SCIENCE & DIPLOMACY

AAAS

Andy Boyd, Jane Gatewood, Stuart Thorson, and Timothy D. V. Dye, “Data Diplomacy,”
Science & Diplomacy, Vol. 8, No. 1 (May 2019).
http://sciencediplomacy.org/article/2019/data-diplomacy

Data Diplomacy

Andy Boyd, Jane Gatewood, Stuart Thorson, and Timothy D. V. Dye

“Data! Data! Data!’ he cried impatiently. ‘I can’t make bricks without clay.”[1]

For data to be put to the best possible uses, the various modalities in which it operates must be aligned. One must also recognize, when thinking about data, that it has generators, subjects, managers, and owners; it can be public or private, copied and changed, shared or not shared, used or not used. Furthermore, the use of data can lead to good or harm, benefits or disadvantages, and can have an impact at the individual, institutional, state, or global level. As such, data can be

Andy Boyd is a data scientist in the Department of Population Health Sciences at the University of Bristol Medical School. He is also technical lead for data linkage at the Avon Longitudinal Study of Parents and Children, for which he designs research infrastructure to facilitate publicly acceptable data science. Through the CLOSER cohort consortium, he lobbies the UK government and data owners to establish a more permissive, yet secure, environment for research data sharing.

Jane Gatewood is vice provost for global engagement at the University of Rochester, where she facilitates the development of bi- and multilateral partnerships for international research, education mobility, and economic development.

Stuart Thorson is Professor Emeritus, Political Science and International Relations, at the Maxwell School of Citizenship and Public Affairs at Syracuse University. His research focuses on the impact of information and communications technologies on governance.

Timothy Dye is a medical anthropologist and social epidemiologist who specializes in applied public health, particularly within marginalized, isolated, and global populations at the University of Rochester School of Medicine and Dentistry. His interests lie in helping communities create and govern data for social good in the global context.

a tool precipitating, driving, or else impeding change. It therefore follows that the concept of data diplomacy would relate to a broad spectrum of interactions among data and diplomats, as well as diplomatic principles, practices, and objectives. Indeed, as a construct, data diplomacy connects and attempts to describe the motivations, methods, values, policies, and other dimensions of negotiations involving data or data use.

In traditional frameworks, diplomacy occurs between nation-states. But the rising availability, even ubiquity, of data for state and nonstate actors at the supra- and subnational levels challenges these traditional notions, aligning with a public diplomacy approach—which has a wider range of stakeholders. By regarding data as a domain for diplomacy, practitioners can more easily consider shifts in traditional geopolitical, organizational, and thematic boundaries, with the resulting contours likely more virtual or conceptual. The role of the citizen also changes markedly in this model. Within international affairs, the diplomat represents the state, which in turn represents citizens. But when the focus turns to data, drivers of change emerge as well in industry, academia, and directly from the public. In this shifting landscape, clarity is rare as to who serves as the "diplomat" negotiating mutually acceptable exchanges. Also affecting public trust in evidence-based policy are data complexity, lack of provenance, interpretation, and monetization. These factors tend to weaken Western social democratic models constructed around the notion of "informed choice," whether relating to democratic elections, autonomy over healthcare and education decisions, or control over personal data use. In turn, the risk increases that data, along with its sites of creation, use, access, and consumption, can become loci of conflict—explaining the need for data diplomacy.[2]

From Diplomacy to Data Diplomacy

Diplomacy, known in broad terms as a nonviolent approach to international relations that relies on negotiation, is conducted by diplomats who represent and protect state and citizen interests while monitoring and promoting cooperative relations with other states.[3] Formal diplomacy, known as Track I, typically occurs at an international level, whereas informal, or Track II, diplomacy can happen in various forums—including sometimes led by credentialed diplomats. Track II efforts allow for a wider range of stakeholders relative to Track I, and they entail less consistency in representation of views and modes of message transmission.

Science diplomacy spans both diplomatic tracks, as elucidated in a 2010 study authored by American Association for the Advancement of Science (AAAS; the publisher of Science & Diplomacy) and the Royal Society, and can be separated into the following subcategories:[4]

- **Science in diplomacy,** where scientific experts enter the diplomatic process, seeking to help identify and address national and global policy issues (e.g., climate change).
- **Diplomacy for science,** where diplomats work together to advance science and international collaborative science programs (e.g., the International Thermonuclear Experimental Reactor).
- **Science for diplomacy,** where scientific interaction and collaboration serve as tools to establish and build on relationships between nations (e.g., Cold War interactions between Western and Soviet scientists, which generated trust and opened communication channels used for wider policy and diplomacy purposes).

Similar subclasses can be applied to the data diplomacy field —thus, "data in diplomacy", "diplomacy for data" and "data for diplomacy" warrant further exploration. Within these concepts, respectively, data can be enlisted to drive or prevent change; data can require formal and informal diplomatic techniques in order to fulfill its potential or to help prevent harms resulting from its use; or data use and control can further the building of diplomatic relations and capacity. While the scientific use of data falls within the permutations of science diplomacy, data and data use transcends science, and today permeates society and the market state, not just the nation-state. The emergence of data as a critical tool is amplified by the growth of data-driven artificial intelligence (AI), within which correlations identified in transactional, behavioral, sensory, health, or any other "big data" areas are increasingly used as the computational basis for algorithms. These affect prices for goods, the nature of healthcare advice disseminated, and the ways police perceive suspects. Increasingly, people may be expressed as the sum of their data-related algorithms, a blurring of their baseline humanness. Academics and industry start-ups are proposing ways to use a person's "digital footprint"—i.e., their lifetime accumulation of data—to build AI representations that outlast their lives, a so-called augmented eternity. In life, such profiles could be used to produce virtual AI copies of people—"intelligent avatars"—enabling them, for instance, to operate professionally in multiple forms at once.[5] This example, extreme as it may sound, is intended to illustrate the rapid progress that has moved "data" well away from being simply a tool within science or administration.

Based on these considerations and retaining AAAS/Royal Society's framework, we propose the following definition of data diplomacy: the harnessing of diplomatic actions and skills by a diverse range of stakeholders to broker and drive forward access to data, as well as widespread use and understanding of data.

This definition accounts for the boundaries of the field as well as the breadth of potential stakeholders. It also recognizes the crucial need for data diplomacy to extend across the data life cycle, from generation to use to societal impact. As such, the definition acknowledges the possible distinctness of data diplomacy from traditional diplomacy. Data diplomacy, namely, may be conducted without Track I diplomats, without a connection to a state's interests, and without a discrete international dimension. This aligns with a 2012 study on how the internet has transformed international relations and empowered previously marginalized non-state political and social groups and individuals.[6] To reduce confusion with other related concepts, other commentators have coined the phrases "digital diplomacy" and "virtual diplomacy" to describe, respectively, modes of communication in diplomacy and means of providing diplomatic functions.[7] By comparison, data diplomacy focuses on the thing—data (including its creation, use, access, and interpretation)—rather than the mechanism or medium of communication. The data can then be interpreted by a range of actors with a variety of goals. To be sure, distinguishing between data and digitization/communication mechanisms masks possible interactions between them within the diplomacy domain. For example, advances in technology could see a "virtual diplomat" operate using data-driven AI.[8]

Few others have formally considered the convergence of data and diplomacy. In 2014, William J. Hybl chaired the U.S. Advisory Commission on Public Diplomacy, which issued a report using the term "data-driven diplomacy" to refer relatively narrowly to data metrics meant to evaluate the impact of diplomacy and diplomats.[9] In 2018, Barbara Rosen Jacobson and her co-authors discussed the emerging use of big data to inform Track I diplomacy, provide insights into other parties' agendas and circumstances, and bring objective information into a negotiation that could form the basis of mutual trust. They describe a three-part typology wherein big data becomes a tool for diplomacy,[10] emerges as a new topic on the diplomacy agenda, and ultimately transforms the diplomatic landscape. Assessing the role of big data in international relations, Andrej Zwitter emphasizes the associated benefits and challenges alike, and identifies how changes in technology are upending power dynamics and rewriting the international order. To prevent harm associated with these shifts, he calls for remedial action.[11] But as this paper has shown thus far, the story doesn't stop with traditional diplomats or even big data. All people are potential actors in data diplomacy, and this diplomacy can be extended to address any type of data.

Data in Diplomacy

This paper's concept of data in diplomacy refers to the infusion of data, and expertise on data, into relations between nation-states or other entities. This definition, however—distinct from AAAS/Royal Society's framework—does not restrict itself to traditional diplomats. In both positive and negative ways, data can affect diplomatic processes, sometimes triggering policy actions. For example, performance measurement using comparative nation-state-level data may help establish acceptable norms, direct aid, secure international credit, or influence negotiations, and in turn can promote the interests of selected actors.[12] Much data in diplomacy will occur as a subgroup of science in diplomacy, such as when using mortality statistics to help drive international healthcare improvements. Other Track I interactions are themselves data focused, such as the figures on human rights violations produced by the non-profit Human Rights Data Analysis Group. The group apply these data to diplomatic efforts to support improvements, including through "name and shame" campaigns in both formal and informal diplomacy,[13] and to assist post-resolution work,[14] including trials—which can be part of informing and delivering a Track I negotiated settlement.

In contrast to such well-defined data collection and use within international relations, one finds unstructured and uncontrolled examples of data in diplomacy. For example, whistle-blowing data disclosures—e.g., Edward Snowden's public revelation of the U.S. National Security Agency's PRISM and metadata surveillance programs—can themselves be considered diplomatic gestures. Snowden's stated motivation was to help redress a perceived injustice in the U.S. claim of collecting citizens' data for "state security" purposes (with connotations of intrusion and possible oppression) rather than transparent "national security" purposes (aimed at protecting the public good).[15] Snowden intended for the data release to lead to policy changes, explaining why his act could fit within data in diplomacy. Moreover, after Snowden's disclosure, the data became useful to other diplomatic actors sympathetic to his views, or whose thinking and actions were influenced by the content of released data extending the life span and impact of his actions. For instance, high-profile phone tapping and the potential harvesting of European Union citizens' data stored by U.S.-based corporations led to calls for the proposed EU General Data Protection Regulation to require explicit citizen consent for all data use. Yet while intended to protect citizens' rights, this move would have had unintended consequences, such as hindering research on health and social issues.[16] A coordinated response was needed from the scientific community to amend the suggested changes in a way that safeguarded research for the public well-being.[17] This response, which resulted in permissive exemptions for public good research, illustrated how diplomacy for data could be enlisted toward favorable societal ends.

Improvements in AI suggest opportunities for "virtual diplomats" to conduct Track I or II virtual diplomacy, through either virtual or augmented reality. This could possibly allow for low-cost, scalable diplomatic functions. But such opportunities could come with risks—e.g., that a virtual diplomat could be "hacked" by actors with malicious intent—and will require robust data governance frameworks. These frameworks, in turn, should ensure fairness in how AI algorithms are framed, recognizing that individuals with different digital footprints may have different experiences and outcomes, with some more positive, others more negative. One salient illustration of these risks came with the alleged micro-targeting of voters during the 2016 U.S. presidential election by Cambridge Analytica using citizen social media records. In this example, citizens' beliefs and behaviors—as defined by their digital footprints—were used to promote a political viewpoint without their knowledge or consent. This practice did not constitute data in diplomacy, at least not within the current diplomacy paradigm, given that the intended outcome was to manipulate rather than achieve some sort of negotiated change. Nevertheless, it does show the risks associated with malicious intent in data-driven interactions, the dangers of non-transparent algorithm development and use, and potential confusion among states and citizens in identifying diplomats, particularly on Track II.

Diplomacy for data will be required to address these concerns, coupled with education and training for practitioners. Indeed, elsewhere, such as in the United Kingdom, allegations of state-sanctioned intervention in elections continue to raise concerns.[18] In response, states are developing cybersecurity infrastructures to ensure resilience. Given the inherently global nature of cyber threats, the success of this approach will have an international dimension wherein Track I data in diplomacy is enlisted for sharing classified data (e.g., IT system vulnerabilities, insights into attacks) and coordinating responses.[19] For example, recently the United States, United Kingdom, and Netherlands have published data illustrating Russian state-sponsored cyberattacks: here the publishing of the data is an act of data in diplomacy in response to cyberattacks, which are in themselves an indirect example of data in diplomacy intended to influence relations and diplomacy.[20]

Data diplomacy, such instances show, is driven by actors at the macro and micro levels, often at the difficult-to-discern cusp of soft and hard power. In the emerging data diplomacy community, a main role would be to educate and train Track I diplomats in data theory, use, and interpretation. In turn, diplomats could be equipped to understand and thereby facilitate protections against people, entities, and states using data in malicious ways on the global, national, or personal level. In such training, rigorously sourced data should always be given the most weight, with less-rigorous data afforded appropriate caution.

Diplomacy for Data

The practice where stakeholders interact to advance data, data use and data interpretation is diplomacy for data. Examples include:

- Standardized data coding languages—e.g., the International Classification of Diseases coding system established by the World Health Organization

- Frameworks within which data can be exchanged for purposes of international comparison—e.g., data sharing for United Nations benchmarking
- Initiatives to include the public in data-related decision making—e.g., the UK's public-inclusion program known as INVOLVE, created by the National Health Service's Institute for Research
- Efforts promoting the value of evidence-based policy, medicine, and scientific insights—e.g., the Pew-MacArthur framework for evidence-based policy[21]

Each of these cases depends on diplomacy on both tracks to promote recognizable frameworks for data generation, access, use and reporting. Relatedly, as this paper has already indicated, recent years have seen radical changes in data volume, variety, and usage, driven in part by increased monetization.[22] The corresponding challenges involve ensuring international legislation and ethical frameworks keep pace with these technological advances and the ways in which these changes impact societal and diplomatic attitudes toward data.

One illustration of the uneven regulatory environment involves AI-controlled autonomous vehicles in the United States, currently regulated by thirty-three states while the National Highway Transportation Safety Administration develops federal guidelines.[23] Here, the need for coordinated legislation is quite obvious, given that vehicles must cross state and international borders. Diplomacy will be required to reconcile technical issues from legal and ethical perspectives. A more fluid example comes in the form of AI personal avatars and "digital companions" being developed by academics and technology firms. Potential tangible benefits provided by avatars could include management of data preferences, provided by a "consenting" personal avatar,[24] whereas an AI digital companion could counter the harms associated with loneliness. But technologies such as digital assistants (e.g., Siri, Alexa, Cortana) remain in their relative infancy and are not immune to providing incorrect or even harmful guidance.[25] Mitigating these potential dangers will be the province of diplomacy for data, which can help safeguard people's rights and minimize damages—e.g., ensuring that AI medical guidance is based on consensus scientific opinion rather than product placement determined by the highest bidder. Diplomacy for data can likewise work to establish regulation and oversight on how AI avatars operate and make certain AI algorithms are transparent and understandable to their developers, regulators, and users.

Also in the purview of diplomacy for data will be protection of user privacy. One cautionary tale here is the Google Glass product, said to have video-recorded public interactions by users. Other data-sharing initiatives, such as the United Kingdom's care.data repository for medical records, have been introduced in insensitive or unclear ways and therefore been regarded as unacceptable by key stakeholders. Care.data was abandoned following public and media criticism of the system's ill-defined governance frameworks. These were not considered to sufficiently protect patient confidentiality, allow informed decision making by the public, and placed family doctors in the role of gatekeepers in a Big Data science initiative. Commentators seeking to understand how to enable public good data science at scale have contended that while care.data was entirely legal, its misalignment with public expectations denied it "social license" to continue.[26] All in all, the potential for adverse public reactions increases when data involves sensitive material, such as mental health, or is used for profit rather than the public good.[27]

In seeking to optimize the overall climate for data access, protection, and use, diplomacy for data includes each stage of the traditional "data pipeline"—from generation to sharing, use, and ultimately archiving or destruction. International resources for data discovery and documentation, such as the Web Observatory, operated by the UK-based Web Science Trust, could help interested parties discover and gain insight into available data.[28] Standards for collecting and reporting data—e.g., CONSORT, which covers randomized controlled trials—will facilitate understanding though internationally consistent standards.[29] In storing data, repositories would rely on processes acceptable to stakeholders[30] and amenable to secure analysis under appropriate conditions—e.g., the Data Safe Haven model,[31] exemplified by the SAIL database containing medical and social records for the Welsh population. To promote responsible data-based input for policy development, the data science community would appeal directly to legislators and regulators, as with the CLOSER longitudinal research consortium's lobbying effort to improve data-sharing provisions within the UK Digital Economy Act 2017.[32] The benefits of this approach would be manifold. Scientists would be able to understand data, assessing error in qualitative and quantitative terms. Subjects for whom data are collected would remain autonomous and protected, through transparent reporting on use of their data and their reserved right to withdraw. Governments would be informed about policy effectiveness and efficiency. Responding to such improvements, the stewards of key health, social and economic records – from governmental departments, health care providers through to industry – might become less risk-averse and more open to data sharing.

Summing up these changes, case studies could illustrate the success of data diplomacy in securely delivering improvements for the public good. Such reporting could also serve as a tool to further the respective spheres of data science and diplomacy for data. In the case of citizen-informed or led science, involvement in this process can engage individuals to become advocates for responsible and proportionate data use. The data diplomacy community could therefore work to enhance bona fide data generation, use, and frameworks while also educating stakeholders about data interpretation and critical analysis of data provenance.

Data for Diplomacy

The practice wherein data experts interact to create a platform for relationships is called data for diplomacy. An example is the UN Global Pulse, an initiative focused on making best use of big data in international humanitarian and development work, which includes a volunteer program for data scientists to contribute to development and humanitarian programs.[33] Networks like these transcend national boundaries and can serve as conduits for diplomacy. But the UN example fits especially neatly within the science for diplomacy domain, with its potential unbounded by a data element.

Yet data itself becomes an occasion for diplomacy in certain circumstances. Sometimes in research studies, the analysed health or social records of patients come from different communities than the researchers, creating a sense of disparity and the potential for misinterpretation or misrepresentation. To address associated concerns, including lack of input and control in the research process, Canadian First Nations communities have developed the OCAP framework; the acronym, referring to data, stands for community Ownership, Control, Access, and Possession. While established within the realm of diplomacy for data,[34] Rob McMahon, Tim LaHache, and Tim Whiteduck describe how data science within the Mohawk community of Kahnawà:ke is supporting self-governance in the education sector. They conclude that "people working in the area of data management are actively engaged in the production, curation, and sharing of the community data assets that support Nation rebuilding and resurgence."[35] In this example, considering data, building the OCAP framework, and subsequently taking control of the data and running First Nation–led data science is seen to support communities, and as such represents data for diplomacy. Here, we must acknowledge that this reflects a shift from the science for diplomacy paradigm, wherein the scientist serves as a conduit for diplomatic activity. In this alternative view, data can be used to empower communities by granting them cultural capital, thereby opening up varying diplomatic possibilities. These possibilities include a community's advocacy for greater autonomy, something distinct from using data as leverage in a diplomatic exchange.

Certain states, fearing the consequences of such potential, have moved to block citizens' access to their data rather than empowering them by making this data available. In other cases, states use data as a tool to manage citizens' behavior. China, for example, is planning a "social credit system" that uses "financial standing, criminal record and social media behavior" metrics to inform citizen-to-citizen interactions and establish a "sincerity culture."[36] The Chinese state has supported the creation of its own social media platforms (e.g., Weibo, WeChat), a move likely motivated by national and commercial interests, awareness of the potential for data to shape public discourse, and a reluctance to let these platforms be controlled or mined by foreign states.

In the data for diplomacy paradigm, a diplomat's role would be to facilitate – using diplomacy for data - the potential for data as a positive soft power mechanism within cultural and societal interactions. While restrictions to data access and use can be tools of oppression, diplomats will need to understand the use of data to foster interactions and relationship building within the science for diplomacy tradition.

Cross-cutting Interactions

Inevitably, the various spheres of data diplomacy described here will overlap. For example, the information management unit of the UN Office for the Coordination of Humanitarian Affairs (OCHA) supports humanitarian operations through the production of targeted data (data in diplomacy).[37] To achieve this aim, they build information management systems and negotiate access to data from data owners across the world (diplomacy for data). OCHA has visualized the information it holds on ReliefWeb (www.reliefweb.int), an online tool to inform and appropriately direct humanitarian aid (data for diplomacy). Earlier, we made the case that Edward Snowden's motivations for the NSA leak could be seen as a form of data in diplomacy, with the released data sparking new and distinct journeys. It led, for example, to the EU revocation of "safe harbor" data-sharing agreements between the United States and EU, given that European citizen data repatriated by U.S. social media companies was vulnerable to NSA data collection (diplomacy for data). Concerns also led the EU Committee on Civil Liberties, Justice and Home Affairs to suggest applying habeas corpus principles to the digital era, thus providing a legal instrument to safeguard individual freedom against arbitrary state action (data for diplomacy).[38]

Challenges in Data Diplomacy

Data stakeholders, as this paper has shown, include anyone producing data, extending across the full data life cycle. They encompass, in an expanded list, data subjects, generators, collectors, managers, processors, owners, aggregators, marketers, users, regulators, and archivists—although most would not associate their data use with the label "diplomacy." Furthermore, in countries with a free press and open access to the internet, decentralization is a notable quality of the data-driven world. One could then argue, in contrast to Track I diplomacy, that any stakeholder can assume the role of "data diplomat." A primary challenge within data diplomacy, therefore, involves who will undertake this role and how they will approach it. Factors in this approach include sustainability, credibility, skills, possible overcommitment, and lack of resources. Considering this is an emerging concept—there is no manual and few resources framing the issues, challenges and solutions through this particular lens—differential skill sets are likely to lead to uneven power dynamics, wherein comparatively inexperienced "diplomats" lack the capacity to fully engage with professionals. This imbalance may strain traditional diplomatic notions of credibility, expertise, and norms.

Diverse stakeholders will need to consider issues relating to ethics, politics, citizens' rights, and fundamental concepts such as the role of evidence-based decision making. To address potential imbalances in skills and power, smaller stakeholders with shared objectives can organize into consortia, yielding subsequent increases in influence and the ability to support specialist resources (i.e., a professional with a focused data diplomacy purview). Similarly, large numbers of individuals could enter collectives that utilize strength in numbers (e.g., the online campaign groups Change.org and 38 Degrees[39]). While these mechanisms provide potential routes for any stakeholder to act as a data diplomat, the (lack of) capacity for all to do so in a meaningful manner remains an open challenge.

In addition, the processes designed to guide data diplomacy must be transparent and faithful to ethical-legal standards. Intentions must be equally scrupulous. Moreover, socio-technical frameworks—such as those encompassed within the data safe haven model—should be designed with input from key stakeholders, including the broader public.[40] While confidentiality is an important notion, achieving absolute privacy may be unrealistic, and stakeholders should recognize that rigorous data science aiming to serve the public good typically carries low levels of risk. This suggests more emphasis should be placed on the potential improvements to human capital that responsible data use brings. To balance these tensions, many data owners require individual consent for data sharing and use. Gaining such consent, however, may be impracticable and risk further marginalization of vulnerable groups. In some cases, individuals may not be given a fair, free, or informed choice as to how their data are used. For instance, requiring Google to give consumers the choice to either accept its data use policy or not partake of its services may be unfeasible, given Google's dominance and adoption by many employers. But such inadequate control and transparency may erode trust in data use, introduce barriers to the use of data, and challenge the data diplomacy process. This explains the need for more multilateral, multifaceted exploration of these issues and complications. The Understanding Patient Data initiative[41] provides an example of multiple stakeholders attempting to hold a national conversation to resolve how health records are used for public good research in the UK whilst implementing the requisite controls needed to maintain trust.

Inaccurate data, data manipulation, and the associated loss of diplomatic capital pose steep challenges for the emerging data diplomacy community. Members should work hard to counter these activities while simultaneously promoting and building trust in reputable data use. An example is the BBC's Reality Check service, which fact-checked political statements during the 2017 UK election. Generally assisting in the endeavor may be increased data transparency. For example, "open access" and "freedom of information" legislation provides citizens, the media, and scientists routes for obtaining credible data to inform personal opinions, media stories, and evidence-based science. For bona fide data users, international mechanisms could provide accreditation through a "data passport" scheme.[42] Under this concept, akin to the U.S. Transportation Security Administration's Trusted Traveller Programs, data users could undergo "prechecks" in their home country to enable data access within a host country. Not all data, however, can likely be kept open, considering that states and other entities must keep certain data classified and intellectual-property data restricted, and that generators in the developing world must safeguard data from exploitation by well-resourced users.

When thinking about the important topic of data veracity, stakeholders must bear in mind that legitimate data should have rich metadata, including provenance, enabling users to gauge authenticity and to audit data access, manipulations, and transformations across the data life cycle. Technological solutions to this challenge, such as the proposal by Google's artificial intelligence unit, DeepMind, to use blockchain algorithms to indelibly embed data-sourcing histories within the data themselves, should be explored.[43] These approaches may enable "trusted data" to be distinguished from "fake data," "authorized use" (by trusted users) from "unauthorized use" (by potentially malicious or noncompetent users), and "objective use" (by independent analysts using the scientific method) from "non-objective use" (by those promoting a particular agenda or product). The benefits of such actions would include enhancing public belief in data-driven thinking (i.e., evidence-based policy) and awareness of the societal good brought about through altruistic sharing of personal information within controlled conditions.

Conclusions

Data diplomacy sits at the nexus of data, governance, and social norms, and seeks to anticipate the impact and influence of future data use. Bearing this in mind, pioneers in the field can prepare mechanisms to facilitate data transactions in ways that are acceptable to a range of stakeholders. Likewise, they can devise an analytical construct to cover the domains of data and diplomacy, thus helping characterize the work of various actors who fit within the rubric of traditional diplomacy as well as outside it. In other words, many people will engage in data diplomacy even as few will define the heart of their profession this way. Given the fast pace of change in the data arena, this framework could allow for steady adaptations to the data diplomacy landscape, wherein each individual's "data journey" reshapes the contours for those to come.

The emergence of data diplomacy also says something larger about the trajectory of diplomacy as a discipline. In a world filled with increasingly complex data, diplomacy plays out more often outside the bounds of the state, and relies less often on the full-time professional diplomat as its principal agent. The need will only grow for new classes of data diplomats who can facilitate data-driven processes while developing safeguards to reassure and protect the public. This paper has intimated next steps for building this field, but much exploration remains to maximize the potential for data to serve the public good. Here we must strive to understand comprehensively how data interacts with diplomacy in all its forms, in turn helping inform diplomatic interactions. We must likewise seek to comprehend how diplomacy can eliminate barriers, again to the benefit of the public. And we must pursue insights into how data can bring communities together, helping foster diplomatic relationships.

One might argue that data diplomacy could continue to exist as a subset of the larger diplomatic field. But this would be a mistake. A failure to view data diplomacy as a distinct category runs the risk of not recognizing its expanding role in society and the potential good or ill it may bring. SD

Acknowledgements:

Dr. Dye's time was supported, in part, by the University of Rochester CTSA award numbers TL1TR002000 and UL1TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article reflects insights from the Worldwide Universities Network (WUN) symposium on data diplomacy, which was convened by Timothy Dye and organized by the University of Rochester. The symposium—following the introduction of the data diplomacy construct at the 2015 Scientific Drivers for Diplomacy conference at the American Association for the Advancement of Science (AAAS; publiser of Science & Diplomacy), in Washington, D.C.—brought together researchers with cross-disciplinary expertise in areas spanning public health and translational research, governance and public policy, and data sharing and management, to investigate this intersection of data and diplomatic engagement.

Delegates included: Andy Boyd (University of Bristol, UK); Daryl Copeland (Guerrilla Diplomacy, Canada); Robert Downs (Columbia University, USA); Timothy Dye (University of Rochester, USA); Jonathan Foster (University of Sheffield, UK); Jane Gatewood (University of Rochester, USA); David Glance (University of Western Australia); Jennifer Lees-Marshment (University of Auckland, New Zealand); Angela Murillo (University of North Carolina, USA); Simon Price (University of Bristol, UK); Stuart Thorson (Syracuse University, USA). The symposium was held at the Academy of Medicine, New York, on October 28–30, 2015.

Endnotes

[1] Doyle, Sir Arthur Conan, "The Adventure of the Copper Beeches", "*Strand Magazine*", June 1892

[2] The following insights are based on contributions from an international workshop attended by interdisciplinary experts in the fields of data and diplomacy. We must acknowledge that the authors of this article and the participants in this

workshop all represent English-speaking, Western, developed-nation perspectives. As such, our findings are necessarily skewed toward those perspectives. The implications of data diplomacy for developing and non-Western perspectives warrant further research.

3. Article 3, Vienna Convention on Diplomatic Relations, United Nations, April 18, 1961, available at: http://legal.un.org/ilc/texts/instruments/english/conventions/9_1_1961.pdf.
4. Royal Society/AAAS, *New Frontiers in Science Diplomacy* (London: Royal Society, 2010), https://royalsociety.org/~/media/Royal_Society_Content/policy/publications/2010/4294969468.pdf.
5. Courtney Humphries, "Digital Immortality: How Your Life's Data Means a Version of You Could Live Forever," *MIT Technology Review*, October 18, 2018, https://www.technologyreview.com/s/612257/digital-version-after-death/amp/.
6. Robert Reardon and Nazli Choucri, "The Role of Cyberspace in International Relations: A View of the Literature," paper prepared for the International Studies Association annual convention, San Diego, April 1, 2012, available at https://nchoucri.mit.edu/sites/default/files/documents/%5BReardon%2C%20Choucri%5D%202012%20The%20Role%20of%20Cyberspace%20in%20International%20Relations.pdf.
7. Brian Hocking and Jan Melissen, *Diplomacy in the Digital Age* (The Hague: Clingendael Institute, 2015), https://www.clingendael.org/sites/default/files/pdfs/Digital_Diplomacy_in_the_Digital%20Age_Clingendael_July2015.pdf.
8. U.S. Advisory Commission on Public Diplomacy, *Data-Driven Public Diplomacy: Progress towards Measuring the Impact of Public Diplomacy and International Broadcasting Activities* (ACPD, September 2014), https://www.state.gov/documents/organization/231945.pdf.
9. Barbara Rosen Jacobson, Katharina E. Höne, and Jovan Kurbalija, *Data Diplomacy: Updating Diplomacy to the Big Data Era* (Geneva: Diplo Foundation, 2018), https://www.diplomacy.edu/datadiplomacy/policyresearch.
10. "Big data" can be summarized as data generated through individual and system/entity interactions with online services as well as physical, commercial, and state-provided services, and through sensors and the "internet of things"—devices connected to the internet and designed to relay substantial volumes of status and transactional data. As such, big data refers to a particular class of data that is generated across national and potentially international populations and used to identify correlations rather than causation. For a thorough description of big data in a diplomatic context, see Andrej Zwitter, "Big Data and International Relations," *Ethics & International Affairs* 29, no. 4 (2015): 377–89.
11. Ibid.
12. Susanne Soederberg, "The Promotion of 'Anglo-American' Corporate Governance in the South: Who Benefits from the New International Standard?" *Third World Quarterly* 24, no. 1 (2003): 7–27.
13. Emilie Marie Hafner-Burton, "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem," *International Organization* 62, no. 4 (October 2008): 689–716.
14. Patrick Ball and Megan Price, "The Statistics of Genocide," *CHANCE*, February 2018, http://chance.amstat.org/2018/02/statistics-of-genocide/.
15. See Alan Rusbridger and Ewen MacAskill, "Edward Snowden Interview," edited transcript, *Guardian*, July 18, 2014, https://www.theguardian.com/world/2014/jul/18/-sp-edward-snowden-nsa-whistleblower-interview-transcript; and William E. Scheuerman, "Whistleblowing as Civil Disobedience: The Case of Edward Snowden," *Philosophy & Social Criticism* 40, no. 7 (2014): 609–28.
16. Beth Thompson, "The Impact of EU Data Regulation on Research," Wellcome Trust, January 29, 2014, https://wellcome.ac.uk/news/impact-eu-data-regulation-research.
17. Jeremy Farrar et al., "Open Letter Re: Amendments to EU Data Protection Regulation," n.d., available at https://wellcome.ac.uk/sites/default/files/WTP055585.pdf.
18. Elizabeth Denham, "The Information Commissioner Opens a Formal Investigation into the Use of Data Analytics for Political Purposes," UK Information Commissioner's Office (blog), May 17, 2017, https://ico.org.uk/about-the-ico/news-and-events/blog-the-information-commissioner-opens-a-formal-investigation-into-the-use-of-data-analytics-for-political-purposes/.
19. The UK's National Cyber Security Centre aims to nullify threats and attackers by globally disseminating data on them. This requires diplomacy for data, given that the information was previously classified and is likely sourced from multiple international government agencies.
20. National Cyber Security Centre, "Reckless Campaign of Cyber Attacks by Russian Military Intelligence Service Exposed," October 3, 2018, https://www.ncsc.gov.uk/news/reckless-campaign-cyber-attacks-russian-military-intelligence-service-exposed.
21. Gary VanLandingham et al., *Evidence-Based Policymaking: A Guide for Effective Government*, Results First Initiative (Pew Charitable Trusts/MacArthur Foundation, November 2014), https://www.pewtrusts.org/~/media/assets/2014/11/evidencebasedpolicymakingaguideforeffectivegovernment.pdf.
22. "Data Is Giving Rise to a New Economy," *Economist*, May 6, 2017, https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy.

23. "Autonomous Vehicles/Self-Driving Vehicles Enacted Legislation," National Conference of State Legislators, March 19, 2019. http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx.
24. M. J. Elliot, "Privacy and AI: Problems or Opportunities?" presentation to National Centre for Research Methods workshop on social science methods and automated data algorithms, London, November 2018. https://www.youtube.com/watch?v=kNSqiHSxn88&list=PLzv58M2GAfm4tbNZJHEh1iPlxpUdav0F8&index=5&t=0s
25. T. W. Bickmore et al., "Patient and Consumer Safety Risks when Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant," *Journal of Medical Internet Research* 20, no. 9 (2018): e11510.
26. Pam Carter, Graeme T. Laurie, and Mary Dixon-Woods, "The Social Licence for Research: Why *care.data* Ran into Trouble," *Journal of Medical Ethics* 41 (2015): 404–9, https://jme.bmj.com/content/medethics/41/5/404.full.pdf.
27. Ipsos MORI, *The One-Way Mirror: Public Attitudes to Commercial Access to Health Data*, report prepared for the Wellcome Trust (London: Ipsos MORI, 2015), https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf.
28. Simon Price et al., "Worldwide Universities Network (WUN) Web Observatory: Applying Lessons from the Web to Transform the Research Data Ecosystem," *17 Companion: Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee. pp. 1665-1667, 2017;* Thanassis Tiropanis et al., "The Web Observatory: A Middle Layer for Broad Data," *Big Data* 2, no. 3 (2014), https://www.liebertpub.com/doi/10.1089/big.2014.0035.
29. Begg C et al., "Improving the quality of reporting of randomized controlled trials: the CONSORT statement." *Jama* 276, no. 8 (1996): 637-9.
30. We acknowledge that "social acceptability" is contextual, and that attitudes will change over time and vary internationally and according to individual beliefs. Despite the associated challenges, social acceptability remains an important aim, and through stakeholder engagement and involvement, identifying broad themes across particular stakeholder groups is possible.
31. Paul R. Burton et al., "Data Safe Havens in Health Research and Healthcare," *Bioinformatics* 31, no. 20 (2015).
32. See "Digital Economy Bill: Written Evidence Submitted by CLOSER," University College London, https://www.closer.ac.uk/wp-content/uploads/Digital-Economy-Bill-CLOSER-Public-Bill-Committee-submission-October-2016.pdf.
33. "Scientific Research on Data for Development and Humanitarian Response," UN Global Pulse, August 29, 2018, https://www.onlinevolunteering.org/en/un-global-pulse/scientific-research-data-development-and-humanitarian-response#Task.
34. Jennifer Espey, *OCAP & Stewardship*, discussion paper (Ottawa: First Nations Statistical Institute, 2002), https://fnigc.ca/sites/default/files/OCAP%20&%20Stewardship.pdf.
35. Rob McMahon, Tim LaHache, and Tim Whiteduck, "Digital Data Management as Indigenous Resurgence in Kahnawà:ke," *International Indigenous Policy Journal* 6, no. 3 (2015): 6, https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1226&context=iipj.
36. Michelle FlorCruz, "China to Use Big Data to Rate Citizens in New 'Social Credit System,'" *International Business Times,* April 28, 2015, https://www.ibtimes.com/china-use-big-data-rate-citizens-new-social-credit-system-1898711; "Establishment of a Social Credit System," China Law Translate, http://www.chinalawtranslate.com/socialcreditsystem/?lang=en.
37. See "Information Management," UN Office for the Coordination of Humanitarian Affairs, https://www.unocha.org/our-work/information-management.
38. See European Parliament, "Report on the U.S. NSA Surveillance Programme, Surveillance Bodies in Various Member States and Their Impact on EU Citizens' Fundamental Rights and on Transatlantic Cooperation in Justice and Home Affairs," February 21, 2014, https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2014-0139+0+DOC+XML+V0//EN.
39. https://home.38degrees.org.uk/
40. See Burton et al., "Data Safe Havens."
41. https://understandingpatientdata.org.uk
42. For the concepts of "bona fide research" and "bona fide researchers," see "Bona Fide Research," in MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies, v01-00, 23 November 2011, ed. Medical Research Council (UK) (2017). https://mrc.ukri.org/documents/pdf/data-sharing-from-population-and-patient-studies/
43. Mustafa Suleyman and Ben Laurie, "Trust, Confidence and Verifiable Data Audit," DeepMind.com, March 9, 2017, https://deepmind.com/blog/trust-confidence-verifiable-data-audit/.