The Big Data Paradox

Juggling data flows, transparency and secrets

CREDIT: PHOTO MCD/J. VAN HELVERT

On more than one occasion the *Militaire Spectator* has paid attention to the digital domain, its characteristics and the role of the military within the digital arena. Recently, articles related to this subject have been published on military operations, defence policy, cyber issues, and big data analytics.[1] It is abundantly clear that the challenges and opportunities for the digital era are of great relevance and concern for the military domain. This article looks into one important aspect of this domain – big data. It will highlight the possibilities and pitfalls of this phenomenon by putting emphasis on the (neglected) paradox at the heart of big data developments. It does so from a theoretical, indeed more philosophical, point of view. Far from lessening its relevancy to military practice, it is argued that such an approach will help understand the dynamic and complex twenty-first century digital arena in which military activities play an important and constituent part.[2]

*Dr. A. Claver**

Big data is a captivating phenomenon in many respects. Data are to this century what oil was to the previous one: a driver of growth and change.[3] The interconnectedness of exponentially growing data flows entails great promises for personal, commercial, as well as governmental use. But is it clear what we mean by big data? There are many perspectives on big data imaginable and at the moment there is no single accepted definition. Big data can be looked upon from a technical, socio-technical as well as governance level, either generic or in detail, and is considered both an opportunity and a threat. Change holds promises, but is by the same token unsettling and intimidating. From one perspective the advent of big data ensures improved transparency from which society will greatly benefit. From another big data forebodes the horrifying perspective of an all-knowing and possibly authoritarian, regime. In other words: big brother will be watching you… soon!

The attractiveness of data flows as a powerful means of security improvement based upon

* Alexander Claver works at the Dutch Ministry of Defence and is currently following the Executive Master Cyber Security at Leiden University. This article has been written in the personal capacity of the author.

1 See for example Paul A.L. Ducheine, 'Defensie in het digitale Domein', in: *Militaire Spectator 186* (4) 2017, p. 152-168. Paul A.L. Ducheine and Kraesten Arnold, 'Besluitvorming bij cyberoperaties, in: *Militaire Spectator 184* (2) 2015, p. 56-70. Erik S.M. Akerboom, 'Cyber security. Samenwerken voor een veilige en vitale cybersamenleving', in: *Militaire Spectator 181* (12) 2012, p. 532-536. Allard D. Dijk, Bas Meulendijks and Frans Absil, 'Lessons Learned from NATO's Cyber Defence exercise Locked Shields 2015', in: *Militaire Spectator 185* (2) 2016, p. 65-74 and Paul C. van Fenema et al. 'Big data analytics en Defensie. Visie en aanpak', in: *Militaire Spectator 184* (9) 2015, p. 374-387.

2 In full agreement with Peer H. de Vries (Brig Gen Ret.), who argued in the *Militaire Spectator* that military practice should be considered from another, philosophical, perspective in order to broaden and deepen insights into one's actions. Peer H. de Vries, 'Filosofie voor Militairen', in: *Militaire Spectator* 184 (10) 2015, pp. 421-428.

3 'Fuel of the Future. Data is giving rise to a new economy', in: *The Economist*, 6 May 2017.

(automated) pattern recognition, analysis, and increased predictive value, is well recognized. This attractiveness from a security point of view has aroused suspicion regarding possible infringements on civil liberties. It is therefore closely monitored by e.g. human rights activists and privacy watchdogs. They fear that fundamental rights will be jeopardized by the increased leverage of the state in tracking its citizens for a range of purposes, e.g. maximizing tax returns, minimizing social benefit payments, countering radicalization, or punishing criminal behaviour. Notwithstanding the validity of these purposes, these advocates stress that the position of the individual versus the state has deteriorated. They plead for more and better safeguards. This includes transparency and oversight when it comes to the handling of big data flows by the state in general, and the security and intelligence community in particular.

To counter radicalism, terrorism and other threats in general, Dutch society seems to value a well-functioning intelligence and security apparatus more than before

Big data developments: increased transparency and secrecy

In 2013 the authors of a short essay cautioned against what they called the utopian rhetoric of big data.[4] Without denying that big data holds major potential for the future they claimed that the benefits of large dataset analysis were overstated. To illustrate their point the authors discussed three, in their opinion understated, values: i.e. individual privacy, identity, and checks on power. The description of these values stressed the presence of self-contradictory traits (i.e. paradoxes) in each of the values discussed.[5] This matches the definition of a paradox as a situation or statement that seems impossible, or is difficult, to understand because it contains opposite facts or characteristics.[6]

This article takes a different approach. It does not focus on aspects of dichotomy, but highlights the complementarity and/or compatibility of seemingly opposed notions. This perspective fits a different definition of a paradox: 'A statement that is seemingly contradictory or opposed to common sense and yet is perhaps true.'[7] This is illustrated by showing that the notions of transparency and secrecy do not exclude one another but comprise two sides of the same coin.

In order to counter radicalism, terrorism and threats in general, Dutch society at present seems to value a well-functioning intelligence and security apparatus more than before. This apparatus consists of a police force and military, complemented and assisted by the proportionate activities (as circumscribed by law) of the two

4 Neil M. Richards and Jonathan H. King, 'Three Paradoxes of Big data', in: *Stanford Law Review Online 66*, 2013, pp. 41-46.

5 Richards and King, 'Three Paradoxes of Big Data'. 1) the Transparency Paradox, which concerns the collection of private information by means of big data operations that are themselves shrouded in secrecy; 2) the Identity Paradox, which emphasizes big data results operations, but ignores the fact that these techniques seek to identify, and therefore work at the expense of individual and collective identity; 3) the Power Paradox, which deals with the characterization of big data transforming society without paying attention to the accompanying power effects favouring large government and corporate entities at the expense of ordinary individuals.

6 *Cambridge Dictionary*. See: https://dictionary.cambridge.org/dictionary/english/paradox.

7 *Merriam Webster Dictionary*. See: https://www.merriam-webster.com/dictionary/paradox.

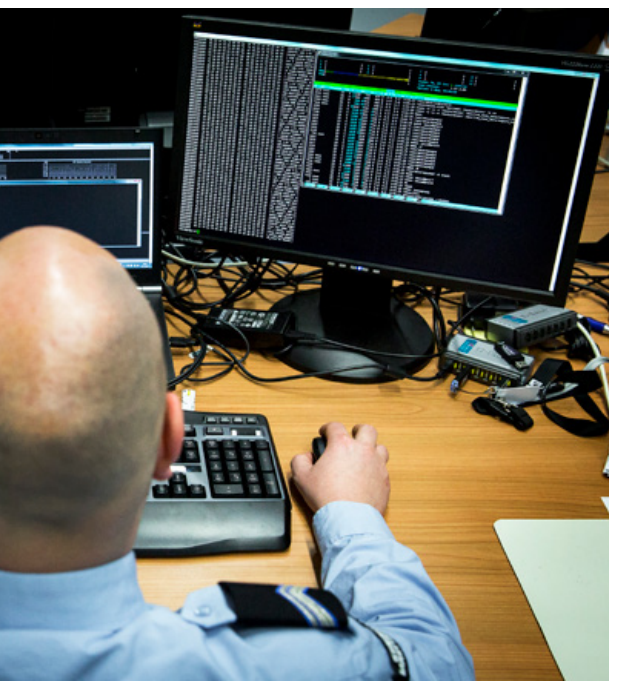

PHOTO MCD/V. KUYPERS

Dutch intelligence and security services: AIVD and MIVD.[8] The toolbox of these organizations naturally includes the full potential of the digital era exemplified in, for example, big data developments. Big data, however, is by definition connected to a free and transparent flow of information. This does not seem to relate well with the behaviour of intelligence and security services.

This article discusses big data developments in connection with the simultaneous need for transparency and secrecy. It will zoom in on the concept of big data whose characteristics need to be understood better. Clearer definition, sharper demarcation, and the use of conceptual modeling will help the current debate wherein the contributors tend to speak different languages. The article will also show that there is an apparent, yet not absolute incompatibility of transparency and secrecy, even though it is commonly perceived and/or framed as such in the public debate. Recognition and awareness of this big data paradox will serve current and future discussions. This is exemplified by the ongoing Dutch debate with regard to a substantial revision of the country's first intelligence law of 2002.

The first section of the article briefly points out the historical roots, definition(s), and main characteristics of big data, including the important question of correlation versus causality surrounding the phenomenon. The second section deals with the Dutch debate regarding big data policy and definition. A conceptual three-layer model of cyberspace is offered to help structure the intelligence law discussion by showing that it is predominantly driven by technical issues (e.g. database design, intercept possibilities, collection, selection and search protocols). Attention is put to the fact that these issues manifest themselves on the socio-technical level (privacy and security issues). The third section addresses the paradox of transparency and secrets, linking it to the important governance level. The final section offers some concluding remarks.

Big data; some characteristics

The first attempts to quantify the growth rate in the volume of data produced have been traced back to the 1940s when the term 'information explosion' was also introduced.[9] Around 1970 computers became inextricably tied to this concept when Gordon E. Moore coined his famous, and still valid, rule of thumb that overall processing power for computers will double every two years (so-called *Moore's Law*).[10] The first studies to estimate the amount of new information created annually worldwide appeared in 2000 and 2003. The researchers involved (including Hal Varian, now chief economist at Google) concluded that the amount of new information created annually in 1999 amounted to 1.5 billion gigabytes and had doubled to 3 billion gigabytes in 2002.[11]

8 AIVD and MIVD are the Dutch acronyms for the civil and military intelligence and security services. AIVD = Algemene Inlichtingen- en Veiligheidsdienst and MIVD = Militaire Inlichtingen- en Veiligheidsdienst.

9 Gil Press, 'A Very Short History of Big data', in: *Forbes*, 9 May 2013.

10 See: http://www.mooreslaw.org/.

11 *How Much Information?*, School of Information Management and Systems, University of California (Berkeley, 2000 and 2003). See: http://groups.ischool.berkeley.edu/archive/how-much-info/ and http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/.

Defining big data

There is no definition of big data agreed upon yet.[12] NASA scientists appear to have coined the notion first in a paper published in 1997.[13] However, it took the term more than a decade to become mainstream, and – ultimately – part of popular culture. The current marketing popularity of big data has little in common with the original scientific description of the information revolution, computer accomplishments, application development (commercial or otherwise), and the possible implications connected to this.[14] Big data today appeals above all to the possibility of entering a new world full of promises, economic opportunities, and profit.[15]

A number of current definitions appear to have in common the focus on the magnitude of the amount of data, measured nowadays in thousands of petabytes (1 petabyte = 1,000 terabytes = 1,000,000 gigabytes), and the associated management problems. This led industry analyst Doug Laney in 2001 to focus on Volume, Variety, and Velocity as the key data management challenges.[16] His well-known '3Vs'-definition of big data is far from outdated: 'Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.'[17]

'Big data is high-Volume, high-Velocity and/or high-Variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation' – Doug Laney

Other definitions tend to focus less on the massive amounts of data and more on the opportunities and challenges they offer[18] on the technical, socio-economic and governance level of cyberspace (see paragraph *Modelling Cyberspace* below.) These definitions point to the importance of what can actually be done with the data and why its size matters.[19] They emphasize the fact that cyberspace data – and the information that can be extracted from it – are giving rise to a new economy.[20] This so-called data economy derives its strength from self-enforcing network effects: 'using data to attract more users, who then generate more data, which help to improve services, which attracts more users.'[21]

Likewise, the Dutch investigative journalist Dimitri Tokmetzis stresses that his informants are not considering data amounts as such. They refer to big data as connected developments in computer technology, consisting of ever more advanced hardware and software enabling the

12 Ernst M.H. Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving', *WRR-rapport, nr. 95* (Amsterdam, Amsterdam University Press, 2016) pp. 33-35. Also: Dimitri Tokmetzis, 'Wat is big data?, in: *De Correspondent*, 11 November 2013. And Gil Press, '12 Big data Definitions. What's Your's?', in: *Forbes*, 3 September 2014.

13 Michael Cox and David Ellsworth, 'Managing Big data for Scientific Visualization.' *ACM SIGGRAPH*, 1 May 1997, 21-38.

14 Though anything but mainstream, the scientific tradition in this respect is not dead. See for an intriguing account of the human and technological limits of computing the mental exercise by Nick Bostrom, 'Are You Living in a Computer Simulation', in: *The Philosophical Quarterly 53* (211) 2003, 243-255.

15 'Data, data everywhere. Special Report: Managing Information', in: *The Economist*, 27 February 2010. 'Fuel of the Future', in: *The Economist*. 'The world's most valuable resource is no longer oil, but data', in: *The Economist*, 6 May 2017.

16 Doug Laney, '3D Data Management: Controlling Data Volume, Velocity, and Variety', Application Delivery Strategies 949 (Stamford, META Group, 2001). See: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. In the next decade Laney continued to work on his concept and expanded it to '12V's: 'Deja VVVu: Others Claiming Gartner's Construct for Big data', in: *Gartner* (January 2012) http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/.

17 See: http://www.gartner.com/it-glossary/big-data/.

18 Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving', p. 33-35.

19 Viktor Mayer-Schönberger and Kenneth Cukier, *Big data. A Revolution That Will Transform How We Live, Work, and Think* (London, John Murray Publishers, 2013). Seth Stephens-Davidowitz, *Everybody Lies: Big data, New Data, and What the Internet Can Tell Us About Who We Really Are* (New York, Harper Collins Publishers, 2017).

20 Marshall W. van Alstyne, Geoffrey G. Parker and Sangeet Paul Choudary, 'Pipelines, Platforms, and the New Rules of Strategy', in: *Harvard Business Review*, 94 (4) 2016, 54-62.

21 'Fuel of the Future', in: *The Economist*.

collection of ever more data, and statistics, attaching meaning to dispersed data flows by relating them to each other.[22]

Correlation versus causality

Attaching meaning to data by relating, or correlating, them to each other touches upon a crucial element of big data and big data usage, one that has not escaped the attention of many authors. Distinguishing between correlation and causation is immensely difficult, and often correlation is mistaken for causation. At its core, however, a correlation merely quantifies the statistical relationship between two data points. When one data point changes, the other is likely to change as well in case of a strong correlation. In case of a weak correlation this change is less likely to occur. When considering correlations attention should be paid to the fact that even strong correlations might occur because of… coincidence.[23]

After all, correlation does not imply causation: it only implies probability. Probabilistic outcomes should, therefore, never be taken at face value, but have to be treated as indications of possible outcomes. As a result, any analysis based on statistical probabilities will, by definition, produce both false positives (e.g. criminalizing innocent people) and false negatives (e.g. allowing security risks to go unnoticed).[24]

Traditionally, analysis was driven by hypotheses, which were validated by collecting and analysing data. Insights were extracted from scarce, static, and poorly relational data sets with a specific question in mind. Scientific understanding today is driven more and more by the (over) abundance of data. When mining these data the main challenge will be how to cope with the variety, messiness, and uncertainty of the generated data set, bearing in mind that much of what is collected does not have a specific question in mind, or is the (unintended) by-product of another activity.[25]

Here we touch upon an important distinction between current big data and the infinitely smaller data sets used before.[26] Contrary to established scientific practice, big data analysis is not about validating hypotheses, but about finding interesting links and identifying patterns that might be relevant. As said, these analyses might provide unexpected correlations and insights, but run the risk of elevating correlations to causations, even though the causality of the linkages found remains uncertain.[27] Ultimately, big data shifts the focus of inquiry from causation to correlations. Formulating a (policy) response will thus depend more on the knowledge that something is happening rather than why it is happening.[28]

Some scholars view this positively[29] and compare the big data revolution to a classic scientific paradigm shift. According to Rob Kitchin, big data analytics enable an entirely new approach to making sense of the world. Rather than testing a theory by analysing relevant data, new data analytics seek to gain insights 'born from the data'.[30] Jim Gray argues that current data techniques and technologies are so different that it's worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration[31] (see table 1).

22 Tokmetzis, 'Wat is big data?', in: *De Correspondent*.

23 Viktor Mayer-Schönberger and Kenneth Cukier, *Big data. A Revolution That Will Transform How We Live, Work, and Think* (London, John Murray Publishers, 2013) pp. 52-53.

24 Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving', p. 38. See also: Dennis Broeders, Erik Schrijvers and Ernst Hirsch Ballin, 'Big data and Security Policies. Serving Security, Protecting Freedom', *WRR-Policy Brief no. 6* (The Hague, WRR, 2017) pp. 6-7. See: https://english.wrr.nl/topics/big-data-privacy-and-security/documents/policy-briefs/2017/01/31/big-data-and-security-policies-serving-security-protecting-freedom.

25 Rob Kitchin, 'Big data, New Epistemologies and Paradigm Shifts', in: *Big data & Society*, April-June 2014, p. 2. Mayer-Schönberger and Cukier, *Big data. A Revolution,* p. 70.

26 See for an interesting view on the lasting importance of small data in the era of big data developments by Rob Kitchin and Tracey P. Lauriault, 'Small Data in the era of big data', in: *GeoJournal* 80, 2015, 463-475.

27 Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving', p. 38.

28 Kevjn Lim, 'Big data and Strategic Intelligence', in: *Intelligence and National Security*, 31 (4) 2016, p. 622.

29 Jonathan Shaw, 'Why Big data is a Big Deal. Information science promises to change the world', in: *Harvard Magazine* March-April 2014. See http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal.

30 Rob Kitchin, 'Big data, new epistemologies and paradigm shifts', in: *Big data & Society*, April-June 2014, p. 2.

31 Toney Hey, Stewart Tansley and Kristin Tolle, (2009) 'Jim Gray on eScience: A transformed scientific method', *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, Microsoft Research, 2009) p. xviii-xix.

Paradigm	Form
1. Experimental science	Empirical method, describing natural phenomena
2. Theoretical science	Using models, generalizations
3. Computational science	Simulating complex phenomena
4. Data-intensive science	Data-exploration: unifying experiment, theory, and simulation

Table 1 Scientific Paradigm Shifts

Source: Compiled and adapted from Hey, Tansley and Tolle 2009; Kitchin 2014

Other scholars are less convinced. Martin Frické argues that so-called data-driven science is a chimera[32]. Methodologically speaking, it merely gathers more data, and does not in itself offer any explanations or theories, solve scientific problems, or aim to do anything of that nature. In his eyes, big data encourages passive data collection, and unsound statistical fiddling. Theory, experimentation, and testing remains needed as ever. The strength of big data lies, above all, in supporting this by providing access to (much) larger sample sizes, permitting cheaper and more extensive testing of theories, and allowing the continuous assessment of theories.[33] Nicholas Krohley admits to a wealth of data, but speaks of a poverty of insight. According to him, the 'fetishization of data' has led to increasingly complex patterns of correlation accompanied by increasing failure to contextualize. He wonders whether an exceedingly complex human environment can be broken down into binary patterns and then reconstructed in a remotely meaningful way?[34]

Definitions and debates aside, the inevitable conclusion so far must be that the information revolution is producing a data-driven society anchored in cyberspace, which will influence people's lives to a continuously increasing extent. For some this is a positive development heralding great promises.[35] Others highlight the negative aspects and warn against harmful consequences.[36]

Big data, cyberspace and secrecy: the Dutch case

Digital developments have neither escaped the Netherlands nor the attention of the Dutch government. The economic and societal potential of big data (e.g. maximizing tax returns, or countering radicalization through profiling) have been realized as well as the vulnerabilities with regard to the personal sphere (e.g. issues of privacy and equal treatment). The Dutch government is actively striving to accomplish a digitalized bureaucracy in the foreseeable future. The notion 'iGovernment' has become an accepted label in this respect.[37] Other clear indications of the government's digital awareness are its efforts at formulating big data policy, both in the private and public sector. The letter to parliament of then Secretary of Economic Affairs Henk Kamp, published in 2014, has been the point of departure with regard to the private sector.[38] Public sector policy regarding big data has been investigated by the Netherlands Scientific Council for Government Policy (Wetenschappelijke Raad voor het Regeringsbeleid or WRR). Being an independent advisory body the WRR

32 Wikipedia on Chimera: 'A monstrous fire-breathing hybrid creature of Lycia (Turkey), composed of the parts of more than one animal.'

33 Martin Frické, 'Big data and its epistemology', in: *Journal of the Association for Information Science and Technology*, 66 (4) 2015, pp. 651–661. Also Renato Dos Santos, 'Big data: Philosophy, Emergence, Crowdledge, and Science Education', in: *Themes in Science & Technology Education*, 8 (2) 2015, pp. 115-127.

34 Nicholas Krohley, 'The Intelligence Cycle is Broken. Here's How to Fix it', in: *Modern War Institute at West Point*, 24 October 2017. See: https://mwi.usma.edu/intelligence-cycle-broken-heres-fix/.

35 An outspoken positively inclined author is former Google data analyst Seth Stephens-Davidowitz, who published *Everybody Lies: Big data, New Data, and What the Internet Can Tell Us About Who We Really Are*.

36 A distinct negatively inclined author is mathematician and former hedge fund data scientist Cathy O'Neil, who related her experience in *Weapons of Math Destruction: How Big data Increases Inequality and Threatens Democracy* (New York, Crown, 2016). Additional background information can be found in the following interview: Gerard Janssen, 'Wiskundige Cathy O'Neil en de Weapons of Math Destruction', in: *Vrij Nederland*, 16 November 2016. See: https://www.vn.nl/cathy-oneil-en-weapons-math-destruction/.

37 Corien Prins et al., 'iGovernment', in: *WRR-Report 86* (Amsterdam, WRR/Amsterdam University Press, 2011). The WRR-website provides additional information (in Dutch) on the iGovernment issue including the official government stance. See: http://www.wrr.nl/publicaties/publicatie/article/ioverheid/.

38 'Kamerbrief over big data en profilering in de private sector. Brief van minister Henk Kamp (EZ) aan de Tweede Kamer over big data en profilering in de private sector, in relatie tot het recht op privacy en het recht op gelijke behandeling', 19 November 2014. See: https://www.rijksoverheid.nl/documenten/kamerstukken/2014/11/19/kamerbrief-over-big-data-en-profilering-in-de-private-sector.

was tasked to advise the government on this matter, which resulted in the publication of several reports in 2016.[39]

The WRR also looked into related cyber matters within the project *Freedom and security in the cyber domain*.[40] This resulted in a number of publications advocating the state's responsibility for the 'public core of the internet'. States need to involve themselves by making sure that the internet core – i.e. the central protocols and infrastructure considered to be public good – are safeguarded from state interference. The project emphasized the interconnectedness of technical, socio-technical and governance elements in the cyber domain and stressed that cyber policy issues are, by necessity, played out internationally and cannot be confined to the national level (for national security considerations) or left to market forces alone. The council, therefore, aimed 'to provide knowledge to assist in developing a coherent foreign policy for the cyber domain, one in which the interests of economic, physical and national security, on the one hand, and political and economic freedom, on the other, are weighed up against one another.'[41]

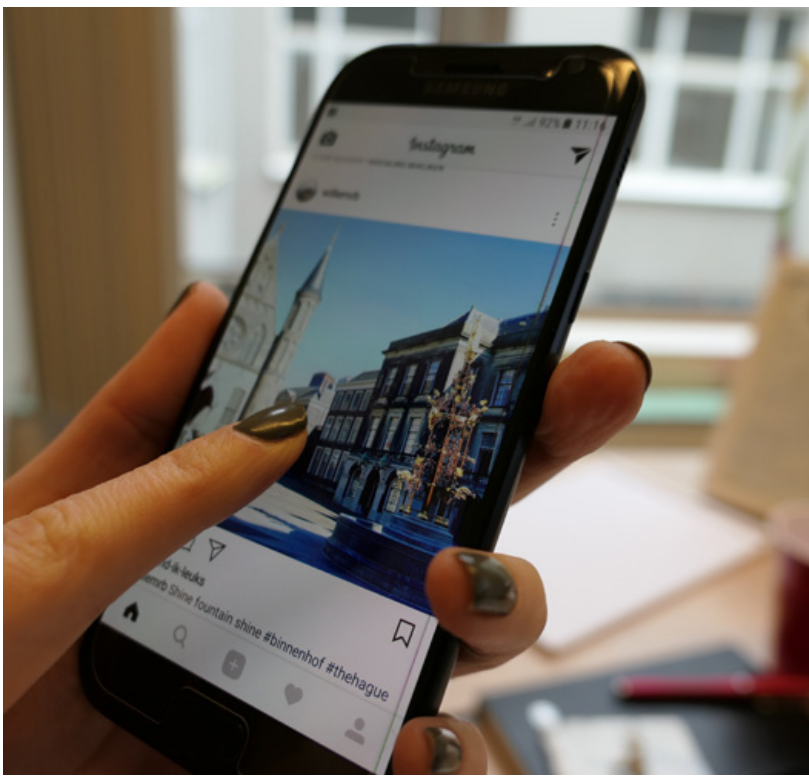

PHOTO RIJKSOVERHEID

People are sharing more and more data in the digital domain by social media

A related topic within the current Dutch public debate is the new intelligence law.[42] Within this debate the earlier mentioned concepts of big data, cyberspace, transparency and secrecy – and by proxy, freedom and security – are linked and hotly contested. The inability so far to find common ground owes much to the failure of clearly demarcating and/or defining the issue(s) at stake. Two examples will suffice to illustrate this.

Defining big data in the Netherlands

The previous section on big data has clearly shown the elusiveness of the notion. Notwithstanding Doug Laney's clear and concise '3Vs'-definition of big data, no *communis opinio* on the subject exists to date. The arduous attempt of the Dutch government to clarify the issue in relation to the revision of the intelligence law merely confirms the fuzziness of the concept and the difficulty of demarcating it.

Within the *Memorie van Toelichting* (Explanatory Notes) concerning the new intelligence law big data is described as follows: '…the phenomenon that manifests itself among others in the fact that the amount of data is growing exponentially, data collections are becoming bigger and more complex as a result of which relevant data can no longer be stored physically or logically in a location or in a system….' [43]

39 Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving'. An English translation of the aforementioned report is: Broeders, et al., 'Big data and Security Policies'.

40 Dennis Broeders et al. 'De Publieke Kern van het Internet. Naar Buitenlands Internetbeleid', in: *WRR-rapport nr. 94* (Amsterdam, Amsterdam University Press, 2015). English translation: Dennis Broeders, 'The Public Core of the Internet. An International Agenda for Internet Governance', *WRR-Policy Brief no. 2* (The Hague, WRR, 2015). See: https://english.wrr.nl/publications/reports/2015/10/01/the-public-core-of-the-internet.

41 Broeders et al. 'De Publieke Kern van het Internet.'

42 See https://zoek.officielebekendmakingen.nl/dossier/34588. See also: https://www.internetconsultatie.nl/wiv/details.

43 Tweede Kamer der Staten-Generaal 2016-2017, 'Regels met betrekking tot de inlichtingen- en veiligheidsdiensten alsmede wijziging van enkele wetten (Wet op de inlichtingen- en veiligheidsdiensten 20..)', in: *Memorie van Toelichting, Kamerstuk 34588-3*, p. 130. See: https://zoek.officielebekendmakingen.nl/kst-34588-17.html (author's translation).

Data	large structured and unstructured data from different sources
Analysis	data-driven, automated searches for correlations, in particular with the potential for analysis of the present (real-time analysis) and the future (predictive analysis)
Actionable Knowledge	analysis should result in actionable knowledge, to be made applicable for decision-making at group or individual level.

Table 2 Big Data Characteristics (WRR)

Source: Hirsch Ballin et al. 2016

This wording – 'hidden' as a subordinate clause on page 130 – is anything but exact. It does not constitute a clear-cut definition and is contrasted on the very same page by referring to a characterization of big data in a WRR-report. The authors of this report hold that the concept of big data is ambiguous. Instead of providing a definition, they therefore chose to focus on what they consider the three main characteristics of big data: data, analysis, and actionable knowledge[44] (see table 2).

Big data here is not seen as a well-defined – or even a definable – concept, but as the dynamic interplay between the three displayed characteristics. According to the authors, this leaves room to discuss the use of data analysis in public policy making.[45] This characterization is subsequently accepted in the *Memorie van Toelichting* with the concluding remark that an interpretation of big data as provided by the WRR is in line with the assumptions of the proposed law revision.[46]

Modelling cyberspace

Big data developments are inextricably connected to cyberspace. But, most people are unable to answer basic questions, such as: What is cyberspace? How is cyberspace being governed? Who are its attackers and what are their motives? How does the (underlying) technology work?, etc.[47] It stands to reason that without the existence of generally accepted answers, clear-cut definitions, and suitable demarcations, it becomes difficult to see eye to eye with each other when perceptions and/or interpretations differ.

A conceptualization of cyberspace is, therefore, urgently needed as will become clear from the debate in the Netherlands regarding the new (revised) intelligence law (see paragraph *Debating Secrecy* below). A promising start in this respect has been the approach of Van den Berg et al. In an award-winning paper, published in 2014, the authors suggest a conceptual model dividing cyberspace into twelve cyber subdomains, arguing that these domains need to be analyzed on three separate, but interconnected layers: a technical, socio-technical, and governance layer[48] (see figure 1).

From the model follows that the traditional inclination to concentrate on and investigate the technical aspects of cyberspace does not suffice. It is imperative that socio-technical and governance aspects are considered as well. Historically, the technical layer focusing on robust communication services and information security has received the most attention. However, global interconnectivity and huge numbers of applications with an easy to use human interface have given rise to a socio-technical layer. Here, people perform a vast range of cyber activities, which translates into the complex interaction of billions of people active in cyberspace with the available IT-systems – i.e. data storing and data processing systems, including to an increasing extent intelligent and autonomous decision-making systems. The governance layer consists of the large and complex number of human actors and organizations that govern both the technical and socio-technical layers.[49]

44 Ernst Hirsch Ballin, et al., 'Big data in een Vrije en Veilige Samenleving', pp. 33-35. [Author's translation].
45 Broeders, Schrijvers and Hirsch Ballin, 'Big data and Security Policies', p. 6.
46 Tweede Kamer der Staten-Generaal 2016-2017, 'Regels met betrekking tot de inlichtingen- en veiligheidsdiensten alsmede wijziging van enkele wetten', in: *Memorie van Toelichting Kamerstuk 34588-3*, p. 130.
47 Jan van den Berg, et al., 'On (the Emergence of) Cyber Security Science and its Challenges for Cyber Security Education', in: *NATO STO/IST-122* and *Cyber Security Academy* (Den Haag, 2014) See: https://www.csacademy.nl/images/MP-IST-122-12-paper-published.pdf, p. 1.
48 Van den Berg, 'On (the Emergence of) Cyber Security Science and its Challenges, p.2.
49 Ibidem.

Debating secrecy

To a substantial degree the intelligence law debate in the Netherlands centers around the technical possibilities of large-scale automated data communication interception and the risks involved in approval of such methods. The privacy versus security argument is at play here. Adhering to its 'iGovernment' principles the Dutch government has put its intelligence law proposal up for online consultation.[50] According to the Ministry of Internal Affairs this resulted in 1,114 responses, evenly divided between confidential responses and responses open to public scrutiny.[51]

The cost incurred by companies, cooperation with foreign services, technical issues and their potential consequences were eagerly debated and subjected to criticism, next to matters of oversight (both ex-ante and ex-post).[52] A number of issues (in random order) received particular attention:

1. Large-scale interception of cable communication;
2. Search through large amounts of data;
3. Automated access and analysis of databases;
4. Obligating companies and organizations to decrypt communication.

Criticism on these issues boiled down to:

1. Matters of necessity, proportionality, and subsidiarity;
2. Question marks concerning privacy goals;
3. The technical impossibility of compliance.

This emphasizes the importance of a well-functioning oversight mechanism, given the fact that intelligence and security services already possess far-ranging powers by law regardless of the actual outcome of the intelligence law revision. Though governance encompasses more than oversight, the third layer within the cyber domain (see figure 1) has at long last appeared on the horizon as an integral part of an indispensable system of checks and balances.

The intelligence law debate, however, has remained focused on the technical and to a

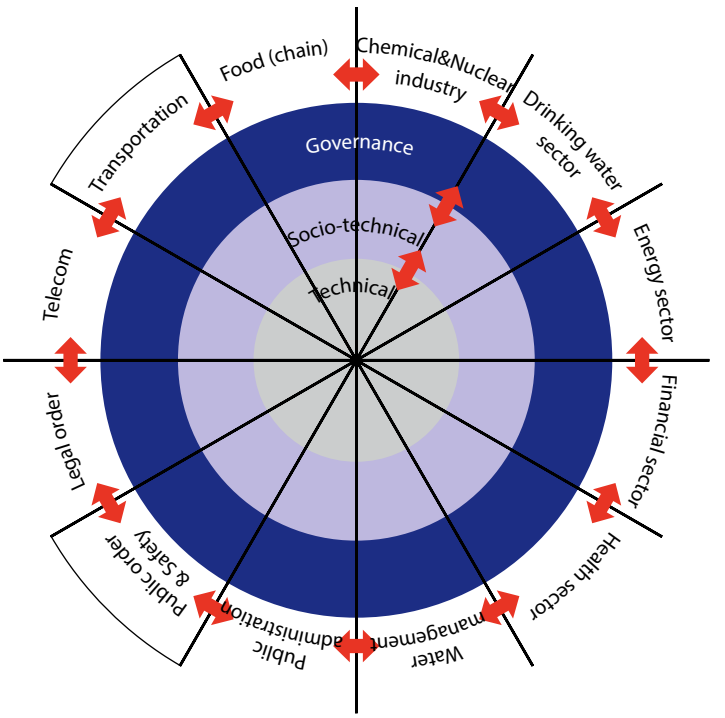

Figure 1 A Conceptualization of cyberspace in layers and (cyber) subdomains

Source: Van den Berg et al. 2014

lesser extent on socio-technical layers, as can be seen from other contributions. For understandable reasons the technically possible interception, collection and storage of huge amounts of data tickles people's imagination. Catchphrases such as 'select before you collect', 'collect before you select' and even 'select while you collect' exemplify the main road taken by most researchers.[53] A look at the reports of the

50 See: https://zoek.officielebekendmakingen.nl/dossier/34588 and https://www.internetconsultatie.nl/wiv/details.

51 Maurits Martijn, 'Wat zijn de wensen van dit kabinet voor de geheime diensten?', in: *De Correspondent*, 11 January 2016 https://decorrespondent.nl/1632/wat-zijn-de-wensen-van-dit-kabinet-voor-de-geheime-diensten/50193792-ce33fa45.

52 https://www.internetconsultatie.nl/wiv/details and Maurits Martijn, 'Vier redenen waarom de nieuwe aftapwet een slecht idee is', in: *De Correspondent*, 12 July 2017 https://decorrespondent.nl/7054/vier-redenen-waarom-de-nieuwe-aftapwet-een-slecht-idee-is/713051614880-1a2bce5c. NB: This article is a later version of the article mentioned in note 44 by the same author. The two interactive articles link to many important contributions concerning the intelligence law debate.

53 Bart Jacobs, 'Select before you collect', in: *Ars Aequi*, Vol. 54 (No. 12) pp. 1006-1009, 2005 and Bart Jacobs, 'Select while you collect. Over de voorgestelde interceptiebevoegdheden voor inlichtingen- en veiligheidsdiensten', in: *Nederlands Juristenblad Vol. 91* (Den Haag, 29 January 2016) p. 256-261.

PHOTO MCD/V. KUYPER

The intelligence law debate in the Netherlands centers around the possibilities of large-scale automated data communication interception and the risks involved in approval of such methods

Dutch Review Committee on the Intelligence and Security Services (Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten or CTIVD) corroborates this. Twelve out of sixty-six investigative reports listed on its website deal with Signals Intelligence (SIGINT), or more precisely communication interception (tapping and/or hacking), storage and analysis.[54] A WRR-working paper on the use of big data by the MIVD and AIVD leads to a similar conclusion.[55]

Paradoxes at play; transparency versus secrets

Is another perspective on the big data phenomenon feasible and can it be of use in, for instance, the debated Dutch intelligence law? A small strain of research has a welcome, more theoretical, bird's eye view of the problem at hand. In his inaugural lecture Dennis Broeders zooms in on the notion of secrets in today's information society and looks at the position of both individuals and the state.[56] He claims that the possibility of keeping individual secrets in today's society is decreasing, whilst the volume of state secrets has increased. At the same time state secrets have become more vulnerable. Broeders identifies the exponential rise and spread of new digital technologies as being at the core of this development.[57]

The vulnerability of state secrets has been aptly dubbed by Peter Swire the declining 'half-life' of

54 http://www.ctivd.nl/onderzoeken.

55 Sascha van Schendel, 'Het Gebruik van Big data door de MIVD en AIVD'. *WRR-working paper 18* (Den Haag, 2016).

56 Dennis Broeders, *Het geheim in de informatiesamenleving*, Oratie Erasmus Universiteit Rotterdam, (Den Haag/Rotterdam, 2015).

57 Broeders, *Het geheim in de informatiesamenleving*, pp. 14-15.

secrets. Swire concentrates on the fact that the average time to disclosure nowadays is decreasing. At the same time the expected costs of disclosure to decision-makers become higher when secrets become known sooner, and might end up on the front-page while they are still in office. The time frame for keeping state secrets has customarily been measured in decades, as evidenced by the basic classification system of the United States that was developed during the Cold War. According to this system an executive agency must declassify its documents after twenty-five years unless an exception applies. The basic mind-set displayed in this time frame is revealing especially when secrets today often get exposed within a few years, months, or even days. The implications are multiplied owing to the continuing effects of *Moore's Law*. The continued improvement in computing power, combined with the infinitely improved possibility of dissemination through the internet, enables huge leaks at marginal costs and with relative ease,[58] as was the case with Assange, Snowden, the Panama or Paradise papers, or any other of a multitude of highly-publicized recent breaches and/or leaks.

So, the possibilities offered by the information revolution enable people to gather information and connect with each other at the speed of light. In its wake transparency has become the norm and is also expected from the state apparatus. However, enthused state bureaucracies are collecting more and more information on the lives of its citizens and show no inclination to keep fewer things secret. Broeders translates this development into a privacy paradox and a transparency paradox. According to him most people claim to be worried about their privacy, but in (digital) practice they behave without concern for these worries. To this must be added the acceptance, or more likely neglect, that many (commercial) databases hide themselves in and behind computer applications that make digital life so comfortable. Citizens are thus becoming ever more transparent for the state, whereas it has become more difficult to identify which government agency possesses what personal information. In other words, the state is becoming less transparent.[59]

'The continued improvement in computing power, combined with the infinitely improved possibility of dissemination through the internet, enables huges leaks at marginal costs and with relative ease' - Peter Swire

The right to have secrets is part of the social contract between citizens and the state. In a recent study Paul Frissen pays attention to the fact that a democratic state (democratische rechtsstaat) keeps secrets in the interest of the welfare, well-being, and (self) development of its citizens. Being of a personal nature these secrets are very much connected to privacy aspects. In addition, state secrets serve the purpose of state security, and the stability of society and the democratic legal order. Frissen emphasizes that state success in the second domain requires the acquisition of secrets. In the Netherlands these secrets are obtained with the help of special powers by ministerial approval without prior judicial review. Obviously, such state activities might seriously infringe upon people's privacy and their liberties and, therefore, require supervision and oversight. This demands the state to be strong and weak at the same time.[60]

58 Peter Swire, 'The declining Half-Life of Secrets. And the future of signals intelligence' in: *New America Cybersecurity Fellows Paper Series - Number 1* (July 2015).

59 Broeders, *Het geheim in de informatiesamenleving*, p. 18-19, 22-25, 29.

60 Paul Frissen, *Het geheim van de laatste staat. Kritiek van de transparantie* (Amsterdam, Uitgeverij Boom, 2016). See in particular the paragraph on paradoxes of secrets (*De paradoxen van het geheim*) in the final chapter (243-250). See also chapters 4 (137-163) and 5 (pp. 165-223) on the secrets of the state and the part played by the intelligence and security services.

PHOTO NFI/J. VISSER

Paul Frissen: 'The state needs to protect the secrets of its citizens, but to be effective it needs secrets of its own, and transparency of its citizens'

Paul Frissen maintains that the freedom of any citizen rests partly on his right to have secrets. In order to protect that right the state is obligated to prevent and counteract any attempts to undermine it. Paradoxically this requires secrecy to a certain extent. Or, in his opinion, the legitimacy of the state to perform secretive acts is also based upon people's right to have secrets. As long as the state protects this, it is entitled to have secrets of its own. In other words, to be legitimate the state needs to protect the secrets of its citizens, but to be effective the state needs secrets of its own, and … as much transparency of its citizens as possible. [61] Owing to big data flows and technology 'transparent citizens' seem within grasp. Frissen distinguishes two elements within the earlier mentioned transparency paradox. First, total transparency as the societal norm precludes the existence of secrets. However, the concept of total transparency is a treacherous misnomer, since it will ultimately rob a person of his individual freedom (e.g. to have scandalous thoughts or despicable opinions; in other words to be allowed to have secrets).[62] Second, it has simultaneously motivated intelligence and security services to try to use information and communication technology by indiscriminately intercepting data flows, discovering and analysing their trends in order to obtain

61 Watch the interview with Paul Frissen, broadcast by NPO 1 television on Sunday 24 January 2016. http://www.vpro.nl/boeken/programmas/boeken/2016/24-januari.html.

62 The disturbing consequences have been eloquently fictionalized by writers like George Orwell in *1984* (1949), or more recently, Dave Eggers in *The Circle* (2013).

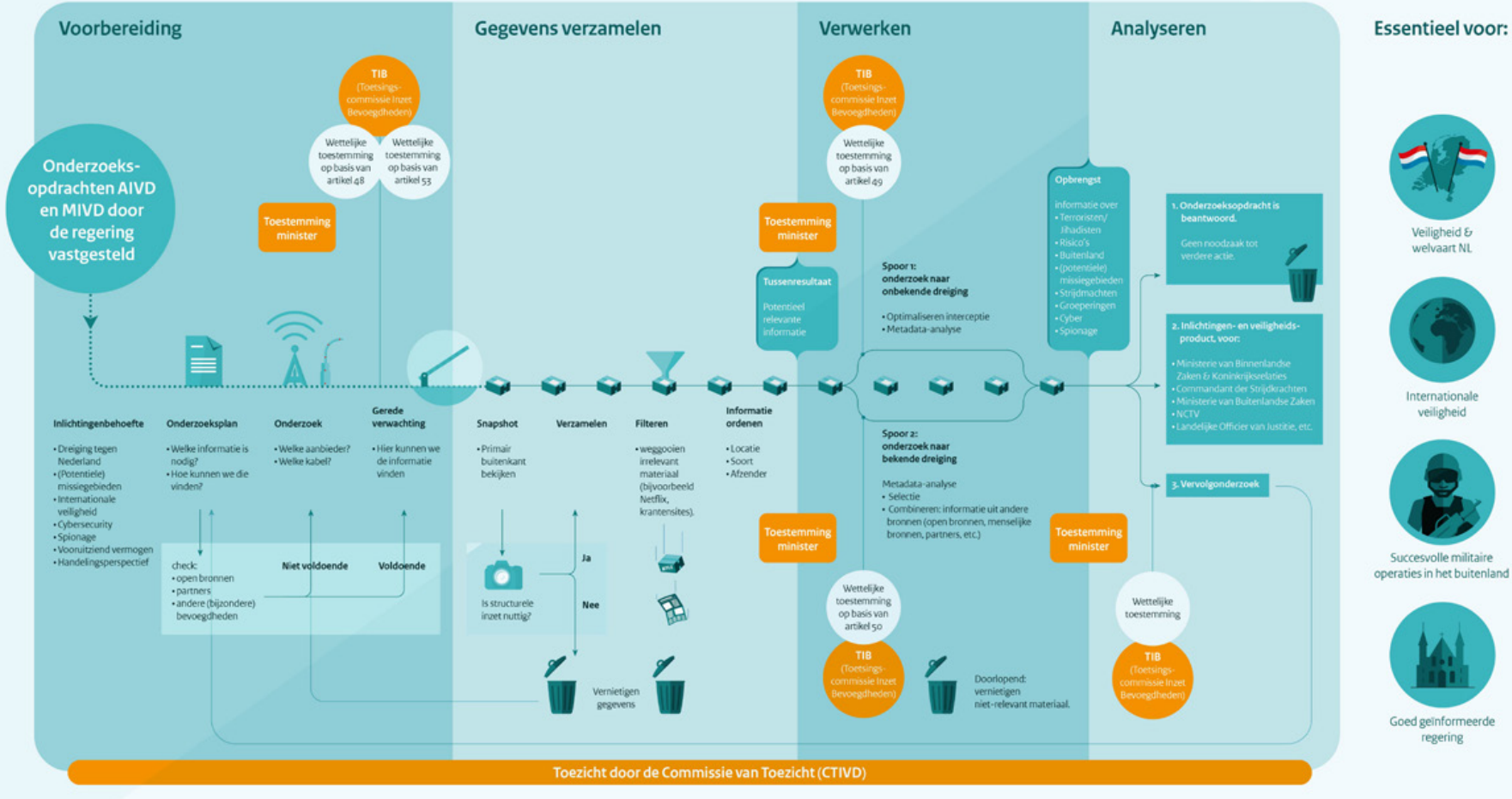

Figure 2 Intelligence law safeguards (incl. TIB and CTIVD; excl. IVD and ARK)

SOURCE: HTTPS://WWW.RIJKSOVERHEID.NL/DOCUMENTEN/BROCHURES/2016/10/28/INFOGRAPHIC-WAARBORGEN-ONDERZOEKSOPDRACHTGERICHTE-INTERCEPTIE

At the core of the matter there is no real dichotomy between transparancy and secrets, or freedom and security. These notions constitute two sides of the same coin, and they very much depend upon each other

predictive value and actionable intelligence. This huge effort tries to protect civil society by preventing for example criminal and terrorist activities.[63]

The usefulness of paradoxes

The question remains whether recognition of the paradoxes can shed light on the complex interplay of big data flows, transparency and secrecy as indicated in the subtitle of this article. How can the described paradoxes be of assistance in this juggling act? First of all, they show that at the core of the matter there is no real dichotomy between transparency and secrets or – indeed – freedom and security. These notions constitute two sides of the same coin, and they very much depend upon each other. Basic understanding of this fact can help bridge differences between freedom of information and security advocates. Secondly, to put this into practice requires steering away from the current inclination to zoom in on technicalities. Much more focused attention is needed on the governance level within the cyber domain. There is no paradox involved when it comes to big data flows and big data technologies on the technical level. Simply put, the data streams and the techniques to mine them exist, as well as the profits (monetary or otherwise) to reap from them. These are not imaginary. They are just there, ready to be used. The paradoxes come into play when the technical capabilities are translated into cyber activities on the socio-economic level. At this level consequences are being felt and conflicts of interest are being fought over. And third, this juggling of interests could benefit from more objective guidance offered by a robust policy and governance framework. A framework with the implicit goal of generating transparency and trust[64] among parties involved. In this governments have a role to play by managing parts of the data economy, such as public infrastructure, and opening up more of their own data vaults (open data).[65]

The need for a governance framework

In the Netherlands a governance framework has not matured yet. The Dutch government set a laudable step on the (international) governance track by assigning the WRR the task to research internet governance, which resulted in an elaborate advice on (inter)national governance and protection of the internet's public core.[66] The government response to the advice was favourable and future policy development is now eagerly awaited.[67]

When it comes to internal and external big data-control, commercial practice is still in its infancy. Pushed by (supra)national governments, watchdogs and consumer organizations, confidentiality and accountability issues are nowadays on the business agenda. Still, the main

63 Interview with Paul Frissen by NPO 1 television.

64 Bibi van den Berg and Esther Keymolen, 'Regulating Security on the Internet: control versus trust', in: *International Review of Law, Computers & Technology 31*:2 (2017) p. 188-205.

65 'Living with technology: The Data Republic', in: *The Economist*, 26 March 2016; 'The world's most valuable resource', *The Economist*, 6 May 2017.

66 Dennis Broeders et al. 'De Publieke Kern van het Internet'.

67 'Kabinetsreactie op AIV-advies 'Het internet, een wereldwijde vrije ruimte met begrensde staatsmacht' en WRR-advies 'De publieke kern van het internet: naar een buitenlands internetbeleid', 19 May 2016. See: https://www.rijksoverheid.nl/documenten/kamerstukken/2016/05/19/kabinetsreactie-op-aiv-advies-het-internet-een-wereldwijde-vrije-ruimtemet-begrensde-staatsmacht-enwrr-advies-de-publieke-kern-van-het-internet-naar-een-buitenlands-internetbeleid.

challenge for the coming years will be the integrity of the ever-growing amounts of data in possession of and used by companies. Recently, the accountancy profession has started to realize that auditing of databases and/or large datasets with regard to privacy violations or integrity issues are indispensable from a governance – i.e. a management – point of view, and hold great promise for the future.[68]

Within the security context a governance/oversight mechanism with regard to the intelligence and security services is in place, and has been strengthened in the revised intelligence law (see figure 2). The Toetsingscommissie Inzet Bevoegdheden (TIB), CTIVD (ex-post), Court of Audit (Algemene Rekenkamer or ARK) and a special parliamentary committee (Commissie voor de Inlichtingen- en Veiligheidsdiensten or CIVD)[69] are tasked for this job. However, their findings are not legally binding, with the notable exception of the TIB from 2018 onwards.[70] In the end those responsible for the intelligence and security services – the Minister of Interior Affairs and/or the Minister of Defence – decide if and when what action is to be taken. The request of the CTIVD for more authoritative powers has not been honoured so far, except for CTIVD-judgement on citizens' complaints, which has become binding.[71] In the meantime the CTIVD is preparing itself for the future oversight challenges connected to the new intelligence law, i.e. bulk interception and big data analytics.[72] In a press statement of April 2017 it announced the start of project *Toezicht 3.0* (*Oversight 3.0*), which aims to investigate the possibilities of effective oversight with regard to the collection, analysis and destruction of large amounts of data.[73]

Conclusion

The introduction stated that big data is connected to the free flow of information and the transparency that seems to flow from it. It was said that this does not relate well with the secretive behaviour of intelligence and security agencies. Evidently there is more to this assertion than would appear at first sight. Transparency and secrets are inextricably linked and do meet in unexpected ways.

At the core of the debate we have found paradoxes to which little attention has been paid within the media stream of technical possibilities and, to a lesser extent, socio-technical applications of big data. Recognition of the big data paradox will serve the purpose of an integral and more balanced perspective on the counter-intuitive, but complementary, notions of transparent and secretive behaviour. If successful, all three levels of the cyber domain will be combined. Excesses on the internet required (and require) governance through regulation in order to benefit from its accomplishments. Big data requires a comparable course of action. This awareness should translate into robust oversight of secret actions and, more generally, into well-organized, state-sponsored, international governance of big data flows and technologies that are revolutionizing our everyday lives. ■

68 Franka Rolvink Couzy, 'Big four duiken in big-datacontrole', in: *Financieel Dagblad*, 30 March 2017.

69 Constant Hijzen, 'Tot het lachen ons vergaat. Over de noodzaak van parlementaire aandacht voor inlichtingen- en veiligheidsdiensten', in: *S&D 70 Nr. 4*, July 2013.

70 It remains, however, to be seen how the binding mandate of the new TIB will relate to overarching(?) ministerial responsibility. See: 'Toezicht in nieuwe wet op de inlichtingendiensten goed regelen', 8 February 2017 in: *Rechtspraak.nl*. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Paginas/Toezicht-in-nieuwe-Wet-op-de-Inlichtingen--en-veiligheidsdiensten-goed-regelen.aspx.

71 'Reactie CTIVD op het concept-wetsvoorstel Wet op de inlichtingen- en veiligheidsdiensten 20XX', Consultatieversie juni 2015. See: http://www.ctivd.nl/documenten/publicaties/2015/08/26/reactie-ctivd-conceptwetsvoorstel and 'Bijlage I Essentiële waarborgen. Zienswijze van de CTIVD Op het wetsvoorstel Wiv 20..' November 2016. See: https://www.ctivd.nl/binaries/ctivd/documenten/publicaties/2016/11/09/bijlage-i/Zienswijze+van+de+CTIVD_Bijlage+I_november+2016.pdf and Maurits Martijn, 'De waakhond van de geheime diensten wil door kunnen bijten, maar heeft er de tanden niet voor', in: *De Correspondent*, 11 January 2016.

72 Hilde Bos-Ollermann, 'New surveillance legislation & intelligence oversight challenges. The Dutch experience', International Intelligence Oversight Forum, 11-12 October 2016. See: https://www.ctivd.nl/documenten/toespraken/2016/10/11/index.

73 'Start Project Toezicht 3.0', CTIVD, 25 April 2017 See: https://www.ctivd.nl/actueel/nieuws/2017/04/25/index-2.