014.

To describe the techniques that collect user data Ghostery uses the term 'third party elements', or in short '3pes'. Ghostery orders and ranks third party elements by indexing them into different types. It does so not according to their technological terms (such as pixels and bugs) but according to what they *do*. Ghostery says third party elements can deliver advertisements (AD), provide research or analytics for website publishers (AN), track user behaviour (T),[20] provide some kind of page function through widgets (W), or disclose data practices involved in delivering an ad (P).

Ghostery's ranking system (Ghostrank) presents the weight of these elements according to their relative presence on the Web – at least on the part of the Web that is visited by Ghostery's user population, because Ghostrank is made possible through the participation of the people who use the tool. The database is constructed by people that opt-in to automatically share their third party encounters with Ghostery's database. In spring 2013

Ghostery had 17 million users and 7 million took part in Ghostery's 'panel' that contributes to the database.[21] The table has the form of the periodic table of elements (http://knowyourelements.com). The higher the relative chance one encounters a specific third party element the higher it is ranked in the table. Therefore by providing visualisations and information during browsing Ghostery makes third party elements not only 'present' but also more accessible for further analysis.

By making the invisible Web visible Ghostery aims to help Internet users to make informed decisions and to give them more control over when they are being tracked and by whom. The behaviours per element are filed in a library. According to Ghostery's parent company, Evidon, the library contains more than '1,600 companies and 4,100 different types of trackers', which makes it, according to them, 'the only comprehensive library of trackers on the internet'.[22] The library provides information about what kind of data are collected (such as geo-location data, IP address, or phone number) by a particular third party element and whether it shares data with (again) other parties. Ghostery also suggests different ways to 'handle' third parties. It offers users the possibility to block all or only some third parties by separately flagging them.

The database is not only of use to privacy-aware individuals. Evidon uses the information to inform online marketing companies about the implementation of their tools and to offer advice about how to comply with privacy rules.[23] Evidon's mission is 'to enable a more transparent, trusted environment for consumers and advertisers'.[24] The company takes part in a larger program managed by a consortium of advertising and marketing associations – the Digital Advertising Alliance (DAA) – which pushes a label that draws a parallel with ethical (food) consumption, referring to the idea of a nutrition label: '[f]or businesses and NGOs, Evidon provides the technological underpinnings that put the AdChoices icon, which functions as a "tracking nutrition label" into ads, as well as reports on trackers and what they are doing on the web.'[25]

Ghostery as an issue device

Now that we have gotten to know Ghostery a bit better we can get back to how to think about 'devices in action'. How does Ghostery (following Ruppert et al.) distribute information and intervention, and what does that inscribe to the issue at hand? Through its database and vocabulary Ghostery mobilises particular concepts and distributes what counts as information

and action. Through Ghostery Web tracking becomes something that can be ordered, something that becomes knowledgeable. The (visual) language of the periodic table is maybe just a metaphor but at the same time it helps framing trackers as components and tracking as an environment. Trackers, instead of consisting of intangible processes, become elements that can be mined themselves.

From science and technology studies we know that ideas of nature can be constitutive in sorting out what belongs to the realm of knowledge and what belongs to the realm of values (and social action).[26] Ghostery is engaged in a similar distribution as well – in addition to the third party environment as something to become 'informed about' one can also learn how to 'cope' with it. By offering a knowledge repository accompanied by an action repertoire of possible 'options' you can detect, block, and pause. There is a common denominator in this action repertoire – 'you'. How to evaluate Web tracking becomes a matter of responsibility on the part of the individual Internet user, who can asses his or her trust relation with different kinds of companies. Tracking becomes something that the info-aware individual can choose to consume or not.

In a text on data communities Harvey et al. use the notion of 'transparency devices' to describe how these communities map things such as government transactions or community conflicts with a set of specific tools for measurement and visualisation; but they also show how these communities, by making things transparent and legible, simultaneously inscribe something to the thing they study.[27] Ghostery does exactly that. Through making Web tracking transparent it enacts tracking as a material thing, as something consisting of components that can be studied and ranked; it subsequently calls an ethics of Web tracking into existence. Web tracking can be 'bettered' through labels, changing consumer behaviour and coalitions between companies. Thus, in addition to looking at community practices we can also analyse processes by which transparency and inscription coincide through devices themselves. Here I refer to the work of Marres who has coined the term 'material participation'.[28] With this term she wants to stress the extent to which objects can facilitate matters of concern, and 'issue articulation' is one way in which this happens. Building upon Marres' work we could say that Ghostery is a device that 'redistributes participation'; by articulating the issue in this way it organises the work and responsibilities relating to how to cope with Web tracking. So, if digital devices materialise social relations, Ghostery materialises an issue and it does so in a very literal sense. Therefore I use the term issue device rather than transparency device to refer to the way in which Ghostery brings Web

tracking to the fore, because I think the performativity of the issue is a relevant point if one is concerned with what the device does to the method.

Ghostery as a research device

Because Ghostery provides certain ordering principles to detect third parties and a typology relating to their activities it has proved to be very useful as a research tool. The Digital Methods Initiative at the University of Amsterdam deploys the ordering principles of existing Web devices for social research. Considering that these devices take part in specific 'device cultures' they can produce situated knowledge that is valuable for understanding contemporary social life.[29] The adagio is to follow the 'language of the medium', or the 'actors', in Latourian jargon.[30] That means instead of using previously established categories from the social sciences that emerged out of other research sites besides the Web, one would stay close to the terms of Web devices and look at how they articulate the connections between various Web objects.

The Tracker Tracker is part of the toolbox of the Digital Methods Initiative. The Tracker Tracker uses a database of pre-defined fingerprints of Web technologies provided by Ghostery and compares those traces with the URLs that are of interest to the researcher. The DMI built upon Ghostery and not on a comparable device such as Lightbeam because the latter was not yet publicly known at the time the Tracker Tracker tool was built, also because Ghostery publishes their lists of trackers and updates them regularly. This enables researchers to analyse specific data sets by making use of Ghostery's classificatory scheme. After inserting a list of URLs into the Tracker Tracker it provides a spreadsheet with all the domain names and the respective names of third party elements that are detected per URL, also adding their type (AD, Analytics, Widget, etc.). Therefore the tool does not only give an indication of the overall presence of third party elements that collect data online but it also enables you to zoom in on the different types of elements and to do a comparative analysis between websites.

Tracker Tracker research has been relatively new and experimental; projects have been done with data sets such as the top-Alexa websites, technology blogs, and political party websites.[31] As work by Gerlitz & Helmond on the top-1000 Alexa websites has shown, the Tracker Tracker can be used to map the connections between websites and the 'data objects' that they share. Such maps provide insight into what they call an 'alternative fabric of the web'. This texture is not based on the hyperlinks through which we

often imagine the Web but on the relations between third party tracking devices and the respective websites at which they are detected.[32] If we look at such networks of websites we get a glimpse of the material relations that provide the conditions for data transactions within the previously mentioned 'Wild West'. Hence, this kind of exploratory research helps us to imagine the contribution of data collectors to what Callon & Muniesa have termed 'calculative spaces' – those arrangements that make things calculable.[33] In line with these kinds of digital methods studies, I looked at the shared third party elements on a particular set of websites, particularly the websites of the Government of the Netherlands.

The Third Party Diary

The context of my case study was a debate in The Netherlands about the Dutch implementation of the EU e-Privacy Directive. Since June 2012 the Dutch law obliges website owners to ask for the consent of Internet users for technologies that access their devices in order to collect or store data – a law which became (badly) known as the 'cookie-law'.[34] A few months later the Government of the Netherlands ('Rijksoverheid') was criticised for failing to obey the law. The debate focused on two main governmental websites: rijksoverheid.nl and government.nl. Both sites were setting cookies. On 9 August 2012 the government announced that they would disable all the cookies on these two websites and that they would further assess whether 'other websites' needed to be adjusted as well.[35]

This discussion provided an incentive for me to dig a bit deeper into this issue. The response by the government made me think about which 'other websites' could be of relevance. Thanks to open data guidelines the whole Website Register of the Government of the Netherlands ('Websiteregister Rijksoverheid') can be found online. This register gives information about approximately 1100 websites that belong to the Dutch government (cities and regional governments are excluded).[36] This data set provided the starting point for my research. The question about which particular tracking devices are allowed (or not) I will leave aside by reformulating the debate in socio-technical terms: can we measure the response of the Dutch government to this issue to by mapping the presence of third parties on these websites?

For four months in 2012 I registered the third parties that collect visitors' data on websites belonging to the Government of the Netherlands. I presented the results in an online logbook titled The Third Party Diary, which gives an impression of third party encounters when visiting the government

online (http://thirdpartydiary.net). The format of the diary was chosen for several reasons. Keeping a diary would be a means to structure the project and feature the results online, as it dealt with a current affair.[37] Another reason was that the research was not a clean and automated process and I did not want to suggest it was – working with this device was in fact pretty messy.[38] As argued by Leistert, digital methods can give the impression of being some kind of disembodied process with respect to the objects of research and the researcher as well.[39] A diary seemed to be a good format to deal with the idea that the outcome of the project was not just through the tool but also through an engagement with the tool.

The methodological steps I took were as follows. I inserted the total list of URLs in the Website Register in the Tracker Tracker tool. The Tracker Tracker output mentions third parties multiple times per domain name when similar elements are detected in different 'patterns'. Therefore these double findings were deleted from the tool's results. I then determined the total list of domain names containing third party elements, the total amount of third party elements, and I randomly checked for false positives and negatives. I repeated the study every month for four months, from August until November 2012. In 2013 the study was taken up again in January and repeated irregularly. The Website Register of the Government of the Netherlands is regularly updated. The latest revision of the register was used as input for the Tracker Tracker tool each time. Below I will present my findings and discuss how this contributes to an understanding of Web tracking practices.

Third party presence

In August 2012, in total, 856 third party elements were detected by 38 different individual third parties (Google Analytics, Webtrends, Facebook Connect, etc.). The figure below is a visualisation of the relative presence of third party elements (the size refers to the amount of third party elements, the colour to the type of activity).

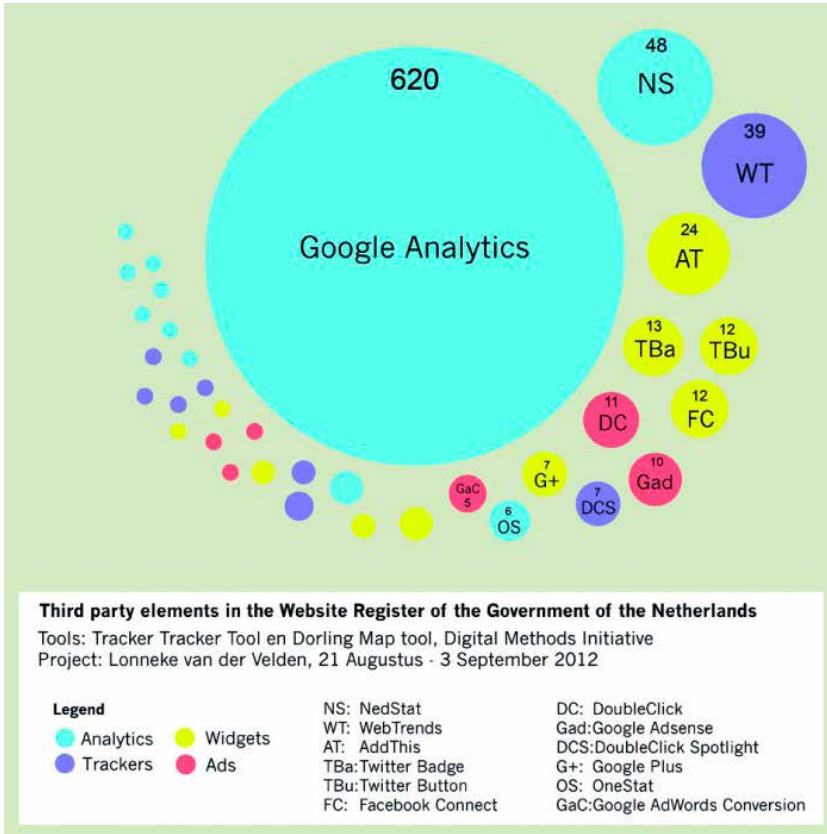

Fig. 2: Third party presence, August 2012. The nodes refer to the different third party elements (3pes) as distinguished by Ghostery (http://www.knowyourelements.com/). Elements that occurred less than five times are not listed by name. The size indicates the amount of 3pes in the Website Register of the Government of the Netherlands and the colour refers to the type of 3pe. The Register contained 110 websites in total.

Several third party elements are operated by the same company, which leads to the conclusion that only 28 companies seem to be involved, of which Google is the biggest (see Figure 3 below) followed by Comscore, Webtrends, Twitter, AddThis, and Facebook. This finding is supported by Hoofnagle et al., who reviewed tracking practices on top websites in 2009 and 2011 and concluded that there is a concentration of a relatively small amount of companies operating a large amount of Web tracking technologies.[40]

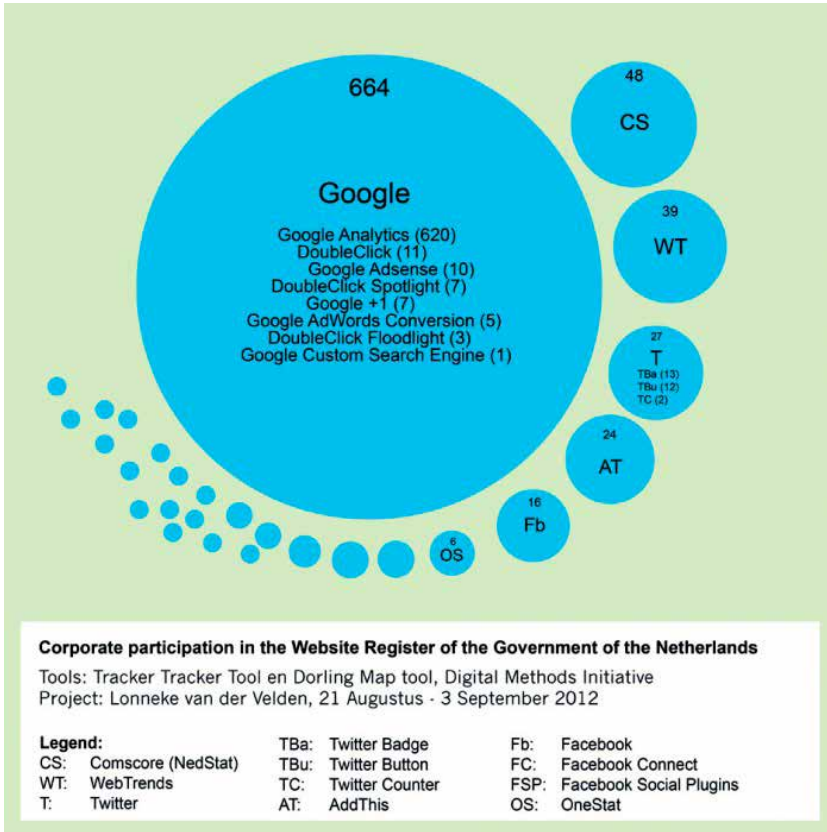

Fig. 3: Corporate participation, August 2012. The nodes refer to third party elements (3pes) in the Website Register of the Government of the Netherlands, as indicated by Ghostery (http://www.knowyourelements.com). Elements that occurred less than five times are not listed by name. The size indicates the share in 3pes companies have in the total amount of 856 3pes. The register contained 110 websites in total.

On average the percentage of websites containing third party elements is always more than half of the website register. The percentage lies higher when taking into account the fact that many domain names are not even active. For instance, in September the Website Register contained 1088 websites of which 913 were active. 658 domain names contained third party elements – that makes 60% of the whole register but 72% of the active domain names. A study by Koot, who simultaneuously investigated the same data set as I did in September 2012 (though using a different approach), points to similar findings. He used software for automated browsing (Mozrepl and Burp Suite) in order to fetch the third party content on the domain

names and to analyse the traffic.[41] He found that 671 domain names of the active URLs contained third party content (73%). Thus, despite Ghostery's detection method not being 100% complete[42] it does come pretty close to the findings of other researchers.

Table 1 gives an overview of the presence of third party elements in the website register for the months August-November 2012, the months directly following the public debate.

Table 1: Results 3pes (August-November 2012).

	Domain names in Website Register	Amount of domain names containing 3pes	Amount of 3pes	Percentage of the Website Register containing 3pes	Percentage of the active domain names containing 3pes	Amount of different types of 3pes	Amount of companies (estimation)
August '12	1110	696	856	60%	n/a	38	28
September '12	1088	658	803	60%	72%	36	26
October '12	1052	588	721	56%	64%	35	27
November '12	1129	598	728	53%	n/a	34	26

Because the government was given an explicit warning in September 2012 by the Independent Post and Telecommunications Authority of the Netherlands (OPTA) to abide by the law, I expected to see a decrease in third parties over time.[43] There was a small drop in October and November but it is hard to say whether that really indicates removal. The decrease might also be due to the fact that the Website Register was updated and now excludes a few redirects that were included in September.[44] In November 2012 the overall percentage of third party elements in the Website Register was still 53%. Hence, over four months the decrease in third party elements was 7%. In fact, when I checked a year later in December 2013 the percentage was back to 63%. We can therefore conclude that after the August 2012 debate about the government tracking their visitors the removal of tracking devices has been limited.

Shared third parties

It is also possible to visualise the connections between websites and third parties. The image below gives an impression of the associations between the third party elements (the collectors of the data) and the websites within

which the elements are located. The output of the Tracker Tracker tool from September 2012 was visualised with Gephi.[45] It shows the massive outreach of Google Analytics; it also shows how certain nodes are surrounded by clusters of websites, for instance the Webtrends cluster on the bottom right. This means that several websites use a Webtrends tracker.

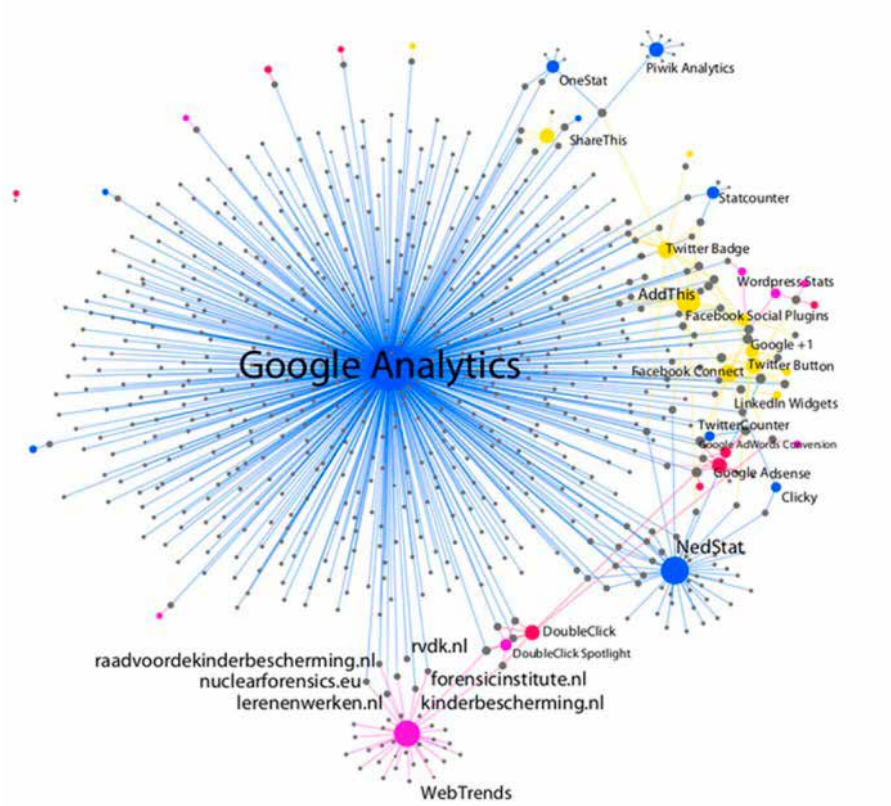

Fig. 4: Gephi visualisation, September 2012. The coloured nodes are trackers. The grey nodes are the domain names. The names of the websites are deleted for reasons of clarity, except for the bottom to illustrate the purpose of the map. For instance, nuclearforensics.eu and forensicinstitute.nl are connected with both WebTrends and Google Analytics.

There are a few interesting insights when zooming in further into that particular cluster. I first manually sorted the results by 3pe-type and name (see Table 2 below).

Table 2: Third party elements sorted by type, September 2012.
(Selection. Complete list available at http://thirdpartydiary.net.)

domain name	type 3pe	name 3pe
werkenbijvtspn.nl	ad	AppNexus
duurzaamdoen.nl	ad	DoubleClick
www.eenovervallermoetzitten.nl	ad	DoubleClick
traineebijdeeu.nl	ad	Google Adsense
psosamenwerken.wordpress.com	ad	Quantcast
www.rijveiligmetmedicijnen.nl	tracker	DoubleClick Spotlight
adviescollegeverloftoetsingtbs.nl	tracker	WebTrends
dienstterugkeerenvertrek.nl	tracker	WebTrends
psosamenwerken.wordpress.com	tracker	ScoreCard Research Beacon
kiesbeter.nl	analytics	Clicky
internetpillen.nl	analytics	Google Analytics
irak.nlambassade.org	analytics	Google Analytics
iran.nlambassade.org	analytics	Google Analytics
iran.nlembassy.org	analytics	Google Analytics
iraq.nlembassy.org	analytics	Google Analytics
ireland.nlembassy.org	analytics	Google Analytics
israel.nlambassade.org	analytics	Google Analytics
israel.nlembassy.org	analytics	Google Analytics
istanbul-tr.nlconsulate.org	analytics	Google Analytics
istanbul.nlconsulaat.org	analytics	Google Analytics
istanbul.nlconsulate.org	analytics	Google Analytics
italie.nlambassade.org	analytics	Google Analytics
pleegzorg.nl	widget	AddThis
hetgezondevoorbeeld.nl	widget	Hyves Widgets

It is here that Ghostery becomes more than a magnifier and shows its microscopic capacities. This way of sorting shows which websites share similar third party elements and how in some cases the use of third party elements corresponds to departmental orderings of the respective ministries. For Table 2 I selected only a sample, but at least 23 sites of the Ministry of Security and Justice were using Webtrends in September 2012, including sites such as the website of the Council for Child Protection (Raad voor de Kinderbescherming), a committee for research into child abuse (Commissie-Samson), and a committee advising on the release of mentally-disordered offenders (Adviescollege Verloftoetsing TBS).

Trying to zoom in even further I picked one website, the website of the Council for Child Protection (kinderbescherming.nl), and received a Webtrends cookie in my browser which included my IP address. The IP address stayed the same when I visited another website of the Ministry of Security and Justice (avtminjus.nl) within the Webtrends cluster. Webtrends only set a new cookie when I emptied my browser. Checking the host of these Webtrends cookies led me to a company called Imetrix, which provides hosting and analytics. Apparently the Ministry of Security and Justice hired this company to take care of a whole set of its websites.[46] This suggests Imetrix collected IP addresses (and maybe more data) categorised in a specific 'departmental' way, through websites that deal with child protection issues and mentally-disordered offenders – issues that fall under the category of 'Security and Justice'. They removed the trackers by the end of 2012.

Another interesting insight from the same data set is that all Dutch embassy websites share Google Analytics. In Ghostery's library one can find a summary of what Google Analytics collects, which includes (according to their terms) anonymised IP addresses, locations, and search queries. This means that this kind of information related to people interested in Dutch embassies is most probably shared with Google's servers. The cluster entails 250 Dutch embassies and consulates. The point here is not only that behavioural data is transferred from governmental websites to third parties, but it is the standardisation in this process that raises interesting questions. Because the government implemented Google Analytics as a *standard* on almost all of the ambassadorial websites the government shared with Google a data set that is in effect organised (as an ambassadorial category), and as the December 2013 results indicate they still did so a year later.

Lessons from The Third Party Diary

The results of the case study raise critical political-economic, legal, and security-related questions. Is the Dutch government, in a sense, a 'miner' for what Leister calls 'Wild West data mining capitalism',[47] by already preparing datasets and giving companies such as Google and Facebook a helping hand in 'audience sorting'? And since we are already familiar with Google Flu Trends as a form of research into flu activity (http://www.google.org/flutrends/) one could imagine what kind of 'trends research' Google could do with ambassadorial data sets. Will 'Visa Request Trends' become the new migration studies? There could be potential legal consequences as well, because data is shared with servers that are under the jurisdiction of

the United States. More concretely, the use of tracking devices can bring along a range of privacy and security problems. Koot's study explains how third party content can provide easy access points for cyber attacks (such as session hijacking and malware infection).[48] Tracking devices can be 'repurposed' too. Since the leaking of the NSA files by Edward Snowden we know that Google cookies are repurposed by the NSA to follow the behaviour of potential targets before the agency installs malware on their computers.[49] These new insights into the use of Web tracking devices show how consumer surveillance and state surveillance coincide.

The case study raises questions with respect to the method as well. Over time a few websites changed their tracking policy and began to ask for explicit consent from the visitor (for example the Education Council of the Netherlands at http://onderwijsraad.nl). This basically means that the Internet user will get a pop-up that asks whether he or she agrees with the use of cookies. Upon agreement the page should load the trackers, or otherwise it should not (ideally speaking). The effect of this change was that some third party elements disappeared from my output. However, this does not mean that third party elements are not operative. Studies have shown that people tend to accept terms of services.[50] Therefore people may consent to and load third parties that were (at the time of the project) not indexed by the Tracker Tracker output of Dutch websites. The disappearance of third party elements is therefore an interesting phenomenon by itself.

Elmer, about a decade ago, argued that cookies should be understood as mechanisms of communication instead of using the flattened definition of 'a piece of text'. According to him the 'data definition' of cookies obscures the process by which this information reaches the hard drive of the computer.[51] Along the same lines, in the example of the webpage above, the loading of third party trackers also depends on a process of negotiation. Moreover, the way websites organise the consent-procedure happens through different programming languages. At the time of study the Tracker Tracker tool did not recognise JavaScript and therefore behaved as an atypical and old-fashioned browser. Some websites will treat this as a 'yes I accept' and others as a 'no'. In other words, the device cannot consent. It is treated differently depending on how the website treats the device.

This brings me to the more reflective question of whether turning Ghostery from issue device into research device mattered for the way Web tracking was presented in the research project. What is, to recall Marres & Weltevrede, the epistemology built into the tool? Does it matter that Ghostery imagines 'tracker data' as components, as a materialised environment, as things that can be mined in turn, and that it distributes 'tracker

allowance' to the realm of individual choice? To a certain extent I think it does. If we follow the device by only focusing on its detection principles we limit ourselves to an elementary understanding of tracking in which it is located in the page source. The Tracker Tracker then operates under the assumption that the activities of third party elements are dictated by the set of sites and their code. However, we cannot assume that in this case.

Since we are dealing with a particular local context in which website owners are encouraged to ask for consent and people have to interact with that code the social or legal-material arrangement is one in which interventions take place before scripts are loaded. In some of these cases, depending on how the website responds to the Tracker Tracker's automated character and the inability of the tool to interact with site content like a regular visitor, it will not show all the trackers the latter would encounter. A negative output from the Tracker Tracker tool cannot be judged 'tracker clean' unless a manual check – by accepting cookies – follows. In other words, in this context 'tracker allowance' turns out to be more complex than individual choice only because Web tracking is dealt with through a complex of state legislation, cookie-walling, and user interaction. This becomes particularly relevant in research projects with smaller and specific data sets. A question of methodological challenge then becomes whether it is feasible for digital methods to enrich the Tracker Tracker in such a way that it captures these processes of negotiation and acceptance. Can 'docility' be built in?[52] At the time of writing an update of the tool is being worked on (in the sense that it now recognises JavaScript).

Lury & Wakeford have compiled a range of studies on devices clustered under the term 'inventive methods'. According to them devices can be inventive when they can 'change the problem to which they are addressed'.[53] In this case study the Tracker Tracker has prompted a reorganisation of the project by provoking new questions: can we capture Web tracking as a more interactive thing? Should and can the tool be changed in order to do that? A more general conclusion for future tracker research could be that the context of the data set matters. One could use digital methods to study 'social life' (in my case this was the state of the issue and institutional-tracking assemblages). However, it is important to ask what kind of new questions a data set brings with respect to the Web objects that we investigate.

Conclusion

In this project Ghostery was shown to operate on a range of levels. As an issue device it brings Web tracking to the fore, and we need a qualitative approach to see that. Ghostery maps and ranks practices of Web tracking and uses a particular vocabulary to make these technologies present and accountable. Ghostery's inscription into the issue is one in which Web tracking becomes a material environment to be coped with.

As a research device it can point out the associations between websites and shared objects and relate to existing studies into the transactions of behavioural data. The Tracker Tracker also allows zooming into clusters of websites and provides empirical data that can feed concrete public affairs. The Government of the Netherlands was shown to intensively participate in the market of behavioural data. We get some insight into how specific data move from one organisation to another, such as from the ministry of Foreign Affairs to Google. It gives few clues about the make-up of these data sets and about which actors participate in this process. The project therefore contributes to a better understanding of the first steps of the process of behavioural targeting. It suggests that orderings by category are already embedded in the process of collecting data due to very mundane and institutional aspects of governmental life. Thus, instead of assuming that data collection is a starting point for further enhancement and profiling processes, practices of categorisation turn out to be already active from the start.

The case study has also interrogated the device. Reflecting upon the way Ghostery imagines its data and taking the device out of its device culture to study a new context has led to the question of how to capture Web tracking as a negotiated practice.

Acknowledgements

I would like to thank the anonymous reviewers for their comments as well as the participants at the workshops organised by the Digital Methods Initiative (DMI) and Goldsmiths College, particularly Noortje Marres. Matthijs Koot helped me with his topical expertise and my colleagues at the DMI, in particular Erik Borra and Emile den Tex, supported me in better understanding the technicalities of the tool.

Notes

1. Rogers 2009. The Tracker Tracker tool in particular was developed in a collaborative project by Yngvil Beyer, Erik Borra, Carolin Gerlitz, Anne Helmond, Koen Martens, Simeona Petkova, JC Plantin, Bernhard Rieder, Esther Weltevrede, and Lonneke van der Velden during the Digital Methods Winter School 2012, 'Interfaces for the Cloud'. Project page: https://wiki.digitalmethods.net/Dmi/DmiWinterSchool2012TrackingTheTrackers.
2. Savage & Burrows 2007; Marres 2012; Ruppert et al. 2013.
3. Marres & Weltevrede 2013, p. 319.
4. Marres 2012; Borra & Rieder 2014. Weltevrede n.d.
5. Marres 2012b.
6. Gitelman & Jackson 2013, p. 3.
7. Ibid.
8. Ruppert et al. 2013, p. 35.
9. Law & Urry 2004; Callon & Muniesa 2005.
10. Ruppert et al. 2013, p. 32.
11. Gitelman & Jackson 2013, p. 4.
12. Marres & Weltervrede 2013.
13. McStay 2013, p. 597.
14. Zuiderveen Borgesius 2013.
15. Raley 2013.
16. Zuiderveen Borgesius 2013; Leistert 2013.
17. McDonald & Cranor 2008, p. 541; Zuiderveen Borgesius 2013.
18. Raley 2013; Van den Berg & Van der Hof 2012.
19. Ghostery. 'How It Works'. http://www.ghostery.com/how-it-works (accessed on 20 January 2014).
20. In a later version Ghostery updated the 'Tracker' to Beacon (B) to prevent confusion with the general term Tracker.
21. Evidon. 'The Evidon Blog'. http://blog.evidon.com/tag/ghostery/ (accessed on 7 March 2013).
22. Evidon. 'Analytics'. http://www.evidon.com/analytics (accessed on 25 January 2014).
23. 'What does Evidon do with Ghostrank information', https://www.ghostery.com/faq#q17 (accessed on 3 April 2014).
24. Evidon. 'Better Advertising Acquires Ghostery'. http://www.evidon.com/blog/better-advertising-acquires-ghostery (accessed on 30 January 2014).
25. Ghostery. 'Frequently Asked Questions'. https://www.ghostery.com/faq (accessed on 25 January 2014).
26. Latour 2004.
27. Harvey et al. 2013.
28. Marres 2012a.
29. Rogers 2013; Rogers et al. 2013.
30. Latour 2005.
31. See the Tracker Tracker project page (https://wiki.digitalmethods.net/Dmi/DmiWinterSchool2012TrackingTheTrackers) and the work of Helmond 2012 on Dutch political party websites.
32. Gerlitz & Helmond 2013, p. 1349.
33. Callon & Muniesa 2005.
34. Because the law is formulated in a broad manner it applies to more tracking technologies than just cookies. 'Telecommunicatiewet, Artikel 11.7a', available at http://wetten.overheid.nl/BWBR0009950/Hoofdstuk11/i111/Artikel117a/geldigheidsdatum_03-09-2012.

35. de Haes 2012.
36. According to the 'Whois' information the domain names listed in the register are not all legally 'owned' by the government. Still, the government presents this list as their responsibility. The Website Register can be found at: http://www.rijksoverheid.nl/onderwerpen/overheidscommunicatie/eisen-aan-websites-rijksoverheid/websiteregister-rijksoverheid (accessed on 25 January 2013).
37. The Dutch news site nu.nl paid attention to the study. See de Winter 2012.
38. It entailed cleaning data and preparing the URLs before even using the tool and going through many error reports. More background to the method can be found at https://wiki.digitalmethods.net/Dmi/ThirdPartyDiary.
39. Leistert 2013.
40. Hoofnagle et al. 2012.
41. Koot 2012.
42. Brock 2010.
43. Wokke 2012.
44. For instance, in September raadvoordekinderbescherming.nl, which was redirecting to kinderbescherming.nl, was excluded in the October update. Therefore third party elements that were previously counted twice were counted only one time in October.
45. The Digital Methods Wiki provides instructions for how to visualise Tracker Tracker data: https://wiki.digitalmethods.net/Dmi/WorkshopTrackingtheTrackers#A_42DMI_Projects_using_the_Track_the_Trackers_tool:_42.
46. Checking the Whois and trace route of the IP address suggests that minjus.sdc.imetrix.nl was physically located in Amsterdam at the hosting company hostingbedrijf Redbee.nl.
47. Leistert 2013.
48. Koot 2012. See also Tran et al.
49. This concerns the 'PREF-cookie', which also comes with Google Analytics. See Soltani & Peterson & Gellman 2013.
50. Rogers 2008. This can be due to terms of services being non-negotiable (King & Jessen 2010).
51. Elmer 2004, p. 130.
52. Rogers 2008.
53. Lury & Wakeford 2012, p. 13.

References

Berg, B. van den and Hof, S. van der. 'What happens to my data? A novel approach to informing users of data', *First Monday*, Vol. 17, No. 7, 27 June 2012.

Borra, E. K. and Rieder, B. 'Programmed Method. Developing a Toolset for Capturing and Analyzing Tweets' in *Aslib proceedings*, edited by K. Weller and A. Bruns, 2014.

Brock, J. 'Credibility Gap: What does Ghostery really see?', *Privacy Choice Blog*, 4 March 2010: http://blog.privacychoice.org/2010/03/04/credibility-gap-what-does-ghostery-really-see/ (accessed on 25 January 2014).

Callon, M. and Muniesa, F. 'Peripheral Vision: Economic Markets as Calculative Collective Devices', *Organization Studies*, Vol. 26, No. 8, 2005: 1229-1250.

Elmer, G. *Profiling machines: Mapping the personal information economy.* Cambridge: MIT Press, 2004.

Gerlitz, C. and Helmond, A. 'The Like Economy: Social Buttons and the Data-intensive Web', *New Media Society*, Vol. 15, No. 8, 2013: 1348-1365.

Gitelman, L. and Jackson, V. 'Introduction' in *'Raw Data' is an oxymoron*, edited by L. Gitelman. Cambridge: MIT Press, 2013.

Haes, A. de. 'Rijksoverheid zet alle cookies uit', *Webwereld*, 9 August 2012: http://webwereld.nl/nieuws/111422/rijksoverheid-zet-alle-cookies-uit.html.

Harvey, P., Reeves, M., and Ruppert, E. 'Anticipating Failure: Transparency devices and their effects', *Journal of Cultural Economy. Special Issue: The Device: The Social Life of Methods*, Vol. 6, No. 3, 2013: 294-312

Helmond, A. 'Trackers gebruikt op de websites van Nederlandse politieke partijen in kaart gebracht', *annehelmond.nl*, 11 June 2012: http://www.annehelmond.nl/2012/06/11/trackers-gebruikt-op-de-websites-van-nederlandse-politieke-partijen-in-kaart-gebracht/.

Hoofnagle, C. J. et al. 'Behavioral Advertising: The Offer You Cannot Refuse', *6 Harvard Law & Policy Review 273; UC Berkeley Public Law Research Paper No. 2137601*, 28 August 2012, available at SSRN: http://ssrn.com/abstract=2137601.

King, N. J. and Jessen, P. W. 'Profiling the mobile customer – Is industry self-regulation adequate to protect consumer privacy when behavioural advertisers target mobile phones? – Part II', *Computer Law & Security Review*, Vol. 26, No. 6, 2010: 595-612.

Koot, M. 'A Survey of Privacy & Security Decreasing Third-Party Content on Dutch Websites', 26 October 2012, available at http://www.madison-gurkha.com/press/2012-10-SurveyOfThird-PartyContent.pdf.

Latour, B. *Politics of nature: How to bring the sciences into democracy*, translated by C. Porter. Cambridge: Harvard University Press, 2004.

_____. *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press, 2005.

Law, J. and Urry, J. 'Enacting the social', *Economy and Society*, Vol. 33, No. 3, 2004: 390-410.

Leistert, O. 'Smell the fish: Digital Disneyland and the right to oblivion', *First Monday*, Vol. 18, No. 3, March 2013.

Lury, C. and Wakeford, N. *Inventive methods: The happening of the social*. Oxon: Routledge, 2014 (orig. in 2012).

Marres, N. 'Redistributing problems of participation', in *Material participation: Technology, the environment and everyday publics*. London: Palgrave Macmillan, 2012a.

_____. 'The redistribution of methods: on intervention in digital social research, broadly conceived', *The Sociological Review*, Vol. 60, 2012b: 139-165.

Marres, M. and Weltevrede, E. 'Scraping the Social? Issues in real-time social research', *Journal of Cultural Economy*, Vol. 6, No. 3, 2013: 313-335.

McDonald, A. M. and Cranor, L. F. 'The cost of reading privacy policies', *I/S: A Journal of Law and Policy for the Information Society*, Vol. 4, No. 3, 2008: 540-565.

McStay, A. 'I Consent: An Analysis of the Cookie Directive and Its Implications for UK Behavioral Advertising', *New Media & Society*, Vol. 14, No. 4, 2013: 596-611.

Raley, R. 'Dataveillance and Countervailence' in *'Raw Data' is an oxymoron*, edited by L. Gitelman. Cambridge: The MIT Press, 2013.

Rogers, R. 'Consumer technology after surveillance theory' in *Mind the screen: Media concepts according to Thomas Elsaesser*, edited by J. Kooijman et al. Amsterdam: Amsterdam University Press, 2008: 288-296.

_____. *The end of the virtual: Digital methods*. Amsterdam: Amsterdam University Press, 2009: 1-38.

_____. *Digital methods*. Cambridge: MIT Press, 2013.

Rogers, R., Weltevrede, E., Niederer, S., and Borra, E. K. 'National Web Studies: The Case of Iran Online' in *A companion to new media dynamics*, edited by J. Hartley, A. Bruns, and J. Burgess. Oxford: Blackwell, 2013: 142-166.

Ruppert, E., Law, J., and Savage, M. 'Reassembling Social Science Methods: The Challenge of Digital Devices', *Theory, Culture & Society*, Vol. 30, No. 4, July 2013: 22-46.

Savage, M. and Burrows, R. 'The Coming Crisis of Empirical Sociology', *Sociology*, Vol. 41, No. 5, 2007: 885-899.

Soltani, A., Peterson, A., and Gellman, B. 'NSA uses Google cookies to pinpoint targets for hacking', *The Washington Post*, 10 December 2013.

Tran, M. et al. 'Tracking the Trackers: Fast and Scalable Dynamic Analysis of Web Content for Privacy Violations', *Proceedings of the 10th international conference on Applied Cryptography and Network Security*, Singapore, June 2012.

Weltevrede, E. 'Introduction: Device-driven research' (working paper) in *Re-purposing digital methods: Exploring the research affordances of platforms and engines* (diss). Digital Methods Initiative, Department of Media Studies, University of Amsterdam, n.d.

Winter, B. de. 'Overheidswebsites sturen gegevens door naar derden', *nu.ul*, 1 November 2012: http://www.nu.nl/internet/2947863/overheidssites-sturen-gegevens-derden.html.

Wokke, A. 'OPTA waarschuwt overheidswebsites voor overtreden cookiewet', *Tweakers.net*, 6 September 2012.

Zuiderveen Borgesius, F. 'Behavioral Targeting: A European Legal Perspective', *IEEE Security & Privacy*, Vol. 11, No. 1, 2013: 82-85.

_____. 'Online Audience Buying', *Unlike Utechnolos Conference*, Institute of Network Cultures 8-10 March 2012, Amsterdam, 9 March 2012. Conference video available at http://vimeo.com/38840197.

About the author

Lonneke van der Velden is a doctoral researcher in the Digital Methods Initiative (DMI) at the University of Amsterdam. She is interested in issues of surveillance, publics, and evidentiary technologies. More particularly, she looks at initiatives that turn surveillance into an object of public scrutiny through the use of digital tools. Van der Velden has a background in Science and Technology Studies and Philosophy and is part of the editorial board of *Krisis*, a Dutch open-access peer-reviewed journal for contemporary philosophy.