Surveillance as public matter

Revisiting sousveillance through devices and leaks

van der Velden, L.C.

4 Transparency devices and leaked data

Sites for radical expertise?

4.1 Introduction: WikiLeaks Making the Surveillance State Public

The publishing platform WikiLeaks often features as an example of a grassroots form of watching over the authorities. WikiLeaks is a platform that became known for providing an online encrypted channel for whistle-blowers to bring their data out while remaining anonymous. Not all, but many of the datasets that WikiLeaks publishes contain information about surveillance, which makes it an interesting case for a study into making surveillance public. WikiLeaks acquisitions range from diplomatic information to corporate surveillance, to sophisticated interception technologies and human infiltration. According to the initiative, instead of intelligence 'about the people' WikiLeaks produces intelligence 'of the people' (Assange quoted in Mortensen 2014, 24). WikiLeaks, according to surveillance critic Andrejevic, is 'turning the tools of the informated elites back upon them' (Andrejevic 2014, 2619). It is a form of 'sousveillance' (Mortensen 2014) or 'undersight' (Van Buuren 2013) that hold governments to account using the threat of publishing leaked data as leverage. Assange himself has theorised that the threat of leaks pushes the state to tighten the information channels of the state, thus critically affecting the way in which the state handles data (Quill 2014, 136). In other words, both WikiLeaks spokespersons and analysts have pointed to a certain sousveillance dimension of WikiLeaks.

In the previous chapter I argued that sousveillance has an analytic and investigatory dimension. In this chapter I similarly take into account the analytic practices of WikiLeaks. At the time of writing, the WikiLeaks website contains 33 separate so-called 'leaks', some of which some contain hundreds of thousands of documents. Together they form a knowledge base covering a wide range of governmental and corporate practices. In this study, I take a closer look at this knowledge base demarcated by a set of important leaks dealing with surveillance.[43] I discuss three datasets, one that is presented by WikiLeaks as emerging from the heart of the 'surveillance state' (Cablegate), one on the corporate sector producing and exporting surveillance technologies (the Spy Files) and one on a corporate intelligence company (Stratfor). These publications are diverse in the sense that they are about different aspects of surveillance (classic governmental surveillance, technological aspects of surveillance, and corporate intelligence) but they are also different in terms of composition.

The material for this chapter was written as a response to early debates about WikiLeaks that tried to grasp the project as an example of total transparency or as a 'data dump' (suggesting that WikiLeaks' publications constitute a rather homogenous and amorphous pile of data and are published in a raw manner without a careful curating policy). I prefer to use the term 'device' instead of 'dump' to signal that particular publications of WikiLeaks undergo particular translations and configurations for the benefit of analysis. As suggested by Latour, I follow the 'theories' of the actors (Latour 2005b, 149-150) and I look at how they try to draw leaked data together. I discuss the devices and analytic interventions by which WikiLeaks makes data public and for which kind of public: in other words, the case study addresses the materiality of transparency. Whereas the previous chapter stressed the translations made by one particular device, this chapter shifts emphasis to the enabling effects of digital devices and analytic practices for the emergence of publics. In other words, it is a story about the performativity of devices (§2.5).

As explained in §2.3.1, this text can be seen as a conversation between concepts and cases. In my analysis of WikiLeaks I build on three theoretical pillars: First, the idea that WikiLeaks establishes a form of radical transparency (Birchall 2014, Heemsbergen 2014). This is the idea that WikiLeaks has been disruptive of traditional channels of disclosure by pointing to alternative methods for making things public (Birchall 2014). Second, in this study, I take inspiration from scholars that have theorised the effects of so called 'transparency devices'. They say transparency itself can be seen as a 'political device' (Barry 2010), meaning that it enacts very specific, and usually civil, kind of auditors. Therefore, digital devices in transparency endeavours are performative of very specific 'data publics' (Ruppert 2015). This is useful because these theories, which can be seen as 'transparency critiques', can be made productive for a study into the devices and data practices

[43] Hence, it does not deal with what became the most controversial issue around WikiLeaks in 2016, which is the extent to which it intermingled with the US elections by publishing e-mails from Hilary Clinton's private e-mail server and from her campaign Chairman John Podesta.

within the radical transparency agenda of WikiLeaks. Third, I reflect upon insights about hackers (Coleman 2013) and the 'protocols' of contemporary network culture (Lovink and Rossiter 2011). As Lovink and Rossiter argue, network culture can be characterised by particular recurrent forms of collaboration. The question is: How does that work in the specific context of leaked data? In answer to this, I propose an agenda for the study of radical data publics.

In trying to attune the theoretical vocabulary to what takes place in practice (Mol 2010, 261), and by taking into account the type of data coming from the leaks and the type of analytic constructions offered by WikiLeaks, I conclude the chapter by arguing that WikiLeaks is, regarding a set of particular leaks, a pioneering site for the production of 'radical expertise'. It is an unauthorised form of knowledge production, it attempts to share this type of literacy, and it might lead to radical political insights or consequences. It is a form of 'expertise' because, despite this radical dimension, the kind of auditor that WikiLeaks enacts is an expert one. If sousveillance is considered in terms of literacy, we should reflect upon whether and how this kind of literacy can be shared and institutionalised.

4.2 WikiLeaks as a radical transparency intervention

WikiLeaks has been a goldmine for academics. For many it was an innovation that needed to be understood. I reflect upon the notion of 'radical transparency' although it offers only one of the many analytic entry points to the study of WikiLeaks. I will briefly highlight a few of these other dimensions: WikiLeaks has been analysed as being a new form of 'journalism' (Benkler 2013; Winseck 2013) and a form of 'activism' (as situated in a broader field of liberation technology, policy initiatives and hacktivism) (Hinz 2013; Milan 2013). Studies into WikiLeaks include its reception in news media (Hindman and Thomas 2014) and historical comparative analyses about the figure of the 'leaker' (Castronovo 2013). Others have focused on its new way of data gathering and distribution by looking into the kind of 'archive' that WikiLeaks is offering (Snickars 2014) and by reflecting upon the containment of information in contemporary 'liquid environments' (Jurgenson and Rey 2014). Also the internal organisation has gained a lot of attention. Books have been written about key figures (Greenberg 2012) and the organisation's ethics and internal struggles (Brooke 2011). People associated with WikiLeaks have contributed to this body of work by publishing about their philosophies (Assange 2014; Assange, Appelbaum and Müller-Maguhn 2013) and by producing collections of articles that deal with the content of their data (WikiLeaks 2015). In other words, there are many dimensions of WikiLeaks that are worth looking into. Lovink and Riemens (2013) produced no less than twelve theses on WikiLeaks, including the thesis that its significance lies in a new kind of politics of exposure. This latter reading of WikiLeaks is my central point of departure. Since my concern is with making surveillance public I build on literature that regards WikiLeaks as a disruptive way of making things public, that is, as a 'radical' transparency intervention (Andrejevic 2014, 2620; Birchall 2014; Heemsbergen 2013; Heemsbergen 2014).

If we look to the original meaning of something being 'transparent', we find that it stems from the Latin word 'transparere' (from trans- 'through' + parere 'appear'). As an adjective, it means '(of a material or article) allowing light to pass through so that objects behind can be distinctly seen' (Oxford Dictionaries 2015). Hence it refers to a state of matter. In the contemporary context, we often associate transparency with information provision. Claire Birchall states that transparency is more than that. She defines it as an 'attitude' or a 'commitment'. It is a:

> commitment to operating in the open, under the scrutiny of customers, stakeholders, citizens, and other interested parties through the publication of any or all of the following: datasets; minutes, transcripts, or live feeds of meetings; accounts; policies; decision-making procedures as well as the decisions themselves; and records of actions taken. (Birchall, 2014, 78)

The notion of transparency and its democratic features are often taken for granted and it is usually presented as a virtue: transparency goes against state secrecy, represents responsible policy and stands for openness about data and research methods. As Lessig (2009) has once remarked in a critical piece about the transparency movement in the *New Republic*: 'How could anyone be against transparency?'

WikiLeaks itself has been mobilising transparency discourses. On its website up until 2015, WikiLeaks explained the aims of the project as follows:

> Publishing improves transparency, and this transparency creates a better society for all people. Better scrutiny leads to reduced corruption and stronger democracies in all society's institutions, including government, corporations and other organisations. A healthy, vibrant and inquisitive journalistic media plays a vital role in achieving these goals. We are part of that media.

WikiLeaks' description contains a classic assumption about transparency: publishing information improves transparency and enhances democracy. This is not to say that WikiLeaks has a classic perspective about the workings of transparency per se. According to

Fenster (2012), WikiLeaks' theories about leaking are more complex, and the transparency vocabulary should probably also be seen as a way of communicating with (liberal) transparency advocates. Also, as touched upon in the introduction of this chapter, Assange's theory of leaking information was that it would effect or even suffocate knowledge flows within the state apparatus (Quill 2014, 136), which comes close to similar (optimistic) expectations of potential awareness effects of sousveillance (Bakir 2010, 157; Brucato 2015).

Whether or not WikiLeaks' transparency discourse is classic or not, it is the trust in the power of information that critics have questioned after the publications by WikiLeaks. Especially *Cablegate*, a WikiLeaks release of hundreds of thousands classified diplomatic cables, led to reflections about the significance of making large amounts of (classified) documents public. These diplomatic cables have been of enormous importance for how WikiLeaks is understood and criticised. For instance, Jodi Dean (2011) warned of expecting too much in terms of public response to information liberation in the context of WikiLeaks. According to her, the problem is not so much that we do not know things, but that we actually do know many things and that society does not act. In other words, the world faces a social problem more than an informational problem (2011). Also, Quill refers to a miscalculation by WikiLeaks with respect to the American public already being implicated 'in the wrongdoings of foreign policy' (2014, 134-135): 'What the cables and videos exposed were facts that they already knew but would rather not have acknowledged' (ibid., 135). Žižek rephrases this positively when he situates WikiLeaks' importance exactly in its confrontation with the (cynical) public's failure to acknowledge that of which it has knowledge: 'The function of WikiLeaks, I claim, in concrete ideological, political situations, is to push us to this point where you can (no longer) pretend not to know' (Brevini et al. 2013, 270). He compares the workings of WikiLeaks with what happens in the fairy tale *The Emperor's New Clothes*: Everyone knew the emperor was naked, but things changed after this was confirmed through public speech. Similarly, WikiLeaks is radical because it is 'changing the very rules how we were allowed to violate the rules' (ibid., 257); it is changing the playing ground for the truth teller.

Along the same lines, many analyses have considered the impact and potentiality of WikiLeaks by focusing not necessarily on its content but on its way of organising and publicising which depart from established institutionalised ways of making things public. Lievrouw calls it a form of 'commons knowledge': 'an example of new challenges to authoritative, institutionalized knowledge—that is, how knowledge is created and circulated and how its value is established, and who gets to decide' (Lievrouw 2014, 2633). Birchall stresses its disruptive effects, as WikiLeaks disrupted a 'hierarchy of revelatory modes' (2014, 84). Following these scholars we can say that WikiLeaks is an 'unauthorised' form of transparency: it publishes things that have been displaced from secret locations and against the will of the ones in charge of their curation. In Birchall's view this is an uncivil and unsettling form of revelation, with consequences for how we think transparency is or should be (materially and socially) organised. As she suggests: '"Radical transparency" might, then, be a holding space for something yet to come: an unsettling, perhaps, of what it means to "see through" and the relationship between data, narrative, information, interpretation, and understanding' (ibid., 85). For her, this involves questions of participation: Who should be authorised to be involved in disclosing matters of concern? Therefore, 'radical' refers to methodologies of making things public: '"Radical" indicates not more (of the same) transparency, but transparency rethought through a resistant, critical methodology' (ibid.).

The notion of radical transparency has also been linked to its (material) design (Heemsbergen 2013), which comes close to the aims of my study. Heemsbergen has studied the change in affordances of WikiLeaks as a 'radical transparency apparatus' (2013, 56). He shows that, although transparency is often associated with liberal democratic theory, it has been an important concept in political philosophies of various kinds. He has mapped transparency literature in its different (political) contexts and discusses how transparency is expected to bring about different political effects in these different environments respectively. For example, transparency is expected to be a regulatory vehicle (within perceptions of liberal democracy), to foster open debate (this is the deliberative version), to enable collaborative production (in notions of reciprocal societies), and to enforce revolutionary stances (in autonomist approaches). Transparency is therefore politically flexible, and, according to Heemsbergen, WikiLeaks has appealed to different varieties of transparency during its history when it modified its design. In another piece, Heemsbergen builds on and adds to the distinctions in editorial 'models' of Wikileaks, as defined by Micah Sifry, when he distinguishes: '(a) a wiki-based anonymous drop box; (b) press-release soliciting advocacy network; (c) distributor of editorialised content; (d) mainstream print media partner' (Heemsbergen 2014, 1347). By shifting in the design and decision-making, WikiLeaks has experimented with different forms of transparency.

4.3 Transparency devices produce specific data publics

Whereas Heemsbergen's concern is with the 'democratic affordances' (1345) of WikiLeaks as a radical transparency apparatus, mine is with its 'analytic affordances'. In emphasising the analytic affordances of WikiLeaks as a radical transparency apparatus, or 'radical transparency device', I make use primarily of work from STS on transparency devices and in particular the work of Andrew Barry and Evelyn Ruppert. I engage with these authors because Barry provides a classic STS argument of why transparency itself can be considered a device: to be more concrete, a 'political device', by engaging with the history of public scientific experiments. Ruppert looks in a similar way at the performative effects of transparency devices in the context of digital data. Moreover, her work deals with a case study that shows how, through the publication of datasets, the state is being made visible, which is, as I will show later on, a concern of WikiLeaks too. Whereas Ruppert speaks about processes of 'witnessing the state', WikiLeaks speaks of 'indexing the empire'.

According to Andrew Barry transparency is a political device, because transparency is performative of certain auditors. He refers to Shapin and Schaffer's famous historical analysis of material, social and literary devices in public scientific experiments (carried out by Robert Boyle), which were directed at the inclusion of reliable witnesses and which constituted a controllable space in which the experiment could be conducted (Shapin and Schaffer 1985). In the contemporary context, Barry (2010) studied the workings of transparency through the 'Extractive Industries Transparency Initiative (EITI)', an initiative that was aimed at making transparent the financial transactions between the government and the oil industry and vice versa. He argues that, similar to the workings of devices in public scientific experiments, transparency and auditing procedures are directed to specific 'reliable' auditors, such as NGO's, oversight organisations, and participants who are expected to have the skills to review in a rational way. In other words, transparency expects a controlled process and a 'civil public'. As he further explains by referring to Gabriel Tarde (in *L'Opinion et la Foule*), reading ordered data contains a soothing promise: 'In this book, Tarde looks forward to the day when the public would read and digest social statistics rather than indulge in the highly contagious imitative and affective forms of behaviour characteristic of street demonstrations' (Barry 2010, §26). He notes a similar expectation of rational behaviour in the transparency device that he studied:

> The Transparency Initiative embodies this political logic. It is expected to provide a technical solution to the management of affect, a preventative cure to the contagious forms of imitative behaviour that Tarde saw in the late 19th century urban crowd (Salmon, 2005). Transparency, in effect, is a device intended to foster the formation of a rational civil society and a rational government, albeit in embryonic form. (Ibid.)

This connection of transparency to civility also resonates in Birchall's analysis of transparency. Birchall (2014) regards transparency as a narrative technique. She describes how transparency can be seen as the civil form of disclosure set against more contentious forms of disclosures. Transparency is perceived to be more 'enlightened' in comparison with forms of disclosures such as rumours, gossips, and so on; the latter are narrative techniques that are considered to be suspect and unruly (Birchall 2014, 80-81).

Evelyn Ruppert also articulates the idea of transparency as a performative device in the context of the digital. She did a study into the 'UK transparency agenda', which is the publication of governmental data sets after the controversy about the UK Parliamentary Expenses Scandal (Ruppert 2015). In 2009 a leak revealed enormous expense claims made by members of the Parliament over several years. This became a huge scandal, and many people were fired within the UK Transparency Agenda (TA). Subsequently, the UK government published a whole range of data sets, such as expenses and business plans, ready for the public to inspect. Ruppert argues that these data sets should not be seen as isolated data-as-such, but they come with arrangements of tools for analysis, 'agencements', that mediate how these data can be read. They bring along a reconfiguration of expertise and enact particular 'data publics' (Ruppert 2015, 8). She states: 'by taking up the issue: the TA actively configures the subjectivities and agencies of these data publics through particular socio-technical arrangements' (ibid., 9). One of the problems in the TA, as considered by Ruppert, is that the public experiences the state as one of rankings and scores, but the visualisations the data public ends up with 'smooth out' (Ruppert 2015, 15) everything that goes behind the TA.

Ruppert explains that to work with the type of datasets published through the Transparency Agenda one needs particular skills and technologies. Therefore, to witness the state, one needs a particular apparatus. Ruppert, just like Barry, refers to the 'witnessing experiments' conducted by Boyle that were described by Shapin and Schaffer (1985). As Ruppert explains, the way by which Boyle managed to make audiences virtually attending the experiments, involved 'material', 'social' and 'literal' technologies. (These included: the material for the air pump, scientific rules and conventions, and detailed reports.) (Ruppert 2015, 2). She translates these insights to the TA setting in order to think about the kind of witnesses (or public) that become constituted through the TA. The TA operates through its particular arrangement, she mentions: ‘material technologies such as computer infrastructures, websites, apps and data; social technologies including rules, data formats, software protocols; and literary technologies such as visualisations, maps, photos, matrices and profiles’ (ibid., 10).

Ruppert draws several conclusions of which I make a selection for the later discussion about WikiLeaks. Less than creating ‘virtual witnesses’ that become enrolled through detailed reports which make the witnesses experience the experiment, as Boyle did, the TA constitutes ‘data publics’ (ibid., 3). These imagined data publics are expected to do their own data analysis. However, the arrangement of tools does not come from nowhere. Some of the tools come out of the journalistic sector, but many come from the corporate sector. Therefore, she argues, the UK open data project reconfigures expertise through mediators that regard data in their economic value. These mediators:

> tend to be the most active organisers and analysers of the data. (…) It is a reconfiguration that is also connected to the purported economic value of open data, which is intended to feed the digital economy through the stimulation of new analytics and applications. Big government data is not only a source of knowledge about the state but also worth billions of pounds. (Ruppert 2015, 11)[44]

Next to the economy of open data, the post-controversial context of the TA has also impact on the type of data public that can emerge from it. As argued in another article, together with Harvey and Reeves, Ruppert emphasises that the UK transparency agenda started as a response to the governmental failure of the corruption scandal. By posing transparency as a response, and by actively calling upon citizens to monitor the state, people become implicated and responsible for the potential moral failure of the state (Harvey et al. 2013, 306). This potentially constitutes ‘hypervigilant’ subjects: ‘In these ways (…) we consider contemporary transparency devices as potentially generative of hypervigilant, suspicious and doubtful witnessing subjects, neoliberal subjects who must keep a watchful eye over the micro doings of the state’ (ibid., 306). Clare Birchall expresses similar concerns about the same UK Transparency Agenda: citizens are expected to be ‘innovative’ with

[44] When one visits the website of the UK Transparency Agenda, one can get a sense of the discourse. Key terms that are associated with openness and transparency are efficiency and money: ‘Openness and transparency can save money, strengthen people’s trust in government and encourage greater public participation in decision-making’ (Gov.UK 2016).

regards to open data sets, and data are expected to be made productive. (Birchall 2014, 83). According to Birchall, this is expressed very clearly in a report commissioned by the U.K. Cabinet Office, stating that 'the release of data is intended to support the development of "social entrepreneurs"'.[45] Birchall says: 'The citizen-auditor is called on to be perpetually vigilant, to be part of "a continuous network" (Deleuze, 1992, 6), a key participant in this new informational capitalist-democracy' (Birchall 2014, 83).

From these transparency critiques, we learn that techniques used to organise transparency are embedded in specific contexts and bring about particular audiences. In the examples of formal transparency procedures the audiences are constituted as civil, vigilant and neoliberal. The arrangement of tools and mediators bring about a reconfiguration of expertise. This allows for the emergence of particular subjectivities and experiences of the state. My aim is to translate those insights to WikiLeaks, as I will explain in more detail in the next section.

4.4 Textures of radical transparency

4.4.1 Open data versus opened data

The remaining part of this chapter deals mainly with the question: How can we study 'radical transparency' with respect to the material that is at stake – leaked data? I do that by discussing various data practices surrounding WikiLeaks. Subsequently, I reflect in a more speculative way upon the question: If we take into account transparency critiques such as the ones articulated by Barry and Ruppert, what happens if we transfer them to the field of leaked data, in which transparency is not part of a legitimised procedure but part of a presumably uncivil and unauthorised disclosure? In other words: if Wikileaks is engaged with a form of radical transparency, and transparency is a political device and performative for the emergence of specific data publics: Through what kind of articulations are data rendered transparent? How does WikiLeaks reconfigure expertise? I try to tackle this issue by looking at WikiLeaks' arrangement of tools by which the organisation facilitates analysis and by posing the following questions: What kinds of data practices contribute to WikiLeaks' transparency agenda? To whom is this transparency device directed? And: What does that imply for

[45] These kinds of expectations are recognisable as well in discourses about the 'Smart City', in which citizens are expected to engage productively with urban data. In the Dutch context this phenomenon is critically examined by Dorien Zandbergen (2014).

the data public? In other words, I aim to make the above mentioned transparency critiques productive for the context of leaked data.

Before looking more closely to some of the datasets WikiLeaks works with, it is worth clarifying the distinction between the examples of the formal transparency procedures mentioned above, and an ‘unauthorised’ transparency device such as WikiLeaks, because the way in which the data are provided and the setting in which the data are analysed are obviously different. First of all, in the case of WikiLeaks, data are displaced. In the example of the UK Transparency Agenda, Ruppert argues that the tools that come with the data are mobilised within a specific setting and that they come with certain expectations. The setting is post controversial and embedded in a governmental-corporate complex. The expectations are that data analyses could prevent future moral failures of the state and that data can be made (economically) productive. However, in the case of leaked data, the data has been taken out of governmental or corporate settings. The working environment, which informs the modes of analysis, differs from governmental or commercial settings, and the tools are produced relatively self-organised collectives. That means that different modes of reasoning and analysis co-produce the witnessing of leaked datasets. This allows for a different reconfiguration of expertise.

Scholars from social movement studies, digital anthropology and internet culture have emphasised the specificity of collaborative tools and modes of knowledge production that emerge out of alternative digital culture. Hacker anthropologist Gabriella Coleman tries to describe the specificity of ‘hacker sociality’ (Coleman 2013, 105) by paying attention to how technical hacker practices connect up the certain values. She points to particular forms of ‘legal’ thinking that emerged out of the hacker scene, when the connections between programming code and legal code became hybridised in a productive way and stabilised in the notion of code as speech (ibid.,169).[46] Using a different vocabulary to describe more explicitly the activist dimension of alternative digital platforms, Stefania Milan writes about ‘radical tech activists’. The term refers to ‘the groups and networks of individuals who provide alternative communication channels on the Internet to activists and citizens, on a voluntary basis and through collective organizing principles’ (Milan 2013, 12) According to her, they share a culture of emancipation and empowerment, since they do not depend on corporate or state-owned infrastructures (ibid.).

One interesting example of a conceptualised form of digital sociality, which links this sense of autonomous practices, as expressed in Milan, with analytic practices, is provided by Barret Brown’s distributed

[46] See ‘jurisgenesis’, explained in chapter three (§3.5.1).

think tank ‘Project PM’ (Operation Pursuant 2012).[47] His ‘Guide to the Establishment of Autonomous Online Entities’ contains a list of more than forty tools for collaborative work, suggestions for strategies, including methods for investigating the intelligence community. He has conceptualised this loosely connected hackathon style of working as a ‘Pursuance’ (Operation Pursuant 2012). Garrido refers to pursuances as part of a range of ICT-enabled ‘modalities of resistance’ (2015, 164):

> Quite recently, hacktivists and hacktivism collectives as well as collaborative networks that crowdsource open data analysis—which journalist Barrett Brown calls ‘pursuances’ (Brown 2012)—have shed light upon the strategies by which the state-corporate nexus deploys espionage and persona management (i.e., using online identities for purposes of astroturfing or disinformation) to infiltrate or hinder the activities of non-profit organizations and sociopolitical activism groups (Masnick 25th November 2013). (Garrido 2015, 163)

Zooming out from the radical tech and hacker scene, Lovink and Rossiter (2011) argue that ‘network culture’ is characterised by a set of specific and recurrent methods (see also §2.1.2). They mean that many alternative platforms for collaboration, meeting spaces, and methods for knowledge production may develop in the niches and margins as opposed to institutionalised and commonly known platforms. These methods for collaborative production, or in some cases, ‘protocols of radical tech activism’, whether they are physical meet-ups or digital platforms, are important because it is from this loosely connected working environment that the tools and interpretative devices emerge for the analysis of leaked data.

A second way in which the case of WikiLeaks differs from the ‘open data’ examples as discussed in the previous section, is the manner in which WikiLeaks acquires the data. Receiving open data is different than ‘opening up’ data. Gurstein highlights that:

> where Open Data has chosen to adopt a collaborative approach to its efforts—working with governments to find ways of “opening up” government data in ways which are presented as being mutually beneficial, WikiLeaks has taken a rather more radical and conflictual approach, forcibly opening information to broader public scrutiny against the wishes of its current owner, the US Government. (Gurstein, 29 December 2010)

Therefore, open data versus ‘opened data’ initiatives have different goals and articulate different expectations. Leaked data do not necessarily come in circumscribed databases and people, therefore, have to be

[47] Brown is a journalist who was indicted on 12 federal charges relating to the 2011 Stratfor hack (Courage 2015).

inventive. They have to deal with different file formats and improvise to make sense of the mess.

4.4.2 Arrangements of tools

How does WikiLeaks bring order to the mess? How does the platform present and organise datasets to facilitate analysis? WikiLeaks is known for publishing datasets of a very large scale. In 2015, WikiLeaks had published 2.325.961 US diplomatic cables and State Department records (Assange 2015, 1). The cables are important for WikiLeaks according to Assange because they form the heart of the (surveillance) state:

> the State Department represents, and even houses, all major elements of US national power. It provides cover for the CIA, buildings for the NSA mass-interception equipment, office space and communications facilities for the FBI, the military, and other government agencies, and staff to act as sales agents and political advisors for the largest US corporations. (Ibid., 4)

Assange's perceives the state as an organic empire (see also Fenster 2012): US Empire has an 'anatomy' (Assange 2015, 3). His theory is that leaks will damage internal flows of information because they induce paranoia among the different parts of the state body and that subsequently the body of the state will be suffocated (Fenster 2012, 24). Assange uses the Renaissance artist-investigator as an analogy: understanding the state means 'dissecting' the body of the state: 'One cannot properly understand an institution like the State Department from the outside, any more than Renaissance artists could discover how animals worked without opening them up and poking about inside' (Assange 2015, 5). He regards the publication of the cables as 'the vivisection of a living empire' (ibid.). Assange explains how the diplomatic communications can be seen as a residue of this living empire: they are bi-products of the state's daily operations. In other words, they provide 'behavioural data' about the state. Moreover, Assange argues that they are not distorted by public relations concerns because they were internal documents and not meant to be public (5).[48] The truth promise of this repository is that these bi-products of daily operations are expected to reveal a reality that would otherwise remain unseen. It is this promise of pattern-recognition that WikiLeaks needs to facilitate to make the data more than a data dump.

[48] It is a truth promise that is similar to the one we sometimes hear about behavioural online data: different from opinion polls in which people can choose and adapt their answers in relation to the researcher or the research, it is harder to make by-product data lie.

So how does WikiLeaks help people witness the state? In "Indexing the Empire", WikiLeaks spokesperson Sarah Harrison provides a kind of manual on how to deal with the data. To help people understand and analyse the big datasets, WikiLeaks offers a set of tools. One example is 'The Public Library of US Diplomacy', or 'PlusD', a large collection of internal documents from the US Department of State (Harrison 2015, 145), which includes Cablegate. WikiLeaks began publishing the searchable archive in 2010 and in 2015 it contained more than 2,3 million documents (ibid.). What Harrison's explanation of PlusD shows is that one needs to learn a specific kind of language to understand the material. This includes understanding the anatomy of the cables, their classification fields and acronyms, and a language of secrecy. The material in this archive is of a specific kind. As Harrison explains, since diplomats have written these texts, the documents should be seen as assuming prior knowledge and as commentary on unfolding events (Harrison 2015, 147). In other words, the style of language assumes a certain common understanding referring to things outside of the text.

To facilitate the analysis of the archive, WikiLeaks has added features for navigation, such as metadata fields that could demarcate datasets. WikiLeaks has also made corrections when the metadata are messy, for instance by adding tags with consistent spelling. In that way they improve the data and enlarge the possible findings one can make when using search software. Indexing through metadata also allows one to make the connections between cables and opens up the possibility to follow some of the story lines. Through these interventions, WikiLeaks is 'adding value' to datasets:

> Some of our hardest work goes toward adding value to datasets and making our publications more accessible and usable. This involves researching the structure of the data, designing and implementing search engines, optimizing metadata, and adding a large number of features to make the data easier to navigate and explore for researchers, journalists, human rights groups, historians, students, and other. (Harrison 2015, 157)

Sorting by metadata categories allows for sorting by tags, for example when one is interested in a certain period or region. One interesting remark in Harrison's piece relates to how data that seems uninteresting in the first instance could become interesting through a detour of sorting other types of data. She mentions that sorting by highest classification could potentially give clues to information that might be found in *other* – unclassified – documents:

> Even if a cable is marked "SECRET//NOFORN," this does not mean that the information contained within it will be more sensational or interesting

> for your purposes than information contained in a document with a lower classification level. (…) But there will often be other cables from that period, possibly at a lower classification level – or even unclassified – that contain important comments by a senior diplomat shedding light on a US perspective on a national issue, or in aggregate disclosing an historically significant or important insight. (Ibid.,151)

In other words, metadata categories could be a starting point for finding what is *not being said* in the classified files, but *might be commented about* in unclassified files. So, what is kept secret is studied by linking a certain set of secret files with related unclassified files.

Cablegate contains diplomatic cables, which conform to a certain format of communication that allows the indexing of metadata. But not all datasets are formatted. They can be composed in different ways and data are assembled in different ways. Therefore they require different kinds of analysis and navigation structures. People engaged with analysing leaked data have to come up with ways to analyse them. For example, *The Spy Files* (2014), another repository published by WikiLeaks is composed differently and leads to other analytic practices. It is an unformatted dataset. The publication is a collection of surveillance technology promotion material. The Spy Files consist of hundreds of documents and they were released in a few tranches from 2011 till 2014. Not only do the Spy Files consist of different kind of files (such as papers, brochures, contracts and video material), they were also assembled at different places. Staff from the NGO *Privacy International* collected the files when going undercover to attend surveillance trade shows. These are gatherings at which tech companies showcase their tools and at which governments and regimes of various descriptions attend. Bringing these different strands of information together, they give insight into the surveillance industry. The files have been visualised with an export map showing where surveillance companies are located and to which countries they sell their interception tools ('WikiLeaks Counter Intelligence Unit (WLCIU) Location Tracking Map'). They have also enabled the 'tracking' of surveillance contractors. These so-called 'targets' are available on the WikiLeaks website (see the 'WikiLeaks Counter Intelligence Unit (WLCIU) Location Tracking' (WikiLeaks 2013)). So whereas Cablegate has produced a particular study of state operations (through cables), the Spyfiles have lead to counter-intelligence practices (through tracing technological devices). The Spyfiles have given activists a sense of people, tools and the locations involved in the spyware industry through the promotion material.

The Spyfiles have informed other surveillance research too. So it is important to not regard the separate leaks as isolated. Surveillance researchers look for relationships and matches. For instance, in 2012 the Citizen Lab (University of Toronto) found traces of malware in e-mails sent to Bahraini pro-democracy activists (Citizen Lab 2012). The malware was hidden in an executable jpg-file. When loaded, the malware would create a directory that was able to harvest data from the computer of the target. Traces of the code indicated it was the 'FinSpy' tool, part of a package called FinFisher: IT Intrusion offered by the company Gamma International. The Spy Files have shown further specifications of the functionalities of FinFisher. It includes tapping somebody's microphone and key logging what the person would type. (For a list of functionalities and explanation of the malware research see the report by The Citizen Lab (2012).) In 2014 a hacked file with information about spyware company Gamma Group was leaked (through a torrent file) and published on the web. It contained various kinds of information such as price lists, names of clients and customer support data. It turned out that they were responding to customer support questions from Bahrain. The hacked dataset also included the names of Bahraini targets, amongst which opposition leaders, human rights lawyers and members of pro-democracy advocacy groups (Internet Protection Lab 2014). They were plotted on a (Google) map to show how Bahrain reached out to targets internationally through internet surveillance. Taken together, the report, the Spyfiles, and the hack provide a larger picture of the tool, its possibilities, its implementations and its reach. Therefore leaked data can enrich other existing data or hypotheses.

Another dataset on surveillance practices, this time again a 'formatted dataset', is the repository of the *The Global Intelligence Files* (WikiLeaks 2012). This was a set of emails from the global intelligence company Stratfor. The Global Intelligence Files exposed the mundane practice of confidential intelligence work by the firm. Stratfor monitored the online activities of various activists. WikiLeaks itself was part of Stratfor's working domain: more than 4.000 e-mails dealt with WikiLeaks or its spokesperson Assange (WikiLeaks 2012). Besides drawing attention to the way (individual) activists were surveilled, WikiLeaks stated that these files give insights into the knowledge flow passing through the company involving governmental, diplomatic, corporate and journalistic actors. Accompanying the files, WikiLeaks provided a glossary to explain the internal terms and codes. However, as this list was incomplete, the public was asked to participate in decoding the files (through the Twitter hashtag #gifind*)* (ibid.). As such, the issue of 'how to read the data' and decoding the files was presented as an exercise for the public.

Analytic tools and analyses do tend to pop up on other websites than WikiLeaks. 'Cablesearch' was a well known a search engine for

Cablegate and information about WikiLeaks' redactions (Heemsbergen 2014, 1352). 'Cabledrum' does a similar thing. This website also published a 'Gifind news index' and WikiLeaks Press (an organisation endorsed by but not tied to WikiLeaks) provided analyses on the Stratfor leak (WikiLeaks Press). At the time of writing, however, Cablesearch and WikiLeaks Press appear to be offline.

4.4.3 The production of radical expertise

So, what kinds of data practices contribute to WikiLeaks' transparency agenda? To whom is this transparency device directed? And: What does that imply for the data public? With respect to the first question a few examples have been discussed. WikiLeaks is adding value to data, thereby allowing persons to sort data and to link secret files with unclassified files. Furthermore, verification strategies are required to give leaked data a more trustworthy status thereby stimulating connections with other data. This could be other leaked data or this could be legally acquired data. Crucially, pieces of data are being fitted together: datasets need to be connected to find matches. Take the Finfisher Malware whose traces ran through the Spyfiles, the Citizen Lab research and the Hacking Team Leak.

What does the above imply for the way subjectivities are shaped in WikiLeaks' data public? The need for verification and especially the awareness of what is *not said* in the documents, such as the subtext in the cables, makes the way in which knowledge is constructed on the basis of leaked data different from the examples that are highlighted in the open data study. With the publication of leaked data there is a continuous awareness of incompleteness that needs to be uncovered. For example, through their sorting practices WikiLeaks politicises data that do not seem to be of interest in the first place. Consider Harrison's remarks about how one can search for relationships where they might not be obvious. When sorting data by secrecy and taking that as a point of departure for looking at unclassified datasets, the assumptions about what is interesting information is reversed, and the sense is that even the less secret files become sources to be carefully dissected. Therefore, as with open data, there is a sense of vigilance that speaks from the prescribed procedures, but it is not directed towards preventing the state from future failure. The sense of vigilance that operates through the analysis of opened data is directed towards the data itself: What is not in the data? And are the data trustworthy in the first place? The ambiguity of, on the one hand setting hope in data liberation and analysis goes together with an awareness of incompletion and this makes working with leaked data fundamentally different. Making data public goes together with an articulation of distrust.

To whom is this transparency device directed? If digital devices take part in the emergence of data subjects, as Ruppert (2015) suggests, and allow for a particular form of witnessing, in the case of WikiLeaks, this subject should take on the role of the *investigator-interrogator* rather than that of the *witness*. One is expected to invest in the data and not to look for the obvious, but one is expected to tinker with the various categories. Therefore WikiLeaks seems to aim for producing data *experts*.

Through their repositories, software and accompanying texts, WikiLeaks is explaining how to read the leaked datasets in a particular way. With explaining the basic steps they contribute to constructing investigatory data literacy. Assange's ideals of scientific journalism are well documented (Fenster 2012). This refers to the inclusion of primary sources as a verification method and an image of the existence of 'true data' (ibid.). Also the manual expresses this image: one is also expected to read the data as a scientist, by not starting in the obvious way, but by learning a particular language, performing a vivisection, doing field work, clustering categories and making connections to find something unexpected. At the same time, the working environment for leaked data is dynamic and experimental. There are clusters of small and temporary research collectives that add tools and knowledge to the repositories of leaked files. Data are added, analyses added, search engines and maps are added. Research collectives are not static: websites that provided insights into leaked files decease. So yes, there is a reconfiguration of expertise, but it is very unstable.

For these reasons, I propose to see WikiLeaks as an experiment or 'pioneering' project with the production of radical expertise. As Hess argues about other (yet less antagonistic) 'pioneering communities' (Hepp 2016): 'the use of media by these pioneering groups creates a horizon of *possibility* to which the everyday media appropriation of others orients itself, or at least can do so' (Hepp 2016, 919). Along the same lines, WikiLeaks has pioneered new forms of reading; it has created a horizon of possibility for radical expertise.

I see the notion connected to that of radical transparency. I noticed the use of the term when privacy advocate Van Daalen defined a 'radical expert' as 'someone who gains deep juridical or technical knowledge, which could potentially lead to, from a certain perspective, radical conclusions' (Van Daalen cited in Martijn 2014; personal translation). However, I find it an appealing term for rethinking practices tied to radical transparency. If the notion of 'radical transparency' refers to disruptive and unauthorised disclosures, 'radical' refers not just to the conclusions following from knowledge, but also to the status of the data and the skills that emerge from working with those data and from within disruptive settings. As argued in the previous section, we should pay attention to the (data) literacy and investigatory skills that emerge from these interventions. Radical expertise can be seen as a (multi-layered) process. Radical expertise is ‘radical’ in the sense that it is the production of knowledge from data people are not authorised to look at, it attempts to share this type of literacy, and it might lead to radical political insights or consequences. It resonates with the, rather radical, risk that is often involved of working with those datasets (Interference Conference 2014). It is a form of ‘expertise’ because, despite this radical dimension, the kind of auditor that WikiLeaks tries to enact must be a scientific one, a knowledgeable one, hence an expert. In other words, WikiLeaks wants to train people to read data that one is not allowed to look at, which is radical. At the same time, it aims to enact a public that treats the data in an expert way. Radicalness and expertise are combined. The term radical expertise contains no false promises of total openness, which is one specific reading of radical transparency, nor suggestions of radical participatory decision-making, which is the other reading of radical transparency (as in Birchall’s notion (2014)). The data are publically available, but the expertise is not, at least, it is not yet, since the way to get there is not very straightforward: tools die and manuals are published in a book and not on the WikiLeaks site. Data liberation and the democratisation of expertise are two different things, and the latter is still work in progress.

4.5 Conclusion: Sousveillance as radical data literacy?

It is worthwhile noting that whereas the website of WikiLeaks described itself as a transparency initiative up until 2015, the current website speaks of itself as a ‘multi-national media organization and associated library’ (WikiLeaks 2016). This represents a shift in vocabulary moving from transparency to a discourse of knowledge production.

The analytic practices that I discussed show how WikiLeaks data are used to make behavioural patterns of the state understandable, rationales of secrecy visible, and find matches with long-standing investigations into surveillance. Transparency originally refers to material allowing light to pass through so that objects behind can be ‘distinctly seen’, but the texture of radical transparency in the case of WikiLeaks can be better described as a data enrichment. It entails the addition of categories, tools and the making of connections to different new datasets. With open governmental data, as described (and criticised) by Ruppert (2015), the data come pre-packaged with analytical tools. In the case of the leaks, neither the data nor the tools are ‘given’, and people have to come up with methods of categorising and ordering the data and explaining to others how to do it. However, these methods are also not unmediated but come from a particular working environment. Rather than leaving the data untouched, WikiLeaks and researchers that make use of its publications are engaged with enriching and translating the data. Rather than having to deal with amorphous data (or a ‘data dump’), the datasets are all very different in form and content, of which I gave three examples, and therefore they require different kinds of treatments and analytic techniques.

The imagined WikiLeaks data public is both expert and radical. WikiLeaks datasets require trained and investigatory subjects: experts. At the same time working with this data is radical because they enrol people in how to read unauthorised data. There is obviously a tension here. To recall Barry (2010), the reading of data in an ‘ordered manner’ has in the past been associated with imaginary visions of a rational and civil public. The enactment of rational data subjects would constitute a public that behaves well. WikiLeaks provides a complicated picture because it expresses something different. It expresses a combination of expertise and radicalness, rationality and incivility. The question is therefore whether and when this combination can be a mobilising force.

Other authors have framed WikiLeaks as a phenomenon of sousveillance because it has turned a watchful eye towards the watchers. What does taking into account WikiLeaks’ datasets and analytic steps imply for sousveillance theory? For sousveillance theory this means that it is not only the interventionist dimension which matters – the question of whether governments, or the surveillance industry, actually feel threatened by WikiLeaks or not – but it is important to know how this form of radical expertise is being shared and whether the analytic techniques develop in a progressive manner over time. Will this remain in the domain of hackers and journalists or could and should these skills be institutionalised in educational or research settings? Will they stabilise as new ‘protocols’ for doing surveillance studies?