Data Diplomacy

Updating diplomacy to the big data era

Barbara Rosen Jacobson,
Katharina E Höne and Jovan Kurbalija

February 2018

DiPLO
www.diplomacy.edu

FORMIN.FINLAND.FI
ULKOASIAINMINISTERIÖ
UTRIKESMINISTERIET

Table of contents

Acknowledgements

Executive summary

Data is often described as a critical resource of modern society, or even the oil of the new economy. Vast amounts of data are generated every day through the use of electronic devices and the Internet. The private sector has begun to harness big data sources to improve their products and services, streamline procedures, and ultimately increase revenues. Big data analysis is said to create insights that were hitherto unavailable. What is the position of diplomats, who rely on data and information in their everyday work, in this changing environment? Some ministries of foreign affairs (MFAs) and international organisations are tentatively exploring the uses of big data for policy planning, knowledge management, development, humanitarian action, and emergency response, recognising the potential benefits. Yet, there is still a large number of perceived obstacles that prevent others from stepping on board the big data train.

This report aims to increase the awareness of the opportunities, limitations, and challenges of the big data trend, and to understand how MFAs could adapt their work, procedures, and organisational structures to the big data era. In this report, we provide a broad overview of the main opportunities of big data in different diplomatic fields and functions, and highlight the key issues that need to be addressed for big data diplomacy to flourish. This framework of possibilities and constraints opens up a diversity of applications and implications that can be explored in further detail, and is meant to inform MFAs that are exploring big data to adapt diplomatic practice to the data-driven era where possible and feasible.

How can big data be understood in the framework of diplomacy?

Characteristics of big data: Although a contested concept, most definitions focus on the size and heterogeneous characteristics of the datasets, and the current speed at which the data is generated, requiring new tools for analysis, compared to more traditional sources of data. This is often related to the so-called four V's of big data – volume, velocity, variety, and veracity.

Output of big data: Other definitions focus on the new kind of information and knowledge that can be produced through big data, often taking the form of trends, patterns, and correlations.

Big data in data diplomacy: This report focuses on automatically generated data of relevance for diplomacy, including online data, sensor data, satellite data, and textual data. Through the generation of trends, patterns, and correlations, these new data sources can contribute to diplomatic functions and areas, while their technical characteristics provide several key considerations that need to be mitigated for big data's effective use in diplomacy.

How can we understand the concept of data diplomacy?

Data diplomacy – tool, topic, environment: Big data interacts with diplomacy in three ways. First, big data can be used as a tool to make diplomacy more efficient, effective, and inclusive. Second, it provides a new topic on the diplomatic agenda and features in international negotiations in areas such as cross-border privacy, e-commerce, and international cybersecurity, to name a few. Third, it is a factor that changes the very environment in which diplomacy operates, potentially shifting geopolitical and geo-economic positions. While this comprehensive approach falls largely outside of the scope of this report, it can be further developed in future research.

Big data as a tool for diplomacy: In this research, we focus on gathering, analysing, and integrating big data in negotiations, reporting, consular protection, and other diplomatic activities, while looking at the limitations and other key considerations that need to be mitigated to allow big data to flourish.

How does big data impact the core functions of diplomacy?

Core functions of diplomacy: We looked at four diplomatic core functions in the context of big data – information gathering and diplomatic reporting, negotiation, communication and public diplomacy, and consular affairs.

Information gathering and reporting: Information gathering is likely to be one of the areas that will be most affected by big data. Big data opens up new sources and new ways of analysis, from social media discourse to government open data and geospatial information, which can feed into policy-making and strategy. The digitalisation of diplomatic reports similarly provides new opportunities to analyse and detect patterns in the conduct of diplomacy. Ultimately, big data can serve to provide new insights, challenge biases, and corroborate information.

Negotiation: Negotiation is a fundamentally human endeavour. Big data can play a role by providing relevant arguments and insights to understand counterparts and support the development of negotiation positions and strategies. These insights might be especially important in the prenegotiation phase. Further, big data insights could provide common ground by adding external information that might be considered more objective, such as satellite images, on which agreements can be built.

Communication and public diplomacy: Generally speaking, the biggest promise of big data in the area of communication lies in the ability to understand patterns and trends in discourse, to tailor messages, and to measure the effectiveness of a communication campaign. Public diplomacy has adapted to the opportunities of digital technologies, and in particular social media, to become more effective in communicating with foreign and domestic publics. Diplomats are increasingly able to use these tools in new ways and improve their understanding of foreign and domestic discourse, as well as the effectiveness of the reach and engagement of their own messages.

Consular affairs: With experience in data management and a role in service delivery, consular affairs might well be the area to benefit from big data. With increased demands from the public in its interaction with governments, consular departments are under pressure to keep up with the latest big data opportunities to optimise their online services. Big data can support consular functions in making use of internal data to improve consular service delivery, using innovative big-data-supported means to locate citizens in need, and using (social) media monitoring to react faster to or even predict crises and the need for consular services.

What is the role played by big data in various diplomatic fields?

Big data and trade: Statistics and data have long been at the core of trade promotion and economic diplomacy. Big data provides new possibilities to monitor and evaluate trade flows, especially with the advent of e-money, e-banking, and e-commerce.

Big data and development: With its ability to track patterns over time and space, big data can be of great value for development. In particular, big data could help track progress towards the Sustainable Development Goals (SDGs), by providing additional input for its indicators. For example, sensor and satellite data can be used for monitoring climate change, financial transactions can shine a light on economic differences between social groups, and social media data can help detect patterns in discrimination. Still, there is a need for a better understanding of how big data can be smoothly integrated into existing monitoring efforts. Big data in the development sector is also explored to assess the needs of beneficiaries, track aid flows, and monitor and evaluate programmes.

Big data and humanitarian affairs: As some forms of big data can become available rapidly, almost in real-time, they may be able to assist in responding to quickly unfolding emergencies and humanitarian action, and feed into early warning systems. The analysis of communication channels, such as mobile phone records and social media, is particularly valuable in detecting abnormalities and engagement on topics that could indicate emergencies. Yet, in these volatile contexts, it is more important than ever to ensure privacy and the proper managing and protection of personal data.

Big data and international law: As societies become increasingly dependent on digital tools and services, they leave behind an array of data, which could be transformed into new forms of accountability and evidence. International courts are now exploring how to use these traces, such as social media messages, e-mails, and geospatial data, to serve international law. These new sources open up new questions on the technical tools that are needed in international law to analyse big data, especially on how to verify the authenticity of such content.

Areas of big data potential: Big data can contribute to a number of diplomatic fields and functions, albeit in different ways. To generalise, we have identified six

ways in which big data could benefit diplomacy and the corresponding fields and functions for which they are most relevant:

- Providing new information and challenge bias (information gathering and reporting).
- Meeting the expectations of government service delivery (consular affairs).
- Better understanding people's perceptions and behaviour (communication, public diplomacy, and negotiation).
- Tracking programmes and progress over time and space (trade and development).
- Tracking developments over short timeframes (humanitarian affairs and emergency response).
- Identifying new forms of evidence and accountability (international law).

How does big data analysis fit within the organisational culture of the MFA?

Big data as a tool, not a panacea: Big data is a tool to support good foreign policy. It does not aim to replace expert knowledge with automatisms.

Need for subject knowledge and context: Experts with subject knowledge, based on years of experience, are needed more than ever in the data-driven era because big data that is not embedded in its proper context can be dangerously misleading.

Emphasis on concrete benefits over technical details: The idea that big data analysis is only accessible to those with the relevant technical or programming skills needs to be countered by presenting the results of big data analysis, as opposed to talking about methodology, and by emphasising the contribution that these results can make to better foreign policy.

Carefully crafted relationships: Carefully crafted relationships between those working quantitatively and those working qualitatively within the MFA are important. Exchange and collaboration should be fostered where possible, while respecting the unique contribution of each.

How can the MFA adapt to meaningfully include big data practices and insights?

Keeping the aim in sight: Decisions about organisational transformation need to be driven by the goal of making better decisions and contributing to better foreign policy.

Establishing a big data unit: The big data unit should have a cross-cutting function and be able to serve a variety of regional and thematic departments in the MFA. It should be relatively small, diverse (consisting of data scientists as well as diplomats), and free to innovate and experiment.

Fostering exchange through big data champions: Appointing data champions within relevant departments and units can promote communication between the big data unit and the wider MFA in order to both disseminate big data insights, and to better understand the concrete needs of various departments and the role big data can play in addressing these.

What kinds of partnership are needed to update the MFA to the data-driven era?

Reasons for entering into partnerships: Partnerships will play an important role due to a lack of internal capacity and the challenges associated with internal retraining or hiring. Partnerships, especially with the business sector, will also facilitate access to otherwise restricted data. Further, building sustainable long-term relationships with relevant institutions, especially in the private sector and academia, can be another driver for partnership.

Outsourcing big data analytics: Several reasons contribute to the potential need for outsourcing big data analytics. In-house knowledge in relation to data science is, in most cases, still lacking. The demand for data scientists on the job market is outpacing the supply of highly qualified professionals. Outsourcing can save resources and be more cost-effective. Outsourcing increases flexibility and can be helpful if longer-term commitments to new units or a changed organisational structure are not yet plausible.

Need for a minimum of in-house capacities: Compared to the private sector, there are important limitations for the MFA related to the utility of outsourcing big data analyses. Especially when it comes to sensitive issues, it is advisable that MFAs develop their own in-house big data analysis capacities.

How can capacity-building needs in big data diplomacy be addressed?

Identifying needs: Existing capacities and capacity gaps need to be carefully identified, keeping in mind that not every diplomat needs to have the same level of familiarity with big data.

Offering different levels of training: With regard to capacity building in data diplomacy at the individual level, we suggest a three-tier structure, which reflects the different levels of expertise required.

- Foundation level: able to assess the challenges and opportunities of big data with a general knowledge of big data diplomacy.
- Practitioner level: able to work with big data tools and techniques to verify and find information.
- Expert level: able to design and implement appropriate big data tools for diplomatic insight.

Bridging worlds: The question of capacity building in big data diplomacy is also a question of enabling and supporting communication between two different worlds – the world of the data scientist and the world of the diplomat.

Enabling the best possible use of big data tools in support of the work of diplomats: Ultimately, the aim of organisational change and capacity building is not to transform diplomats into data scientists. Rather, the aim of all such efforts should be to highlight where big data can make a contribution to diplomatic practice and to support the work of diplomats through making the best possible use of the available tools.

What key considerations need to be mitigated for diplomacy to capture the potential of big data?

Big data can provide important opportunities for the MFA, from knowledge management and information gathering to monitoring programmes, understanding discourses, and delivering services. These opportunities are captured through organisational measures, such as creating a unit, forging partnerships, or engaging in capacity development. Yet, there are a number of practical considerations that need to be tackled in order for the MFA to make the best use of big data: data access, quality, interpretation, protection, and security.

Data access: While some big data can be found or generated within the MFA, most big data reside without. There is an increasing amount of open data, publicly available, that is ready to be analysed. This can be a good starting point to engage in data diplomacy. Similarly, open data within the MFA, such as unclassified texts and documents, can be obtained relatively easily.

Yet, most big data, such as mobile phone data or data from sensors, are held by the private sector and is confidential, classified, or only made available against a fee. Obtaining this data requires entering into partnerships, which can be complex to negotiate. For such partnerships to be effective, there is a need for transparency about the motivations, policies, and regulations of both the MFA and the data provider; a framework of accountability; and a fair value exchange between the two parties.

Data quality: Advantages related to the size of a big dataset also engender important challenges. These datasets are often messy and difficult to manage. Key concerns relate to big data's complexity, completeness, timeliness, accuracy, relevance, and usability. Before embarking on a data diplomacy project, it is important to assess the quality of the data and to decide whether it is feasible to proceed and worth the effort.

Data interpretation: Just like any other analysis, the study and interpretation of big data can be prone to important biases. For big data, there is a particular need to be aware of the risk of confusing correlation with causation. A big dataset with many variables can detect many inter-relations, yet often not with a causal character. In addition, there needs to be an awareness of selection biases, as big data often only includes those who are using digital devices, and risks creating gender, generational, or wealth gaps. Finally, the interpretation of data can be tainted by political incentives, as there is usually more than one way to frame and visualise the outcomes of a study.

Data protection: The collection, management, and analysis of big data needs adequate data protection provisions to avoid violating privacy rights and to protect the reputation of the MFA. Relying on open data, minimising the data that is collected, and only sharing or obtaining aggregate data could help to keep privacy risks to a minimum and to maintain the trust of the people whose data is used.

Data security: To avoid data breaches, the MFA needs to invest in technical infrastructure and in training staff on the proper management of their own and others' data. Data security could involve securing the data location, the data format, and the data design.

Through advancing an understanding of what big data can do for diplomacy, how it can be put to use by MFAs, and which aspects need to be kept in mind when using big data, this report aims to ultimately serve as a toolkit for data diplomacy.

Introduction

Diplomacy, the management of international relations, is tasked to continuously adapt to an ever-changing world. And today's world seems to increasingly evolve around data. Data is often perceived as an unstoppable force for innovation, a defining element of modern society. It has even been called 'the new oil'; the world's most valuable resource.[1] It has been lauded for its promises and condemned for its threats. As these promises and perils are often exaggerated, they make it difficult to determine how important data really is for society, beyond the rhetorics of alarmists and optimists. Yet, no one can question the reality of data. Our mobile phones collect GPS signals wherever we go and 452 000 Tweets are posted every 60 seconds.[2] Satellites orbiting the Earth generate a continuous stream of geospatial images. Smart watches literally keep our finger on the pulse. And this world, where data has moved from scarcity to abundance, is also occupied by the diplomat, who depends on reliable and readily available information. With this in mind, this report attempts to answer the question: What is the potential of big data for diplomacy?

Data, and especially big data, have begun to shape our personal, political, and economic lives. From their very different perspectives on big data, various sectors of society have begun to look into harnessing big data or to critically engage with the potential of big data as a means of governance and control. If data is the building block of knowledge, and if knowledge is power, then it is no surprise that big data has been dubbed society's new oil. In a world where Google might know more about the world's citizens than their own governments (and might even know us better than we know ourselves[3]), power relations are changing swiftly, as is the realm of diplomacy.

We distinguish three ways in which diplomacy interacts with big data:
- Big data as a tool for diplomacy.
- Big data as a new topic on the diplomatic agenda.
- Big data as a technological development that changes the very environment in which diplomacy is conducted.

While we will look more closely into these three dimensions in Chapter 2, the focus of this report lies almost exclusively on the first interplay: big data as a tool for diplomacy.

To better understand the promises of big data, it could be useful to learn from the private sector. It is this sector in particular that heralds the potential of big data as a tool to improve products and services, streamline processes, offer more tailored customer experiences, and, more generally, make better decisions. Big data almost appears like a much awaited magic solution. Discussions related to big data in the private sector remind us of the promises of big data:
- Better decision-making based on evidence.
- Better service provision based on the clearer identification of needs.
- More streamlined business operations.
- Insights into 'unknowns' that can only be achieved on the basis of big data.

While the application of big data analysis to business processes is relatively well developed, the same cannot be said about big data in diplomacy. Diplomacy as a practice has always been cautious towards new technological developments and has always been a comparatively late adopter. Nonetheless, while staying true to its core functions, diplomacy has been able to adapt to and effectively integrate new technology into its practice. There is no reason to believe that this will be different with regard to big data and diplomacy. Yet, this development is very much in its infancy, with only a few MFAs having begun to explore big data analysis as a tool.

While the promises of big data analysis for the private sector appear very straightforward and convincing, we need to keep in mind that the practice of diplomacy differs from economic activities in a number of crucial ways.[4]
- Diplomatic services differ from businesses in their function. Companies are established with a focus on profit while diplomacy promotes national interests and aims to maintain international order.
- The work of diplomats lacks a feedback loop provided by market mechanisms or customer responses. More

generally, it is not easy to define criteria according to which to measure the performance of diplomats.

- 'Processes in diplomacy are not simple instruments to an end, but have importance in themselves. Sometimes the diplomatic solution to a problem lies in the process' and not immediately in an outcome.[5]
- Comparing the private sector and diplomatic services, we can see that there is a different relation to and perception of time at play.

These differences impact the way and the extent to which big data tools and trends, as identified by the private sector, are applicable to diplomatic practice. In other words, we cannot transfer the available lessons from the private sector directly to the practice of diplomacy. Rather, the utility of big data has to be assessed on its own merits. In doing so, it is important to stress that traditional diplomatic knowledge will not become obsolete. On the contrary, we argue that traditional insights are as important as they have always been to provide context and meaning for the results of big data analysis.

At the same time, new data sources will inevitably need to be picked up by diplomats, if old data sources cease to exist or prove not to be the optimal tool for a given situation or question. For example, a public diplomacy country brand might be more accurately depicted by the millions of social media users than by studying the more traditional news outlets. Mobile phone data might facilitate locating people in need in times of crisis at a faster and more accurate rate than traditional means. GPS patterns could predict the spread of disease or migration movement, and the analysis of online conversations could feed into early warning systems.

Not only can it be useful to look across sectors, it might also be worth looking across time. Big data is usually seen as something inherently new, innovative, and often, disruptive. Enabled by the Internet and digital technology, we are not only generating more data than ever before, but we have also developed the capacity to store and analyse data to an unprecedented extent. With new data come new methods, new ways of interacting with information. For diplomacy, there could be a newfound focus on correlation, new pressures on privacy, and new partnerships with the private sector. For example, a diplomat tasked with monitoring the development of migration patterns will draw on different tools, skills, and processes when relying on GPS data rather than survey data. And yet, in the end, data is data. Diplomats have always dealt with data, in different shapes and forms, and have a rich history in data and information management.

While big data is a relatively new topic, the application of new technologies to diplomacy is as old as the profession itself. In light of new technology, especially information and communication technology (ICT), the question of how diplomacy needs to adapt to a changing environment has always been important. Initial scepticism, such as Lord Palmerston's pronouncement of the 'end of diplomacy' when faced with the telegraph, have been quickly replaced by a more pragmatic approach that has focused on adapting diplomatic practice in the light of new challenges and opportunities.[6]

For example, in 1996, Kurbalija suggested that '[t]he wider use of IT within Ministries of Foreign Affairs … should produce in the near future a new IT-influenced diplomatic interaction symbolised by the exchange of communication through dedicated e-mail'.[7] Just like the telegraph and the telephone before, e-mail has become part of communication routines to such an extent that it is now used without much reflection. As such, discussions on big data in diplomatic practice are related to the long-standing question of the impact of new technology on diplomatic practice.

The advent of the telegraph, the telephone, and other ICTs has impacted diplomatic practice in many ways, but the most profound way might well be the way in which diplomats communicate. Big data's influence might be slightly different, and will mainly be felt in relation to the effectiveness and efficiency of diplomacy, as well as the generation of new insights.

In this report, we explore big data as a tool for diplomacy in four steps, corresponding to its four chapters. In Chapter 1, we delve into the core concepts related to data diplomacy. As data diplomacy is a rather new concept – in fact, it has not yet been systematically looked at – there are many ways in which it can be defined and interpreted. To make matters more complicated, there is similar terminological confusion around the concept of big data. Even though there is an abundance of resources about big data, the plethora of different definitions and approaches has not yet matured, leading to many possible ways to frame the concept.

In Chapter 2, we ask how big data can be used in diplomacy and international affairs. In answering, we look at the core functions of diplomacy, identified as information gathering and diplomatic reporting, negotiation, communication and public diplomacy, and consular affairs, and explore existing big data practices and the potential for future practices in each case. We then apply the same style of discussion to various areas of diplomacy,

including trade, development, emergency response and humanitarian action, and international law.

In Chapter 3, we take a more inward look. Here, we focus on the organisational considerations of ministries of foreign affairs (MFAs). How should MFAs adapt if they are to engage in data diplomacy? Can we learn from the private sector's data-driven companies? We also look at the skills needed for diplomats to understand data (and for data scientists to understand diplomacy) and propose concrete solutions on how to best integrate data units in the ministry.

Having established an understanding of the opportunities of big data and ways for the MFA to meet them, Chapter 4 considers how to overcome some of the key obstacles to data diplomacy, zooming in on access to data, data quality, data interpretation, data protection, and data security. Not only do these topics cover the key considerations and concerns regarding the use of big data in general, such as questions of privacy, they also serve as a useful reminder

of the kind of issues that MFAs who use big data tools need to address.

We have designed the report in such a way that individual chapters and sections can, largely, be read on their own without any loss of meaning. In other words, while the report represents a comprehensive and step-by-step approach to exploring big data in diplomacy, it can also serve as a quick reference guide to some of the key topics. The report is informed by a combination of literature reviews, interviews, and consultations with experts (Annex I).

It highlights a number of challenges and opportunities for big data in diplomacy, and demonstrates one constant when it comes to the practice of diplomacy. As so often in the history of a profession that is presumably as old as humanity itself, the answer lies in the middle. In the service of its core mission, diplomacy can adapt, as skilled negotiators do, with firm flexibility, remaining rigid on its objectives and flexible on the ways to achieve them.

Notes

[1] *The Economist* (2017) The world's most valuable resource is no longer oil, but data. May 6. Available at https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [accessed 12 December 2017].

[2] Desjardins J (2017) What happens in an internet minute in 2017? *World Economic Forum*, 31 August. Available at https://www.weforum.org/agenda/2017/08/what-happens-in-an-internet-minute-in-2017 [accessed 12 December 2017].

[3] Carmichael J (2014) Google knows you better than you know yourself. *The Atlantic*, 19 August. Available at https://www.theatlantic.com/technology/archive/2014/08/google-knows-you-better-than-you-know-yourself/378608/ [accessed 27 January 2018].

[4] Kurbalija J (2002) *Knowledge management and diplomacy*. Kurbalija J [ed] *Knowledge and Diplomacy.* Msida, Malta: DiploFoundation. Available at https://www.diplomacy.edu/resources/general/knowledge-management-and-diplomacy [accessed 12 December 2017].

[5] Ibid.

[6] Kurbalija J (2013) The impact of the Internet and ICT on contemporary diplomacy. In Kerr P & Wiseman G [eds] *Diplomacy in a Globalizing World. Theories and Practices*. Oxford, UK: Oxford University Press, pp. 141–159.

[7] Kurbalija J (1996) Information technology and diplomacy in a changing environment. *Diplomatic Studies Programme. Discussion Papers*, No 20, p. 37. Leicester, UK: Centre for the Study of Diplomacy.

1. The key concepts of data diplomacy

My god, this is the end of diplomacy.
Lord Palmerston, British Prime Minister in the mid-nineteenth century[1]

Diplomats must harvest and adapt the best innovations from elsewhere.
Big Data will require us to reshape how we find and use information; how we deliver a service;
and how we network and influence.
United Kingdom Foreign and Commonwealth Office[2]

While there is a plethora of research and practical experience related to the use of big data in the private sector, much less is available with regard to the various areas of government. Gradually, we are seeing more thinking, research, and practice related to big data in public policy-making, public service provision, and the organisation and work of the public sector. Within the public sector, the eagerness to adopt big data varies. Those ministries engaged in service delivery to citizens, such as ministries of internal affairs, or in economic, biological, environmental, or energy-related research, seem to be more advanced in taking steps towards adopting big data. Diplomacy is usually not found at the forefront of this big data battlefield, of which the frontlines are usually comprised of those who can afford to take, and are used to taking, risks. Diplomats – and of course there are exceptions – rather find themselves towards the back, observing failures and achievements, and only moving into the field after careful analysis.

To assist and speed up this process, which is needed in a time characterised by fast-paced technological developments, this first chapter clarifies the meaning of big data

and its potential role in diplomacy and the work of the ministry of foreign affairs (MFA). These clarificatory steps are necessary precisely because there is little previous research on big data and diplomacy. Similarly, there is no agreed upon definition of big data. As such, we need to specify terms and concepts and put in place the basis for our more in-depth engagement with the functions of diplomacy and the organisational culture of the MFA in the following chapters.

The first part of the chapter develops the concept of data diplomacy, its importance for the future of diplomatic practice, and our approach to and perspective on data diplomacy. In the following, we explore the emerging concept of data diplomacy, as well as discussions on the definition of big data, its interplay with knowledge and information, the basics of its analysis and visualisation, and the relevance of algorithms for diplomats. We do this in an effort to dispel some of the myths associated with big data and its potential and to facilitate a level-headed discussion that serves as the basis for discussion in the following chapters.

1.1 An introduction to data diplomacy

Although some MFAs have started to look into the interplay between data and diplomacy, there is very little systematic engagement with big data in diplomatic practice. Yet, as big data gradually permeates into every field of science, academic discussions on big data and diplomacy are slowly emerging. One such scarce effort is a definition offered by Timothy Dye at a 2015 conference of the American Association for the Advancement of Science:

Data diplomacy is an emerging construct that integrates concepts from data science, technology, and computing with social science, international relations, and diplomatic negotiation, and in some cases, offers a new diplomatic tool that facilitates global (and local) relationships.[3]

While the emphasis on the need for interdisciplinary collaboration and the potential of data as a tool for diplomacy are important, this definition opens up a wide spectrum of interpretations. To make sense of the interplay between data and diplomacy in more concrete ways, we can look at data as a new technology affecting diplomacy in three dimensions:[4]

- as a tool for diplomacy;
- as a topic for diplomacy; and
- as something that changes the very environment in which diplomacy operates.

This three-part typology applies to any aspect of new technology in diplomacy, and it is now clearly expressed in relation to digital technology. Many have looked into the ways in which the Internet can be a tool to make diplomacy more effective, efficient, and inclusive, captured in the concept of e-diplomacy. At the same time, it brings new topics to diplomatic agendas, whether it relates to international negotiations on cybersecurity, the promotion of multilingualism online, the establishment of a global Internet infrastructure, or the discussion of cross-border e-commerce rules; all of this is captured by the term Internet governance. Finally, the Internet has affected the environment in which diplomacy is practiced, causing shifts in geopolitics and geo-economics, from the physical position of the main Internet cables that carry data and the data centres that host data, to the consideration of data as a national asset, not unlike the importance of oil for the traditional economy.

The same framework can be applied to data: How can it be used to advance diplomacy? How is its regulation discussed internationally? How does it affect the environment in which diplomats operate?

For example, data diplomacy includes the use of social media data to better understand foreign sentiments in the service of public diplomacy. However, data is also the topic of international discussions, such as data sharing between countries, international standards related to data, or the protection of personal data across borders. Finally, as data is now considered 'the new oil',[5] it is dramatically shifting power dynamics, placing significant leverage on those countries and actors that collect, store, and control data. Zooming in on the capabilities of MFAs, a 'data divide' might even arise, between those MFAs that are more familiar with new digital environments and the analysis of new kinds of data, and those that are left behind in this data revolution.[6]

While the study of all three aspects of data diplomacy is necessary to generate a full understanding of how diplomacy is affected by data, this report zooms in on one of its pillars: data as a tool for diplomacy. We define this as gathering, analysing and integrating data in negotiations, reporting, consular protection and other diplomatic activities by using big data, artificial intelligence (AI) and other digital driven technologies. In addition, rather than focusing on the broad concept of 'data', this report is concerned with 'big data', generated by digital technology in ever-greater magnitudes. While big data lacks a globally agreed-upon definition, we chose to look at:

- Data generated automatically by digital devices: for example GPS data from mobile phones.
- Online information: data that is available on the Internet, such as social media or website data.
- Geospatial data: satellite data and remote sensor data.

In addition, while not always considered to be 'big' data, we also examine the growing number of textual documents stored online. Texts are at the heart of diplomacy, and text-mining can be particularly advantageous for diplomats.

Since such data is continuously generated in exponentially growing amounts, it creates correlations, new insights into behaviours, and a greater understanding of global patterns, which might ultimately affect the way in which information and knowledge is generated and perceived in the MFA.

Discussion around big data, whether applied in business settings, the public sector, or by civil society, has become somewhat hyped. It is surrounded by grand hopes and big aspirations, with some claiming that big data represents 'the key to economic development'.[7] At the opposite end of the spectrum, there are those pointing at the grave dangers of big data and how it could spell the end of privacy, democracy, and even free will.[8] In this report, we opt for a balanced approach, exploring the practical opportunities of big data in different areas of diplomatic activity, as well as the limitations of big data, and the challenges related to its use.

Given the demonstrated value of big data in the private sector, we need to wonder to what extent the promises of big data are applicable to the practice of diplomacy. The questions that need to be asked remain largely the same, no matter what technological innovation we are referring to:

- What areas of diplomatic practice can benefit from the application of big data?

- What are the concrete benefits and what needs to change in the conduct of individuals and the organisational structure in order to harness them?

- How can we ensure that the benefits outweigh the costs and that any changes are taking place in a sustainable way?

1.2 The basics of big data

To better understand the application of big data to diplomacy, we explore three key themes related to big data. First, it is important to distinguish between data, information, and knowledge. This distinction allows differentiation between the accumulation of vasts amounts of data and their transformation into information and knowledge. We then move into the concept of big data, and the basics of its analysis, as well as an introduction on the algorithms that are crucial for the analysis of big data.

1.2.1 Data in a historical perspective

To better understand the concept of data, it is be important to consider that data has always been around, and been used by diplomats. In the third century BC, all human knowledge was believed to be stored in the Library of Alexandria. Since then, data has multiplied in many different ways. If we would now write the sum of human knowledge on CDs, and stack them up, 'the CDs would form five separate piles that would all reach to the moon.'[9]

Although the growth of data is unprecedented, the phenomenon of data is age-old. When conceptualising data as a record of an activity, event, message, or other type of human behaviour, we can see it in prehistoric cave-paintings and the hieroglyphs of the ancient Egyptians. Ever since the first census, data collection and analyses have been vital to enhancing the functioning of society. Advances in calculus, probability theory, and statistics in the seventeenth and eighteenth centuries opened up new possibilities to determine social developments. Early modern data science was already used in the 1800s to map cholera clusters in London and to boost productivity in American factories.[10]

Throughout time, we have improved our methods to collect, store, and analyse these records of human behaviour. From the interpretation of these data records, we gather information, and after analysing information, we gain knowledge. As knowledge is the key resource of a diplomat, data – as the building blocks of knowledge – has always played a prominent role in diplomacy.

While data has always existed, its form and quantity has changed immensely in recent decades. The big data phenomenon emerged with increasingly sophisticated ways to store and process data, and to track the behaviours and conditions of people and places. At the same time, it has become easier for people to contribute to and access data, primarily through the rise of the Internet. As early as 1975, the first Very Large Databases conference, held in Framingham, MA, USA, discussed how to manage census data. In the 1990s, practitioners were already using massive scanner data to compute optimal pricing schedules, and since the 1990s, automatically produced data – capturing people's economic behaviour – has been used to analyse consumer sentiment.[11]

Yet, in recent years, the price of sensors and other technological tools has become significantly lower, making it cheaper to collect a large amount of data. In parallel, the cost of storing data has declined. These technological changes coincide with a societal change, as people are providing an increasing amount of information about themselves online.[12]

With the growing sophistication of data gathering, storage, and analysis, tech communities have taken the interpretation of data to a new level. The mysteries around data science, and the many interesting computations made by data scientists, have baffled many who are unfamiliar with data-mining skills, and have contributed to the rise of the data science hype. Diplomatic institutions, which, for good reasons, usually take more time to adapt to new trends, need to begin exploring the potential of data science for their craft while keeping both feet firmly on the ground. Nevertheless, there is an enormous opportunity to discover if these new data technologies can be applied to diplomatic activities.

1.2.2 Data, information, knowledge

Knowledge and information are at the core of diplomacy. Whenever new technologies are developed, concerns on their impact on knowledge and information in diplomacy often follow, whether it concerns the printing press, the telegraph, the telephone, or the Internet.[13] This might be particularly relevant for data and big data, considering their intricate relationship with the concepts of knowledge and information.

Tim Berners-Lee – the inventor of the World Wide Web – is often quoted as having said, 'Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.'[14] Taking this statement into account, we might conceive of data as the building blocks of our knowledge. Although raw data might not offer many insights in itself, bringing it into a useful order and aggregating it, we can create information. By analysing information, we can gain knowledge, and in certain cases, we can turn this into wisdom.

This triad of data, information, and knowledge is often illustrated as a pyramid with data at the bottom and knowledge at the top. Data refers to numbers, letters, and images. Information describes data that is structured in a useful way, where meaning is added to the data through aggregation or abstraction. Knowledge is built on data and information and represents a theoretical or practical understanding of a subject matter or situation which is highly aggregated and problem-driven.

The lines between data, information, and knowledge are blurred in practice and subject to interpretation. Hence, when we use the terms data and big data in this report, we often already imply the further steps implicated in turning data into information and ultimately into knowledge. What is important to note is that data goes through a number of processes of abstraction and interpretation before it becomes something upon which decisions and actions are based.

In a complex and often chaotic world, knowledge, properly managed, is the key for diplomatic decision-making. In other words, 'knowledge – a combination of information, training, experience and intuition – is what enables a diplomat to act appropriately in unpredictable situations.'[15] Knowledge in diplomacy comes in a variety of forms:[16]

- General knowledge gathered in the course of regular education.
- Knowledge of particular subjects, such as international relations and international law, gathered through specialised diplomatic training.
- Knowledge gained through experience.
- Tacit knowledge of how to react in particular situations.
- Knowledge of procedures.

If these are the traditional forms of knowledge in diplomatic practice, knowledge based on big data analysis adds an additional element. Through the ability of big data analysis to highlight large-scale patterns, a new form of insight is becoming possible.

1.2.3 Data science and big data

In this section we give an overview of data science and big data. In the first part, we engage with existing definitions of big data and put forward our own perspective. We then follow with an introduction to how big data analysis could work in practice, and how it differs from more traditional forms of analysis, such as statistics. Afterwards, we provide a short overview of the role of data visualisation in the context of big data and creating insights for diplomatic practice. Finally, we look at some of the main algorithms that interact with big data, and which questions they could answer that are of relevance for diplomats.

1.2.3.1 Towards a definition of big data

One of the main difficulties in developing the concept of big data diplomacy is that big data can mean many things for many different people. The widespread disagreements and turf battles fought on how to define big data are closely related to the popularity of, and hype around, big data. Promoting products, services and research around 'big data' cleverly plays into this hype, and extends the definition, as everyone wants to be associated with big data. This has lead to uneasiness among those who operate *really* big datasets, who claim that the term big data has extended beyond proportion, to include 'large data' that can be processed and analysed with traditional tools. This leads some to conclude that big data is only present in a few industries, such as the Internet of Things, telecommunication companies, and Internet companies, which all analyse massive amounts of real-time data, whereas others are just free-riding on the hype.[17]

There are roughly two approaches that can be taken when defining big data. First, some people define big data by looking at the data itself. They differentiate 'big' data from small, normal or large data in relation to the size of the dataset. However, the size of the dataset might not be a sufficient indicator; it will only lead to rather arbitrary lines between what is big data and what is not. This has led people to explore other attributes of big data, for example that it has 'heterogeneous characteristics';[18] it can take many different forms. Big data is also often generated automatically, almost in real-time.[19] It is often semi-structured (containing descriptors only) or unstructured, which adds complexity to the ways in which it can be analysed. These characteristics often lead people to claim that big data is only concerned with those databases that are so big that they cannot be processed and analysed with traditional tools.

Big data is often referred to in relation to the widely quoted 'V's', originally proposed by Laney in 2001. Laney put forward three elements to describe big data V's: volume (the size of the datasets), velocity (the speed at which big data is generated), and variety (the many different forms that big data takes).[20] Later, this definition was updated with a fourth V: veracity (the complexities related to the analysis of big data and related questions of accuracy).

A second approach considers big data according to the kind of analysis that is performed on them. Rather than looking at the characteristics of the dataset, they look at the new kind of information and knowledge that can be produced. For example, a *Foreign Affairs* article describes data as 'the idea that we can learn from a large body of information things that we could not comprehend when we used only smaller amounts'. They point to the utility of big data in the identification of patterns and correlations. A drawback of this approach is that it is difficult to understand when data is 'just' data, and when it is considered to be 'big data'; patterns and correlations can be found in almost all datasets, and big data is therefore not necessarily something new and revolutionary; rather, it is a new way of combining data for new insights.

The definition adopted by the *Oxford English Online Dictionary* attempts to merge both approaches, and defines big data as 'extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.' As for this report, we needed a definition that is sufficiently concrete to practically relate it to diplomacy, yet sufficiently broad to be able to generate comprehensive conclusions and observe general effects. Therefore, chose to only look at big data that is generated automatically, usually for a purpose different than the analysis for which it is used in diplomacy, and we then look at how this data can serve diplomacy in creating new insights. The data sources that we particularly focus on are:

- **Data exhaust:** passively collected data generated by people's use of digital services (e.g. phone records, purchases, web searches).
- **Online information**: web content (e.g. social media interactions, news articles, website content).
- **Physical sensors:** satellite or infrared imagery (e.g. changing landscapes, traffic patterns).
- **Textual data:** the large number of texts, reports, messages, and transcripts in digital format, produced by MFAs and other institutions in international affairs. Although this kind of data might be beyond the scope of certain technical definitions of big data, there is are

particular benefits for diplomats in conducting text-mining analyses on such documents.

1.2.3.2 Big data analysis

Rather than using the data as a starting point, the best way to start any data analysis is with a question. These questions should clarify the purpose of using the data, the kinds of behaviours that are to be studied, and the scope of the analysis. For example: How can assistance be provided most effectively after a natural disaster? Which cities are in danger of flooding due to rising sea levels? How positively do citizens feel about foreign countries?

To illustrate how a diplomat might be using data analysis, we will describe the hypothetical case of a diplomat in the economic diplomacy unit of Alphaland. Alphaland will organise a trade mission to Betastan, and the diplomat is tasked to find out how the citizens of Betastan perceive Alphaland, and which segments of the population positively or negatively perceive the country. Betastan has a high Internet penetration and Twitter is one of the most popular social media outlets, so our diplomat decides to analyse conversations containing #Alphaland on Twitter, from users in Betastan.

Once the question has been defined, the next step is to collect the data. The abundance of data that is being generated suggests that there is no challenge in the lack of available data. Nevertheless, it can be difficult to understand which types of datasets can be used, and to gain access to privatised databases.

In our example, the diplomat will rely on Twitter's Application Programming Interfaces (APIs), which allow users to interact with Twitter services and data. For example, Twitter's Search API allows access to past tweets, although it will only display the last 3200 tweets of each user, or the last 5 000 tweets per keyword. Twitter's streaming API monitors tweets as they become available, although Twitter only provides a sample. The Twitter Firehose API is comprehensive and makes all tweets available, but is very costly. Our diplomat choses to rely on Twitter's search API, as she is only interested in the current situation, and will monitor the mentions of #Alphaland over the last seven days. Our diplomat gets in touch with a data scientist from the data diplomacy unit (Chapter 3) of the MFA, who helps her collect the relevant data – tweets containing #Alphaland – in a programming language, such as 'Python' or 'R'.

After access to the data has been established – which is analysed in further detail in Chapter 4 – the data needs to be effectively organised so that it can be analysed. Cleaning data can involve deleting duplicates, errors, checking incomplete data, etc. This can be done using automated procedures or manually, depending on the problem and the available knowledge and technologies. Once the data has been cleaned, it needs to be analysed. By analysing data, the data scientist is able to observe patterns to better understand behaviours. Data scientists usually start by conducting an exploratory data analysis, to get a grasp of what the data contains. Exploratory studies could also lead to more data cleaning by identifying gaps or errors in the dataset. Next, the data is modelled and visualised according to the needs of the study.

In our example, the data scientist cleans the data, using commands in Python, and filters the data, so that only tweets with a location inside Betastan are included. Following a discussion with the diplomat, the data scientists choses the types of analyses that might be most relevant. For example, which hashtags were used most often in the combination of #Alphaland? Which regions within Betastan tweet most about #Alphaland? Which users are most active in tweeting about #Alphaland? These insights are relevant for the diplomat, as she can now better determine the priorities and areas of potential for the trade mission. She verifies the information with other sources she finds online, and compiles a report that contains visualisations of the analysis, to submit to her supervisor.

As a result of the massive amount of digitally stored data, big data analysis has several important advantages. For example, as data scientists collect and process a lot of data, instead of using a small amount of samples, margins of error decrease to a minimum. In other words, more data equals less statistical error.

Big data is also different than more traditional kinds of data analyses, as it often does not aim to discover a cause, but rather a correlation. By analysing massive amounts of information about particular events and everything that might be related to them, analysts can look for patterns. In short, 'big data helps answer what, not why, and often that's good enough.'[21] However, in diplomacy, knowing 'what' may not be sufficient since understanding 'why' something is happening could lead towards deeper causes, which essential for dealing with foreign policy situations properly.

Despite big data's many advantages, it is not the ultimate, bias-free statistical solution. Many of the challenges related to traditional statistical analyses remain, and are discussed in Chapter 4. For example, having high-quality, reliable data remains of utmost importance to make sure that the analysis is valid. It is therefore very important to consider how the data was collected and processed. How was the data 'cleaned' and screened before it was used to make the analysis? The enormous quantity of big data means that ensuring the quality of the data is even more complex than for ordinary data analyses.

Even with the best data imaginable, we will still have trouble in accurately predicting future events, and 'we may never be confident that we have understood the pattern to the point where we believe we can isolate, control, or understand the causes.'[22] In the area of diplomacy, Hocking and Melissen describe these challenges in the following way:

> The growth of 'datafication' means that, almost imperceptibly, size permits the acceptance of inaccuracy. There is a discernible trend in the direction of causality being replaced by correlation, and a risk of trivialization of the distinction between the two. It is important to underline such crucial differences, as it is to bear in mind that 'big data' cannot be used to make a prognosis of future developments. The potential for policy lies in the capacity of 'big data' to detect certain patterns in human behaviour and the characteristics of groups of people.[23]

In Hocking and Melissen's argument, 'inaccuracy' most likely refers to the messiness and unstructured data of big datasets, which inevitably contain small inaccuracies. As these small inaccuracies are usually buried in datasets containing millions of (accurate) entries, they will often automatically be evened out. Large inaccuracies, such as a lack of fit between research interest or focus and collected data, or inconsistency in how the data is collected, should be avoided, as this can lead to very misleading conclusions (Section 4.2). In sum, big data that is of high quality and computed intelligently, can provide enormous benefits. However, the key to successful analyses cannot be found in simply having the largest possible amount of data; it lies in gathering and processing the right kind of data for the intended research objectives.

1.2.3.3 Data visualisation

Data visualisation is tasked with making the rather abstract results of big data analysis more tangible. The old adage of 'seeing is believing' gets an even more poignant

meaning in a world where knowledge is built on the level of abstraction of big data analysis. If the underlying data is meant to move people, motivate individual action, or stimulate change in an organisation, data and the results of its analysis need to be presented in such a way as to be graspable in an intuitive way.

Data visualisation has become a cornerstone of discussions in connection with the data-driven era, highlighting the need to present data in a manner which can aid understanding and facilitate decision-making.[24] It is useful to distinguish between visualisations which are intended to illustrate and explain results, and those that are drawn up to support the interpretation of data.[25] Visualisation is not only a matter of communicating big data findings to a wider public or utilising the findings as part of awareness raising. Data visualisation can be 'concerned with exploring data and information' in order 'to gain insight into an information space' to aid 'knowledge discovery'.[26]

In other words, data visualisation is an intermediary step in the move from data and information to knowledge. Yet, this comes with a caveat. Data visualisation can require further interpretation and translation work. It is important to keep in mind that any visualisation shows us one aspect of the data, while potentially hiding others. This is true for visualisations based on 'small' data as well as those based on big data. We need to be careful not to assume that big data visualisations are necessarily more accurate because they are based on larger data sets.

1.2.4 Algorithms

It is important to remember that the value of big data is created not through its accumulation, but through its analysis. While access and adequate management of data are important, it is the interpretation of data that generates knowledge. Increasingly, big data feeds into algorithms to ultimately generate AI.

Put simply, an algorithm is a description of how to do something, for example how to cook a meal or how to perform a computation. It is a step-by-step procedure for solving a problem.[27] In the world of big data, algorithms

become especially important as ways of making sense of data by sorting, recognising connections, and finding patterns.

When it comes to decision-making on the basis of algorithms, concerns have been raised about the implicit assumptions, which are at the bottom of computational algorithms. In contrast to a cooking recipe, algorithms for big data can only be understood by specialists. For non-specialists, it is often not possible to gain insight into the assumptions behind an algorithm. It is not without a hint of irony that the algorithm used by the Cambridge University Psychometrics Centre to predict individual psychological traits from digital footprints of human behaviour is called *Magic Sauce*. Concerns regarding algorithms relate to their potential for doing harm, ethical considerations, the impact of AI-driven automation on the job market, and algorithmic biases that disadvantage along socio-economic, racial, or gender lines.[29]

When compared to human decision-making on the basis of intuition, algorithms for big data might seem less biased. Yet, they operate on the basis of the assumptions about the world that are made as part of the specific description that underlies the algorithm. When it comes to big data for diplomacy, it is crucial to involve diplomats in algorithm design and to maintain transparent communication about the elements of the algorithm and related assumptions.

A full examination of the potential of AI goes beyond the scope of this research. Nevertheless, just like any other profession, diplomacy will most likely be faced with a certain degree of automation of its processes in the future. While traditional diplomacy, relying on empathy and creativity, is unlikely to be taken over by AI-powered robots – not least because there is an interest in keeping a degree of human agency in such negotiations – there are other processes that could be affected by AI. These could include certain processes in the consular department or the writing of basic texts and reports, which is an area that is currently experimented with in the field of journalism.

Different types of algorithms

Algorithms drive the analysis of big data. In essence, an algorithm is a set of rules that takes data as input and delivers a specific answer as output. This overview, based on Brandon Rohrer's classification of algorithms, lets us identify five types of algorithms.[28] From the description below, we can see how each kind of algorithm is linked to a different sort of question and a different sort of research interest.

1) Two-class or multi-class classification: Is it A or B? Is it A, or B, or C, or D?
The two-class classification algorithm lets you ask questions that have two possible outcomes as answer. For example: Does this satellite image show a missile? Is it better to focus on foreign direct investment in Tanzania or Taiwan? Will this user click on the top most link on the website of the MFA? A variation of this is the multi-class classification. At its core, it asks the question: is it A, or B, or C, or D? A multi-class classification can answer questions, within predefined parameters, such as: What is the topic of this newspaper article? What is the sentiment of this tweet about this policy issue?

2) Anomaly detection: Is this weird?
This type of algorithm looks for data points that are different, unexpected, or unlike the others. It operates without knowing exactly what the deviation is that one is looking for. It is useful when the anomaly is rare, or when researchers have not had a chance to collect many examples of the anomaly yet. Questions that this algorithm can answer include: Is this Internet message from a foreign leader typical? Is the trading pattern different from the country's past trade behaviour? Based on e-mail communication received, is the negotiator acting different than in previous negotiations?

3) Regression: How much/How many?
This algorithm does not provide a class or category as output, but a number. They predict output values based on input values. For example, a regression algorithm can provide answers to the following questions: What will the GDP of Spain be in the first quarter of 2018? How many Twitter followers will the MFA's or the ambassador's account gain next week? How many migrants are expected to seek asylum in the next three months?

4) Unsupervised learning: How is data organised?
This algorithm teases out the structure of data and presents an intuitive way of dividing it into smaller segments. It is performing clustering, a.k.a. chunking, grouping, bunching, or segmentation. At the core of the algorithm is a distance metric, a definition of closeness or similarity. The idea is that by dividing data into more intuitive chunks, a human analyst can more easily make sense of it and take further steps for analysis based on the new insights. Questions for unsupervised learning are: Is there a way of dividing these policy documents into smaller topic groups? Which types of social media users generally agree with the messages of the MFA? Which beneficiaries of consular support are likely to need the same kind of service?

5) Reinforcement machine learning: Can a machine decide?
This type of algorithm makes automated decisions and is best used in automated systems that make a lot of small decisions. The algorithm gathers data on the go and learns by trial and error. As such, it needs feedback on each decision. Typical questions that this algorithm can deal with are: Where should I place this information on the webpage of the MFA so that the viewer is most likely to click on it? Should I vote in favour of, against, or abstain from this proposal, based on my countries' context and the negotiation history?

1.3 Chapter summary

This chapter explored the core concepts related to data diplomacy. The novelty of the concept opens up many approaches and frames. The terminological confusion around the concept of big data, from general disagreement on its elements to the turf-battles between generalists and big data 'purists', adds another layer of complexity. In an attempt to make our definitions both comprehensive and concrete, we propose a three-part approach to data diplomacy, comprised of the use of data as a tool for diplomacy; data as a topic on the diplomatic agenda; and the ways in which data changes the geopolitical environment in which diplomacy is situated. This report focuses only on the first dimension, and further zooms in on the ways in which big data can be gathered, analysed and integrated into negotiations, policy planning, and other diplomatic activities.

For diplomats, data in itself is not new. In fact, data is as old as diplomacy itself. Nevertheless, the current 'big data era' does add new elements to the possibilities of data, such as the speed and volume at which it is generated, the cost-effectiveness of storing and analysing new forms of data, and the possibilities of linking big data sources. As a result, the power of big data lies in the way in which it is able to identify patterns, new insights, and unforeseen relations between variables. It emphasises correlation, rather than causation, which both provides opportunities and challenges for diplomacy. The generation of such complex insights can be visualised with new tools, although there should always be a consideration of the fact that visualisations almost always reduce the nuance of an analysis and might generalise the outcomes of a big data study. Finally, big data feeds into algorithms that can provide answers to a wide range of questions of relevance for diplomacy.

Notes

1 Quoted in Kurbalija J (2013) The impact of the Internet and ICT on contemporary diplomacy. In Kerr P & Wiseman G [eds] *Diplomacy in a Globalizing World. Theories and Practices.* Oxford, UK: Oxford University Press, p. 141.

2 United Kingdom Foreign and Commonwealth Office (2016) *Future FCO Report,* p. 9. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/521916/Future_FCO_Report.pdf [accessed 12 December 2017].

3 American Association for the Advancement of Science (2015) Scientific Drivers of Diplomacy. *Science Diplomacy 2015 Conference*, p. 19. Available at https://www.aaas.org/sites/default/files/Summary%20of%20Science%20Diplomacy%202015-Scientific%20Drivers%20for%20Diplomacy.pdf [accessed 12 December 2017].

4 Kurbalija J (2013) The impact of the Internet and ICT on contemporary diplomacy. In Kerr P & Wiseman G [eds] *Diplomacy in a Globalizing World. Theories and Practices.* Oxford, UK: Oxford University Press, pp. 141–159.

5 The Economist (2017) The world's most valuable resource is no longer oil, but data. May 6. Available at https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [accessed 12 December 2017].

6 Manor I (2016) Between two digital divides. *Exploring digital diplomacy blog*, 27 January. Available at https://digdipblog.com/2016/01/27/between-two-digital-divides/ [accessed 12 December 2017].

7 Green J (2013) Big data. The key to economic development. *Wired*. Available at https://www.wired.com/insights/2013/03/big-data-the-key-to-economic-development/ [accessed 16 January 2018].

8 Rimbey B (2017) Michal Kosinski: the end of privacy. *Insights by Stanford Business*. Available at https://www.gsb.stanford.edu/insights/michal-kosinski-end-privacy [accessed 12 December 2017]. Helbing D et al. (2017) Will democracy survive big data and artificial intelligence? *Scientific American*. Available at https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/ [accessed 12 December 2017]. Harari VN (2016) Yuval Noah Harari on big data, Google and the end of free will. Financial Times, August 26. Available at https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5d-d5a28c [accessed 12 December 2017].

9 Cukier KN & Mayer-Schoenberg V (2013) The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*, May/June issue. Available at https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data [accessed 12 December 2017].

10 Podesta J (2014) Big data. Seizing opportunities, preserving values. *Executive Office of the President*. Available at https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [accessed 12 December 2017].

11 The World Bank (2014) *Big Data in Action for Development*. Available at http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf [accessed 12 December 2017].

12 Stanton J (2012) *Introduction to Data Science*. Syracuse, NY: Syracuse University. Available at https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf [accessed 12 December 2017].

13 Kurbalija J (2007) E-diplomacy: The Challenge for Ministries of Foreign Affairs. In Rana K & Kurbalija J (eds.), *Foreign Ministries: Managing Diplomatic Networks and Optimizing Value*. Geneva, Switzerland: Diplo Books, pp. 305–338.

14 Berners-Lee quoted in Stanton J (2012) *Introduction to Data Science*. Syracuse, NY: Syracuse University, p. 9. Available at https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf [accessed 12 December 2017].

15 Kurbalija J (2002) Knowledge management and diplomacy. In Kurbalija J [ed] *Knowledge and Diplomacy*. Msida, Malta: DiploFoundation. Available at https://www.diplomacy.edu/resources/general/knowledge-management-and-diplomacy [accessed 12 December 2017].

16 Ibid.

17 Groenfeldt T (2012) 'Big Data' often isn't really, says Bain. *Forbes*, 4 January. Available at https://www.forbes.com/sites/tomgroenfeldt/2012/01/04/big-data-often-isnt-really-says-bain/#1e9c45e37e6e [accessed 27 January 2018].

18 International Telecommunications Union (2015) *Big data - Cloud computing based requirements and capabilities.* Recommendation ITU-T Y.3600. Available at http://www.itu.int/itu-t/recommendations/rec.aspx?rec=12584 [accessed 12 December 2017].

19 Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group*, 6 February. Available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [accessed 12 December 2017].

20 Ibid.

21 Cukier KN & Mayer-Schoenberg V (2013) The rise of big data: How it's changing the way we think about the world. *Foreign Affairs,* May/June issue. Available at https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data [accessed 12 December 2017].

22 Stanton J (2012) I*ntroduction to Data Science,* Syracuse, NY: Syracuse University, p 62. Available at https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf [accessed 12 December 2017].

23 Hocking B & Melissen M (2015) *Diplomacy in the Digital Age*. Clingendael Report, p. 16. Available at https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf [accessed 12 December 2017].

24 Chen C *et al.* [eds] (2008) *Handbook of data visualisation*. Berlin and Heidelberg, DE: Springer.

25 Ibid.

26 Chen M *et al.* (2009) Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications* 29(1), pp. 12–19.

27 Neapolitan R & Naimipour K (2011) *Foundations of Algorithms*, 4th ed. London, UK: Jones and Bartlett.

28 Rohrer B (2015) Five questions data science answers. Brandon Rohrer's blog, 31 De*cember. Available at* https://brohrer.github.io/five_questions_data_science_answers.html [accessed 31 January 2018].

29 Tutt A (2016) An FDA for Algorithms. *69 Administrative Law Review 83,* pp. 84–123. Available at https://ssrn.com/abstract=2747994 [accessed 12 December 2017]). Jaume-Palasí L & Spielkamp M *(2017)* Ethics and algorithmic processes for decision making and decision support. *Algorithm Watch Working Paper, No. 2.* Available at https://algorithmwatch.org/wp-content/uploads/2017/06/Ethik_und_algo_EN_final.pdf [accessed 12 December 2017]. Furman J *et al.* (2016) Artificial intelligence, automation, and the economy. Executive Office of the President. Available at https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf [accessed 12 December 2017]. Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

2. The use of big data in international affairs

Much of the work of international relations is based on anecdote, sentiment and judgment– 'qualities,' not 'quantities'...But we ought to use more tools in the toolkit–because we now can. We need to tap data, and use our discrimination alongside it.
Kenneth Cukier, author of *Big Data: A Revolution That Will Transform How We Live, Work, and Think*[1]

More information does not always make for better analysis; and big data analysis depends on algorithms which the users of the analytical product may not understand.
Shaun Riordan, Senior Visiting Fellow of Clingendael and independent geopolitical analyst[2]

To understand how big data can serve diplomacy, we can begin by thinking about the benefits that big data is said to have for the private sector and other ministries. In these areas, big data is used as input to make better informed, evidence-based decisions. The service industry uses it to better identify the needs of consumers; organisations use it to streamline their operations; and researchers use it to gather insights into 'unknowns' that remain out of reach with more conventional methods of statistics.

It is clear that these promises cannot directly be mirrored onto the realm of diplomacy. Diplomacy is a specialised activity, rooted in tradition and protocol, which sustains the relations between states and other actors. The promises of big data do not translate one to one and the analogy with the private sector remains a deeply imperfect one. As we outlined in the introduction, diplomatic services and businesses differ in their functions: the work of diplomats lacks feedback loops provided by market mechanisms; diplomacy is less focused on immediate results; there is also a different perception of time at play. In the context of this chapter, two points need to be emphasised:

- **The qualitative nature of core functions**: Business activities offer greater opportunities for easily quantifiable targets. However, some diplomatic functions, such as reporting and negotiating, are qualitative at heart. Big data might be able to support these activities with insights, but does not seem to warrant fundamental changes or breakthroughs in what can be known.
- **The role of service orientation**: Diplomacy lacks the degree of service orientation that characterises businesses that work with big data. While the promise of better service provision might find application with regard to consular affairs, few other areas of diplomacy follow this logic of customer-service orientation.

A careful balance needs to be struck between exploring the potential of big data for diplomacy, and falling prey to the big data myth. When we talk about data diplomacy, we have to resist any temptation to replace sound political judgement based on experience and interpersonal contacts, with what seems like a neutral and immutable picture of the world as presented by big data. We need to keep in mind that big data is not as objective as it is often perceived to be; big data, as well as its analysis, can be tainted by biases and assumptions. Such concerns are, for example, now prominently discussed in the area of national security and intelligence.[3]

Big data can feed into policy planning, testing the assumptions of the diplomatic institution and challenging the biases of diplomats and foreign service officials. Further, big data is relevant for specific areas of diplomacy, such as development, climate change, and humanitarian affairs. As such, diplomats need to be aware of the role played by big data in these areas. Big data can play an even more central role when it comes to service provision in the area of consular affairs and designing better communication campaigns, especially on social media. In sum, the full spectrum of big data insights should, at least in principle, be applicable to diplomacy.

If we want to understand the impact of big data on the practice of diplomacy and, more specifically, the relevance of big data as a tool for diplomacy, we should start from an analysis of the core functions of diplomacy:[4]
- Information gathering and diplomatic reporting
- Negotiation
- Communication and public diplomacy
- Consular affairs

Big data, albeit to differing degrees, is of relevance to four of these functions and we will explore them in greater detail further down.

In parallel to looking at these functions of diplomacy, we also need to be aware of the areas of diplomacy and diplomatic negotiation in which data and big data now play a specific role. We can think of the relevance of scientific findings for climate change negotiations, or the impact that algorithms have on global financial markets. In this regard, we have identified the following areas and topics of diplomacy, which are sometimes referred to as diplomacies in their own right, as highly relevant in relation to big data discussions:

- Trade
- Development co-operation
- Humanitarian affairs and emergency response
- International law

We begin this chapter by looking at the core functions of diplomacy and then go on to explore various areas of diplomacy in relation to big data. The aim is to be able to pinpoint more precisely where the opportunities of big data in the practice of diplomacy are located.

2.1 Big data and the core functions of diplomacy

Scholars of diplomacy tend to define the practice of diplomacy as 'the dialogue between states';[5] a specialised activity carried out by select officials who enjoy special privileges and immunities. Harold Nicholson, a famous early twentieth-century diplomat and scholar, described diplomacy as 'the management of relations between independent States by the process of negotiation'.[6] Its purpose is to enable states to secure the objectives of their foreign policies. It achieves this mainly by communication between professional diplomatic agents and other officials designed to secure agreements.[7] With the emergence of new actors and new processes at the international level, this traditional perspective has changed, to a certain extent, to feature non-traditional diplomats and more inclusive processes guided by multistakeholder principles. New technologies, the Internet first and foremost, are among the driving forces behind this trend. We now highlight some of the opportunities and challenges of big data for the traditional practice of diplomacy between state officials, focusing on four of the core functions of diplomacy – information gathering and diplomatic reporting, negotiation, communication and public diplomacy, and consular affairs.[8]

Big data has a role to play in all of these areas, although it remains to be seen whether big data will truly revolutionise the core of diplomatic activities. While big data can provide insights that are not accessible on the basis of more traditional data, these insights are unlikely to fundamentally change the practice of diplomacy. Rather, big data can supplement qualitative assessment and more traditional data. While the core functions of diplomacy will most likely remain intact, big data can make these functions more efficient and better informed, challenging biases and assumptions, supplementing the knowledge and human judgment of the diplomat.

2.1.1 Information gathering and diplomatic reporting

Information gathering and diplomatic reporting are of central importance for diplomacy. On the basis of these activities, MFAs are enabled to make informed decisions with regard to strategy, policy, and communication and negotiation with counterparts. Big data is important in so far as it supports these functions, providing better evidence as the basis of strategy and policy decisions.

Big data can contribute to information gathering and diplomatic reporting in at least four ways.
- Provide a broader picture that supplements qualitative insights.
- Corroborate existing information.
- Challenge persistent assumptions.
- Generate insights that are otherwise not available.

The utility of big data in these areas is demonstrated by the newly established Open Source Unit of the UK Foreign and Commonwealth Office (FCO), which was set up to analyse open data to make foreign policy better informed. This unit has started to feed big data analyses into policy planning, and has found, on multiple occasions, that its insights challenged traditional assumptions in policy-making, corroborated existing information, and provided a broader picture. In addition, the unit is exploring ways to generate new insights through the analysis of non-traditional sources, such as examining social media to understand foreign support for extremist groups; looking at geospatial data to understand conflict dynamics in war zones; and creating a media dashboard to keep track of the latest developments related to hurricanes in the Caribbean.[9] This view is also underscored by the US State Department, which claims that 'harnessing the power of

voluminous and complex datasets (big data) can offer more meaningful insights required for informed decisions, problem solving and risk analysis'.[10]

2.1.1.1 Social media analysis

One of the big data sources with the greatest utility for the MFA is social media, largely due to the relative ease with which it can be accessed and used for analysis, as well as its convincing power derived from the millions of tweets, posts, and other content that it is based on. In fact, in one minute, there are 900 000 Facebook logins worldwide, almost 500 000 Tweets sent, and almost 50 000 posts uploaded on Instagram.[11]

What big data analysis on social media looks like:
- **Opinion mining** is a form of natural language processing which focuses on extracting opinions from texts written in natural language based on automatic systems.[12]
- **Sentiment analysis** is a combination of natural language processing, computational linguistics, and text analytics to extract 'subjective information in source material'.[13]
- **Network analysis** is based on Twitter data. A network of key participants in a debate on a specific topic is built which illustrates connections between participants. Key nodes in the system are assumed to be well-connected individuals, who can connect diverse groups and are important for a topic to gain traction in the discussion.
- **Hashtag tracking** follows the use of a hashtag on Twitter, or across social media platforms. Looking at it spread through a network and the quantitative changes in its use over time can provide insights into the development of a debate and the prominence that a topic has gained among the public.

Social media analysis in practice: Uncovering links between countries

An interesting example of using social media analysis to understand the relations between countries is the Mapping the World's Friendships project. 'Countries are sorted by a combination of how many Facebook friendships there are between countries, and the total number of Facebook friendships there are in that country'.[14] The resulting maps show emotional, social, economic, and historical links among countries. Diasporas as well as past colonial relationships also become visible. For example, the Marshall Islands have strong relations with the immediate geographic environment, but have the strongest connection with the USA (Figure 1), which can be explained by the past history of the two countries.

Figure 1: Mapping the world's friendships – relations between the Marshall Islands and the US[A][15]

This analysis can contribute to a number of diplomatic insights, including:
- An understanding of the key issues in public debates at home and abroad for a given topic.
- An understanding of the opinion leaders and influencers at home and abroad for a given topic.
- Monitoring social movements at home and abroad, especially in times of crisis and unrest.

Head of the Open Source Unit of the UK FCO, Graham Nelson, highlighted how this kind of social media analysis is used to check, and where needed correct, existing assumptions about key topics of debate and key opinion leaders.[16] For example, his Open Source Unit used social media analysis to better understand the British public debate about Britain's exit from the European Union (EU) and was able to correct assumptions about lines of

debate and key influences. As such, social media can offer a picture of parts of society that diplomats might not have access to or are not typically involved in. This comes with the caveat that any social media analysis can only give insights into those parts of a society that are connected to the Internet and actively use social media and as such it can be biased towards certain age groups, levels of income, and education. That having been said, traditional forms of data gathering might be affected by similar challenges of representation, as they depend on sample sizes that are minimal compared to massive amount of analysable data that social media platforms provide. While there might be no perfect method of analysis, it is encouraged to combine social media data with other sources of information to provide the most accurate picture possible.

2.1.1.2 Text mining and pattern tracing across diplomatic documents

While social media is one of the most prominent examples of applying big data analytics to the area of information gathering for political and diplomatic decision-making, big data is by far not limited to social media.

Text is central to diplomatic activities. Almost any diplomatic activity results in text, ranging from formal negotiated treaties via diplomatic notes, to reports from diplomatic meetings. Speeches and verbatim transcripts from conferences can also be considered in this context. Some of these documents are public like treaties. Others, such as diplomatic reports, are not accessible, except in the case of major leaks as it was the case with Wikileaks' CableGate. With text mining it is possible to analyse vast amounts of texts in order to identify patterns and, ultimately, get deeper diplomatic insights.

The vast amounts of texts available, are often held in an unstructured way. In order to make them accessible to big data analysis, texts need to be brought to a form that allows them to be compared; this includes taking care of spelling variations and other issues that prevent comparability, categorising the content of documents by adding tags, and providing metadata. These tasks require time and cannot be fully automated. This in itself might be a barrier to making quick use of the available insights. Looking towards the future, however, repositories of diplomatic texts for big data analysis can be built.

Text-mining approaches depend on the availability of verbatim records for a given conference and are most useful when transcripts of serial conferences are consulted, allowing comparisons over time. An example of the exploration of text-mining is the Emerging Language of Internet Diplomacy project, which looked for reference frameworks, concepts and approaches, terminology, and patterns of communication used at the Internet Governance Forum (IGF).[17] The project has been able, for example, to highlight the relevance of key topics on Internet governance and illustrate shifts in their relative relevance over time. It has also been able to trace the use of prefixes – such as cyber, net, e-, and online – and highlight changes over time. These insights can be important to understanding larger shifts in the debates and to reacting to them appropriately. Because text-mining approaches can take all the records from a given conference into account, they can be an important supplement to the insights of individual diplomats, who can only be in one place at a time.

The possibility of text-mining the outputs of diplomatic conferences, especially verbatim records, should also be taken into account. Transcripts of discussion, where available, can be used for data analysis to:
- Identify word frequency and make inferences about relevance.
- Find key words and identify their association with other words and phrases.
- Identify changes in the use of key terms over time.
- Measure the relevance of a specific issue in the discussion and how this changes over time.

2.1.1.3 Finding more patterns: past voting behaviour and policy decisions

Big data analysis typically involves the search for patterns. In the area of multilateral diplomacy, an interesting application is the analysis of voting patterns and the search for correlations across different forums and institutions. Combined with other quantitative and qualitative information, as well as diplomatic experience, this can lead to being able to better support a negotiation strategy by building alliances and anticipating responses.

In diplomacy, decisions are frequently made. Countries vote, issue statements and side on important policy issues. This leaves an important 'diplomatic trace', which could be analysed by big data analysis. One example is the analysis of voting patterns in the United Nations General Assembly and Security Council, which has in fact been conducted by the US State Department since 1984, as 'a country's behavior at the United Nations is always relevant to its bilateral relationship'.[18] The US State Department calculates the number of times for which both the USA and the other country casted similar votes on certain

policy issues. Moreover, it also recognises that voting records in the UN represent only one dimension of the relationship with another country, which is also tainted by other political and economic issues. Global Affairs Canada is exploring similar tools.[19] Tha Harvard Dataverse project contains a dataset of roll-call votes in the UN General Assembly 1946–2015.[20] This data is also analysed for affinity of nations' scores and ideal point estimates.[21] The more voting and policy decision data is collected, the more useful insights it could generate by identifying patterns in countries' behaviour.

2.1.1.4 Diplomatic reporting

Traditionally, diplomatic reporting is narrative-based, which means that the inclusion of quantitative elements might be met with resistance. However, this is an obstacle related to organisational culture, rather than an inherent and insurmountable characteristic of diplomatic practice.[22] The quality of a report is often determined by the perceptiveness and experience of the diplomat. Arguments are linguistic rather than data-driven. Statistics, surveys, and other data sources often take a backseat. Such quantitative elements frequently feature as appendices, as a way of supporting the main points of the report.[23] As such, it seems that there is little place for big data. Yet, it is worth highlighting that big data can add value to diplomatic reporting in a supporting role. The point is not to replace traditional insights, but to bring big data analysis and traditional insights together in a useful way that enhances the quality of insights and builds foreign policy knowledge. It is also useful to recall the distinction between data, information, and knowledge introduced in Chapter 1. Big data analysis is transformed into knowledge when brought together with insights based on experience. It is at this point that diplomatic reporting produces knowledge about a subject or situation. Similar to how some diplomatic reports feature statistics on the host country regarding GDP, employment, or education, insights generated by big data analysis, for example sentiments on Twitter regarding a specific policy of the sending state, can be featured in diplomatic reports to support the main argument. In addition, textual analysis of large amounts of past reports can reveal insights that can be fed into current reports about a specific country or topic, thus harnessing the already existing knowledge within the ministry.

Beyond the narrative nature of diplomatic reporting, there are good reasons for the reluctance to include big data insights into diplomatic reports. Among them are:[24]

• Lack of predisposition and 'culture' regarding the inclusion of quantitative data.

• Lack of relevant skills to validate big data analyses.
• Lack of suitable applications and ways to access data analyses.

However, big data, understood in a wider sense, can also help with making the knowledge in diplomatic reports more accessible. Interviewees suggested that the focus of big data analysis can also be turned inward to harness the institutional knowledge already available within the MFA.[25] Making diplomatic reports and other writing that the MFA produces easily accessible, searchable, and available for large-scale data analysis to extract insight, could be extremely valuable in harnessing diplomatic knowledge by way of creating a fuller picture of past and present reporting. Loss of valuable knowledge needs to be prevented. The collection and transformation of reports into a searchable format can be a starting point for many MFAs. In addition, textual analysis can be conducted on the basis of the many reports entering the MFA, to identify trends in global policy discussions and bilateral relations.

2.1.2 Negotiation

Without a doubt, negotiation is a fundamentally human activity that requires human insights, interpretations, and relationship building. Thus, it is one of the least prone activities for digitalisation. Digitisation of negotiations would only be possible if emotions and subtle cultural references can be mastered by artificial intelligence (AI). Hence, while it is very unlikely that computers will replace human negotiators in the foreseeable future, big data and AI will nevertheless have an impact on how and what we negotiate. Already, the discussed functions of information gathering and diplomatic reporting are at the core of researching the positions of negotiation counterparts and developing negotiation tactics and strategies.These are crucial at all stages of the negotiation process. As such, the often (almost) real-time nature of big data means that information on counterparts and relevant developments are not only collected constantly, but can impact negotiations in all their phases. In the following part, we look at the relationship between big data and negotiations from two perspectives.

First, we look at how big data helps pre-negotiations by understanding how countries represented in the negotiations view the negotiated issues. In particular, it is important in deciding whether a topic is ripe for negotiations as well as understanding the 'red lines' in negotiations which the other side cannot pass. Second, we highlight how big data analysis can provide a common ground and objective

criteria on the basis of which negotiators can more easily come to an agreement.

2.1.2.1 Know yourself, know the others – in the context of negotiations

Big data analysis of social media content can give insights into the attitudes of domestic audiences at home and abroad. This ability to use social media data to get a fuller picture of domestic discussions, discussion leaders, and sentiments, both at home and in the countries of counterparts, can play an important role in making better foreign policy and developing better negotiation strategies. One the one hand, this can support developing communication campaigns about ongoing negotiations for home audiences, in cases where negotiations are surrounded by a level of public interest and scrutiny. On the other hand, it can give clues about the kinds of discussions counterparts are facing at home. Reluctance to use certain terminology or reluctance to be seen making certain concessions on the side of counterparts might be related to domestic discussions and as such, this information can very valuable. For example, for the UK, mapping out social media discussion about Brexit has been identified as a way to map potential sensitivities in the Brexit negotiations.[27] Further, when it is impossible to have personnel on the ground, such as in the ongoing conflict in Syria, big data

analysis of social media and geospatial sources can provide insights at a distance that can be used to corroborate other findings.[28]

Big data analysis can also contribute to a fuller picture of past reactions and positions of counterparts with regard to the topic under negotiation or similar issues. Simon Pomel from Global Affairs Canada, supported the idea that a data-driven approach can inform negotiation tactics and strategies. He emphasised the potential of looking for patterns by analysing past responses in similar situations or with regard to similar issues.[29] This can be especially useful in multilateral settings as already elaborated. However, this way of using big data analysis remains largely unexplored.

Finally, there are machine-learning tools available, based on big data and text mining, to support negotiations   so-called negotiation support systems. In fact, they have been around since as early as the 1980s and are tasked with supporting one of the parties, or supporting the negotiation process as a whole. Based on the data they receive as input, they can advise on a variety of matters, even the seemingly human aspect of negotiations, such as trust building, flexibility, and fairness.[30] However, the fact that ever since they appeared in the 1980s, they have not been widely adopted, demonstrates the distrust towards such systems in completely grasping the element of human judgment that is at the heart of negotiations.

2.1.2.2 Objective criteria and common ground

In the seminal *Getting to Yes*, Fisher and Ury outline four elements of what they call principled negotiations.[31] Using objective criteria is one of the elements. It encompasses the idea that the outcome of the negotiation is based on a standard that is considered fair by all sides and developed independently of their influence.[32] Examples of such standards include market value, an expert opinion, a custom, or a law. Results of certain big data analyses can be counted as objective criteria. In conflict situations, data provided by the United Nations Operational Satellite Application Programme (UNOSAT) of the United Nations Institute of Training and Research (UNITAR) could play such a role. The geospatial data, such as satellite images of settlements or movements of parties involved in a conflict, collected and analysed by UNOSAT is heralded as an effective tool in emergency response. However, it might also serve the function of objective criteria. If accepted by all parties, such findings can prepare a common ground upon which an agreement can be built. This last point is important to stress. First, objective criteria are

not necessarily neutral and can contain implicit biases. Second, their value as part of negotiations depends on whether or not they are acceptable to all parties involved.

2.1.3 Communication and public diplomacy

The communication function of diplomacy (including public diplomacy) benefits, in a general sense, from improved information gathering and diplomatic reporting. The strongest potential of big data analysis lies in:

- an understanding of the discourse, perceptions, and behaviour of those that are communicated with;
- the ability to tailor messages to recipients; and
- the potential to measure the effectiveness of a communication message or campaign and, based on that, decide on resource allocation.

With regard to the first point, the old adage 'know your audience' takes on new meaning in the area of big data. Social media analysis in particular can help in better selecting a target audience and crafting a message that speaks to their key concerns and needs. In fact, many social media platforms provide the option to define a message according to demographic or interest group. On social media itself, the analysis of hashtags can help diplomats to connect with the desired audience and to speak to key concerns. Yet, the quality of the message also depends on well-trained personnel, and the effectiveness of communication will depend on the way in which the communication department can combine the new possibilities of technology with the art of crafting the right message. Big data results can be a useful guide, but cannot replace trained communications personnel.

More interestingly, with regard to the value of big data analysis, is the second point mentioned. In times of decreasing public funding and budget constraints faced by MFAs, big data analysis of social media can contribute to measuring the effectiveness of a communication message. It can answer questions such as: Has the desired audience been reached and what kind of impact does the message have? Further, it can help craft communication campaigns with greater impact and as such contribute to more efficient resource allocation.[33]

Public diplomacy relies heavily on the availability of data. Big data could revolutionise public diplomacy conducted on social media in four ways:

- Analysing the discourse and sentiments of foreign and domestic populations
- Identifying influencers on social media

- Adapting social media messages to target groups
- Monitoring the effectiveness of a public diplomacy campaign on social media

With the aim of improving foreign public perceptions, the first step is analysing the country image abroad, and how it is formed. Public diplomacy depends on listening, and listening can be achieved by measuring public opinion and analysing foreign media.

Social media is playing an increasingly important part in public diplomacy, as it offers the potential to connect more easily to a large number of people. It offers a vast amount of data that can be gathered and analysed. For example, sentiment analysis can be conducted on social media data to analyse attitudes towards particular issues, regions, or countries. Network analysis could point towards influencers and agenda-setters, and together with social media profiling, this method could support the creation of targeted messages.

Social media analysis is becoming a lucrative niche for consultants and platforms that analyse social media popularity over time, providing good-looking graphics on social media conversations around particular topics in different geographic areas, designed for businesses and other interested actors. Yet, the question remains: How does a ministry know whether it is having a real impact on its target audience? Can it be purely observed through Facebook likes and Twitter followers? How is a country or government discussed on social media, and are its users representative of a population? Numbers alone might not provide a full picture, as most followers might not actually read the messages on social media, or interact with the social media accounts. The complex communication across platforms cannot always easily be captured. 'The study of Public Diplomacy on social media is still struggling to find an appropriate research method that is able

to capture the complexity of the social media communication and assess engagement and participation.'[34]

This realisation has led to some scepticism towards the use of social media data for monitoring public diplomacy, especially when its analysis is conducted by external consultants. Two scholars even argue that public diplomacy's 'embrace of increasingly sophisticated analytical technologies – as they reflect and further fetishize policy relations and preferred narratives – is entrenching something very different from a foreign policy truly focused on peace, security and development.'[35]

Despite its limitations and shortcomings, social media analysis is probably one of the lowest-hanging fruits in understanding how a country, government, or ministry is perceived. The US State Department uses real-time monitoring of social media, through its Bureau of Public Affairs' Rapid Response Unit. It employs a team that reacts to social media developments that could have an effect on US national interests.[36] The UK FCO's Open Source Unit has analysed European sentiments towards Brexit using social media to map out conversations and sensitivities.[37] For Global Affairs Canada, social media analytics is the area that shows most potential for the integration of big data into diplomacy.[38]

For example, Global Affairs Canada mapped its 'digital diplomacy universe' by examining data from all its social media accounts on Twitter and Facebook, including all its interactions with other users. It geocoded all this data to have a better understanding of where its messages resonated, and found that it had most interactions with Brazil, India, Mexico, Myanmar, and the Philippines.[39] It should be noted, however, that both in the UK FCO and Global Affairs Canada, social media data and sentiment analysis is used to complement traditional data, rather than as a stand-alone piece of information.

The post-truth era: Is data still convincing?

Not only can big data be used to analyse social media trends, but outcomes of big data studies at the MFA could also be communicated through social media. However, it seems that the use of data to substantiate arguments is becoming less popular. The 2016 US Presidential Election, and particularly the failure of the polls to accurately predict its outcome, generated a steady stream of skepticism towards statistics and data analysis. Recent studies claim that the general public in the USA and the UK are increasingly distrustful of data and statistics provided by governments. For example, in October 2016 – a month before the US elections – a survey found that 43.6% of US citizens somewhat or completely distrusted 'the data about the economy that is reported by the federal government.'[40] In the UK, 55% think that it is either definitely or probably true that the 'UK Government is hiding the truth about the number of immigrants living here.'[41]

The common approach of using statistics and data to strengthen reasoning and convince someone of the validity of your argument sounds logical and commonplace. Whether it concerns informing the public as part of public diplomacy activities or a fellow diplomat about a diplomatic negotiation, the logic is 'If we can just give people the *real facts* then they will be better informed.'[42] However, and this might be particularly relevant in public diplomacy, a growing body of research demonstrates that this approach often fails, leading to 'the myth of myth-busting.'[43]

The combination of a potentially innate tendency to be swayed more by stories than by facts and the growing mistrust of data presented by government agencies might result in a situation in which facts become powerless in informing a citizenry. When there is a mistrust in facts and figures, an approach based on data might not be very convincing. Mitigating this challenge could involve relying on convincing visualisations or supporting the data with anecdotal evidence, integrating data into storytelling and bringing the data closer to the experiences of those with whom the data is communicated.

2.1.4 Consular affairs

One of the areas of diplomacy that might particularly benefit from big data is consular affairs. Compared to other departments within the MFA, and even compared to other government agencies, consular departments are usually relatively advanced in dealing with data. In fact, the Bureau of Consular Affairs în the USA even earned a special mention in the Government Big Data Awards in 2011.[44]

There are at least three reasons for the match between big data and consular affairs. First, consular work is the most service-oriented aspect of diplomatic activity. Second, consular activities already result in some of the most detailed records within the whole of the diplomatic service. Third, there are good reasons to be motivated to further improve consular services on the basis of big data analysis. Nelson stressed that in consular affairs, diplomacy is, in a sense at the frontline, and is dealing with customers when they are often at their most vulnerable.[45] Further, the services that the MFA is able to provide are directly linked to its perceived legitimacy at home.

It is important to stress that we do not suggest that big data can somehow replace the personal touch of consular affairs and replace the importance of personal visits and personal contact with consular officers. However, we are suggesting that big data holds the promise of better services and more targeted services that are driven by the identification of actual needs or foresight or emerging needs. As we discussed earlier, the big data trend as far as it is related to the private sector is, in part, concerned with improving services, customer experiences, and the efficiency which with these are provided. As such, there is considerable scope for the application of existing business solutions to consular services. For instance, big data can help analyse which group of citizens often use their services, or which group has trouble accessing e-government tools. It can predict at which points their websites get more or less traffic, and by which citizens. In response, consular departments are better able to tailor their services to the particular needs of certain groups in society, for example.

The citizen-turned-customer is a development that MFAs need to react to by fostering areas of greater service orientation, while also managing citizen's expectations. 'Governments are the victims of their own success and citizens often seek government assistance as a first rather than last resort.'[46] Given that most government departments are facing resource constraints, MFAs need to manage expectations through awareness campaigns that urge greater personal responsibility of travellers, while using the available technical tools, such as big data, to improve the effectiveness of their services.

While consular affairs have a relatively long history of dealing with datasets, the adoption of big data tools by consulates has also arisen from the pressure of the general public, which has raised its expectations when engaging with consulates. The challenge to meet these rising expectations was summarised during the Global Consular Forum in 2013:

> 'More' defines the consular landscape: more travelers, more overseas workers, more scrutiny, more complex case work, more emergencies, more exotic locations, and more expectations of a timely and personalized service. Technology is a major new factor, empowering governments, but also energizing clients more.[47]

To meet these challenges, we suggest three areas of focus:
- Making use of internal data to improve consular service delivery.
- Using innovative big-data supported means to locate citizens in need.
- Using (social) media monitoring to react faster to, or even predict, crises and needs for consular services.

First, consular services already have a plethora of internal data that, when processed accordingly and used in connection with suitable algorithms, can harbour new insights to improve service provision. Internal data for improved consular services includes:
- An overview of the number of tasks undertaken by the consular service and their time- and personnel-intensity.
- An overview of the kinds of requests and demands consular officers are faced with.
- A map of previous crisis situations and the response provided by consular affairs in terms of timing and time- and personnel-intensity.

With regard to internal data, algorithms can be employed to learn from past experiences and make suggestions for the timely and effective deployment of personnel. For example, it might be possible to build on a typology of crisis situations to know when and how many additional personnel are needed for support in response to a crisis. This can also include the deployment of staff from other departments to fill gaps in consular service provisions as the need arises. The prediction of increased personnel

demands would be the ultimate goal of such optimisations. Similarly, algorithms can also support the critical analysis of past situations and help pinpoint lessons-learned. In addition, making existing data comparable and connecting hitherto unconnected databases to gain new insights will be an important task for improved consular services.

Second, locating citizens in need is a major task of consular services. This is often done on the basis of registration systems. Expatriates and travellers to regions with travel warnings often have the option to self-register with their MFA. Registrations increase during times of crisis. But these self-registrations often give an inaccurate picture, because not everyone registers and those registered often fail to signal when they move away or when their circumstances change. Technology can help to simplify registration systems by creating user-friendly smartphone applications and using social media tools to make it easier to keep the registrations updated.[48]

Beyond that, big data tools can also be used to map the location and status of those in need of assistance and can help to create a more comprehensive map of locations and needs, especially in times of crisis. For example, similar to how mobile phone data is used after major natural disasters to pinpoint areas of high concentration of survivors, consular services can use this kind of data as a proxy that helps them focus their activities. Geo-spatial data can fulfil a similar role in highlighting areas of need. One can imagine 'futuristic scenarios with not only people but also their belongings being tracked down by means of GPS tracking'.[49] Examples from the development and humanitarian sector, especially the innovations explored by UN Global Pulse are good indicators of possible uses of big data in consular affairs (Sections 2.3 and 2.4).

Third, in addition to monitoring tools, big data analysis also harbours the promise of providing forecasting tools to anticipate crises and increased demands for consular assistance. For example, social media analysis that monitors social and political unrest, especially when used in combination with geo-location data, can be useful in anticipating the needs for consular services in a country or region. This is helpful in updating consular advice, but also in making provisions for having personnel available. It remains to be seen to what extent such big data tools can be built and effectively employed. The potential promise of faster reaction times and better service delivery in times of crisis is considerable.

Having said this, consular affairs is also an area of diplomacy which holds some of the most sensitive data, at least as far as the privacy of individual citizens is concerned. Questions of data security and anonymisation, especially before data is shared with external partners for analysis, are therefore most crucial here.

2.2 Big data and trade

Trade diplomacy depends a lot on data and statistics. Compared to other units in an MFA, the trade department is usually more used to collecting, processing, and managing data. In Global Affairs Canada, which addresses foreign policy, trade, and development, the trade unit is the most advanced and data-savvy.[50] Prior to the 'big data era', it was already used to aggregating large datasets, building human capacity around data science, and shifting towards a more data-informed culture. Senior management expects proposals and advice backed by empirical data with good baselines, and the ability to track progress.[51]

Building on this already-acquired data experience, big data provides new applications to international trade. The growth of e-banking and e-money systems, such as m-pesa, provide new opportunities to measure money flows and related effects on financial inclusion, income inequality, and economic growth. Consumer sentiments could be analysed using social media data or Google Trends. E-commerce platforms provide indicators of global price levels. Overall, the availability of real-time information on the state of the market could help to maintain stability in the global economy.[52]

A number of government agencies and regional and international organisations have started to look into how big data could practically be used for the benefit of trade. The Japanese Ministry of Economy, Trade, and Industry established a Strategic Council for Creating Data-Driven Innovation in 2014.[53] In 2016, the Asia-Pacific Economic Cooperation (APEC) launched the Advancing Big Data Application in Trade project, proposed by Taiwan, to 'explore the possibilities and benefits of a data-driven method for improving cooperation and information sharing in the field of international trade.'[54] A recent study by the World Customs Organization studied the application of big data to customs, such as e-commerce, the cross-border movement of cargoes, and sensor data derived by future applications of the Internet of Things.[55]

Monitoring international money flows – SWIFT

One of the more concrete examples of how international money flows can be monitored is through the analysis of SWIFT (Society for Worldwide Interbank Financial Telecommunication) data. SWIFT provides secure financial messaging services and is the dominant system used by financial institutions worldwide. SWIFT provides aggregated data for each type of message, the number of messages sent, and the value of the payments. By making this data available, it can be applied to serve several purposes, such as estimating GDP growth, monitoring the use of currencies in global and regional financial transactions, and nowcasting trade flows. Yet, there are limitations to SWIFT data: they need to be purchased, and there are limits on the download size of the data. Processing the data can be complex, and any publication of the data needs to be approved by the SWIFT.[56]

One of the international actors best positioned for the use of big data is the Universal Postal Union (UPU), which

estimates that there will be 'over 30 billion historical international postal tracking records by 2020', hourly updated and geo-located. At the moment, it has almost 400 million electronic customs declarations on its big data platform, as well as 75% of postal shipments.[57] With this data, the UPU is able to monitor postal flows, and ultimately provide important insight into trade and development (Figure 2).
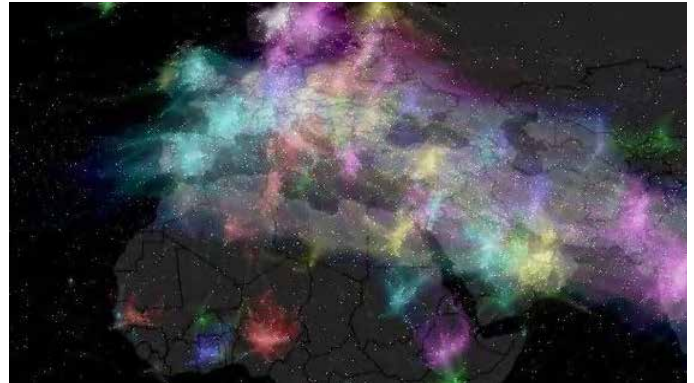
Figure 2: A visualisation of international postal flows in 2014 based on big data from the UPU, created by UN Global Pulse[58]

2.3 Development data

The way in which development diplomacy is incorporated into the activities of an MFA differs per country. While some ministries have merged development and foreign affairs, in others the two areas remain separate and are dealt with in a separate agency or ministry. In this section, we outline the contribution that big data can make to the development sector, regardless of the specific organisational structure. Big data could contribute to a more accurate picture of the situation and needs in remote places, both over long and short time spans. On the one hand, this allows for better informed, real-time decision-making, and on the other, for the monitoring of progress over time. As a result, big data has extensive potential to improve development efforts.

Like traditional sources of data, big data for development aims to give insight into human wellbeing and development. It supplements traditional statistical analysis and insights gained from interpersonal engagements on the ground. The promise of big data in development is to 'uncover hidden patterns, unknown correlations and other useful information.'[59] We can also say that on the one hand, we have 'digital patterns or signatures' that reveal shifts in behaviour or mobility, and on the other, we have 'digital smoke signals' that can serve as early warnings.[60]

In some cases, big data insights can be used to offer more cost-effective ways of gaining insight into, for example, the movement of people, which can be an indicator for migration or the spread of disease. Further, on the basis of big data, census data, which is not available in some development contexts, can be approximated.[61]

2.3.1 Big data for the Sustainable Development Goalss

However, it is important to realise that the role of big data in development has to be understood as part of a larger trend within development co-operation. This trend is best exemplified by looking at the shifts in relation to the importance of data that took place on the way from the Millennium Development Goals (MDGs) to the Sustainable Development Goals (SDGs). The importance of data for evaluation and measurement of progress was highlighted as one of the key lessons-learned on the road to the SDGs and the 2030 development agenda.

The 2015 MDG Report links this emphasis on quantitative data to four main reasons: to 'galvanize development efforts', to 'track performance and improve accountability', to allow 'governments at national and subnational

levels to effectively focus their development policies, programmes and interventions', and to foster 'evidence-based decision-making.'[62]

The current focus on the need for data to track progress towards the SDGs is reflected in the 230 indicators that are meant to provide data to monitor the 169 targets of the 17 goals. It goes without saying that this has created an enormous need for the collection and analysis of data. There is a growing recognition that new forms of data could help address this challenge, such as through earth observation data, crowdsourced data, mobile data, and open data. For example, satellite images could help monitor the proportion of the rural population who live within 2 km of an all-season road (SDG 9: industry, innovation and infrastructure) through software that identifies road coverage and population settlements.[63] Financial transactions can reveal economic differences between men and women (SDG 5: gender equality), sensors connected to water pumps can help monitor access to water (SDG 6: clean water and sanitation), content from local radio or television programmes can be transformed to text and analysed to detect patterns in discrimination (SDG 10: reduced inequality), deforestation can be monitored through satellite imagery (SDG 13: climate action), and illegal fishing activities could be monitored through maritime vessel tracking data (SDG 14: life below water).[64]

Other sources of big data for development include google searches and tweets. For example, to get an insight into the affordability of food and to contribute to SDG 2 (zero hunger), food prices are able to be monitored through Twitter analysis, creating a model that is very closely correlated with the official price of food. As a result, social media signals could be used as a proxy for food price statistics and as an early-warning system for price fluctuations, made available almost immediately, in comparison with official data that tends to be released with a considerable time lag.[65]

The recognition that big data might contribute to the rising data needs for monitoring the SDGs has also reached the level of policy-makers, who increasingly discuss its potential at the High-Level Political Forum on Sustainable Development, which is tasked to provide political leadership, guidance and recommendations for sustainable development. In fact, this point even reached the level of UN Secretary-General, who claimed that the UN development system 'must ramp up its ability to manage collected data and turn it into insights.'[66] That said, there is still a lack of clarity on the ways in which big data could effectively boost the SDG monitoring effort. Discussions about big data at this forum often do not specify exactly

how big data can best be integrated into the monitoring framework, how to address capacity gaps on big data in national statistical institutes, as well as some of the privacy and ethical considerations related to collecting data at increasingly disaggregated levels.[67]

2.3.2 Big data for development projects

Besides the use of big data to monitor global developments, big data is also used to inform development projects. One of the most prominent examples of big data use for development is mobile phone data, also referred to as call detail records (CDRs). UN Global Pulse gave a description of big data for development in five elements:[68]
- Digitally generated
- Passively produced
- Automatically collected
- Geographically or temporally trackable
- Continuously analysed

We can easily see how these elements relate to mobile phone data. The data is created digitally, produced through the interaction with online and mobile services, collected without human intervention, can be traced back to a location a call time and duration, and can be analysed in real time.

This, according to UN Global Pulse enables the measurement of:[69]
- Mobility – understanding the patterns of movement within a community.
- Social interaction – gaining insights into the geographical distribution of social contacts and approximating census data in places where this is not available.
- Economic activity – estimating household income based on monthly airtime expenses.

Insights derived from these measurements can play an important role in early warning, emergency response, health, socio-economic development, and transportation and infrastructure. For example, farmers in developing countries might be less informed about conditions of the soil, weather, and topography. Sensors could collect and analyse this data, and inform farmers at a continuous basis, for example by sending information to their mobile phones. Analysing patterns of online job search queries could provide inside developments on the labour market.[70]

Google Flu Trends is another prominent, yet somewhat controversial example. Search queries on Google related to flu symptoms were used to estimate 'the current level of weekly influenza activity ... with a reporting lag of about

one day.'[71] The argument is that 'the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms.'[72] In other words, an assumption was made that people using the Google search engine to search for flu symptoms is correlated with the occurrence of flu. This, of course, comes with the caveat that the availability and accuracy of results depends on Internet access and the use of search engines. It only works in areas where people use Google as their search engine frequently.

But more importantly, Google Flu Trends has become a cautionary tale for big data analysis. First, the algorithm suffered from 'over-fitting' by including searches that seemed related but were in fact not. 'Google's algorithm was vulnerable to overfitting to seasonal terms unrelated to the flu, like "high school basketball".'[73] This problem is related to 'big data hubris', where researchers can slip into the assumption that more is always better. Second, the algorithm did not take changes in search behaviour triggered by changes in the Google search engine itself into account. For example, as the google search engine was changed to include 'suggested posts' and 'health-related add-ons', search results changed.[74] However, the algorithm did not account for these changes in the search engine environment and the consequent change in observable behaviour. This second problem serves as a good reminder that the algorithm cannot be static but needs to be adaptable.

2.3.3 Big data for tracking aid flows, monitoring and evaluation

Development aid and co-operation involves the channeling of large money flows around the world, and there has long been a focus on capturing how much aid is channeled where. These flows are already being measured and captured by traditional data, and led to extensive databases, such as *AidFlows*, an initiative of the OECD, the World Bank, the Asian Development Bank, and the Inter-American Development Bank. In addition, the UN Office for the Coordination of Humanitarian Affairs (UNOCHA) operates the *Financial Tracking Service*, which visualises financial contributions, funding flows and appeals. While this institution was already established in 1992, it now increasingly automates the collection of data, although it still triangulates all input before uploading it to the database. There are also initiatives outside of the aid industry that attempt to monitor aid flows, such as *AidData*, which focuses mainly on geo-referencing data, and the *International Aid Transparency Initiative*, which provides standards for data on development co-operation.

In addition to tracking aid flows, big data can also feed into monitoring and evaluation (M&E) activities, which enables development and humanitarian aid providers to analyse the effectiveness of their programmes, learn from experience, and be more accountable. M&E depends on the collection of reliable data, which can often be challenging in the highly dynamic and insecure environments of humanitarian and development aid. Big data allows for the identification of new proxy indicators, potentially providing a more nuanced view of beneficiaries. As big data is continuously generated, it could also provide a more up-to-date picture of developments on the ground. Yet, despite the potential benefits of big data in M&E efforts, this data will always have to be checked and validated on the level of the beneficiary, as misguided outcomes could create mistrust, especially in fragile contexts and among vulnerable populations.[75]

2.4 Emergency response and humanitarian action

MFAs and related ministries often have to respond quickly to unfolding emergencies, whether it is for the identification of nationals in foreign countries or the delivery of aid to affected populations. In emergency response operations, success largely depends on the availability of timely information. In fact, big data might be able to identify needs 'with a level of speed and precision that have never been achieved before.'[76] At the same time, the use of big data in this sector for crisis response and humanitarian action 'in its intellectual and operational infancy', is usually limited to pilots and studies.[77]

The use of big data for emergency response largely depends on the kind of emergency that is taking place. The increased sophistication of tools to measure the environment, such as seismographs, satellites, and drones, continues to provide important mechanisms for monitoring the environmental aspects and early warning indicators of natural disasters. Recently, these tools have been complemented with more 'social' data, such as data derived from mobile phone usage or social media. Conversely, in conflict situations, a field where socio-economic indicators have traditionally dominated, there is an increased awareness of the utility of remote-sensing applications, such as monitoring building damage in conflict areas or

monitoring and mapping refugee camps in migration crises. A smart combination of these resources might be able to provide a clearer picture of upcoming, imminent, or recent emergencies.

It should be noted, however, that big data is generally of limited use in detecting broader trends in international relations, due to the limited number of occurrences of events, such as the outbreak of war or terrorist attacks, bearing in mind that big data and the machine learning that is necessary to predict upcoming events is usually based on millions of records.[78] At the same time, although its predictive power is limited, mining texts such as WikiLeaks and articles on international affairs could provide valuable insights into general patterns in international relations.

Figure 3: The average number of mobile phones moving in Haiti after the 2010 cholera outbreak. Thicker, darker lines indicate a larger number of travellers[80]

Figure 3 shows a map of population movements in Haiti, measured by 2.9 million anonymised mobile phone SIM cards, after the outbreak of the 2010 cholera epidemic in Haiti. Cholera is highly contagious, and the disease rapidly spread across the country. The study showed that mobile phone records were able to predict the spread of the epidemic, and this kind of data could therefore be key in preparing for, and responding to, future disease outbreaks.[81]

In another example, UN Global Pulse and the World Food Programme examined how people communicated during severe floodings in Mexico using mobile phone activity data combined with remote sensing data (Figure 4).[82] Comparing this data with periods of lower rainfall levels reveals abnormal activity patterns among the most

affected populations, as most calls were made from areas that were most affected by the flooding. The hope is that, ultimately, real-time mobile phone data could be used as an early warning indicator and emergency response mechanism. The research also showed that, in this case study, mobile phone data was highly representative of the population.[83]
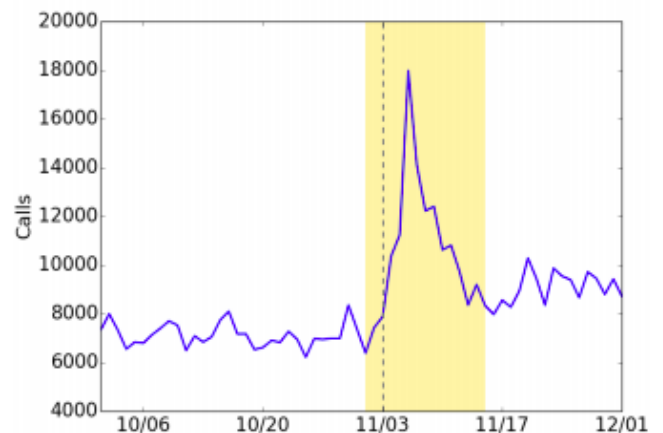
The plot above shows aggregated mobile activity for the cell phone towers closest to the floods. The yellow area indicates the time period of the floods.

Figure 4: Using mobile phone activity for disaster management during floods[84]

With the logic that people communicate more intensively during emergencies than otherwise, other communication channels have also been examined. UN Global Pulse studied whether social media data could support emergency response in the case of forest and peat fires, which revealed that there were common patterns between Twitter activity and the hotspots of the fires, although it also recommended for Twitter data to be combined with additional sources, such as satellite data, call records and mobility traces.[85] In another use of the social media platform, a study found that Twitter messages in Haiti could have been used to predict 2010 cholera outbreaks 'two weeks earlier than they were detected'.[86]

Similarly, the utility of Twitter for earthquake detection has been explored by the US Geological Survey, a government agency in charge of monitoring earthquake activity. It faced the challenge of not being able to cover the entire US territory with its 2 000 real-time earthquake sensors. In its search for a tool with a nationwide cover, it discovered that Twitter data could more quickly detect earthquakes, typically in less than two minutes, and with only two false triggers in a five-month period.[87] Yet, as it can only detects those earthquakes where human impact is felt, it needs to be used in combination with other monitoring tools.

2.4.2 Availability and access to data

These studies and pilot programmes have demonstrated great potential for using data from communication channels, from phone records to social media data, for the identification of affected populations. However, these channels do not come without certain challenges. The popularity of Twitter data for analyses like these is most likely related to the relative ease of obtaining Twitter data compared to the data of other social media platforms, as more of its data is made openly available, not least due to the public nature of Tweets compared to Facebook posts and messages, for example.

Obtaining call records might be even more difficult, as this data is highly sensitive and held in the databases of mobile operators. CDRs can include communication history, location history, and billing data. At the same time, phone records might be more representative of the population than social media data. A comprehensive framework for data sharing has not yet been developed. Acquiring this data could be expensive, and the complexity of the data might make it difficult to use.[88] Nevertheless, several partnerships between mobile operators and government agencies and international organisations or NGOs is gradually evolving. For example, the Romanian MFA launched a system of SMS alerts for Romanians travelling abroad in 2012, co-operating with mobile phone companies Cosmote, Orange, RCS & RDS, and Vodafone. Although not specifically designed for analysing mobile phone records, the agreement ensured that the MFA would not have access to databases containing the personal information of subscribers.[89]

2.4.3 Reliability and representativeness of communication data

Besides limitations on availability, the use of communication data is also constrained by its representativeness. The popularity of communication channels varies over time and space. The use of Twitter data for analyses in Indonesia and the USA is not surprising when considering that together with India, they formed the top three countries with the largest number of Twitter users in 2016.[90] Using Twitter in areas where the platform is less widely used among different demographics might lead to skewed results. In addition to varying preferences for social media platforms across countries, their use is also limited by the availability of the Internet. While social media penetration is more than 50% in North, Central, and South America, as well as Western Europe, East Asia, and Oceania, this number is around 30-50% in the Middle East, Eastern Europe,

and Southeast Asia. The regions that have so far been least penetrated by social media are South Asia (15%), Africa (14%), and Central Asia (7%).[91]

Differences in the use of social media and other methods of communication not only exist across borders and regions, but also across time. Relying on one source of communication for the collection of early warning indicators or the monitoring of a situation over time could be challenging when considering the fact that the use of such tools varies throughout the years. This is particularly evident for social media; whereas the early 2000s were characterised by AOL and MySpace, such platforms were later replaced by Twitter and Facebook. It is difficult to predict the lifespan of social media platforms and the trends for their future use. Even more traditional, seemingly stable tools, such as phones are experiencing fast-paced developments. The adoption of smartphones is growing faster than the adoption of more traditional mobile phones, and the way in which these devices are used is changing as well. In sum, when monitoring data on communication activity, whether based on mobile or social media data, there needs to be a continuous consideration of their general use in society, as well as corrections for their changing use and popularity over time.

2.4.4 Privacy considerations

Sharing and collecting big data for emergency response might generate many benefits. The stakes are also higher when data protection mechanisms are insufficiently taken into account. It is important to consider that populations that are affected by crises can be 'helped as well as harmed by the use of data'.[92] Therefore, as a guiding principle during emergency response, data should never be used only because it is available. Following the Ebola crisis of 2014, where CDRs were used for the first time to monitor population movements and thereby predict the spread of the disease, the international community was criticised for its lack of care for privacy provisions. There was a lack of data protection and data-sharing standards, as well as of anonymisation mechanisms.[93] Yet, this did not prevent 'many of the world's most important and trusted institutions from taking irresponsible, at best, and illegal, at worst, risks with some of the world's most sensitive data'.[94]

2.5 The use of big data in International law

As individuals, institutions and governments leave behind an increasing array of digital traces through their use of digital tools, big data can contribute to international law by enhancing accountability and providing new sources of evidence. With data and information taking new forms, and more and more evidence having a digital format, practitioners of international law have to cope with the evolving nature of evidence. International tribunals and the International Criminal Court have identified the challenge of 'collecting, storing, and analyzing large quantities of information that could potentially be presented as evidence', in particular digital items.[95] The sheer volume of such data could exceed the capacity of courts, and problems could arise if confidential data is obtained outside the proper legal framework, or has been interfered with by an intermediary. In addition, it can be difficult for international lawyers and prosecutors to present such information in a way that is accessible and understandable.[96]

Big data can also help monitoring human rights violations. In 2017, the UN Office of the High Commissioner for Human Rights (OHCHR) announced a partnership with Microsoft on big data. Microsoft will develop a platform aimed to support human rights staff by aggregating information on countries and rights violations in real time through the analysis of internal data, external public data, and social media data.[97] Using big data analytics and AI, the platform can help verify human rights violations by cross-checking datasets and identifying additional clues, and it can also help monitor the situation in areas where staff cannot directly be employed.[98]

2.5.1 Social media data in international courts

The potential future evidence derived from social media could be both an enormous opportunity, as well as a challenge, for international courts. This was concluded at a workshop at Berkeley with the participation of international tribunals and the International Criminal Court in 2012. With the large variety of data – photographs, videos, messages, documents, e-mails – the technical challenges of accurately processing and analysing it as evidence could grow. In addition, social media data can be relatively easily interfered with, raising questions of how to verify the authenticity of such content.[99]

Five years after the workshop in Berkeley, the International Criminal Court seems to have started to move into the realm of social media. In August 2017, it issued an arrest warrant for Mahmoud Mustafa Busayf Al-Werfalli, a Libyan national, for alleged war crimes. Social media is explicitly mentioned among the evidence that the arrest warrant is based on, as all crimes are confirmed by videos posted on social media platforms.[100]

However, the increased reliance on web-based materials, such as posts on social media, carries important risks. The most obvious one might well be that such posts could be deleted, or made unavailable by the platform. There is an important challenge for courts to systematically preserve information from open sources and archive photos, video material, and texts from these platforms.[101]

2.5.2 Geospatial data in international courts

Satellite images are used by a number of actors to document human rights violations. As satellite data is often perceived as objective, and as it is able to visually depict a situation, rather than describe it in words, it is an increasingly attractive advocacy tool. In 2013, Amnesty International used satellite images of political prison camps in North Korea to better understand the situation in these camps.[102] Amnesty International obtained the images from DigitalGlobe, a provider of Earth imagery. Subsequently, it examined the material on the changes over time (2011–2013) on a number of variables, including perimeter fencing, checkpoints, building properties, economic activity, infrastructure, and food facilities. Based on the images, the study could make claims about the growth of the camp's population (construction of new housing structures), the camp's surveillance and control mechanisms, and the kind of forced labour activities (the decrease in forested areas around the camp indicating logging activities). Importantly, Amnesty International chose to complement the satellite images with testimonies, to better contextualise the findings.

Satellite images are also used by other human rights advocates. For example, Google Earth and the United States Holocaust Memorial Museum (2009) partnered to produce satellite images of the destruction in Darfur. Furthermore, the Satellite Sentinel Project was set up in 2010 to bring attention to the mass atrocities in Sudan. Just like Amnesty International, the project received its imagery from DigitalGlobe, which co-operated with the project to 'capture imagery of possible threats to civilians,

detect bombed and razed villages, or note other evidence of pending mass violence.'[103] The latter initiative, however, received criticism for generating unintended, adverse, consequences.[104] For example, the project was perceived to be biased in favour of South Sudan. Furthermore, the satellite images might have provided cartographies of violence that could be used by hostile parties to map out vulnerable areas.[105]

In international courts, satellite images were first used in the Srebrenica trials at the International Criminal Tribunal for the former Yugoslavia, and with the growing ease of obtaining remote sensing data, they seem to be used increasingly more often. Yet, as is the case with the inter-action between diplomats and data scientists, there is a gap in understanding between the international law com-munity and analysts of geospatial information. Whereas the former sector needs technical knowledge of satellite imagery to judge their application to legal cases, the latter needs to learn about the legal principles and procedures that the satellite data needs to comply with.[106]

Although satellite data seems to be a perfect solution for unbiased, neutral evidence that can be brought to a human rights court or an international tribunal, several studies point out that there are significant challenges related to using this data in international courts.[107] One of the issues is that satellite data can be altered, and these alterations are difficult to prove. Therefore, before going to court, the data needs to be tested for its accuracy, objectivity, and authenticity. It gets even more complicated when

considering that there are no uniform standards to verify and validate remote-sensing data.

Furthermore, while the images themselves can seem objective, their interpretation could be highly political. The processing and interpretation of remote-sensing data requires the help of an expert, and cannot easily be done by the judge himself. Therefore, this work is usually per-formed by a third-party analyst. As these analysts often include humanitarian and human rights advocacy NGOs and international agencies, they might have an interest in presenting the data according to their own agenda.[108]

Generally, in the absence of uniform regulations, there seem to be no substantive legal obstacles to the use of remote sensing data, and satellite images. Nevertheless, they can usually only be submitted in combination with other – more conventional – types of evidence. In one case related to the maritime boundary between Ivory Coast and Ghana, which appeared at the International Tribunal for the Law of the Sea, satellite pictures were used to show 'traces of pollution'. The court ultimately decided that:

> Analysis of satellite imageries like any other remote sensing images is not sufficient to reach a conclusion on the existence of a perceived or targeted phenomenon. That is why all remote sensing data are confirmed by direct observa-tion through ground inspection, drilling or sam-ple collection. Therefore, remote sensing data such as analyzed satellite image cannot be relied upon as credible evidence of pollution.[109]

2.6 Chapter summary

Big data has the potential to contribute to diplomacy in a variety of different ways. For the core functions of diplo-macy – information gathering and diplomatic reporting, negotiation, communication and public diplomacy, and consular affairs – big data has the potential to contribute with insights and make certain processes more effective and efficient. However, the relevance of big data in this area is highly dependent on the degree to which these pro-cesses are guided primarily by human attributes, such as interpersonal relations, empathy, experience, and expert knowledge. In all cases, big data supports processes and activities but does not alter them fundamentally.

In terms of information gathering, the combination of the proliferating amount of data and information publicly available on the Internet and the rise in the capabilities

and cost-effectiveness of computing power has given diplomats the opportunity to explore new areas of infor-mation that were previously unlocked, whether it is the public discourse on social media or the many texts and documents that are published online. The same comput-ing systems are able to better analyse the large amount of information that is already within MFAs, such as reports compiled over the years that are now digitised, provid-ing new opportunities to capture insights on the basis of diplomatic reporting, for example. In short, we argue that big data can support information gathering by provid-ing a broader picture that supplements other qualitative insights, corroborating existing information, challeng-ing persistent assumptions, and generating insights that are otherwise not available. Big data-supported ways of information gathering include social media analysis, text

mining of diplomatic reports and other written material, and analysing past voting behaviour and policy decisions. While diplomatic reports are quantitative in nature, the information gathered on the basis of big data can serve a supporting function and be included similarly to how some reports make use of statistical insights.

We argue that the potential of big data for supporting negotiations builds on information gathering. Big data insights, based for example on social media analysis or analysis of media discourses, can provide a fuller picture of domestic discussions and the related sentiments at home and abroad. As such, big data plays an important role in improving foreign policy and developing better negotiation strategies. In addition, we argue that big data insights, especially satellite data, might provide a common ground on which negotiations can progress towards an agreement. This is not to say that satellite images are always objectively given or that their analysis is not political, but if all parties agree on the same interpretation, they can serve as a common ground.

Big data has the potential for developing better targeted communication messages and measuring the effectiveness of communication campaigns – especially as it relates to social media. Better targeted communication is also important for resource allocation and making decisions about where to focus scarce resources. Public diplomacy has made particular strides in not only engaging with social media, but also with the corresponding analysis. It might hold broader lessons for using big data for communication purposes.

Consular affairs, with its plethora of data already collected internal data, has considerable potential to benefit from big data. In order to benefit from the potential of big data, we suggest that consular services make use of internal data to improve consular service delivery, use innovative big-data supported means to locate citizens in need, and use (social) media monitoring to react faster to, or even predict, crisis and needs for consular services

When moving the application of diplomacy to specific domains, it becomes clear that those areas that are already experienced in the management of data, such as trade units, are most likely to be the first to benefit from big data, as there is already an awareness of the value of data, as well as a certain degree of technical and human capacity. Quantitative data has always played a significant role in trade, and this awareness of its value has often been deeply integrated into trade units and their management. With mobile money, e-banking, and e-commerce, it is opening up to new sources to measure the state of the economy.

Similarly to trade units, the areas of development, humanitarian affairs, and emergency response have always had to face challenges to attain data and information across time and geographical distance, or in quickly developing crises. With the quick pace through which some big data becomes available, and with the ability of big data to identify patterns and trends over time and space, big data has opened up new possibilities, for example to track and monitor development, and to better target the delivery of humanitarian aid. In addition, it can feed into early warning systems or heighten situational awareness, whether through social media monitoring or the analysis of geospatial data.

Finally, big data opens up new possibilities for accountability and evidence in international law. Data derived from social media or satellites allows for new ways to gather evidence, such as in the case of international courts. At the same time, it is of paramount importance that this data is objective and can be trusted. Courts should not move into the area of digital data unless they are absolutely certain of the data, which requires new technical skills and knowledge on the interpretation of new data sources.

Notes

1   Quoted in Segal A (2013) Big Data: An Interview with Kenneth Cukier and Viktor Mayer-Schonberger. Council on Foreign Relations blog, 4 April. Available at https://www.cfr.org/blog/big-data-interview-kenneth-cukier-and-viktor-mayer-schonberger [accessed 31 January 2018].
2   Riordan S (2017) We Need to Talk About Algorithms. *BideDao blog*, 5 November. Available at http://www.shaunriordan.com/?p=544 [accessed 12 December 2017].
3   Van Puyvelde D *et al*. (2017) Beyond the buzzword: big data and national security decision-making. *International Affairs* 93(6), pp.1397–1416.

4   The list of functions is based on Article 3 of the Vienna Convention on Diplomatic Relations (1961). The VCDR lists representation, protection of nationals and consular assistance, negotiations, information gathering, and the promotion of friendly relations, including public diplomacy, as the core functions of diplomacy. We have excluded representation from this list as big data has very few, if any, applications in this area. See also Wight M & Butterfield H [eds] (1969) *Diplomatic Investigations. Essays in the theory of international politics*. London, UK: Allen & Unwin. Neumann IB (2008) Diplomacy and Globalisation. In Cooper AF, Hocking B, Maley W [eds.], *Global*

Governance and Diplomacy. Worlds Apart? Houndmills, Basingstoke, UK: Palgrave Macmillan, pp.15–28.

5   Pigman GA (2010) *Contemporary Diplomacy: Representation and Communication in a Globalized World.* Cambridge, MA: Polity, p. 10

6   Freeman CWJ (1993) *Diplomat's Dictionary. Washington, DC: National Defense University Press. p. 101*

7   Berridge GR (2015) *Diplomacy. Theory and Practice,* 5th ed. Houndmills, Basingstoke, UK: Palgrave Macmillan.

8   As mentioned, representation is one of the functions of diplomacy listed in the VCDR. However, we will not explore the representation function, as big data seems of little relevance in this area.

9   DiploFoundation (2017) *Data diplomacy: Big data for foreign policy. Summary of a half-day event at the Ministry of Foreign Affairs of Finland.* Available at https://www.diplomacy.edu/resources/general/data-diplomacy-big-data [accessed 12 December 2017].

10   United States State Department (2016) *Department IT Strategic Goals and Objectives.* Available at https://www.state.gov/m/irm/itplan/264054.htm [accessed 12 December 2017].

11   Desjardins J (2017) What happens in an internet minute in 2017? *World Economic Forum*, 31 August. Available at https://www.weforum.org/agenda/2017/08/what-happens-in-an-internet-minute-in-2017 [accessed 12 December 2017].

12   Batrinca B & Treleaven PC (2015) Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY* 30(1), pp 89–116.

13   Batrinca B & Treleaven PC (2015) Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY* 30(1), pp 89–116.

14   Stamen (no date) Mapping the World's Friendship Project. Available at https://stamen.com/work/facebook-mapping-the-worlds-friendships/ [accessed 31 January 2018].

15   Stamen (no date) Mapping the World's Friendship Project. Available at https://stamen.com/work/facebook-mapping-the-worlds-friendships/ [accessed 31 January 2018].

16   Nelson G (interview 6 April 2017).

17   DiploFoundation (no date) *Emerging Language of Internet Diplomacy.* Available at https://www.diplomacy.edu/IGFLanguage [accessed 12 December 2017].

18   United States State Department (2017) Voting practices in the United Nations 2016. Report to Congress Submitted Pursuant to Public Laws 101–246 and 108–447, August 2017. Available at https://www.state.gov/documents/organization/273686.pdf [accessed 12 December 2017].

19   Pomel S (interview 12 July 2017).

20   Voeten E (2012) Data and Analyses of Voting in the UN General Assembly. Available at https://ssrn.com/abstract=2111149 [accessed 31 January 2018].

21   Bailey MA *et al.* (2017) Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution* 61(2), pp. 430–456.

22   Nelson G (interview 6 April 2017).

23   Saveska V (2017) Has diplomatic reporting shifted from narrative to data? DiploFoundation blog, 15 June. Available at https://www.diplomacy.edu/blog/has-diplomatic-reporting-shifted-narrative-data [accessed 12 December 2017].

24   Rosen Jacobson B & Höne KE [eds] (2017) Data diplomacy: Mapping the Field. Summary Report of the Geneva Data Diplomacy Roundtable, April 2017, p. 9. Available at https://www.diplomacy.edu/sites/default/files/DataDiplomacyreport.pdf [accessed 12 December 2017].]

25   Sini Paukkunen, quoted in DiploFoundation (2017) *Data diplomacy: Big data for foreign policy. Summary of a half-day event at the Ministry of Foreign Affairs of Finland.* Available at https://www.diplomacy.edu/resources/general/data-diplomacy-big-data [accessed 12 December 2017]. Pomel S (interview 12 July 2017).

26   Burri M (2017) The regulation of data flows through trade agreements. Georgetown Jou*rnal of International Law 48(1), pp.407–448.*

27   Nelson G (interview 6 April 2017).

28   Ibid.

29   Pomel S (interview 12 July 2017).

30   Vetschera R *et al.* (2013) Electronic negotiation systems. In James J Cochran [ed] *Wiley Encyclopedia of Operations Research and Management Science.* John Wiley & Sons, Inc. Available at https://www.semanticscholar.org/paper/Electronic-Negotiation-Systems-Vetschera-Koeszegi/a84d3cf96e9595c115b084c6a2f-daee1a70071f4/pdf [accessed 16 January 2017].

31   Fisher R & Ury W (2012) *Getting to Yes. Negotiating an Agreement Without Giving In*. London, UK: Random House.

32   Fisher R & Ury W (2012) *Getting to Yes. Negotiating an Agreement Without Giving In*. London, UK: Random House.

33   Nelson G (interview 6 April 2017).

34   Di Martino L (2016) Cross-platform communication and digital methods: an example. *Personal blog,* 9 July. Available at https://luigidimartino.com/2016/07/09/cross-platform-communication-and-digital-methods-an-example/ [accessed 12 December 2017].

35   Bean H & Comor E (2017) Data-driven public diplomacy: A critical and reflexive assessment. *All Azimuth*, 0(0), pp. 1–16.

36   Hanson F (2012) Public Diplomacy @ State. *E-International Relations*, 26 October. Available at http://www.e-ir.info/2012/10/26/public-diplomacy-state/ [accessed 12 December 2017].

37   Nelson G (interview 6 April 2017).

38   Pomel S (interview 12 July 2017).

39   Bhushan A (2015) Data dive into Canada's digital diplomacy #IODC 2015. *Canadian International Development Platform*, 28 May. Available at http://cidpnsi.ca/data-dive-into-canadas-digital-diplomacy-iodc-2015/ [accessed 12 December 2017].

40   Edison Research (2016) Marketplace Survey Wave Four - *Banner 1, October 2016. Available at* https://cms.marketplace.org/sites/default/files/anxiety-index-data.pdf [accessed 12 December 2017].

41   Faulkner Rogers J (2015) Are conspiracy theories for (political) losers? YouGov, 13 February. Availab*le at* https://yougov.co.uk/news/2015/02/13/are-conspiracy-theories-political-losers/ [accessed 12 December 2017].

42   Katwala S, Ballinger S & Rhodes M (2014) How to talk about migration. London, UK: Br*itish Future. Available at* http://www.british-future.org/wp-content/uploads/2014/11/How-To-Talk-About-Immigration-FINAL.pdf [accessed 12 December 2017].

43   Ibid.

44   Gourley B (2011) Government Big Data Award Special Mention: US State Department, Bureau of Consular Affairs. *Govloop*, 25 November. Available at https://www.govloop.com/community/blog/government-big-data-award-special-mention-us-department-of-state-bureau-of-counselor-affairs-2/ [accessed 12 December 2017].

45   Nelson G (interview 6 April 2017).

46   Wilton Park (2013) *Contemporary consular practice: trends and challenges*. Available at https://www.wiltonpark.org.uk/wp-content/uploads/WP1258-Report1.pdf [accessed 31 January 2018].

47   Haynal G, Welsh M, Century L & Tyler S (2013) *The Consular Function in the 21st Century*. Munk School of Global Affairs, University of Toronto, 27 March.

48   Wilton Park (2013) Contemporary consular practice: trends and challenges. Available at https://www.wiltonpark.org.uk/wp-content/uploads/WP1258-Report1.pdf [accessed 31 January 2018].

49   Hocking B & Melissen M (2015) Diplomacy in the Digital Age. *Clingendael Report*, p. 42. Available at https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf [accessed 12 December 2017].

50   Pomel S (interview 12 July 2017).

51   Ibid.

52   Hammer CL *et al.* (2017) Big Data: Potential, challenges and statistical implications. *IMF Staff Discussion Note,* September 2017.

53   Japan's Ministry of Economy, Trade and Industry (2014) METI to establish a Strategic Council for Creating Data-Driven Innovation. *News Releases*, 9 June. Available at http://www.meti.go.jp/english/press/2014/0609_02.html [accessed 12 December 2017].

54   Hung-lin Y in Taiwan Bureau of Foreign Trade (2016) Taiwan-proposed big data initiative adopted by APEC. *Speeches and News*, 11 August. Available at https://www.trade.gov.tw/english/Pages/Detail.aspx?nodeID=650&pid=572183&dl_DateRange=all&txt_SD=&txt_ED=&txt_Keyword=&Pageid=0 [accessed 12 December 2017].

55   Okazaki Y (2017) Applications of Big Data for customs - How it can support risk management capabilities. *WCO Research Paper*, no. 39. Available at http://www.wcoomd.org/~/media/wco/public/global/pdf/topics/research/research-paper-series/39_okazaki_big-data.pdf [accessed 12 December 2017].

56   Ibid.

57   Anson J (2017) UPU postal big data and trade. *4th International Conference on Big Data for Official Statistics*, 8 - 10 November,

Bogota, Colombia. Available at https://unstats.un.org/unsd/bigdata/conferences/2017/presentations/day3/session1/p-sessionB/2%20-%20Data%20Lake%20-%20Jose%20Anson%20-%20UPU.pdf [accessed 31 January 2018].

58 Universal Postal Union (2016) Postal big data offers key to nations' wellbeing. *UPU News*, 3 June. Available at http://news.upu.int/no_cache/nd/postal-big-data-offers-key-to-nations-wellbeing/ [accessed 31 January 2018].

59 United Nations Global Pulse (2013) *Big data for development. A primer*, p. 4. Available at http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf [accessed 12 December 2017].

60 Letouzé E (2012) Can Big Data From Cellphones Help Prevent Conflict? *IPI Global Observatory*. Available at https://theglobalobservatory.org/2012/11/can-big-data-from-cellphones-help-prevent-conflict/ [accessed 12 December 2017].

61 United Nations Global Pulse (2013) Mobile phone network data for development. Available at http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf [accessed 12 December 2017].

62 United Nations (2015) *The Millennium Development Goals Report 2015*, p. 10. Available at http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf [accessed 12 December 2017].

63 Ascosta Y & Moreno SL (2017) Exploring earth observations to monitor SDG indicators. *4th International Conference on Big Data for Official Statistics*, 8 - 10 November, Bogota, Colombia. Available at https://unstats.un.org/unsd/bigdata/conferences/2017/presentations/day2/session2/case1/6%20-%20Yineth%20Acosta%20and%20Sandra%20Moreno%20-%20DANE,%20Colombia.pdf.

64 United Nations (no date) *Big data for sustainable development*. Available at http://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html [accessed 12 December 2017]

65 Pulse Lab Jakarta (2017) Using big data for statistics to track the SDGs. *Medium*, 26 October. Available at: https://medium.com/pulse-lab-jakarta/tracking-the-sdgs-using-big-data-dad0ad351f2e [accessed 12 December 2017].

66 United Nations Economic and Social Council (2017) Repositioning the UN development system to deliver on the 2030 Agenda - Ensuring a better future for all: Report of the Secretary-General. A/72/124.

67 Rosen Jacobson B (2017) Leaving no one behind in the data revolution: An analysis of four years of High-Level Political Forum reports. *Diplo Policy Papers and Briefs*, no. 5. Available at https://www.diplomacy.edu/sites/default/files/Policy_papers_briefs_05_BRJ.pdf [accessed 12 December 2017].

68 United Nations Global Pulse (2013) Mobile phone network data for development. Available at http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf [accessed 12 December 2017].

69 Ibid.

70 Ali A *et al.* (2016) Big data for development: applications and techniques. *Big Data Analytics* 1(2), pp. 1–-24.

71 Ginsberg J *et al.* (2009) *Detecting influenza epidemics using search engine query data*. Available at http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html [accessed 16 January 2018].

72 Ibid.

73 Lazar D & Kennedy R (2015) What we can learn from the epic failure of Google Flu Trends. *Wired*, 10 January. Available at https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/ [accessed 31 January 2018].

74 Ibid.

75 Mukarji O (2016) Aid agencies' use of big data in human-centred design for monitoring and evaluation. *GCSP Strategic Security Analysis*, no. 8. Available at http://www.gcsp.ch/News-Knowledge/Publications/Aid-Agencies-Use-of-Big-Data-in-Human-centred-Design-for-Monitoring-and-Evaluation [accessed 27 January 2018].

76 World Food Programme (2015) WFP and UN Global Pulse show how big data can save lives and fight hunger. *News*, 8 April. Available at https://www.wfp.org/news/news-release/wfp-and-un-global-pulse-show-how-big-data-can-save-lives-and-fight-hunger [accessed 12 December 2017].

77 Data-Pop Alliance (2015) *Big Data for Climate Change and Disaster Resilience: Realising the Benefits for Developing Countries.* Available at http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-for-Resilience-2015-Report.pdf [accessed 12 December 2017].

78 Stuenkel O (2016) Big data: What does it mean for international relations? *Post-Western World*, 6 March. Available at http://www.postwesternworld.com/2016/03/06/mean-international-relations/ [accessed 12 December 2017].

79 United Nations Global Pulse & GSMA (2017) *The State of Mobile Data for Social Good Report*. Available at http://unglobalpulse.org/sites/default/files/MobileDataforSocialGoodReport_29June.pdf [accessed 12 December 2017].

80 Bengtsson L *et al.* (2015) Using mobile phone data to predict the spatial spread of cholera. *Scientific Reports* 5, no. 8923. Available at https://www.nature.com/articles/srep08923 [accessed 31 January 2018].

81 Ibid.

82 Pastor-Escuredo D *et al.* (2014) Flooding through the lens of mobile phone activity. *IEEE Global Humanitarian Technology Conference (GHTC)*, 2014 IEEE, pp. 279-286.

83 United Nations Global Pulse & GSMA (2017) *The State of Mobile Data for Social Good Report*. Available at http://unglobalpulse.org/sites/default/files/MobileDataforSocialGoodReport_29June.pdf [accessed 12 December 2017].

84 United Nations Global Pulse (2014) *Using Mobile Phone Activity for Disaster Management During Floods*. Available at http://unglobalpulse.org/sites/default/files/Mobile_flooding_WFP_Final.pdf [accessed 30 December 2018].

85 United Nations Global Pulse (2014) Feasibility study: Supporting forest and peat fire management using social media. *Global Pulse Project Series*, no. 10.

86 United Nations Office for the Coordination of Humanitarian Affairs (2012) *Humanitarianism in the Network Age*. New York, NY: United Nations Publication.

87 Earle PS *et al.* (2011) Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), pp. 708-715.

88 Data-Pop Alliance (2015) *Big Data for Climate Change and Disaster Resilience: Realising the Benefits for Developing Countries.* Available at http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-for-Resilience-2015-Report.pdf [accessed 12 December 2017].

89 Romanian Ministry of Foreign Affairs (2012) Launch of MFA system of SMS alert for Romanians travelling abroad. *Press release*, 29 October. Available at https://mae.ro/en/node/16336 [accessed 12 December 2017].

90 Statista (2016) Number of active Twitter users in leading markets as of May 2016 (in millions). *Social Media & User-Generated Content*, July. Available at https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 12 December 2017].

91 We are social (2017) *Digital in 2017: Global overview*. Available at https://wearesocial.com/special-reports/digital-in-2017-global-overview [accessed 12 December 2017].

92 United Nations Office for the Coordination of Humanitarian Affairs (2016) Building data responsibility into humanitarian action. *OCHA Policy and Studies Series*, May 2016, No. 18, p. 2. Available at https://docs.unocha.org/sites/dms/Documents/TB18_Data%20Responsibility_Online.pdf [accessed 12 December 2017].

93 Ibid.

94 McDonald M (2016) *Ebola: A Big Data Disaster*. India: The Centre for Internet and Society, p. 2. Available at http://cis-india.org/papers/ebola-a-big-data-disaster [accessed 12 December 2017].

95 Human Rights Center (2012) *Beyond reasonable doubt: Using scientific evidence to advance prosecution at the International Criminal Court*. Workshop report, p. 9, 23-24 October 2012.

96 Ibid.

97 United Nations Office of the High Commissioner for Human Rights (2017) Technology for human rights: UN Human Rights Office announced landmark partnership with Microsoft. *News and Events*, 16 May. Available at http://www.ohchr.org/EN/NewsEvents/

Pages/DisplayNews.aspx?NewsID=21620&LangID=E [accessed 30 January].

98 Microsoft (2017) Technology helps the UN advance the protection of human rights in new ways. *Microsoft Features.* Available at **https://news.microsoft.com/features/technology-helps-un-advance-protection-human-rights-new-ways/** [accessed 30 January].

99 Human Rights Center (2012) *Beyond reasonable doubt: Using scientific evidence to advance prosecution at the International Criminal Court.* Workshop report, 23-24 October 2012.

100 International Criminal Court (2017) *Situation in Libya in the Case of The Prosecutor v. Mahmous Mustafa Busayf Al-Werfalli.* ICC-01/11-01/17. Available at **https://www.icc-cpi.int/CourtRecords/CR2017_05031.PDF** [accessed 12 December 2017].

101 Koenig A (2017) Harnessing social media as evidence of grave international crimes. *Human Rights Center, UC Berkeley,* 23 October. Available at **https://medium.com/humanrightscenter/harnessing-social-media-as-evidence-of-grave-international-crimes-d7f3e86240d** [accessed 12 December 2017].

102 Amnesty International (2013) *North Korea: New satellite images show continued investment in the infrastructure of repression.* London: Amnesty International Publications. Available at **https://www.amnesty.org/en/documents/ASA24/010/2013/en/** [accessed 12 December 2017].

103 Satellite Sentinel Project (no date) *Our Story.* Available at **http://www.satsentinel.org/our-story** [accessed 12 December 2017].

104 Raymond NA *et al.* (2013) While We Watched: Assessing the Impact of the Satellite Sentinel Project. *Georgetown Journal of International Affairs,* Summer/Fall, pp.185-191.

105 Taylor L (2015) Violence in the age of visibility. *Personal blog,* 4 November. Available at **https://linnettaylor.wordpress.com/2015/11/04/violence-in-the-age-of-visibility/** [accessed 12 December 2017].

106 Bjørgo E (interview 6 April 2017).

107 Nunez AC (2012) Admissibility of remote sensing evidence before international and regional tribunals. *Amnesty International Innovations in Human Rights Monitoring.* Available at **http://www.amnestyusa.org/pdfs/RemoteSensingAsEvidencePaper.pdf** [accessed 12 December 2017]. Kroker P (2014) Emerging Issues Facing the Use of Remote Sensing Evidence for International Criminal Justice. *Harvard Humanitarian Initiative,* December. Available at **http://hhi.harvard.edu/publications/emerging-issues-facing-use-remote-sensing-evidence-international-criminal-justice-0;** [accessed 12 December 2017].

108 Ibid.

109 International Tribunal for the Law of the Sea (ITLOS) (2015) *Dispute concerning delimitation of the maritime boundary between Ghana and Cote d'Ivoire in the Atlantic Ocean.* Case no. 23, 20 March. Available at **https://www.itlos.org/fileadmin/itlos/documents/cases/case_no.23_prov_meas/Volume_III/S-EPA/Statement_of_EPA__20_Mar._2015_.pdf** [accessed 12 December 2017].

3. Organisational considerations for the MFA

[Diplomacy] has always adapted to changing circumstances,
including technological changes, and will continue to do so.
Sini Paukkunen, Head of Policy Planning and Research at the Ministry of Foreign Affairs of Finland[1]

Within the heightened debates on how organisations, especially in the private sector, can transform in order to meet the challenges, but also reap the benefits, of the data-driven era, the term 'data-driven organisations' has emerged. Building on this debate, this chapter clarifies what a data-driven organisation is and to what extent ministries of foreign affairs (MFAs) should orient their organisational considerations accordingly. While, as pointed out, the data-driven era touches on all aspects of society, it is crucial to recognise the specificities of the MFA and to carefully investigate the extent tw which this debate applies to its organisational culture and structure.

We begin this chapter by discussing characteristics of data-driven organisations. We argue that the call for data-driven organisations has to be understood as part of a longer history of adapting organisations to the opportunities and challenges of new information and communication technologies (ICTs) and promoting better knowledge management. In a second step, we explore the organisational culture of MFAs and potential organisational changes needed to integrate big data analysis. We propose the implementation of big data units, which are small in size and are given the space to innovate in order to explore the potential of big data for diplomacy. In the case of big data analysis, partnerships with external organisations will be inevitable and so we explore the shapes that such partnerships can take and point to some key considerations when creating these partnerships. The final section of this chapter hones in on capacity-building opportunities for MFAs in the area of big data, which will be needed at various levels for MFAs to adapt to a changing international environment and new technological possibilities.

3.1 Data-driven organisation?

To be data-driven means that decision-making is based on the collection and analysis of data. Transforming an existing organisation into a data-driven one is usually motivated by a desire to make better decisions, such as improve the quality of a service or product and improve internal workflows, which leads to greater success of the organisation. The focus on specific types of data and big data is determined by the type and ultimate aims of the organisation. This means that tools and techniques can differ from one data-driven organisation to the next.

3.1.1 Features of data-driven organisations and their relevance for MFAs

Some common features are often discussed in connection with data-driven organisations. In the following, we list them and highlight their applicability to MFAs:

- Data-driven organisations typically collect data themselves. MFAs need to consider the extent to which they are able to engage in data collection, especially when it comes to big data. This will depend on available resources and the feasibility of entering into relevant partnerships. In many cases of big data discussed in this report – for example, mobile phone data and satellite images – data will have been collected externally.
- A variety of people within the organisation have access to data and work with data. This has low applicability to MFAs, as there is likely only a small number of people required and qualified to access and analyse big data.
- The organisation will take care to attract and retain people with the skills to collect, analyse, and interpret data. MFAs will have to choose whether to outsource big data analyses, or to try to attract skilled personnel. Attracting skilled personnel might be more difficult

for MFAs than for the private sector, which often has a competitive advantage over the public sector, related to the salary and benefits they can offer. However, for a small unit within the MFA which focuses its work on data analysis attracting personnel with relevant skills is crucial. We explore this in of the subsequent sections in this chapter (Section 3.2.2).

- Collected data should be timely, relevant, as accurate and as unbiased as possible, and trustworthy. This is very important for MFAs, as basing policies and decisions on inaccurate, biased analyses can have adverse political consequences. Data held by the organisation is aggregated where possible and relational databases and queryable filters are used. Making aggregated data easily accessible and searchable is especially important in MFAs. Diplomats who want to make use of it can easily support their analyses and reports with the available data.

- A data-driven organisation embraces an evidence-based culture in all areas, especially when it comes to decision- and policy-making. The whole organisation participates in this culture, not only data specialists. While there needs to be an awareness of big data's potential across all departments of the MFA, there also needs to be a critical understanding of the limits of big data. Further, MFAs are characterised by an organisational culture which gives priority to narrative analysis. We do not suggest that this should change, rather that big data can usefully supplement such analysis.

- Senior management explicitly supports data awareness and data-driven practices. In those areas in which big data can contribute to the work of the MFA, this is important as MFAs tend to rely on strong hierarchical forms of organisation.

- The organisational culture should embrace learning and experimenting, especially in areas such as big data where many of the tools are either under constant development or are yet to be explored and developed. This is important in the early phases of exploring the use of big data. It is highly relevant for those units engaging with big data, such as the proposed big data unit, but of only medium relevance for the MFA as a whole.

- The organisation will typically encourage work across teams and work within diverse teams which allows for different perspectives to emerge and to be discussed. Teams should consist of both data scientists and those with expert knowledge in a field other than data science. This has limited applicability for the MFA as a whole. While we do not suggest that all departments should be structured in this way, the proposed big data unit would benefit from such an

arrangement and for this unit such considerations are of high importance.

With the realisation that solely basing decisions on data can be counterproductive – as big data is simply not perfect – organisations are less frequently adopting the term 'data-driven', and replacing it by the concept of 'data-informed'. This is a label that MFAs can embrace as they explore the possibilities and challenges of the data-driven era for the practice of diplomacy.

3.1.2 Types of roles in data-driven organisations

Generally, there are five types of roles that can be considered as essential in a data-driven organisation: data architects, data engineers, data scientists, business translators, and data visualisers. These roles and their distinctions are based on the experience of the private sector, but have relevance to MFAs. Data architects and data engineers who design and build data systems and products are not mandatory in every data-driven organisation unless the organisation plans to produce their own products in the area of big data analysis. Data scientists who analyse data to develop practical insights for the organisation are needed but their scarcity has become a challenge in the labour market. These business translators are able to convert analytics into insight and to ask the right questions vis à vis the data scientists to derive the right insights. Another important function in a data-driven organisation is visualising data. This skill allows the distillation of complex data into formats that can be understood intuitively and are more easily digestible for non-data scientists. Visualisation can also support the creation of dashboards for reporting. To what extent these roles should be duplicated in MFAs depends on the degree to which data collection, storage, and analysis will be handled in house and to what extent they will be outsourced. The role of business translator has special relevance for MFAs and such a role should exist when engaging in aspects of big data. People in this role can act as mediators between what the MFA needs and what big data analysis can deliver.

3.1.3 Organisational goals and knowledge management

Having outlined these basic characteristics and roles related to data-driven organisations, there are two points that are absolutely crucial to keep in mind. First, the aim to be data-driven only makes sense in relation to the specific goals that an organisation wants to achieve. As such, from the very outset, it is clear that businesses and MFAs differ with regard to the degree to which a data-driven focus can

and should be implemented. The aim of the organisation – in the case of the MFA to contribute to better foreign policy and its implementation – should be the driving force and determining factor behind any organisational transformation. The ideal characteristics of a data-driven organisation as outlined can only ever apply partially. This has in part to do with the organisational culture of the MFA but it is also determined by the fact that many areas of the work of the MFA, such as diplomatic reporting and negotiation as outlined, are qualitative in nature, and data and big data can only ever serve in a supporting function.

Second, the push towards the ideal of data-driven organisations should be understood against the background of two general developments. On the one hand, there is the push towards greater evidence-based decision-making in the public sector, as exemplified for example in the global development discourse and the Sustainable Development Goals (SDGs) in particular.

On the other hand, we can also read this in the context of discussions about knowledge management. The drive towards better knowledge management for various kinds of organisations is nothing new. Knowledge is used to define agendas and frame discourses. Effective knowledge management can therefore significantly add to the potential power in the context of a negotiation. And as such, data diplomacy is closely intertwined with knowledge management. In fact, Deputy Prime Minister of Belgium and Minister of Development Cooperation, Digital Agenda, Telecom and Postal Services, Alexander De Croo, even suggested that knowledge management will be 'revolutionized by big data analytics', as diplomats find themselves 'at the crossroad of various information sources and one of their added values is to synthesize them.'[2]

Knowledge management as a term and an organisational practice gained traction in the 1990s. It generally describes 'the process of creating value from an organization's intangible assets.'[3] It incorporates techniques and processes from artificial intelligence (AI) and developing knowledge-based systems, software engineering, human resource management, and management of organisational behaviour.

The digitalisation and exponential growth of data pose questions related to the efficiency and legitimacy of knowledge management. For example, how will diplomats effectively scan through the massive amount of information that is available online, and pick out the most relevant and reliable sources? How can the most be made of the great amount of data that is gathered by the MFA itself? As Hocking and Melissen point out, 'Gathering information may be easier for foreign ministries: processing and analysing it will be much more complex',[4] and it 'needs someone who can synthesise a vast range of information and extract coherence out of chaos'.[5]

The drive towards better knowledge management in organisations is typically animated by four main aims:[6]
- Intelligent access to information.
- Automation of procedures through the use of workflow.
- Automation of routine activities.
- Development of knowledge as an institutional resource.

These elements of knowledge management in organisations apply just as well to the era of big data. They support the storage, retrieval, and analysis of big data. At the same time, knowledge management procedures and activities also need to be updated to the specific requirements of using big data (Chapter 4). Putting the idea of data-driven organisations into this larger context of knowledge management is helpful for keeping in mind the ultimate focus of improving organisations on the basis of better using existing and easily available data.

3.2 Organisational culture and organisational adaptations

In this section, we explore the organisational culture of MFAs in relation to possible constraints on the implementation of big data analysis. Based on this, we offer suggestions for organisational adaptations in order to better incorporate (big) data science into the work of MFAs.

3.2.1 Organisational culture

There are elements of the organisational culture of MFAs that need to be carefully navigated and negotiated, if big data analysis is to be incorporated into the work of the MFA. Generally speaking, our interviewees and participants of the Geneva and Helsinki workshops, agreed that MFAs are based on a culture that values narrative analysis and focuses on written reports that build on the persuasiveness of carefully crafted language.[7] Arguments are often based on narrative, not on data and quantitative data is used less prominently. If it is used, it is usually employed in a support function.

Another obstacle for exploring the use of big data for foreign policy insight can originate from a lack of self-reflexivity on the part of those involved with, and well-versed in, big data analysis. There is a danger of, even implicitly, treating big data as a panacea that could replace the need for expert insight. Rising fears that big data might be used to replace expert knowledge can prove an obstacle to adopting big data approaches. In part, such concerns and related scepticism towards the use of big data within the MFA need to be addressed through careful communication about big data. As such:

- Big data is a tool to support good foreign policy. It does not aim to replace expert knowledge with automatisms.
- The idea that big data analysis is only accessible to those with the relevant technical or programming skills needs to be countered by presenting results, as opposed to talking about methodology, and emphasising the contribution these results can make to better foreign policy.
- It should be stressed that experts with subject knowledge, based on years of experience, are needed more than ever in the data-driven era because big data that is not embedded in its proper context can be dangerously misleading.

There already seems to be an increasing recognition of the importance of bringing in quantitative analysis and building on evidence-based decision-making. To support change towards incorporating big data insights into diplomacy and foreign policy-making, we need to focus on:

- carefully crafted relationships between those working quantitatively and those working qualitatively within the MFA.
- suitable changes in the organisational structure.
- appropriate communication regarding the opportunities and challenges of big data.

We explore these three points in more detail in the next section. To begin with, in order to successfully include big data insights into the work of the MFA, there needs to be a commitment from senior management.[8] Results of explorations of big data for diplomacy are, for the most part, uncertain. Because there is no clearly definable return on investment, political will and a commitment at a senior level to engage in such 'experiments' are crucial. The value of big data analysis needs to be clearly recognised at that level and communicated accordingly. Having said that, it seems that the best way to convince senior management and indeed all members of an MFA about the importance of adapting big data tools is to showcase the usefulness of these tools in concrete ways and highlight their application and value for specific questions or foreign policy challenges.[9]

Secrecy and the 'need to know' principle

MFAs have a lot of internal data which is generated by daily diplomatic activities. Access to this data depends on organisational procedures and cultures. Traditionally, diplomatic services followed the 'need to know' principle with information accessible only by those who should be concerned with specific data. If the data internal to the MFA is used for big data analysis, the principle of 'need to know' also needs to be discussed and potentially re-negotiated in the era of big data.

3.2.2 Possible organisational adaptations to support communication and cross-disciplinary teamwork

Changes to incorporate big data insights into the work of the MFA need to focus on collaboration, communication, and organisational adaptations. We argue that small organisational changes can foster teamwork across disciplines and communication within the MFA about the potential of big data insights. While we are aware that organisational changes always have to be explored with caution to avoid duplication of effort, we believe that a middle ground of minor organisational changes can benefit the integration of big data insights. These changes need to be driven by the aim to avoid duplication and to foster communication and collaboration.

We have two key suggestions for small organisational adaptations that can be implemented in order to begin to explore big data applications for diplomatic practice and foreign policy.

- **Big data unit:** Create a small, innovative unit within the MFA that can explore possible big data applications.
- **Big data champions:** Foster connections to all areas of the MFA by appointing 'big data champions' within relevant departments and units.

The **big data unit** proposed here draws on examples of similar units in the UK Foreign Commonwealth Office (FCO) and the Norwegian MFA. These units are both small, being around 10 people strong, and were established relatively recently. Such a unit allows for concentrating the exploration of big data insights for diplomacy in one place and avoids the duplication of efforts across units and departments. Those interested in the potential of big data have one central unit as a point of focus, a unit that can share experiences and provide results of big data analysis across the MFA. In organisational terms, the big data unit

needs to be understood as having a cross-cutting function. It should be able to serve a variety of regional and thematic departments in the MFA.

Building on the experiences in the UK and Norwegian examples and given that we are in the early stages of exploring the use of big data in diplomacy, we argue that such units should be given permission to experiment and explore possible applications freely without being bound to certain outcomes or results. In other words, failure should be acceptable and needs to be seen as part of the exploration process. Unit members need to be able to trust and follow their intuition. The unit should be able to operate in a creative and agile way.[10] It should be explorative and innovative.[11]

The unit should bring data scientists and diplomats together, which can bring both challenges and opportunities. The case of the Norwegian unit is a good example. It faced the challenge that the data analyst had never been involved in diplomacy before, while the diplomats had not yet engaged in data science. Establishing a common language or framework of understanding takes time. This poses challenges for capacity building and communication within the units. Nevertheless, this diversity is key in order to support the unit's innovative potential, as it is only through the combination of expert knowledge and data insights that big data can truly inform foreign policy and diplomatic activity.

The aim of the unit should be to contribute to solving specific problems and addressing specific needs in the MFA. The work needs to be driven by specific questions of relevance for diplomatic services and foreign policy, not by what is technologically possible. In other words, their first task is to adopt what is technologically possible with big data analysis to specific processes or questions in order to highlight the potential, but also the limits, of big data analysis. In a second step, bigger projects can be built up on the basis of this initial exploration.

Key characteristics of big data units in MFAs in a nutshell
- Small: comprising around 10 people.
- Diverse: bringing data scientists and diplomats together.
- Innovative: operating with a degree of freedom to explore big data applications.
- Problem- and needs-focused: addressing specific questions or identified needs.
- Cross-cutting: working closely with data champions in the different departments of the MFA

As important as the unit itself are the connections that it can draw with other units and departments in the MFA. It is crucial for big data units to stay in touch with the specific needs and questions that arise from the day-to-day work in other areas of the MFA. Thus it is recommended for other relevant departments to appoint a person who can serve as a **big data champion**. In doing so, the big data unit will have concrete counterparts in various departments whom it can contact. These big data champions do not need to have a background in big data. Rather, their function is to communicate needs and questions to the unit and feed big data solutions and applications back into the work of the respective departments and units. In short, the aim of establishing big data champion is to
- make sure that the right questions are asked.
- enable two-way communication between the big data unit and other parts of the MFA and the creation of synergy.
- foster exchange between traditional expertise and data insights.

Effective communication is particularly important when data collection is conducted in collaboration with other units of the MFA. When data is collected from different units and geographic locations, there is a high risk that the data is not standardised; it might be time-consuming to clean and process this data. Often, clear definitions and guidelines are not sufficient, and such teams need complete guidance in order to ensure that the data can be harmonised.[12]

While communication between the big data unit and the big data champion can and should happen on an ad hoc basis, there should also be meetings scheduled at regular intervals to share new and broader insights, and to maintain an up-to-date understanding of what is technologically feasible. Investment in a big data unit and the technology related to big data tools is only a first step. What will determine the success of bringing big data into the MFA is the ability to use the data to see if those tools actually work.[13] For this, the work of the big data unit needs to be communicated and made available to the relevant departments. The most effective way to raise awareness about the potential of big data is to begin with pilot studies that show the practical uses of big data tools for diplomacy.[14] In other words, the adage 'show, don't tell' applies to raising awareness of the potential of big data. For this, the proposed big data unit can play a crucial role.

3.3 Potential for partnerships

Partnerships with other other ministries and across government, academia, and the private sector are another core consideration for big-data-driven foreign policy insight. For example, the UK FCO participates in a cross-government data science network, and collaborates with the Office for National Statistics, the Alan Turing Institute, and universities.[15] To begin with, it is worth distinguishing between four drivers for entering into partnerships:

- A lack of internal capacity and the challenges associated internal retraining or hiring.
- An opportunity to build internal capacities through partnerships.
- A chance to gain access to data otherwise not available.
- A desire to build sustainable partnerships with relevant institutions to complement skill sets and avoid duplication of activities.

A lack of internal capacity can be addressed by (a) commissioning relevant studies on an ad hoc basis and (b) outsourcing big data analytics. **Commissioning single studies** on an ad hoc basis and on specific topics can be a first step to getting a sense of the potential of big data insights for foreign policy. Universities can be particularly valuable partners in this. Such pilot studies can be used to communicate the value of big data for diplomacy and foreign policy. However, it is difficult to leap from commissioned studies to integrating big data analysis more closely into the work of the MFA.

Outsourcing big data analytics can be a solution, if a more regular supply of big data insights is desired. Similarly to how some companies outsourced IT-related services in the 1990s, big data analytics has, to an extent, the potential to be outsourced. Unlike the commissioning of single studies, this includes a mid-term commitment of resources. Like the commissioning of single studies, it avoids making a commitment to hiring and retraining personnel or changing the organisational structure of the MFA by creating a new unit. It is likely that at least a certain part of the big data analysis process is outsourced, whether it is the collection, storage, management, analysis, or visualisation of the data.

Yet, compared to the private sector, there are important limitations for the MFA related to the utility of outsourcing big data analyses. The following set of concerns is important to keep in mind when looking at questions of outsourcing data science:

- In-house knowledge in relation to data science is still lacking in most MFAs.
- At the same time, the demand for data scientists on the job market is outpacing the supply of highly qualified professionals who can work in a diverse environment in which applications for data science need to be developed first.
- Outsourcing data analysis can save resources and be more cost-effective in the short run.
- Outsourcing to the private sector allows to harness the speed of innovation due to competition among private sector providers.
- It increases flexibility and can be helpful if longer-term commitments to new units or a changed organisational structure are not yet plausible.
- However, in the long run, developing in-house capacities, if feasible given available resources, will be more cost-effective than employing third parties and using external platforms for big data analysis.
- As such, bringing tasks that have been previously commissioned with external providers in-house is a task for the long-term development of data analysis capabilities of MFAs.
- In addition, it is worth highlighting that when it comes to sensitive issues, it is always advisable that MFAs develop their own in-house big data analysis capacities.

Relationships related to outsourcing big data analytics need to be carefully crafted. This is true for the private sector but it is of even greater importance in the context of MFAs. Many of the issues diplomats deal with on a day-to-day basis do not present the easy optimisation tasks that can be found in the activities of the private sector. To begin with, a clear idea of what insights are desired is crucial. In other words: What is the MFA looking for? What insights can support the work of diplomats? What are the challenges associated with this? These questions need to be carefully defined if partnerships, and the outsourcing of data analytics in particular, are to succeed. However, at this stage, it is precisely an absence of a clear definition of the potential of big data for diplomacy that needs to be addressed in the first place. In order to develop potential areas for big data analysis in diplomacy, close collaboration between diplomats and data scientists is needed. This is difficult to achieve by outsourcing and is best done in a small and innovative in-house unit as described in the previous section (Section 3.2.2).

In addition to commissioning studies and outsourcing data analytics, a third aspect – **building sustainable partnerships** for data insight – should also be carefully considered. The private sector, academia, and international organisations should be considered for such partnerships. As we have seen, a number of international organisations and related institutions, such as the United Nations Operational Satellite Application Programme (UNOSAT) and UN Global Pulse, are exploring and using big data insights. Similarly, research done at universities in the area of big data – applying big data to specific cases, or analysing the contribution of big data to society and government functions – needs to be kept in mind. Working with other ministries and government institutions can also support the work of the MFA in the area of big data. Building such sustainable partnerships will be particular to each national case, the potential of these institutions to work with MFAs, and the capacity of MFAs to support and sustain such partnerships.

Further, gaining access to data otherwise not available can be one of the strongest reasons to enter into partnerships, especially with the private sector. The private sector might be inclined to share data that is otherwise restricted when commercial and competitive interests are not touched. As such, the MFA might be able to access data otherwise restricted through carefully crafting relationships with the private sector. Additional considerations in this regard are outlined in Chapter 4.

Last but not least, partnerships for training and capacity building should be considered. Drawing on more experienced big data users, especially in the private sector and academia, to provide in-house and on-the-job training is needed and can usefully combine the development of sustainable partnerships with building internal capacities.

3.4 Capacity building

Data diplomacy requires data skills. Even if an MFA choses to outsource all its studies, diplomats commissioning these studies still need to be aware of what is possible and what is not. Broadly understood, capacity building can take place at three levels:

- Capacity building at the **institutional and policy environment** level describes the establishment of adequate institutions, laws, and policies.
- Capacity building at the **organisational level** includes establishing efficient structures, processes, and procedures within the organisation.
- Capacity building at the **individual level** focuses on developing adequate skills, knowledge, and competencies.

Exploring the institutional and policy environment level is a crucial task for diplomacy. Among other things, this concerns agreeing on international recommendations, regulations, policies, and law. The European General Data Protection Regulation (GDPR), is a good example of this. However, as outlined in Chapter 1, big data as a topic of diplomacy is beyond the scope of this report, which focuses on big data as a tool for diplomacy. The next level of capacity building, capacity building at the organisational level, was addressed in Section 3.2 as part of the suggested organisational changes, especially in the form of the big data unit and the big data champion. Hence, here we are focusing on capacity building at the individual level.

If we are interested in building capacities, it is useful to distinguish between **hard capacities** and **soft capacities**. Hard capacities are those that are technical, functional, tangible, and visible. On the individual level, these include technical skills, knowledge of appropriate methodologies, and explicit subject knowledge. In the case of big data, hard capacities include the skills to maintain databases and to develop algorithms. Soft capacities are social, relational, intangible, and invisible. On the individual level, soft capacities include implicit knowledge and experience; relational skills such as negotiation, teamwork, and conflict resolution; and problem-solving skills. In the case of big data, soft skills are important for adding a holistic understanding and interpretation of big data analysis. They are also crucial for supporting work in diverse teams and communicating big data insights. It is, however, often difficult to draw a line between tasks which require the application of hard skills and those which require soft skills; in fact, much of the work related to interpretation and communication of big data relies on both types of capacities (either of an individual or of a team).

What are the capacities that big data units in MFAs should have?

The UNECE outlines seven competencies that big data teams should poses, which presents a mixture of hard and soft capacities (Figure 5). It will be important to foster these within the big data unit and to keep these competencies in mind when making hiring decisions. Not all members of the big data unit need to posses all seven competencies. The focus should be on combining team members with different skill-sets and levels of experience and to foster these competencies at the team level.[16]

- **Team work:** 'Ability to work collaboratively with others, developing and maintaining good working relationships and sharing information and knowledge.'
- **Interpersonal and communication skills:** 'Ability to communicate with others in a fluent, logical, clear and convincing manner together with an ability to engage effectively with a wide range of stakeholders.'
- **Delivery of results:** 'Ability to deliver outcomes on time and to a high standard and ensure that goals are achieved.'
- **Innovation and contextual awareness**: 'Ability to observe environmental factors and exploit them for the work environment together with the ability to develop new ideas, concepts and solutions outside of established patterns.'
- **Specialist knowledge and expertise**: 'Possess appropriate specialist knowledge and expertise to work effectively.'
- **Statistical and IT skills:** 'Possess detailed knowledge and understanding of statistical methodology and concepts, ability to extract key messages or underlying trends within data and possess the IT skills relevant to statistical production and analysis.'
- **Data analytical and visualisation skills:** 'Ability to work with structured and unstructured data and combine data processing techniques to achieve outcomes, possess knowledge and understanding of data visualisation techniques relevant to big data.'

It is easy to see that diplomats and big data scientists bring different kinds and combinations of hard and soft skills to the table. In terms of big data for diplomacy, data scientists tend to have relevant hard skills more easily available, whereas diplomats already possess a range of

Figure 5: Big data team-level competency[17]

soft skills that are important for integrating big data into the work of the MFA. In terms of big data diplomacy capacity building, the needs of these two groups differ.

Based on an interview with Graham Nelson, Head of the Open Source Unit of the UK FCO, we propose a three-tier approach to capacity building in big data for diplomats. Individual capacities should be built at a foundation, practitioner, and expert level.[21] At the foundation level, awareness of what big data can accomplish as well as related questions of data security and privacy should be raised. This level of training should enable diplomats to assess challenges and opportunities of big data for a given situation. The practitioner level should focus on developing skills that allow the use of big data tools and techniques to find and verify information. At the expert level, the training provided should enable the design and implementation of appropriate big data tools for diplomatic insight. This third tier of individual capacity building focuses on those with prior knowledge of big data analysis or related fields and aligns these skills with the specific tasks and questions of diplomatic practice.

The selection of the appropriate level of capacity building for each individual should be assessed based on existing individual skills, as well as on the desired level of in-depth knowledge of data analytics related to the individual's position in the MFA. To become a fully-trained data scientist can take up to five years,[22] and this clearly cannot

be the aim of in-house training in big data for diplomats. Rather, it will be crucial to build on existing individual capacities, and in some cases, to recruit relevant and trained personnel.

Data scientists and T-shaped skills

Data science has become somewhat of a hype. When it comes to capacity building at the individual level, there are two particular challenges that this hype generates:[18]

- **Rock stars and gods**: Excessive hype results in expected miracles and unreasonably high expectations.
- **Apples and oranges**: The lack of awareness about the large variety of data scientists leads organisations to waste time on recruitment processes, unable to clearly communicate the desired and required skills of their future data scientists.

Further, looking at the skills of data scientists, we find that a breadth of skills combined with highly specialised expertise in one area is often required for a specific position. The best data scientists have developed what is often referred to as T-shaped skills.[19] On the one hand, they have a breadth of skills represented by the horizontal line of the T. On the other hand, they have a deeply specialised expertise, which is represented by the vertical line of the T. In many cases, this means having a broad subject knowledge combined with a depth of expertise related to one area of data analysis. Further, data scientists need to have a disposition that invites interdisciplinary work and facilitates work within diverse teams. Their T-shaped skills should naturally facilitate such collaboration.

Analysing the profiles and skills of different kinds of data scientists, Harris et al. argue that – contrary to the practice in many organisations – data scientists can best work in teams. Yet, evidence suggests that they are often weakly integrated with the rest of the organisation. The tendency to leave the data scientist alone to to do their work, without proper supervision or integration, is often part of the employer's idea of hiring a 'god' who can do it al[l].20

Existing training programmes in big data analysis are often designed with the private sector and industry in mind. This can be a problem when looking to outsource training. Because of this shortcoming in current training and capacity building, it is recommended that MFAs develop **bespoke training and capacity-building programmes**. For example, Global Affairs Canada, the Canadian MFA, used a hybrid solution consisting of in-house training, partnering with a university, and leveraging existing expertise in the office of statistics to build a bespoke training programme which was initially attended by 12 participants, lasting for 20 weeks.[23]

When making decisions about training and capacity-building courses and programmes, the key considerations include:

- **Time:** How much time is available for staff to be trained (keeping in mind that different members of staff will require different kinds of training as outlined)? How much time can in-house staff spend on sharing experiences and training colleagues?
- **Analysis of existing capacities:** It is important to carefully take stock of existing in-house capacities. This includes staff who are already familiar with big data analysis, staff with a prior background in a STEM subject (Science, Technology, Engineering, and Maths), and staff regularly working with statistical tools.
- **Analysis of the desired level of knowledge about big data:** As suggested, training in big data for diplomacy should take place at different levels with the aim of generating different levels of expertise in big data. It will be important to carefully define which positions in the MFA will require which level of expertise with regard to big data.
- **Crafted partnerships:** Partnerships with both the private sector and academia can support capacity-building efforts by drawing on expertise outside the MFA.
- **Offline, online, blended learning:** The choice of training format should depend on training objectives. Whether the aim is to transmit basic facts about the forms of big data and data security, or to discuss and develop the potential of big data for diplomacy makes a big difference in terms of what training format is best suited. Online learning can be a very convenient form of on-the-job training, but it needs to be carefully designed if it is to go beyond simply transmitting information. Collaborative learning and critical thinking are more easily achieved through blended learning and *in situ* training. Blended learning, combining online and *in situ* training, has the advantage that some parts of the training can be fit around busy schedules without the need to take extended time off work.

Regardless of the specific form, individual capacity-building needs to take into account that diplomats typically come from education paths in law, economics, and the social sciences. Only a minority will have a background in a STEM subject. This makes capacity building all the more important, yet it also presents a challenge – especially when capacity building is offered to groups with mixed backgrounds.

In this sense, the question of capacity building is a question of enabling and supporting communication between two different worlds: the social world of the diplomat and the science-based world of the data scientist.[24] Historically, the work of diplomats is narrative-driven. This might lead to cultural resistance to greater integration of data analysis into the day-to-day work in some MFAs. While not an insurmountable barrier, this resistance needs to be addressed through careful awareness raising and capacity building.[25] As we have seen in Chapter 1, data scientists themselves are often asked to have multiple backgrounds and skills to bridge this very gap. Yet, in the context of the MFA, it is not advisable to solely rely on this; capacity building on both sides is needed. However, diplomats do not need to become data scientists, nor do data scientists need to become diplomats. Rather, the aim of all capacity-building efforts should be to bring these two professional communities together and allow for smooth collaboration between the two. In this way, it will be easier to highlight where big data can make a contribution to diplomatic practice and to support the work of diplomats through making the best possible use of the available tools.

3.5 Chapter summary

It is clear that the existing organisational culture of MFAs needs to be carefully considered and respected. Big data analysis should be seen as a way of supporting organisational goals and the existing culture of the MFA. It should never be an end in itself. Keeping this in mind, this chapter has made very concrete suggestions for adapting MFAs to the data-driven era.

We suggest agile big data units and the appointment of big data champions in other relevant units and departments of the MFA. These two suggestions are driven by, on the one hand, the need to explore the potential of big data in diplomacy in the first place and to allow for space to innovate in this area, and, on the other hand, to facilitate an exchange about the potential of big data and the needs of various departments.

Further, entering into partnerships for big data diplomacy will be important, if not inevitable. This can take the form of ad hoc commissions of relevant studies or outsourcing of big data analytics. Commissioning ad hoc studies and outsourcing big data analytics can address the lack of in-house capacities easily while saving resources, being cost-effective, and avoiding longer-term commitments. However, some in-house capacities for big data analysis should be maintained or developed, not least because of the sensitive nature of some of the data that MFAs work with. In some cases, partnerships with the private sector will be the only option to gain access to otherwise restricted data. Further, building sustainable partnerships for data insight should be carefully considered. The private sector, academia, and international organisations should be considered for such partnerships.

In terms of training and individual capacity building, it is important to keep in mind that MFAs are likely to have to design bespoke training in big data diplomacy and should make use of partnerships, especially with the private sector and with academia, to accomplish this. With regard to capacity building in data diplomacy at the individual level, we suggest a three-tier structure, consisting of foundation, practitioner, and expert level. This takes into account the different degrees to which diplomats need to be familiar with and able to use big data insights and tools. The

question of capacity building in big data diplomacy is also a question of enabling and supporting communication between two different worlds: the world of the data scientist and the world of the diplomat. Facilitating an exchange across this divide will be an important task for diplomacy in the data-driven era.

Ultimately, the aim of organisational changes and capacity building as addressed in this chapter is not to transform diplomats into data scientists. Rather, the aim of all such efforts should be to highlight where big data can make a contribution to diplomatic practice and to support the work of diplomats through making the best possible use of the available tools.

Notes

1. Quoted in DiploFoundation (2017) *Data diplomacy: Big data for foreign policy. Summary of a half-day event at the Ministry of Foreign Affairs of Finland.* Available at https://www.diplomacy.edu/resources/general/data-diplomacy-big-data [accessed 12 December 2017].
2. De Croo A (2015) *Diplomacy in the digital age. Introduction* [speech]. 20 November, Brussels. Available at https://www.alexanderdecroo.be/diplomacy-in-the-digital-age-introduction-brussels/ [accessed 12 December 2017].
3. Liebowitz J [ed] (1999) *Knowledge Management Handbook*. Boca Raton, FL: CRC Press.
4. Hocking B & Melissen M (2015) *Diplomacy in the Digital Age. Clingendael Report*, p. 55. Available at https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf [accessed 12 December 2017].
5. Webb A (2009) Public diplomacy: Meeting new challenges. *Report of Wilton Park Conference 902*, 6-8 October 2009. Wilton Park, UK, p. 6
6. Kurbalija J (2002) Knowledge management and diplomacy. In Kurbalija J [ed] *Knowledge and Diplomacy*. Msida, Malta: DiploFoundation. Available at https://www.diplomacy.edu/resources/general/knowledge-management-and-diplomacy [accessed 12 December 2017].
7. Kurbalija J & Slavik H [eds] (2001) *Language and Diplomacy*. Msida, Malta: DiploFoundation. Available at https://www.diplomacy.edu/resources/books/language-and-diplomacy [accessed 12 December 2017].
8. Pomel S (interview 12 July 2017).
9. Fosland M and Martinsen O-M (interview 9 October 2017).
10. Nelson G (interview 6 April 2017).
11. Fosland M and Martinsen O-M (interview 9 October 2017).
12. Alerksoussi R (interview 7 April 2017).
13. Atherton K (2007) When big data went to war - and lost. *Politico*, 11 October. Available at https://www.politico.com/agenda/story/2017/10/11/counter-ied-warfare-data-project-000541 [accessed 12 December 2017].
14. Fosland M and Martinsen O-M (interview 9 October 2017).
15. Nelson G (interview 6 April 2017).
16. The following quotations are taken from UNECE (no date) Big data team level competency. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.54/2016/Big_Data_Team_Nov_2015.pdf [accessed 12 December 2017].
17. Based on UNECE (no date) Big data team level competency. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.54/2016/Big_Data_Team_Nov_2015.pdf [accessed 12 December 2017].
18. Harris HD et al. (*2013) Analyzing the Analyzer. Sebastopol, CA: O'Reilly, pp. 5-6. Available at* http://cdn.oreillystatic.com/oreilly/radarreport/0636920029014/Analyzing_the_Analyzers.pdf [accessed 12 December 2017].
19. Ibid.
20. Ibid. p. 25.
21. Nelson G (interview 6 April 2017).
22. Cisco (2016) Preparing for the Data Science-Driven Era. Available at https://www.cisco.com/c/en/us/solutions/collateral/enterprise/cisco-on-cisco/i-dc-09022015-preparing-for-data.html [accessed 12 December 2017].
23. Pomel S (interview 12 July 2017).
24. Höne KE & Kurbalija J (2018). Accelerating basic science in an intergovernmental framework: learning from CERN's science diplomacy. Global Policy [forthcoming].
25. Nelson G (interview 6 April 2017).

4. Key aspects concerning the use of big data

Despite claims that it is natural and raw, data is the result of the decisions, priorities, interests and values of numerous actors. In the race to exploit the hidden potential of Big Data in business, government and academia to tell us truths about societies, we risk making errors of interpretation and understanding if we don't attend to these questions of how data is socially produced.
Professor Evelyn S. Ruppert, University of London[1]

Protection of privacy is the key element in making data for social good projects successful. It ensures that stakeholders are accountable in their data practices and that beneficiaries – most of the time consumers – have trust and faith in the value of their data being used for social good with minimum risks.
United Nations Global Pulse & GSMA[2]

The previous chapters have outlined some of the opportunities that big data can provide for the core functions of diplomacy, as well as for public diplomacy, consular affairs, trade, development, humanitarian affairs, and international law. In addition, they have clarified some of the institutional challenges and opportunities related to integrating big data analysis into the MFA.

In addition, it is important to have a clear picture of the specific challenges and limitations that arise when working with big data. In this chapter, we look at five such broad challenges:

- Access to data
- Data quality
- Data interpretation
- Data protection
- Data security

In order to understand what is and what is not possible when adopting a data-informed strategy for diplomacy, these broad challenges deserve careful consideration.

4.1 Access to data

The starting point for any big data analysis is to obtain access to data. While this is a seemingly obvious and straightforward idea, this might in fact be one of the largest obstacles to overcome. The location of the data is often highly dependent on the purpose of the analysis. We have identified four categories where data might be found, which together form a data access framework.

Data access framework

	Internal	External
Open	Data accessible by anyone within the organisation	Publicly accessible data
Closed	Data that has been classified within the organisation	Data inaccessible to the general public (including diplomats)

Diplomats engaging in big data analyses can either collect data from scratch or find it within the organisation (*internal*), or rely on data that has been gathered outside of the organisation (*external*). In addition, data can be openly accessible in the organisation or for the public (*open*), or there could be obstacles to access (*closed*). These boundaries take the form of restrictions if the data is confidential, sensitive, or personal; because the data is commodified or protected by intellectual property rights; or because there is a particular interest for the organisation or unit to keep the data classified and undisclosed. The boundaries between the cells in our data access framework are not necessarily rigid. There is data lingering in between the internal and external structures of the MFA, for example data collected by subcontractors, although contractual agreements often indicate the ownership of the data.

4.1.1 External open data

Open data is data 'that can be freely used, re-used and redistributed by anyone'.[3] An ever-growing amount of data is made publicly available, especially data that is accessible through the Web, such as websites and the texts and documents uploaded to them. There are also more and more organisations that have decided to make as much of their data as possible publicly available, including ministries and governments. The Organisation for Economic Co-operation and Development (OECD) created the *OURdata Index* on Open, Useful, Reusable Government Data, with South Korea, France, Japan, the UK, and Mexico forming the top five in their efforts in providing available and accessible public sector data.[4]
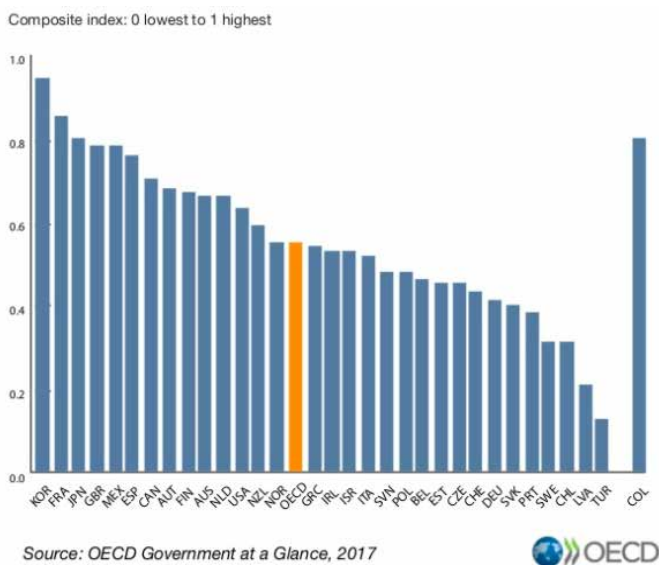

Figure 6: OURdata index – Open-Useful-Reusable Government Data Index[5]

In addition, some large Internet companies are providing free, real-time data and analyses of the activity on their platforms. For example, Twitter provides a subset of the Tweets published available for free for others to analyse. Google Trends provides analyses on the prevalence of search terms, broken down by subtopic and region. Outside the private sector, there are initiatives that might be even more relevant to MFAs, such as the Humanitarian Data Exchange (HDX), operated by the UN Office for the Coordination of Humanitarian Affairs (OCHA), which contains a large amount of humanitarian data. The HDX has particular potential during times of crises, and was widely consulted during the 2014 Ebola outbreak and the 2015 Nepal earthquake. In these instances, data was quickly made available in usable formats and used by many humanitarian actors and policy-makers.[6]

Until the advent of Google Maps and online navigation tools, geospatial data was an expensive resource that could only be obtained upon request. Yet, tools such as Google Maps have changed this completely, making large datasets of geospatial data available to the public. Organisations such as the National Aeronautics and Space Administration (NASA) have also joined the open data revolution and are making maps available, as well as analyses by the United Nations Operational Satellite Applications Programme (UNOSAT). In addition, a large number of crowdsourced maps have started to appear, with OpenStreetmap arguably the most successful to date.

The analysis of open data could be a good starting point for MFAs that are considering moving into big data analysis. In fact, this has been the approach of the UK Foreign Commonwealth Office (FCO) Open Source Unit, recognising the enormous wealth of information that is available online that has not yet been systematically tapped into for foreign policy analysis.

While online information, geospatial data, and crowd-sourced data have all been riding the waves of the open data trend, data exhaust – data that is automatically collected by sensors – is still relatively inaccessible. Part of the reason is technical: The automatically collected data is compiled in enormous databases that are too large to simply be retrieved from a website. In addition, this type of data is usually collected by the private sector, which could have a particular interest in keeping the data secure, or could decide to provide this data against a fee. In addition, there might be important privacy considerations, for example when it relates to call detail records (CDRs) or GPS locations monitored by smartphones.

4.1.2 Internal open data

MFAs gather an extensive amount of data and produce a large number of resources daily. Just as there is a wealth of data outside the MFA that has not been extensively touched so far, there might also be a great availability of data inside the MFA that has not been systematically looked at, although this will usually not be considered *big* data.

The UK FCO identified this shortcoming in its *Future FCO Report*, in which the office wrote:

> 90% of data was created in the last two years. Business was woken up to the transformational potential of Big Data. The FCO has not. The FCO is not yet in a position to 'mine' even its own internal data for insight, which means we miss important patterns and trends.[7]

Other ministries identified similar gaps that they are now attempting to fill. For example, Global Affairs Canada, the Canadian MFA, is looking into the possibility of natural language processing (NLP) to identify trends in the evaluation data at their Development Cooperation department. Through NLP, the ministry might be able to improve searches for relevant information, as well as to synthesise information into better insights.[8] Similar techniques could also be developed for public consultations opened by the MFA. When Global Affairs Canada launched a consultation on new international assistance policy in 2016, it received 10 000 submissions, ranging from e-mails to 20-page papers. Examining all consultations manually takes considerable time and resources, while such information can now be processed and synthesised using NLP. In a pilot project, Global Affairs Canada partnered with IBM Watson to conduct word and sentiment analysis to identify trends and correlations, disaggregated by theme or geography.[9]

4.1.3 External closed data

There are a number of reasons why data holders (entities that are in the possession of big data) might choose not to disclose the data that they have produced or control. One of the key considerations is related to confidential, sensitive, and/or personal data. As a result of privacy considerations, this data often cannot even be shared without anonymisation due to legal frameworks, such as national or regional data protection laws. Another reason for leaving data undisclosed is when it is not in the interest of the dataholder to share the data. It is often important to understand the reason for leaving certain information classified, as the MFA's potential use of this data might not be harming a data holder's competitive advantage. For that reason, partnerships between the private sector and the public sector as well as non-profit organisations might still be feasible despite the data being closed. Finally, data holders might simply choose not to make their data publicly available for free as part of their business model, or to cover the cost of generating the data.

Categories of closed data

Confidential, sensitive, and personal data	Data that cannot be shared due to privacy considerations, or only when anonymised
Classified data	Data that is not shared as it is not in the interest of the dataholder to do so
Commoditised data	Data that is only shared against a fee
Data available upon request	Data that is shared with interested parties, but not made public by default

Most of the world's data is located in databases that are restricted to the public, whether it concerns mobile phone, social media, financial, or sensor data. A survey of national statistical offices and international organisations, conducted by the United Nations Statistics Division (UNSD) and the United Nations Economic Commission for Europe (UNECE), found that 'while most respondents recognize the challenges related to IT, skills, legislation and methodology, most argue that the biggest challenge for Big Data projects is the limited access to potential datasets.'[10] In fact, 'access to the data source can become the principal risk factor to the success or failure of a Big Data project.'[11] In addition, the UNSD and UNECE survey clarified that it might be particularly difficult to obtain access to data sources with an international scope, which might be an added obstacle for research conducted at or for an MFA.

It should be noted that there are a number of data holders that are willing to enter into agreements to provide their anonymised data, or even engage in 'data philanthropy': providing free access to otherwise restricted data to the public sector, research institutes, and non-profit organisations.[12]

Yet, the number of data philanthropists is still relatively small, and accessing external closed big data usually requires entering into partnerships with the relevant data holders. As we have seen in Chapter 3, the creation of these partnerships can be very complex. According to Timo Koskimäki, there is a lack of clear guidelines of what such a partnership should look like, which could result in a reluctance to engage in such agreements with the private sector.[13]

Relying on data obtained or analysed by partners carries certain risks. For example, public institutions can exercise little control over the way in which data is collected and processed, and whether this happens ethically and responsibly. Potential mishandling of data at the hands of the partner organisation could harm the reputation of the ministry. In addition, if data is obtained for longer-term objectives, there is a continuous risk that the company will cease to collect the data in the future. If the contract is breached, data collection is stopped, or access to the data is denied, how can data continuity and a time-series be ensured?[14]

So, what are the key ingredients for a productive and responsible partnership for data sharing? Microsoft identified three principles that could form the basis of a 'trust framework' for data sharing:[15]

1. **Transparency**: Each party needs to be open about their motivations, policies, and regulatory constraints in relation to data collection, storage, sharing, use, and publication.

2. **Accountability**: The rights and interests of 'data subjects' need to be protected with accountability controls regarding data provenance, chain of custody, and algorithmic/analytical transparency.

3. **Fair value exchange**: The partnership needs to demonstrate a fair value exchange between the data providers (or data subjects) and those who use their data.

In the end, there should be a consideration for whether the time and resources spent on accessing data will pay off. Especially in time-constrained situations, the search for more data may not be the right priority, and data access efforts could divert attention away from focusing on a sound analysis of the plethora of data that is already publicly available.[16]

4.1.4 Internal closed data

Finally, there is data that is shielded from certain units within the ministry, whether it is for privacy reasons or to protect the interests of the MFA. Again, it is important to understand the reasons behind the lack of disclosure, as there might be solutions to obtain the data for analysis, such as through aggregation or anonymisation. Nevertheless, as is the case with public undisclosed data, it is important to consider whether this data is really needed, or whether the analysis can be conducted by using different, less sensitive, data sources.

The reluctance to share data across the ministry can also be a sign of siloed departments. Some units within the ministry might not be as smoothly integrated in the rest of the ministry, and could be reluctant to share their data. These concerns might be related to privacy and data protection concerns. One solution is to remove all personal identifiers before data is shared with other departments in the ministry.

The Norwegian MFA has recently opened a new unit tasked with better using the information that is already available in the MFA for data analysis and text-mining, such as reports, memos, and speeches, through a combination of improving knowledge management, ensuring data quality, and conducting data analytics.[17] Yet, their approach faces challenges when it comes to security restrictions and classified documents.

4.2 Data quality

While access to data is the first significant hurdle to overcome, a second appears as soon as this access is granted. Big data is often messy and incomplete, and cannot be used straightaway, or without considering limitations and potential biases. Several factors explain the general low quality of big data. For example, the data collected might not be representative, or it relies on self-reported information that is absent or false (e.g. on social media); there can be problems related the integration of multiple data sources into one dataset; the data can be analysed according to incorrect models based on false assumptions; and it can be costly to keep data up-to-date.

The question of data quality should not be considered lightly, especially if MFAs are to base their policies in part on data-driven analyses. Data of low quality can result in 'biased algorithms, spurious correlations, errors, an underestimation of the legal, social and ethical implications, [and] the risk of data being used for discriminatory or fraudulent purposes',[18] and it is therefore key to consider the quality of the data, before even starting to analyse it. This is not only important to ensure the quality of the output of the analysis, but also to assess the technological feasibility of big data analysis. Messy, unstructured data that needs significant effort to store, clean, and harmonise, might not be worth the investment. Yet, according to the UNSD/UNECE survey, more than two-thirds of national statistical institutes and international organisations had not yet defined a quality assessment framework for big data in 2015.[19] There are several indicators that could be used to assess the quality of big data.

4.2.1 Complexity

One of the most obvious challenges in working with big data is its complexity. Complexity arises in various dimensions, including the structure of the data (which makes it difficult to integrate data tables into a unified dataset) and the format of the data (e.g. discrepancies in standards used to store the data). In MFAs, data often derives from multiple sources, countries, and contexts. For example, in development co-operation, ministries sometimes use different metrics for monitoring different projects, which are then unable to be effectively aggregated.[20] Such

challenges could be mitigated by setting clear guidelines, or unifying data on a single platform.[21] One measure to clear up some of big data's complexity is to introduce standards, at least within the MFA, related to the collection and analysis of data, in order to make the data easier to link. The MFA could also look into standards developed in the wider sector. For example, in the development sector, the International Aid Transparency Initiative has developed standards for aid-related data. This can also improve the consistency of the dataset over time.

4.2.2 Completeness

Not only is big data complex, it is often incomplete as well. This might be surprising, as the 'big' in big data suggests that completeness is not of major concern. Yet, there could be gaps in data collection or data that relies on self-reporting. For example, social media platforms and other Internet accounts often provide the possibility to fill out personal data, such as gender, place of birth, or even relationship data, some of which a respondent could choose not to fill out, or fill out incorrectly.

4.2.3 Timeliness

While some datasets might be available in (near) real-time, this is not the case for all, and any data analysis needs to take the timeliness of the data into account, as well as the time it takes to analyse the data. This is particularly challenging in quickly evolving situations. For example, while satellite data can be made available quickly, it takes some time to create maps, and once a map has been created, it might already be outdated.

4.2.4 Accuracy

While big data is able to measure phenomena with great specificity, it might not actually capture the information that is needed. Data that is not representative could lead to biased results (Section 4.3). If the data does not accurately measure the objectives of the research, its analysis will inevitably reflect these errors and lead to incorrect conclusions.

4.2.5 Relevance

Usually, data has initially not been created or obtained for the specific needs of the MFA, especially when this data is generated externally. Although big data has the reputation of being raw and unfiltered, it is largely the result of 'decisions, priorities, interests and values of numerous actors',[22] such as decisions on which variables to include in a dataset, and how these variables are measured. It is important to understand the relevance of certain datasets and to understand how it can support the decisions and on what scale.

4.2.6 Usability

Given the difficulties arising when working with big data and the complexity of the dataset, part of data quality is also its usability. Is it even possible for the MFA to analyse the data without spending a disproportionate amount of resources on organising and interpreting the data? A good framework to evaluate big data quality has been established by the United Nations Economic Commission for Europe, which provides a framework with 13 indicators, including factors to consider for each indicator listed.[23]

4.3 Data interpretation

Data becomes meaningful only through interpretation. Algorithms are able to pick up patterns and analyse trends (Section 1.2.4). Based on this, in most cases human interpretation is needed in order to transform algorithmic results into decisions. In this section, we point to three considerations to keep in mind when interpreting data and the results of data analysis: the distinction between correlation and causation, issues surrounding selection bias, and the question of data politicisation.

4.3.1 Correlation vs. causation

Unfortunately, big data is not a perfectly unbiased source of insight. For example, in the search for meaningful results, there is often a lack of consideration for the difference between correlation and causation: the fluctuation of trends along the same values does not mean that they are related in any meaningful way. Big data is able to identify many (meaningless) correlations due to the large quantity of data under consideration. The potential of big data for foreign policy is therefore related to its ability 'to detect certain patterns in human behaviour and the characteristics of groups of people', rather than causality and prediction.[24]

Misinterpreted big data analyses could result in misguided policies. For example, an analysis found a spurious correlation between areas of building damage and areas of

SMS streams after an earthquake in Haiti, suggesting that SMS streams are indicators of areas where buildings are damaged. In fact, the text messages and damaged buildings were much more related to building density, and once controlling for this factor, it turned out that there was a slight *negative* correlation between text messages and building damage.[25]

4.3.2 Selection bias

Seeing the enormous amount of information in big data sets, it is tempting to believe that selection bias is brought to a minimum. The analysis seems liberated from the inherent risks related to sample size and selection. Unfortunately, big datasets can be very unrepresentative. While traditional statistics carefully determine their sampling frameworks and parameters to ensure a representative subset of the population, big data often only relates to whoever uses a service, application, or device. For example, studies using mobile phone data could over-represent young and wealthy users of mobile technology, especially in areas where mobile penetration is low. Social media analysis could over-represent segments of the population which actively use a particular social media platform. A large number of reported injuries on Twitter could point at areas where most people are affected by an attack or disaster, or it could identify the area where most people access and use Twitter.[26]

Policies based on unrepresentative datasets risk making misguided prioritisations or failing to target those who need the intervention the most. For example, disaster response based on call data records and phone signals could send a disproportionate amount of aid to the relatively affluent – and sometimes less affected and less vulnerable – areas.[27] As a result, there is a risk for 'differential treatment of and indirect discrimination against groups of people with similar characteristics',[28] and this could reflect existing cleavages in society, such as divisions around ethnicities and the digital gender gap.

4.3.3 Politicised data

When data studies are conducted in politically sensitive contexts, there might be an active willingness to interfere with the data. For example, UN Global Pulse warned that when public violence is measured through SMS streams, the perpetrators could be deliberately attempting to suppress reporting through text messages. The data will over-represent the areas where perpetrators are unable to suppress mobile phone use. According to UN Global Pulse 'we'll have many more "false negative" zones where

there seems to be no violence, but there's simply no SMS traffic,'[29] As the United Nations Children's Fund wrote in a recent report, 'the same traits that make data powerful – revealing inequities, highlighting systemic weaknesses, or unmasking public discontent – also make data political.'[30]

Political choices and sensitivities not only interfere with the process of data collection, but also with their analysis and interpretation. By categorising and structuring data, the data scientist chooses 'one data state to represent the world from among many incommensurable possibilities',[31] imposing a sociocultural frame on the data that often represents the dominant social order. In this sense, data science is considered a constructive practice by some, rather than a practice based on scientific objectivity.

Challenges related to the quality and interpretation of data also exist with regard to satellite data analysis. Not only do such images need to be verified, their interpretation requires more technical skills than people often assume. This subjectivity is also present in the visualisation of data. Due to their very nature in making complex data more understandable and often simplified, data visualisation often inherently conceals more complex realities (Section 1.2.3.3).[32] There is a need to critically assess such visualisations, especially when they are used to serve ideological agendas: Why is this the data that has been chosen to be visualised, and why in these categories?

Many MFAs might not have the necessary resources to properly conduct complex big data analyses, without risking the misinterpretation of data. Or they lack the ability to recognise potentially low data quality. One of the ways in which these challenges can be mitigated is to partner with an organisation that is able to assist in processing and analysing data. For example, Global Affairs Canada collaborates with IBM, which provides its tools for free trials. The combination of capacity building and partnership could provide a fruitful solution until the necessary skills have been developed in the MFA (for further considerations about partnerships, Section 3.3) .

Bearing in mind the quality concerns of big data, as well as potential pitfalls in their interpretation, big data analyses should always be complemented with more traditional forms of data and insights. A common misperception is that data is a panacea, replacing the need for expert insight. Yet, data does not bring full knowledge; it provides a hypothesis or raises questions. It can be a tool to improve traditional foreign policy analysis, not replace it.

4.4 Data protection

As the amount of data handled by the MFA increases, so do concerns and questions about data protection. Data protection matters to safeguard the privacy of individuals, communities, and specific groups and to protect from surveillance and discriminatory targeting. With growing concerns about these issues, MFAs should work to uphold the highest principles of data protection in their policies and practices. In the following, we discuss general data protection issues and concerns related to community identifiable information and discrimination. We also give an overview of the existing legal frameworks. Finally, we discuss cases in which the right to privacy is detrimentally opposed to the public interest.

4.4.1 General data protection issues

One of the biggest risks related to big data is that it is used for extensive surveillance and discriminatory monitoring of citizens. In fact, the MFA should be careful not to step into the shoes of intelligence agencies if it chooses to collect a large amount of data, and should continue to conduct its analyses in compliance with international law, especially human rights law, and in accordance with national or regional legal privacy frameworks. In *Diplomacy in the Digital Age*, Hocking and Melissen write about big data:

> This young field is fraught with risks of inappropriate use, for instance when large swathes of information are used in a deterministic fashion for 'profiling' of individuals and groups. For some, a danger is the gradual triumph of data over politics as governments come to accept the immutability of huge swathes of information over political debate and policy choice, and the application of common sense to human affairs.[33]

In recent years, the effectiveness of anonymisation has been questioned, as it has become possible to reveal personal identities by combining data from different sources. Data breaches infringe not only on individuals' privacy rights, but also risk damaging the reputation and trust of citizens in the MFA. It is therefore of critical importance that big data analyses are conducted responsibly, with the necessary privacy provisions in place.

Privacy risks have always existed when dealing with data and statistics. Yet, the advent of big data has significantly increased privacy concerns, with the growing accuracy of identifying individuals and the many details that can be collected about them. Sometimes, this happens unintentionally. With massive amounts of automatically generated data, it often takes more time and resources to filter and remove unnecessary data than to store it indefinitely.[34] Yet, safeguarding privacy is also keeping sensitive information to a minimum, and needs to be accompanied with transparency and restraint related to data collection. Privacy risks could also be minimised by retaining data for a defined period, after which the personal data is removed. This should also be encouraged or enshrined in agreements with partner organisations that are handling the data.

Another solution to minimise privacy concerns is to base analyses mainly on publicly available data sources. This approach is taken by the UK FCO's Open Source Unit. As there is already a massive amount of data that can be obtained relatively risk-free, this can be a good starting point while keeping privacy concerns low.

Partnerships bring both privacy opportunities and risks. If data is collected and processed by the partner organisation, this can help avoid issues of privacy and confidentiality at the MFA, which will only receive aggregated data. For example, the government of Oman only accepts data from Mobile Network Operators that has been made 'non-personal', which means that any direct identifiers have been removed.[35] At the same time, by allowing partners to work with big data, a degree of control over the process is relinquished.

4.4.2 Community identifiable information and discrimination

Yet, even seemingly non-confidential open data might be sensitive when adopting a 'community-based' approach to privacy, which is taken by organisations such as the International Committee of the Red Cross. They claim that even aggregated or anonymised open data, or open data that does not directly include personal data at all (such as satellite data) can infringe on the privacy of communities and specific groups. Big data tends to make generalisations on segments of the population, or identify their general location. For vulnerable groups, this demographically or community identifiable information could increase their risk to be discriminated against or targeted.

These discriminatory practices can knowingly occur, but they could also accidently slip into data analysis. When

policies are designed on the basis of big data and machine learning, for example to identify risk populations, their outcomes can range from 'tangible and material harms, such as financial loss, to less tangible harms, such as intrusion into private life and reputational damage.'[36] In addition, due to potentially unrepresentative data – arising from the exclusion of people who are not connected to digital technologies – the needs of the most vulnerable people risk being discarded. Discrimination can occur as a result of the design and use of big data technologies, as well as a purposeful way to exclude vulnerable segments of the population.

4.4.3 Legal framework

All states are obliged to comply with international law, including human rights, which includes the right to privacy, and big data cannot be used outside a proper legal framework and without the respect for the right to privacy. According to Article 12 of the Universal Declaration of Human Rights, 'No one shall be subject to arbitrary interference with his privacy, family, home or correspondence.'[37]

Yet, despite this global framework, the specific legal safeguards for privacy might differ among MFAs. The operationalisation of legal data protection requirements could be complex, not least due to the large number of actors that are often involved in the collection, use, storage, and sharing of data: Which actor is responsible and accountable for what?

The development of data protection regulation has rapidly risen during the last couple of years. The EU is arguably the region with the strongest legal privacy provisions. On top of the Council of Europe's Convention 108 on the Protection of Individuals with regard to Automatic Processing of Personal Data, the EU agreed on a new, legally binding privacy framework. In May 2018, the General Data Protection Regulation (GDPR) will become enforceable, which is likely to significantly affect the way in which personal data can be collected and analysed, not only by the private sector, but also by the government.

In most countries, a privacy framework has been created for the purpose of traditional statistics rather than big data. However, many countries lack specific provisions related to data protection in the age of big data. Yet, according to the 2015 UNSD/UNECE survey, most organisations go beyond the legal requirements to ensure privacy, not in the least due to the potential risk to their public image if these protections are not in place. This attitude also seems to be taken by Global Affairs Canada, where for every social media analysis, a mandatory privacy impact assessment is required. Yet, it can sometimes be challenging to find the right balance between preserving the integrity of the data, while at the same time allowing for a modernised approach to data.[38]

4.4.4 Data protection vs. public interest

It can be challenging to ensure the right to privacy in scenarios in which personal data is needed to address crisis situations. When crises are extreme, such as in the case of mass displacement, epidemics or natural disasters, privacy concerns usually take a back seat. During these scenarios, it is considered important to rapidly collect relevant data and share it with all stakeholders involved, to design an appropriate response

Yet, when this threat diminishes over time, concerns about privacy often increase. For example, during the Ebola crisis, a large amount of personal data was collected by aid workers to contain the spread of the disease. When the crisis dissipated with personal data remaining disclosed, 'some survivors were stigmatized and left vulnerable.'[39] Finding the right balance between privacy concerns and human security is not always easy. 'The positives and negatives are not always clear and often exist in tension with one another, particularly when involving vulnerable populations.'[40] This balance is particularly difficult to strike in such contexts of conflict: while big data could make a true difference in identifying people in need, risks related to data breaches and personal security are higher.[41]

There is a prominent example that further illustrates the complex balance between privacy and public interest. In 2015, many refugees traversed Europe in the search for safety from conflicts in Syria, Iraq, and Eritrea. Many of them relied on their smartphones for information about the route, to connect with family, and to understand the legal procedures on their way. The data collected on refugees, derived from call records, text messages, money transfers, social media activities, and WiFi connections, had the potential to identify smugglers and better understand the needs of the refugees. At the same time, the collection of refugees' data resulted in 'perceived and real fears around data collection', which could 'drive them off the grid' and make them invisible to officials.[42] Not only does this make it more difficult to provide them with aid and assistance, it makes them more vulnerable to engaging with criminal enterprises. As a result, it is very important for both governments and refugee agencies to 'establish trust when collecting data from refugees.'[43]

Ensuring citizens' trust in the data that is used is not only important for the proper collection of their data, but also to ensure the collection and use of the data over longer periods of time. To better address this dichotomy and to establish trust among the population from which the data is collected, it is important to be transparent about the purpose of the data collection, how the data is collected and processed, and whether the data is distributed to third parties. In general, the data collected and analysed should be minimised to cover only the level needed for the intended purpose.[44]

4.5 Data security

According to a recent study by the Internet Society, the number, size, and cost of data breaches continue to increase. They have hit private sector companies and government agencies alike. In June 2015, 21.5 million records were stolen from the US Office of Personnel Management (OPM). These included social security numbers, addresses, and even detailed financial and personal information that was collected for security clearance, such as the fingerprints of 5.6 million employees. At the time, the OPM used old systems that were vulnerable to cyber-attacks, and had not encrypted the data that it stored. In addition, the OPM had not removed the data of former employees, which increased the impact of the breach as well.[45]

There have been many breaches in MFAs that are less well-known. The Thai MFA was attacked in 2016 by hacktivist group *Anonymous*, leaking the personal details of more than 3 000 employees.[46] The same group leaked one terabyte of documents from Kenya's Ministry of Foreign Affairs and International Trade in April 2016.[47] The Czech MFA was reportedly hacked in January 2017 and one investigative outlet claimed that 'thousands of files were downloaded from email inboxes of the Czech Foreign Minister and his Undersecretaries.'[48] The Czech MFA confirmed the breach, but denied the breach of classified information. With the combination of the increased amount of data that is available at the MFA and the growing sophistication with which it can be breached by malicious actors, data security has become of vital importance for an MFA.

4.5.1 Technological solutions

When considering how to ensure data security, the most obvious response is usually related to technology: Which technological solutions need to be employed to ensure the security of the MFA's data?

4.5.1.1 Securing the data location

The location where data is stored affects privacy and data security. For sensitive data and data that is not in the public domain, it is advisable that data be hosted internally, and not on external servers. When data is already made public, external hosts and partners can be considered. According to the UNSD/UNECE survey, the lack of reliable partners who are able to securely host sensitive data is a particular bottleneck for the public sector, as it 'will certainly slow the adoption of complex Big Data tools … relative to the private sector.'[49]

The reluctance to host data externally is evident in the Danish MFA. In 2013, it judged that the market for cloud solutions was still too immature to be adopted by the ministry. To ensure the security of the MFA's data, it 'generally keeps its own data within its own four walls' and decided not to invest majorly in the public cloud yet.[50]

Besides general considerations of where to locate data, particularly sensitive data could be isolated and segmented from other data collected or possessed by the MFA. When the overriding concern relates to the integrity or continuity of the data, risks could be reduced by producing replications of the data and storing them off-site, domestically or overseas. Several MFAs have already adopted this strategy, such as the US State Department and the Estonian government.[51] Estonia has even announced the creation of a 'data embassy' in Luxembourg, which will 'store the copies of the most critical and confidential data', and which will enjoy the same protection and immunity as traditional embassies.[52]

4.5.1.2 Securing the data format

Besides securing the location of the data, the format of the data could also be protected. End-to-end encryption can be an important tool to make sure that the data is not legible to those who illegitimately obtain it. In essence, by encrypting data, the data is translated into another form or code, and its original meaning can only be accessed by

those with a decryption key or password. Encryption can have a wide range of applications, from databases and e-mails, to Internet connections and the cloud. Encryption is currently the most popular way to secure data. In fact, the UK Information Commissioner is of the opinion that when data breaches occur without the adequate protection of encryption, regulatory action can be taken against the entity where the breach took place.[53] Although it is an effective way to mitigate some of the challenges related to data security, it should go hand in hand with other data security solutions, and in particular a minimum level of awareness by staff, for example related to which data, information, or message to encrypt and how this is done.

4.5.1.3 Securing the data design

Privacy-by-design is an approach to data protection that emphasises the importance of building data security into the design of the data architecture, such as the design of new IT systems, the development of data policies, or when engaging in a partnership that involves sharing big data. Adopting such an approach would help identify data protection problems at an early stage, raise awareness about data protection across an organisation, and lead to a lower risk of gathering confidential data if it is not absolutely necessary for the analysis.[54]

4.5.2 Countering behavioural challenges

To ensure the privacy of data and prevent breaches, there is an important need to invest in data security, not only at a technical level, but also in the behaviour and conduct around data at the organisation itself. Many data breaches are, in fact, preventable, and can be the result of accidental disclosure of data or the loss of a device containing sensitive

data.[55] Security and privacy breaches created by human behaviour 'will vary from benign to accidental to malicious'.[56]

To address behavioural challenges to data security, training and capacity building are needed for anyone who engages with data within the MFA, as well as potential partners outside of it. Such capacity building could also be part of organisational measures, such as setting up an information security management system with guidelines and policies on IT security, e-mail security, IT equipment usage, information classification, document destruction, and a contingency plan.[57]

> **8 tips for minimising privacy and security risks:**
>
> - Keep sensitive and personal data to a minimum.
> - Make the processing of personal data transparent to those whose data is used.
> - Ensure the purpose of the use of personal data is legitimate and proportionate.
> - Encrypt the data that is stored.
> - Restrict access to data to only those directly involved.
> - Destroy data when the purpose for which it was collected and held is no longer applicable.
> - When working with partners:
>   - Make sure the partners are reliable and have sound incentives.
>   - When data is processed by partners, the MFA should ensure to only receive aggregated data.
> - Keep sensitive data on secure internal servers, or, if available, reliable external hosts.

4.6 Chapter summary

While big data can generate important opportunities, it has its limitations and challenges, which need to be kept in mind for anyone embarking on data diplomacy. Access to datasets is the first challenge to be overcome, especially considering that many of the most valuable data sources are held by the private sector. To mitigate this challenge, the MFA could think about relying on open data for its insights, or entering into partnerships with the entities that gather and control the data. Trust between the MFA and the partner is of utmost importance, especially when managing sensitive data, as both parties would be affected in the case of data breaches or other mishandling of the data. Challenges related to access are important,

and should be taken into account when considering whether the cost, effort, and risks related to obtaining data pay off after the data has been analysed.

Similar considerations can be made for data quality, as datasets can be particularly messy, and it takes time to properly prepare, manage, and clean the data so that it can be used for analysis. Data quality indicators to keep in mind relate to big data's complexity, completeness, timeliness, accuracy, relevance, and usability.

To get the right insights from the data, it is important not to fall into methodological traps. The many variables and

data points that could signal convincing correlations should not be mistaken for causal relations. In addition, the data might over- and under-represent certain groups in society, especially when it is based on the analysis of digital tools, such as social media or smartphones. Basing policies on skewed data analysis outcomes could result in ineffective programmes that fail to address the needs of the MFA.

Finally, when engaging big data, the MFA needs to provide proper mechanisms for data protection and security. The design of the data analysis can be adapted to minimise data protection challenges, for example by avoiding the collection and storage of more data than necessary, by relying on open data, or by ensuring that clear data

protection and accountability guidelines are enshrined in partnership agreements. Data should be kept secure by a combination of technical measures, which secure the location, format, and design of the data architecture, and the training of staff members in order to counter behavioural challenges related to data security.

These key aspects are important to integrate into any data diplomacy project. At the same time, they should not discourage efforts to start engaging with big data. The challenges apply differently to every project and can be mitigated by making smart choices about the design of the big data analysis.

Notes

1 Ruppert ES (2015) Who owns big data? *Discover Society*. Available at https://discoversociety.org/2015/07/30/who-owns-big-data/ [accessed 12 December 2017].

2 United Nations Global Pulse & GSMA (2017) *The State of Mobile Data for Social Good Report*, p. 14. Available at http://unglobalpulse.org/sites/default/files/MobileDataforSocialGoodReport_29June.pdf [accessed 12 December 2017].

3 Open Knowledge International (no date). *The Open Data Handbook*. Available at http://opendatahandbook.org/guide/en/ [accessed 12 December 2017].

4 OECD (2017) *OURdata Index: Open, Useful, Reusable Government Data*. Available at http://www.oecd.org/gov/digital-government/open-government-data.htm [accessed 12 December 2017].

5 OECD (2017) *OURdata Index: Open, Useful, Reusable Government Data (2017)*. Available at http://www.oecd.org/gov/digital-government/open-government-data.htm [accessed 12 December 2017].

6 DiploFoundation (2018) Data Talks October 2017: Data protection and open data [forthcoming].

7 United Kingdom Foreign and Commonwealth Office (2016) *Future FCO Report*. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/521916/Future_FCO_Report.pdf [accessed 12 December 2017].

8 Pomel S (interview 12 July 2017).

9 Pomel S (interview 12 July 2017).

10 United Nations Statistical Commission (2015) *Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data*. Statistical Commission Forty-sixth session, p. 3. Available at https://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf [accessed 12 December 2017].

11 Ibid.

12 Robert Kirkpatrick (2016) The importance of big data partnerships for sustainable development *Blog*, 31 May. Available at https://www.unglobalpulse.org/big-data-partnerships-for-sustainable-development [accessed 12 December 2017].

13 As quoted in DiploFoundation (2017) *Data diplomacy: Big data for foreign policy. Summary of a half-day event at the Ministry of Foreign Affairs of Finland*, p. 2. Available at https://www.diplomacy.edu/sites/default/files/Data_Diplomacy_Big_data.pdf [accessed 30 January 2018.

14 Kitchin R (2015) What does big data mean for official statistics? *Discover Society*, 30 July. Available at https://discoversociety.org/2015/07/30/what-does-big-data-mean-for-official-statistics/ [accessed 12 December 2017]. Baldacci E *et al.* (2016) *Big data and macroeconomic nowcasting: from data access to modelling*. Luxembourg: Publications Office of the European Union.

15 Zwiefel B (2016) *Trusted Data Collaboratives for Safe and Sustainable Water*. International Conference on Big Data for Official Statistics, Dublin, Ireland, 30 August–1 September. Available at https://unstats.un.org/unsd/bigdata/conferences/2016/presentations/day%203/Brant%20Zwiefel.pdf [accessed 12 December 2017].

16 United Nations Office for the Coordination of Humanitarian Affairs (2012) *Humanitarianism in the Network Age*. New York, NY: United Nations Publication.

17 Fosland M and Martinsen O-M (interview 9 October 2017).

18 European Parliament (2017) *European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)). P8_TA(2017)0076*. Available at http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0076&language=EN&ring=A8-2017-0044 [accessed 12 December 2017].

19 United Nations Statistical Commission (2015) *Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data*. Statistical Commission Forty-sixth session. Available at https://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf [accessed 12 December 2017].

20 Pomel S (interview 12 July 2017).

21 Alerksoussi R (interview 7 April 2017).

22 Ruppert ES (2015) Who owns big data? *Discover Society*, 30 July. Available at https://discoversociety.org/2015/07/30/who-owns-big-data/ [accessed 12 December 2017].

23 United Nations Economic Commission for Europe (2014) *A suggested Framework for the Quality of Big Data: Deliverables of the UNECE Big Data Quality Task Team*.

24 Hocking B & Melissen M (2015) *Diplomacy in the Digital Age*. *Clingendael Report*, p. 16. Available at https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf [accessed 12 December 2017].

25 United Nations Global Pulse (2012) *Big Data for Development: Challenges & Opportunities*. New York, NY: Global Pulse. Available at: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf [accessed 12 December 2017].

26 United Nations Office for the Coordination of Humanitarian Affairs (2012) *Humanitarianism in the Network Age*. New York, NY: United Nations Publication.

27 Data-Pop Alliance (2015) *Big Data for Climate Change and Disaster Resilience: Realising the Benefits for Developing Countries*, pp. 25–-26. Available at http://datapopalliance.org/wp-content/uploads/2015/11/Big-Data-for-Resilience-2015-Report.pdf [accessed 12 December 2017].

28 European Parliament (2017) *European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)). P8_TA(2017)0076*. Available at http://www.europarl.europa.eu/sides/getDoc.

do?type=TA&reference=P8-TA-2017-0076&language=EN&ring=A8-2017-0044 [accessed 12 December 2017].

29 United Nations Global Pulse (2012) Big Data for Development: Challenges & Opportunities. p. 27. New York, NY: Global Pulse. Available at http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf [accessed 12 December 2017].

30 United Nations Children's Fund (2017) *Data for Children: Strategic Framework*, p. 8. Available at https://data.unicef.org/wp-content/uploads/2017/04/Data-for-Children-Strategic-Framework-UNICEF.pdf [accessed 12 December 2017].

31 Johnson JA (2015) How data does political things: The processes of encoding and decoding data are never neutral. *LSE Blogs*, 7 October. Available at http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/07/how-data-does-political-things/ [accessed 12 December 2017].

32 Boehnert J (2015) Viewpoint: The politics of data visualisation. *Discover Society*, 3 August. Available at https://discoversociety.org/2015/08/03/viewpoint-the-politics-of-data-visualisation/ [accessed 12 December 2017].

33 Hocking B & Melissen M (2015) *Diplomacy in the Digital Age. Clingendael Report*, p. 16. Available at https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf [accessed 12 December 2017].

34 Medine D (2016) Making the case for privacy for the poor. *CGAP Blog*, 15 November. Available at http://www.cgap.org/blog/making-case-privacy-poor [accessed 12 December 2017].

35 Almufarji A (2016) Mobile positioning data pilot project for official statistics in Oman. *International Conference on Big Data for Official Statistics*, 30 August - 1 September, Dublin, Ireland. Available at https://unstats.un.org/unsd/bigdata/conferences/2016/presentations/day%201/Ahmed%20Almufarji.pdf [accessed 12 December 2017].

36 United States Federal Trade Commission (2016) Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues. *FTC Report, January 2016*. Available at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf [accessed 12 December 2017]. Podesta J (2014) Big data. Seizing opportunities, preserving values. *Executive Office of the President*, p. 51. Available at https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [accessed 12 December 2017].

37 United Nations General Assembly (1948) *Universal declaration of human rights* (217 [III] A) Available at http://www.un.org/en/universal-declaration-human-rights/ [accessed 12 December 2017].

38 Pomel S (interview 12 July 2017).

39 Latonero M & Gold Z (2015) Data, Human Rights & Human Security. *Data & Society Research Institute*, 22 June, p. 5. Available at http://www.datasociety.net/pubs/dhr/Data-HumanRights-primer2015.pdf [accessed 12 December 2017].

40 Latonero M, cited in Wren K (2015) Big Data and Human Rights, a new and sometimes awkward relationship. *AAAS*, 28 January. Available at http://www.aaas.org/news/big-data-and-human-rights-new-and-sometimes-awkward-relationship [accessed 12 December 2017].

41 Mancini F & O'Reilly M (2013) Conclusion: New Technology in Conflict Prevention. In *New Technology and the Prevention of Violence and Conflict*, Francesco Mancini (ed.). New York, NY: International Peace Institute.

42 Latonero M (2015) For Refugees, a Digital Passage to Europe. *Thomson Reuters Foundation*, 27 December. Available at http://news.trust.org//item/20151227124555-blem7/ [accessed 12 December 2017].

43 Ibid.

44 International Conference of Data Protection and Privacy Commissioners (2015) *Resolution on Privacy and International Humanitarian Action*. Amsterdam, 27 October. Available at https://icdppc.org/wp-content/uploads/2015/02/Resolution-on-Privacy-and-International-Humanitarian-Action.pdf [accessed 12 December 2017].

45 Internet Society (2016) *Global Internet Report 2016*. Available at https://www.internetsociety.org/globalinternetreport/2016/wp-content/uploads/2016/11/ISOC_GIR_2016-v1.pdf [accessed 12 December 2017].

46 Amir W (2016) Data breach: Anonymous hacks Thai Navy, Ministry of Foreign Affairs. *HackRead*, 22 December. Available at https://www.hackread.com/anonymous-hacks-thailand-navy-foreign-affairs/ [accessed 12 December 2017].

47 Obulutsa G (2016) Hackers leak stolen Kenyan foreign ministry documents. *Reuters*, 28 April. Available at https://www.reuters.com/article/us-cyber-kenya/hackers-leak-stolen-kenyan-foreign-ministry-documents-idUSKCN0XP2K5 [accessed 12 December 2017].

48 Janda J (2017) Czech Foreign Ministry hacked. *European Values*, 31 January. Available at http://www.europeanvalues.net/czech-mfa-hacked/ [accessed 12 December 2017].

49 United Nations Statistical Commission (2015) Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data. *Statistical Commission Forty-sixth session, p. 10, 3-6 March 2015*. Available at https://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf [accessed 12 December 2017].

50 Ministry of Foreign Affairs of Denmark (2012) IT strategy 2013–2016, p. 13. Available at http://um.dk/en/about-us/organisation/it-strategy-2013-2016 [accessed 12 December 2017].

51 United States State Department (2016) Department IT Strategic Goals and Objectives. Available at https://www.state.gov/m/irm/itplan/264054.htm [accessed 12 December 2017].

52 E-Estonia (2017) Estonia to open the world's first data embassy in Luxembourg. *E-Estonia News*, June 2017. Available at https://e-estonia.com/estonia-to-open-the-worlds-first-data-embassy-in-luxembourg/ [accessed 12 December 2017].

53 United Kingdom Information Commissioner's Office (2017) *The Guide to Data Protection*. Available at https://ico.org.uk/media/for-organisations/guide-to-data-protection-2-9.pdf [accessed 12 December 2017].

54 Ibid.

55 Internet Society (2016) *Global Internet Report 2016*. Available at https://www.internetsociety.org/globalinternetreport/2016/wp-content/uploads/2016/11/ISOC_GIR_2016-v1.pdf [accessed 12 December 2017].

56 International Standardization Organization (2014) *ISO/IEC JTC 1 Information technology: Big Data - Preliminary Report 2014*. Geneva: ISO. Available at https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/big_data_report-jtc1.pdf [accessed 12 December 2017].

57 International Committee of the Red Cross (2017) *Handbook on Data Protection in Humanitarian Action*. Geneva: International Committee of the Red Cross, pp. 31—-35.

5. Conclusion

This report has provided an overview of the opportunities, organisational considerations, and key aspects to bear in mind to leverage the use of big data to make diplomacy more effective, efficient, and inclusive, through better targeted policies, more efficient knowledge management, and the inclusion of perceptions of those who have previously not been heard. Big data will have an important role to play in the future of diplomacy, although it will be most valuable when accompanied by more traditional sources and expert knowledge.

5.1 Capturing opportunities, recognising limitations, mitigating challenges

Relying on different sources that were previously not available, big data is able to **challenge biases, corroborate information,** and **provide new insights** for diplomacy. As a consequence, it can lead to better-informed foreign policy, away from the assumptions of individual decision-makers and inclusive of the great amount of information and knowledge that can be captured from online sources, texts, and sensors.

In addition, big data is able to contribute to various areas of diplomacy in different ways. It can be used by consular departments to **meet the growing expectations of government service delivery**, by analysing user behaviour on consular websites and analysing their needs, as well as by identifying nationals in foreign places in times of emergency. In addition, big data has the potential to provide **deeper insights into human perception and behaviour**, which could be captured by, for example, social media analysis and communication channels. An improved understanding of what people think and how they feel could provide substantial input into both public diplomacy and negotiation processes, enabling diplomats to better understand foreign and domestic discourse on particular issues, to identify influencers, and to target their message to specific audiences.

Big data can be useful in **identifying patterns and trends over time and space**, which could particularly benefit activities in trade and development. In the trade sector, big data offers new ways to monitor and evaluate trade flows, especially as with the adoption of e-money, e-banking, and e-commerce. In the development sector, big data can be put to use to contribute to measuring progress towards the Sustainable Development Goals (SDGs), to create more accurate needs assessments, and to monitor and evaluate aid flows and development projects.

In addition, big data can contribute to emergency response and humanitarian action, as data is continuously and automatically generated, and as such it can **track developments over short timeframes**. As a result, it is able to feed into early warning indicators, and it can heighten situational awareness and identify the affected populations. The analysis of communication channels and satellite imagery is particularly important in this regard.

Finally, as society at large, and the individuals within it, increasingly rely on the Internet, they leave behind digital traces of their daily activities. As a result, new forms of **evidence and accountability** emerge that can be used by international courts and arbitration systems, from social media data to satellite images and mobile phone records.

In order for ministries of foreign affairs (MFAs) to capture the benefits of big data, certain organisational considerations need to be kept in mind. In order to **maintain flexibility while being able to feed big data insights into multiple areas of diplomacy,** some MFAs have started to develop units tasked with experimenting with new data sources to provide insights across the full range of diplomatic functions. Data diplomacy can also be developed on a more ad hoc basis, for example through partnerships with academia or the private sector, when in-house data analysis capacity is not available.

If desired and possible, given potential resource constraints, the MFA could choose to train in-house data scientists who can bridge the world of diplomacy and the realm of data analytics, or attract data scientists to work for the ministry. Ultimately, if MFAs are to engage in data diplomacy, it is important for diplomats or foreign service officers to **develop at least a minimum understanding of**

what big data can do, and what the limitations and challenges of big data are when choosing to work with it. This is important for communicating with data scientists, as organisations risk losing time and effort in the miscommunication between the two professional communities.

The **limitations and challenges of big data** are connected to five key issues: data access, quality, interpretation, protection, and security. All five aspects need to be taken into account when starting a big data project at the MFA, whether in-house or in partnership. Relying on publicly available data can mitigate access challenges, while obtaining external datasets held by the private sector often involves the creation of complex partnerships. At the same time, undisclosed data might be more valuable, such as call detail records (CDRs) or geolocation. A cost-benefit consideration should be made in relation to data quality; big datasets are usually messy, complex, incomplete, and not created for the purpose of analysis

by an MFA. Will the time and effort it takes to clean this data pay off after its analysis? In addition, close attention needs to be paid when analysing the data to avoid traps in its interpretation: Is the data representative of the whole population, or only a subset (e.g. only those using Twitter)? Do the identified relationships in the data represent causality? To what extent is data analysed and presented in certain ways and not in others to serve political agendas?

Finally, whenever dealing with personal data, the **MFA has an obligation to protect and secure the data**. It can take preventative measures to ensure that it is dealing with a minimum amount of data and to anonymise and aggregate datasets. In addition, it needs to take measures to ensure data security, by securing the location, format, and design of the data. Finally, all staff managing the data need to be made aware of security concerns and understand the measures that need to be taken to keep the data safe.

5.2 Finding the right balance for effective data diplomacy

Ultimately, whether diplomacy is able to successfully adopt big data tools is likely to depend on finding the right balance between a number of dichotomies. While the private sector has made substantial strides towards adapting big data tools, the lessons from that sector are only partially applicable to diplomatic practice. The real potential for big data will only be captured on understanding which lessons can be transferred from the private sector and which insights are confined to it; which data sources can and cannot be used; and how data analysis should interact with expert insights. Diplomats should also bear in mind that, even though big data can provide (near) real-time insights, this speed might not be matched by the pace of policy-making. Finally, as has been the case in the adaptation to any new technology, diplomats should consider which diplomatic activities might be affected by the big data era, and which elements will remain more or less the same.

5.2.1 The private and public sector: Detecting transferable lessons and avoiding false optimism

The private sector is traditionally the first to mainstream new technologies in its activities, and this is certainly true for big data. Driven by competition, there is often no choice but to experiment with new innovations to stay ahead of the curve. Diplomacy might have a lot to learn from the private sector, which uses big data to drive decision-making, to fine-tune its service delivery by identifying needs

more precisely, and to streamline business operations by quantifying the performance of employees.

Unfortunately, not all these lessons are transferable:
- While companies are driven by profit, diplomats promote national interests and international order. Diplomats will only adopt big data tools if it serves this mission.
- The speed of decision-making in MFAs is usually slower than in the private sector, as MFAs will typically avoid making any rushed decisions, ensuring to corroborate any big data findings with other information.
- Compared to the private sector, diplomats are less engaged in service delivery, with the exception of the consular department.
- The work of a diplomat might not be as easy to measure as the performance of a private sector employee, given the qualitative and fluctuating nature of a diplomat's work.

Another difference relates to the role played by the two sectors in gathering data. Most of what is traditionally called big data, such as mobile phone data or data from sensors, is held by the private sector. At least for now, it is unlikely that MFAs will engage in the collection of big data, which means that they are dependent on the private sector to access data, creating an additional barrier for the use of big data in diplomatic activities. Moreover, the

private sector might have a competitive advantage when it comes to attracting data scientists, who are increasingly sought after in today's big data era.

Many diplomats to whom we have spoken about this research adopt this scepticism, perceiving big data as something that is confined to the realms of companies, service providers, and researchers. They often see it as something outside of the capacity of the diplomat, a high wall to climb for something that is not necessary to obtain.

Yet, this wall might not be as high as perceived. Bearing the similarities and differences with the private sector in mind, there are relatively simple ways to get started with big data for diplomacy:

- **Start with simple analysis, with the use of tools, or in partnership** with research institutes or the private sector, to demonstrate the potential of big data in practice.
- **Start with the analysis of open data**, which is easier to obtain, and usually contains less sensitive information.
- **Consider the creation of a small unit**, comprised of data scientists and diplomats, that has the freedom to experiment with open data, internal data, and ultimately, big data from third parties.

As a reminder, there is so much data 'out there', ready to be mined for insights, that it would be a waste not to engage in data science for diplomacy.

5.2.2 Promises and perils: Is big data always worth the cost?

While big data has the possibility to generate important new insights, it often comes with a cost. Financial resources might be spent on recruiting or re-skilling staff or on accessing datasets, and the question arises whether this investment is always worth it. Non-financial costs can also be incurred: Do the benefits of analysing personal data outweigh potential privacy risks? Does the large quantity of big data outweigh the potential lack of quality of this data, and the effort it takes to clean and organise the data, as it is often messy and unstructured?

Such questions might prevent diplomats from adopting big data tools, as the obstacles seem too high compared to unclear and unproven benefits. However, there are a number of considerations to make before such concerns lead to a big data project being abandoned:

- When big data sets are only available against a fee, it is necessary to consider whether similar data, or other proxy indicators, can be obtained through more cost-efficient ways. There is often a lot of open, publicly available data that might fit the purpose of the analysis as well.
- Recruitment and re-skilling might cost resources, but attracting data scientists or providing staff with a basic knowledge of what is and is not possible with big data might be an inevitability of current and future times. Ultimately, with skilled staff members, data analyses might even be more cost-efficient than relying solely on statistical procedures.
- As big data is usually not collected within the MFA, data analyses often involve the transfer of data from an organisation to the MFA. When considering analysing big data that contains personal identifiers, a privacy impact assessment is always necessary, and is even an obligation under the EU's upcoming General Data Protection Regulation (GDPR). Based on this assessment – of which many examples can be found online – an MFA will have a better understanding of whether or not it can proceed with the analysis.
- More data does not always mean better data. And when data is particularly messy, unstructured, and incomplete, it might be good to reconsider whether this data is really needed for the purpose it is meant to serve. The answer to this question will rely on the expertise that is available in the MFA to manage unstructured data, the necessity of the research, and the availability of alternative sources of data and information.

5.2.3 Big data velocity and the speed of policy-making: Race cars in slow lanes

One of the key differences between the public and private sector is the speed at which behaviour and organisations are adapting and the speed at which decisions are made. Diplomacy can be fast-paced and diplomats need to navigate different speeds between running sprints in case of crises and have the stamina for a marathon during lengthy and detailed negotiations. Yet, adaptations to innovative practices and technological innovation usually take time as far as the practice of diplomacy is concerned. Fast-paced developments in the area of big data analysis pose a challenge in this regard.

Further, we can find an additional dichotomy in the difference between the speed at which big data insights are generated and the speed at which policy can be made. The velocity of big data has great advantages when it comes to reacting to crises and humanitarian catastrophes. However, policy-making happens at a much slower

pace and might cause tensions between the speed at which insights are generated and the speed at which policy-makers can react. Synchronising the speed by which data becomes available with the pace of decision-making might be one of the biggest challenges in diplomatic services, and this pace has often not substantially changed. For policy-makers 'instant dissemination of information about events both far and near is proving to be as much a bane as a bounty.'[1]

5.2.4 Diplomacy between continuity and change, data and expertise

Diplomacy is a practice steeped in tradition and, for good reason, takes time to adapt to technological change. Pronouncements of the 'end of diplomacy' in the face of technological changes, such as the telegraph in the nineteenth century, have proven to be overreactions. Over time, diplomacy, like all social practices, has adopted to technological changes. Although there are, as of yet, only a few MFAs who have actively begun to explore the potential of big data, it is safe to say that the practice of diplomacy will adopt to the big data era in ways that will support its core functions.

Generally, it is not useful to look at technology and human insights as mutually exclusive. Rather, it is useful to highlight areas in which the two can support each other. This is especially important when it comes to diplomatic practice. The guiding questions for adopting big data into the practice of diplomacy should always be: How can this improve the main goals and mission of diplomatic practice? How can this improve foreign policy-making and support negotiations?

Many of the core functions of diplomacy – such as information gathering and diplomatic reporting, negotiation, communication, and consular affairs – depend on human insights and human connection. Big data is not meant to replace this. Similarly, the organisational culture of the MFA values narrative analysis and human insights over quantitative analysis. The task for the coming period is to carefully show how these core functions of diplomacy can be supported by big data insights and how the organisational culture of MFAs can make room for these insights. For example sentiment analysis on social media can usefully be employed to supplement traditional data analysis, statistics, and expert insight. In some cases, it can correct bias or misperceptions. The key is to strike a careful balance between what big data analysis can add without negating or appearing to negate the value of traditional data and human expertise.

We should also keep the limits of most big data analyses in mind. While being able to show correlations, big data analysis cannot establish causation. In other words, we can find patterns and connections, but big data will not give us reasons for why things are this way, related for example to history, or social and cultural issues. It is insight based on human expertise and experience through diplomatic practice that can and should properly contextualise big data analysis

5.3 Future frontiers for data diplomacy research

This report provides a general overview of the realistic potential of big data for diplomacy. At the same time, it opens up many avenues of further research that could be explored.

First, to 'walk the talk' of data diplomacy, it will be important to develop and publish case studies of big data analyses by MFAs that highlight what worked and what could have worked better. Such case studies will further highlight how big data can be of use for diplomacy, and could form the basis for the exploration of big data at MFAs that are not yet engaging with big data.

Second, in addition to big data, there are many other, new, sources of data that interact with diplomacy. Exploring the use of crowdsourced data for diplomacy could lead to interesting insights and can forge sustainable connections with populations at home and abroad. Furthermore, it could be fruitful to analyse the utility of open data published by the MFA; how much transparency can diplomats permit, and at what point does it become beneficial to keep information undisclosed? Moreover, how will more traditional data sources, statistics, and surveys, change to meet current demands?

Third, with an eye on fast-paced technological developments, there is a need for more comprehensive insights on the ways in which artificial intelligence (AI) will interact with diplomacy. In fact, the interplay between AI, big data, and the Internet of Things (IoT) will become an important topic for research in diplomatic practice. As the global economy will be increasingly characterised by automation, this raises the question of which elements of diplomacy will become automated, and which components are

so essentially human that they cannot be easily replaced by machines. The practice of traditional diplomacy and negotiation, which greatly depends on empathy and creativity, is unlikely to be replaced by machine learning in the near future. Yet organisational processes, consular affairs, reporting, and many other sub-components of the daily work of a diplomat might be affected in the years to come.

Fourth, the use of big data to advance diplomacy is only a small part of the broader conceptual framework of data diplomacy. Big data is beginning to be added to diplomatic agendas, from digital trade to cybersecurity, from data protection to international standards. There is an increasing push for international regulation on the potential negative side effects of big data, such as its chilling effects on privacy and data protection and the risk of algorithmic discrimination. The lack of Internet connectivity in large parts of the world, coupled with a growing reliance on big data for policy discussions, might risk leaving populations behind. If our reality is only informed by what is measurable, and if what is measurable can only be captured by digital technology (such as social media posts and GPS signals), we are excluding those parts of the population who are not yet connected to the Internet, and whose experiences cannot be integrated into policy-making. It is usually difficult to keep international policy up-to-date relative to the quick pace of innovation and rapid changes in society. What kind of regulations are needed that ensure the innovative potential of big data for society, while mitigating some of its potential adverse effects?

Fifth, with data having become an increasingly valuable and strategic resource, it might affect geopolitics and global power dynamics. The control over, and access to, submarine cables transferring data and data centres will give certain countries strategic leverage. In addition, relations between diplomats and the Internet industry will become increasingly complex, especially considering that tech giants will often have more data about a country's citizens than its own government. How will big data transform the world as we know it, and what will be the role of diplomats within this changing environment?

These are important questions to consider for diplomats and MFAs as they navigate the future of diplomacy between new technologies and traditional practice. This report has provided an overview of possibilities and constraints, and created a framework in which some of these issues can be examined in greater depth. Ultimately, diplomacy is not a static profession; it is adaptable, flexible, and malleable to meet the needs of the ever-changing international relations that diplomats attempt to manage.

Notes

1 Solomon R, quoted in Kurbalija J (2007) E-diplomacy: The Challenge for Ministries of Foreign Affairs. In. Rana K & Kurbalija J (eds.), *Foreign Ministries: Managing Diplomatic Networks and Optimizing Value*. Geneva, Switzerland: Diplo Books, p. 381.

Annex: Methodology

There is a considerable amount of academic research and policy advice on big data in relation to the private sector and a growing literature on the use of big data in the public sector, especially with regard to public service provision. Yet, as we approach the topic of data and diplomacy, there are no established frameworks or categories to build on.

For this report, we therefore followed a two-pronged approach. First, we conducted in-depth desk research of the available and growing literature on big data. We did not, however, limit ourselves to foreign affairs or diplomacy. Rather, we used existing academic research and policy advice in the business and public sector and transferred lessons learned there to the area of diplomacy and foreign policy as far as applicable. In addition, we looked at the emerging ways in which big data is used in the wider ecosystem of international affairs, such as by international organisations and civil society in their development, humanitarian, and peacebuilding efforts. We used these examples while bearing in mind the commonalities and differences with the activities conducted at the ministry of foreign affairs (MFA). Second, we conducted interviews with key personnel in MFAs and international organisations who have begun reflecting on or even working with big data.

Given the nascent nature of existing research in big data and diplomacy, we decided to focus on qualitative methods for the interviews. This allowed us to explore various interpretations and approaches and avoid prematurely closing possible awenues of investigation. We conducted qualitative non-standardised interviews with selected officials working in MFAs and international organisations. With this approach, we gave participants 'the opportunity to present their individual understandings and experiences.'[1] In practical terms, this meant that we avoided starting from or suggesting fixed definitions of data diplomacy and big data. Similarly, we used semi-structured interviews that began with an initial set of fixed questions and then became increasingly open-ended, allowing for tailoring to the specific background of the interviewees and their experiences.

With regard to the interview process, ethical considerations and precautions included informed consent, confidentiality, and awareness of the consequences and further use of the interview material.[2] As we approached each interviewee, we included a precise overview of the project, its purpose, and the intended use of the results. We gave interviewees the option to remain anonymous and informed them about access to and usage of their data and interview material and the precautions we took to limit such access. In terms of consequences, we informed participants about the context and use of the study, noted the Ministry of Foreign Affairs of Finland as the commissioning organisation, and highlighted potential consequences of the interviews. In addition to the interviews, two workshops organised over the course of 2017 in Geneva, **Data diplomacy: mapping the field**, and Helsinki, **Data diplomacy: big data for foreign policy**, added insights from practitioners and served as a first testing-ground for our understanding of data diplomacy.

Generally speaking, we noted that it is extremely important to stay open to various interpretations of the role of big data in diplomacy and foreign affairs. We encountered a variety of positions which were in part determined by the specific nature of the organisation or job description of the interviewee. Further, approaches to big data in diplomacy and foreign affairs were also fundamentally shaped by the interviewees' backgrounds, whether rooted in diplomacy or data science.

Notes

1   King, N & Horrocks C (2010). *Interviews in Qualitative Research*. Thousand Oaks, CA: Sage, p. 3
2   Kvale S (1996) *InterViews. An Introduction to Qualitative Research Interviewing*. Thousand Oaks, CA: Sage.

For more information, go to

www.diplomacy.edu