Data Diplomacy: Mapping the Field

Summary Report of the Geneva Data Diplomacy Roundtable

April 2017

DiPLO

MINISTRY FOR FOREIGN AFFAIRS OF FINLAND

This summary report provides a general overview of the discussions at the Data Diplomacy Roundtable: Mapping the Field – a brainstorming event that took place on 5 April 2017 – on the role of (big) data in international affairs and diplomacy. The roundtable was organised in the context of the Data Diplomacy research project, commissioned by the Ministry of Foreign Affairs of Finland. The final report of this project will be presented in December 2017.

DiploFoundation
7bis Avenue de la Paix
1202 Geneva, Switzerland

Publication date: April 2017

Editors: Barbara Rosen Jacobson and Katharina Höne

Contributors to this report: Aye Mya Nyein, Barbara Rosen Jacobson, Dane Burkholder, Jelena Dinčić, Jovan Kurbalija, Katharina Höne, Marco Lotti, Marissa Lopresti

For further information, contact DiploFoundation at data@diplomacy.edu

DATA DIPLOMACY: MAPPING THE FIELD

Executive summary

Data opportunities:

New forms of data, most notably big data, present many opportunities for diplomacy and international affairs in their various activities.

For ministries of foreign affairs, big data has so far mostly been used for consular affairs, for which the management of extensive databases is key. Increasingly, citizens are expecting a user-friendly, digitised, service delivery from their governments, and consular departments are under pressure to keep their services up to date. From e-visas to monitoring nationals abroad, and timely assistance in emergency settings, big data can be of great value.

Furthermore, big data can be particularly helpful in strategic planning and policy research. For example, in bilateral relations, data can help with mapping their political, economic, cultural and other aspects. In multilateral relations, data can provide a comprehensive survey on the legal status of international agreements. Smart searches through databases of signatures and ratifications of international agreements can provide patterns on countries' position on specific international legal issues. Data can also complement diplomatic reporting in substantiating arguments and challenging assumptions and bias. In public diplomacy, data obtained from social media platforms can serve as a basis for sentiment analysis towards particular issues, regions, or countries, as well as to measure the impact of information campaigns. Furthermore, network analysis can point towards influencers and agenda-setters, together with social media profiling this can support the creation of targeted messages.

Big data has substantial promises in development and humanitarian aid work. One of the most important areas in this regard is its utility as an early warning system and prevention of crises; this area is being explored by both ministries of foreign affairs and humanitarian organisations. In addition, big data can help deliver aid to targeted beneficiaries more effectively, particularly with the use of satellite imagery, and can assist in monitoring the impact of international aid. Finally, big data can be useful in monitoring global trends, and in particular – the progress towards the Sustainable Development Goals.

Data limitations:

While big data can be extremely helpful for international affairs, one should be realistic about its limitations.

First, although big data can pinpoint trends, patterns, and correlations, it has limited predictive power. The volatility of international affairs can hardly be summarised into all-explanatory formulas, and causal patterns are generally hard to uncover.

Second, data analysis can be heavily confounded by confirmation bias. Large datasets can be analysed in a multitude of approaches, and their outcomes can be interpreted in an equal diversity of directions. It is often possible to present data in a way that confirms pre-existing ideas.

Third, big data does not always paint an accurate picture of society, as it usually over-represents those who have access to the Internet and digital devices. This representative bias might become even more prominent when analysing data from certain online platforms, such as social media websites, as the analysis will over-represent certain demographics that are particularly active in framing online discussions.

Data challenges:

Whether data can be useful for a particular purpose is further defined by the possibility of mitigating a number of important challenges, which give rise to the following pertinent questions:

- Is access to data free, or does it need to be purchased or negotiated?
- Is the data reliable and of good quality, and is its source trustworthy? If derived from different sources, can the data be compared?
- How to ensure the privacy of personal data? Has the data been obtained with assumed, or informed consent?
- Is there sufficient behavioural awareness and technical capacity to store data in a secure way?

Data management:

When embarking on a strategy towards becoming an organisation that is more conducive to big data, it is important to consider whether data analysis can be conducted with in-house capacity, or whether (part of) it needs to be outsourced. If this is the case, it is important to take a close look at how privacy, security, and quality are ensured by the external data analysers. Partnerships with the private sector, either to obtain the data, or to process, analyse, and interpret it, raise questions regarding the sustainability and reliability of differing and changing interests within private sector companies.

Furthermore, data science is a highly interdisciplinary endeavour. Apart from the technical know-how needed to understand how to work with the data itself, there is a need for specialised policy knowledge in order to pose the right research questions, assist in interpreting data, and provide the right context in which the data is presented. Therefore, the organisational culture should be supportive and understanding of the benefits and challenges brought by data, especially among those who are expected to work more closely together with technical teams that process the data.

Recommendations:

1. Capacity development and awareness raising are needed in organisations for an overall understanding of how to work with data and how to keep it secure.
2. To make optimal use of existing data in an organisation, its findability and management need to be enhanced so that they can be more easily accessed by those who need them.
3. Data should always be embedded in its proper context and combined with traditional expertise. With the right combination of data analysis and expert knowledge, assumptions can be tested and biases averted. It is also important to refer to institutions or official channels to cross-check and verify information.
4. Technical ways to better secure data, such as encryption and blockchain technologies need to be looked into.
5. Where possible, the organisation should have clear open data policies and make data openly accessible within the limits of what is reasonable, given privacy and security concerns.
6. Data consistency across countries to enhance comparability needs to be emphasised through the standardisation of data collection and formats, and through compatible legal frameworks.
7. When outsourcing data collection or relying on publicly available data, credibility of data collection and ownership need to be ensured.

INDIVIDUAL SESSION REPORTS

Part I:
Data as a tool for diplomacy

1. How do we use data to assess foreign policy?

Discussion lead: Graham Nelson (UK FCO, Head of Open Source Unit)

This discussion addressed the utility of data, and big data in particular, for foreign policy making and diplomacy. The participants pointed at a number of areas of particular utility, including early warning, consular affairs, aid work, and possibilities to map bilateral relations, multilateral resolutions, and networks of the MFA.

Big data can be particularly useful to substantiate arguments, challenge assumptions and bias in diplomatic reporting, and verify or challenge diplomatic judgements. Furthermore, when looking at data in a broader perspective, an extensive amount of data generated by the MFA remains underutilised, including more traditional data and records. The participants discussed the importance of a more effective 'findability' and 'searchability' of existing data on common platforms. Today, there is a need to look at this old challenge in new ways, while building on existing expertise.

Data analysis is often outsourced, at least to a certain degree, by MFAs, although some ministries have in-house expertise for data analysis as well. Whether outsourced or not, the organisational culture needs to be supportive of big data, and capacity development across the MFA is needed for a better overall understanding of how to work with data.

Nevertheless, many challenges and limitations remain. Data is not always easily available, and when it is available, it needs to be used with the proper consent of those whose data is being obtained. Data is not always useful in predicting situations, and its analysis and presentation are often confounded by confirmation bias, as it can often be interpreted in many different ways. Therefore, it is very important to embed the data within its proper context, to be clear about whether claims from the data can be made with certainty, and to combine data insights with traditional expertise.

2. What are the promises and challenges of (big) data in humanitarian affairs?

Discussion lead: Nadine Graas (ICRC, Information Environment Strategy Associate)

The way of generating data has significantly changed in the past years. As more and more people have access to the Internet, an ever-increasing amount of data is generated. However, this phenomenon will be characterised by big regional disparities as access to the Internet can vary significantly from region to region. Hence, the information that will be generated and subsequently collected, will be biased towards those who have access to the Internet.

Within this roundtable, the discussion mainly touched upon seven issues:

1. Prevention: Going beyond the use of big data to respond to the humanitarian crisis, big data should also be analysed and used for prevention of such disasters. Although this use of big

data could provide timely information, it also presents additional challenges regarding the collection of such information.

2. Availability of evidence: the focus rests mainly on satellite images whose accuracy allows for precise predictions and documentation of famines, droughts and migration flows. The advantage of this type of resource lies in its objectivity, on the condition that the satellite images are properly analysed and presented with complementary evidence obtained from other data sources.
3. Social Media Monitoring: similar to the use of satellite images, the monitoring of social media channels could potentially be very helpful in collecting timely information. An example is the study conducted by the UN Research Institute for Social Development, which identifies mental issues from some users' posts on Facebook. Applied to the humanitarian field, this method could help develop automatic responses to people involved in a specific crisis, therefore directing them to the relevant institutions/structures for help.
4. Monitoring and evaluation: The use of big data can also help in monitoring the impact that specific humanitarian programmes or humanitarian organisations have in the field.
5. Addressing the veracity of data: One of the main challenges of (big) data is to have reliable information. It is necessary to refer to specific institutions or official channels that can provide cross-checked and verified information.
6. The use of big data and its compatibility with existing legal frameworks: there is a discrepancy between the legal protection available and the way this information would be stored. We need to ensure that data is secured.
7. The role of the private sector: to what extent can humanitarian organisations partner with private ones in order to collect, preserve and analyse data? The storage of this information in clouds entails a high security risk, which is particularly important to address if the information contains personal data or sensitive elements. Additionally, partnering with the private sector can pose reputational risks in case of changing interests and policies in such organisations.

3. Role of data in public diplomacy and international rankings

Discussion lead: Thierry Schwarz (Former Director, Political & Economic Department, Asia-Europe Foundation)

This discussion demonstrated that international rankings matter in many countries worldwide. The importance of ranking requires proper reflection on the way how rankings are created. Rankings matter a great deal to diplomats and organisations in Geneva. For example, it is not unusual for diplomats in Geneva to be asked by their respective capitals to explore the specific position of their countries, as provided in the rankings of international actors based in Geneva. During discussions, international organisations indicated that they have also they received requests from governments to check the formulas and algorithms that are at the basis of rankings.

The main problem is the inevitable bias of each ranking. A brainstorming discussion pointed to a few potential solutions and scenarios. It was agreed that ranking formulas should be transparent. A few participants stressed 'ranking customisation', where anyone can create ranking by combining and weighting variables in easily accessible datasets. Once data is provided, rankings could easily be mashed. Rankings can 'compete' for attention and the most reliable rankings would emerge as the leading ones. A few participants were concerned that this 'competition' aspect among rankings could lead to information overload and thus create more confusion than clarity. It was suggested that one authority could be trusted with selecting ranking criteria in a transparent and inclusive way. The more legitimate the actor is, the more legitimate the ranking. International organisations were indicated as the most legitimate actors to decide on the formulas and criteria for ranking.

Participants warned about all-encompassing ranking which includes too many criteria and datasets. It was suggested that there should be specialised rankings with the possibility for a 'federation' of rankings. In particular, small organisations should be encouraged to contribute with specialised rankings in their 'niche areas' of expertise.

4. What role does data play in development and the Sustainable Development Goals (SDGs) in particular?

Discussion lead: Raymond Saner (Diplomacy Dialogue, Director)

The discussion focused on the role of data for the SDGs. Participants reflected on the road towards the SDG indicator framework (in the form of the Inter-agency and Expert Group on SDG Indicators and the High Level Political Forum), the role of the UN Statistical Commission, the World Bank, and the OECD. Given their capacities, the World Bank and the OECD were highlighted as playing a particularly important role in acting as custodians of the SDGs data.

However, the discussion also stressed that there is significant diversity with regard to institutions and initiatives that are setting out to collect data on the SDGs. Within the UN system, a variety of agencies are set to monitor aspects of the SDGs that fall within their portfolio of work. A UN system-wide initiative is the Global SDG Indicators Database, which acts as a focal point for data compiled through the UN system. In addition, non-governmental actors such as the IISD and the World Resource Institute, have begun collecting data on SDG progress independent of national initiatives.

Participants in this session largely agreed that this diversity with regard to collecting data on the SDGs and measuring progress is to be welcomed. Different ways of data collection and interpreting it are important to allow various voices to be heard and to have a critical conversation on methods and outcomes. This diversity also avoids a situation in which SDG-related data is sanctioned and approved by national statistical offices alone.

While there is a strong emphasis within the SDG framework on national implementation and data collection, the discussion also stressed that sub-national implementation of the SDGs and related activities will be crucial for their success.

The question of where SDG-related data is or will be stored came up during the session. While no clear answer emerged, there was consensus that the question of data storage and ownership is second to ensuring open access to this data. Open access to data was highlighted as one of the most important elements in ensuring the successful implementation of the SDGs in an inclusive and open manner.

This question of open access also raised concerns about uneven development and varying levels of capacities to collect and use data appropriately. Beyond the collection of data, an important factor is the ability of a country to gain knowledge for better policy-making from this data. Capacity building for data diplomacy was stressed as crucially important.

A final set of questions and points for discussion looked at the veracity of data. Participants argued that it will be crucial to have a consistent set of rules for data collection and to offer appropriate training to ensure this. The participants called for mechanisms to agree on data collection methods and to ensure the credibility of data collected. Both of these points were highlighted as a key task for data diplomacy.

In closing, the need to balance the goal of getting as much access to data as possible, while at the same time having to consider ensuring the security of sensitive data, was mentioned. The SDGs principles of inclusiveness, participation and transparency should hence be treated holistically and mechanisms of co-operation be agreed on jointly by the state and non-state stakeholders involved in SDGs implementation.

Part II:
Data management for diplomacy

1. How can and should we ensure data security in an MFA or international organisation?

Discussion lead: Antonio Gambardella (Fongit, Director)

This discussion evolved around the question, 'how can and should we ensure data security in a Ministry of Foreign Affairs (MFA) or international organisation?' This topic was approached through two lenses: behavioural and technological.

The discussion first addressed the behavioural aspect of data security. More than 70% of phishing and malware attacks are successful, which implies that individuals within organisations are still ignorant about how to avoid these traps. Therefore, changing the behaviour within an institution can be an extremely helpful way to improve cybersecurity, especially as ignorance is a 'low-hanging fruit' that can be addressed easily.

For example, the UN recently began a compulsory training system for all employees to learn about identifying phishing attacks and how to report them to their IT department. This has increased general awareness within the organisation. Likewise, the use of technology such as double-identification software can limit human error. However, participants agreed that it is important to find a balance between closing internal systems to increase security on the one hand, and maintaining high levels of communication within the organisation on the other. If security measures are too restrictive, the staff will resort to communicating outside the system through potentially riskier platforms.

The discussion then moved towards cybersecurity from a technological perspective, initially focusing on two main distinctions: static information (where the data is stored, such as an organisation's physical server or the Cloud), and communication. In terms of static data, participants expressed concern that security measures emphasising 'in-house' servers threaten to create silos within the Internet community rather than increasing global interconnectedness.

However, because big data involves massive amounts of real-time decisions and measurements, data analysis is moving away from the Cloud and towards the 'edges', thereby increasing the complexity of security efforts. The source of software used by organisations was also an important concern; although 'off-the-shelf' software designed by Google or Microsoft is often the easiest to use and the cheapest, it could create potential problems for an MFA seeking to avoid dependence on mostly US corporations.

In terms of communication, two concrete solutions were suggested: 'have no data to begin with' while properly encrypting new data or use distributed data as an alternative to centralised systems. One such method discussed was blockchain technology, which distributes information to many different computers within a network. There was also a discussion about blockchains with encryption as a future solution, although this idea had previously received mixed reviews within the Internet community.

Ultimately however, participants in the discussion concluded that cybersecurity is a never-ending cat-and-mouse race involving trade-offs between security and cost. Therefore, although new technology can help, the easiest solutions often address an organisation's internal behaviour.

2. What are the challenges for assembling national data for global insights?

Discussion lead: Rania Alerksoussi (Coordinator of the IFRC Databank)

The discussion touched on 5 main themes: data collection, diversity, trust, privacy, and the legal framework for (big) data. The diversity of participants' backgrounds - from development to IT - allowed for an interesting discussion with many unique perspectives, but it also highlighted that there is a need to bridge the difference in languages and concepts.

Data collection and data diversity are the most important challenges in assembling national data for global insight because different parts of the world format data differently, and are interested in collecting different kinds of data in a variety of ways. For national data to be available for global use, data collection needs to be as consistent as possible across countries. The discussion stressed that harmonising data and getting standardised data are key challenges for the IFRC when working with national societies. At the same time, participants emphasised the benefit of working with data that is already collected, rather than engaging in additional data collection which might lead to a serious doubling of efforts.

Trust also proved to be a key challenge. Participants focused on questions such as: how do you know that you are receiving data from a trustworthy source and how do you ensure the quality of the information? Ideas on how to try and prevent the collection of poor quality data were discussed; suggestions included having a set of authorised data reporters, and inter-organisational cross-checking to ensure quality.

Other challenges are privacy and legality. The issue of privacy concerns both individuals within a community and the community as a whole. Concerns were raised over the fact that organisations operate and share data based on assumed consent rather than informed consent. Participants agreed that it is important to work towards informed consent or explicit agreement from local areas/communities. From a legal point of view, many organisations are still formulating their own data policies and will face additional challenges with the introduction of more formal legal frameworks such as the EU initiative on data protection.

3. What is the promise of (big) data and artificial intelligence for diplomatic reporting?

Discussion lead: Jovan Kurbalija (DiploFoundation, Director)

The participants stressed that the method of analysis and the style used in diplomatic reporting are more qualitative than quantitative. This roundtable focused on three main topics:

1. Diplomatic reporting: most of the reports are mainly narrative-based and qualitative. Quantitative data is presented only in forms of attachments and does not represent the big part of the reports.
2. Use of big data: big data can be used for the analysis of diplomatic reporting, for example in the case of sentiment analysis in conference transcripts. However, there is still reticence towards the use of big data. Many reasons were provided as possible explanations: (i) security concerns; (ii) lack of predisposition and 'culture' regarding the inclusion of quantitative data; (iii) lack of relevant skills; and (iv) lack of technical platforms.
3. Artificial Intelligence (AI): is it possible to elaborate an algorithm that, basing its decision on previous decision patterns, could replace the diplomats? There are some algorithms that can easily predict the content of a discussion based on the calculations of previous patterns. However, they key issue here is the level of probability or accuracy: such algorithms can do so by producing a given output with about a 70% level of accuracy. However, it remains to be seen what level of (in)accuracy can be tolerated by diplomatic services.

Overall, the group concluded (despite a few sceptics) that AI cannot replace diplomats due to practical limitations such as technical limitations and disruptive social dynamics that would trouble the sequence elaborated by the algorithm.

4. What is that big data cannot tell, solve or predict?

Discussion lead: Kars Aznavour (ICRC, Data Analytics Advisor)

The question was approached from the angle of how people who make informed choices are indispensable throughout the process of data analysis. The discussion began by looking at a number of ways in which data is collected, used, and analysed at the International Committee of the Red Cross (ICRC).. It was argued that people with technical data analysis skills need to work with specialists who have background knowledge about the humanitarian context, in order to to make the analyses according to the ICRC's mandate and principles.

Social media analysis can be very useful. For example, Twitter data analysis can help identify humanitarian needs, trends and influencers. However, data coming from social media has to be treated with special caution, as it is not representative of the general population. When doing a social media analysis, data analysts need to be aware of bias, know the methodology used, and have background knowledge about the topics analysed.

Participants asked about the types of data analysed by the ICRC in order to respond to questions that reflect the needs of communities served. The importance of in-house data analysis for confidentiality was highlighted. This was followed by a discussion on the usefulness of open data originating from governments and other sources.

The discussion then focused on various limitations of data and potentially misleading outcomes of analyses. It was emphasised that working with data required critical thinking, contextualisation, verification and a supplementary narrative in order to increase usefulness.

For more information, go to

www.diplomacy.edu

DiPLO

MINISTRY FOR FOREIGN
AFFAIRS OF FINLAND