REPORT

Big Data for Development and Humanitarian Action: Towards Responsible Governance

Global Pulse Privacy Advisory Group Meetings 2015 – 2016

GLOBAL PULSE

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This report was developed by UN Global Pulse with the support of Leiden University. UN Global Pulse would like to thank Leiden University for their support in hosting the meeting of the Global Pulse Privacy Advisory Group in The Hague in October 2015.

Contact: dataprivacy@unglobalpulse.org

LIST OF ABBREVIATIONS

APEC	Asia-Pacific Economic Cooperation
BBVA	Banco Bilbao Vizcaya Argentaria
ECOWAS	Economic Community of West African States
GSMA	Group Spécial Mobile Association
ICRC	International Committee of the Red Cross
ILC	International Law Commission
IOM	International Organization for Migration
ITU	International Telecommunication Union
OECD	Organisation for Economic Co-Operation and Development
PAG or The Group	Global Pulse Data Privacy Advisory Group
PET	Privacy Enhancing Technologies
SDGs	Sustainable Development Goals
UNDG	United Nations Development Group
UNDP	United Nations Development Programme
UNECE	United Nations Economic Commissioner for Europe
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNOHCHR	Office of the United Nations High Commissioner for Human Rights
UN OCHA	United Nations Office for the Coordination of Humanitarian Affairs
USAID	United States Agency for International Development
VPN	Virtual Private Network
WFP	World Food Programme
WHO	World Health Organization

Introduction

1. Opportunities and Challenges in Big Data for Development and Humanitarian Action

As the 2030 Agenda moves into its implementation phase, new sources of socio-economic and behavioral data, collected by the private sector as part of their business offerings, will be needed to understand whether the Sustainable Development Goals (SDGs) are being achieved around the world, and to help ensure that no one is left behind. Applications of mobile network data, financial transaction data, postal flows data, and data collected from social media platforms show particular promise.[1] In recognition of the role that these emerging data sources are playing, together with innovations in technology and analysis, the UN Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development published a report in 2014 entitled "A World that Counts."[2] The Report called for experimentation in data-driven innovation. It is clear that – in an increasingly interconnected world – data can and must be used for public good.

However, current frameworks for data and privacy protection have not kept pace with advances in technology, and often do not account for big data. Big data collected by the private sector as part of their commercial services makes compliance in the context of repurposing those data sources for development and humanitarian action unclear (e.g., original purpose of data collection or purposes other than business purposes). For example, in the case of the 2014 Ebola outbreak in West Africa, many argue that big data could have been useful in improving response efforts. At the time, accessing de-identified call detail records – or merely aggregated subscriber mobility patterns - from mobile networks, proved extremely difficult due to the lack of consensus on data protection and privacy standards. It is in this and similar contexts that privacy groups and researchers highlight the need for clearer guidance on how new data sources can be processed safely without hindering the important objectives of humanitarian and development response.

2. UN Global Pulse and the Data Privacy Advisory Group

Global Pulse is an innovation initiative of the UN Secretary-General charged with accelerating discovery, development, and adoption of big data analytics as a public good. To unlock the value of data safely and responsibly, Global Pulse, in 2015, established a data privacy programme,[3] part of which involves ongoing research into privacy protective uses of big data for humanitarian and development purposes. To ensure an inclusive and transparent process, Global Pulse established the Data Privacy Advisory Group. The PAG is comprised of experts from the private sector, public sector, academia, and civil society. The Group engages in a regular dialogue to strengthen the overall understanding of how today's privacy expertise can be applied to big data and its use for development and humanitarian action.

Global Pulse has developed a set of privacy principles for applications of big data, and separately, in collaboration with the PAG, has drafted a two-phase "Risk, Harms and Benefits Assessment" tool, which includes the Guidelines and Risk Mitigation checklist to help practitioners assess the proportionality of the risks, harms, and utility in a data driven project.

During the first years of membership, the Group has actively participated in discussions through a series of organised calls, and convened for an in-person two-day expert meeting in October 2015 in The Hague. The meeting was hosted on the eve of the 37th International Data Protection and Privacy Commissioners Conference in Amsterdam to discuss questions of data governance and responsible use of big data.

This report presents a summary of the main topics discussed by the PAG in general, which were mainly summarized during the 2015 PAG meeting. It also describes some of the outcomes that came out of the PAG meeting. The main issues included: the fragmentation of the regulatory landscape, potential harms of data collection and use, assessment of justified purpose for data projects, and effectiveness of and the need for consent, as well as issues related to data security, accountability, and project transparency.

[1] UN Global Pulse, "Project Series", available from: http://unglobalpulse.org/blog/big-data-development-action-global-pulse-project-series; UN Global Pulse, *Mobile Phone Network Data for Development*, (UN Global Pulse, October 2013), available from: http://www.unglobalpulse.org/Mobile_Phone_Network_Data-for-Dev.
[2] United Nations Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development, *A World That Counts: Mobilising the Data Revolution for Sustainable Development* (6 November 2014), available from: http://www.undatarevolution.org/report/.
[3] More information is available from: http://www.unglobalpulse.org/privacy.

2.1 October 2015 PAG Meeting Agenda Overview

The morning session of Day One of the meeting, held at the Living Lab at Leiden University's Center for Innovation in The Hague, started with a closed session of the PAG and a few special expert guests from the private sector and regulatory community. Participants discussed a number of topics as part of several breakout groups, led by designated PAG experts.

The afternoon session included a public-facing event, which took place at the Mauritshuis Museum, where attendees included representatives from international organisations, government authorities, academia, civil society, private sector and privacy and data protection commissioners. The first panel, "Beyond the Hype: From Potential to Impact," focused on the role big data can play in shaping decision-making processes for humanitarian action and sustainable development through sharing lessons learned. The second panel, "Big Data Governance in Humanitarian and Development Fields," looked at solving challenges through sharing examples of responsible data governance mechanisms and practices used by various public and private organisations. The third panel, "Big Data Challenges & Opportunities from the Regulatory Perspective," presented an opportunity to discuss differences and similarities between commercial and humanitarian and development applications of big data, how these differences are highlighted by the current data privacy and data protection frameworks, and explored possible global solutions.

The afternoon session also included remarks from Mr. John Edwards (Privacy Commissioner of New Zealand and Chair of the International Conference of Data Protection and Privacy Commissioners), who spoke on the importance of international enforcement cooperation and noted the critical need for humanitarian and development organisations to develop data protection policies. Mr. Jacob Kohnstamm (President, Dutch Data Protection Authority and the host of the International Conference of Data Protection and Privacy Commissioners), spoke on the importance of regulators cooperating to identify best practices on de-identification, noting the theme of the 37th Conference was "building bridges globally."

Day Two of the meeting took place at The Peace Palace in The Hague. It allowed for a deeper investigation of the most common issues that come up in data driven projects in the context of development and humanitarian action. Limited only to PAG members and several expert guests from the relevant development and humanitarian agencies, this session built on the progress of Day One. With the addition of a breakout session on the Concept for Data Aggregation Guidelines, led by Prof. Khaled El-Emam (Canada Research Chair, University of Ottawa, Canada), Day Two continued the conversation in breakout sessions, the same as the morning sessions of Day One.

Additionally, PAG members were asked to review the Risk, Harms and Benefits Assessment tool, through the prism of three different projects using mobile, social media, and public radio data:

- Testing the Risk Assessment on a Mobile Data Project
- Testing the Risk Assessment on a Radio Mining Project
- Testing the Risk Assessment and assessing risks, harms and benefits of a Social Media Project

Main challenges discussed by the PAG highlighted in this report are:

- How to reduce fragmentation of the regulatory landscape?
- How to evaluate proportionality and mitigate the risks and potential harms, and assess the positive impacts associated with data use?
- How to justify the purpose of a given data initiative?
- How to determine when consent is required and how to act when it is not possible to obtain consent?
- How to address the risks of re-identification?
- How to ensure data security?
- How to provide adequate openness?
- How to address public-private data collaborations?

The groups' discussion included further developments of a Risk, Harms and Benefits Assessment tool, Big Data Privacy and Data Protection Guidelines, a Big Data Classification Scheme, Data Aggregation Guidelines, and Terms and Conditions for Data Access and Use in Public-Private Data Collaborations.[4]

Overview of Key Issues

3. Fragmentation of Data Privacy and Data Protection Landscape

Digital communication channels have been developed for global use, and Internet access is now increasingly recognised as a right for all citizens.[5] Digital data and technology have proven their value for global development and humanitarian action. However, one of the biggest challenges as practitioners become ever more dependent on digital approaches is the lack of global consistency in the interpretation and application of data protection and privacy norms and standards. While there have been many calls for a stronger international legal framework for data protection, the international regulatory system governing data practice is still highly fragmented,[6] contributing to significant uncertainty and tension across jurisdictions and between stakeholders.

In 2006, the ILC explored the need and possibility of creating a global framework on Data Privacy and Data Protection. One of its conclusions was that there are many commonalities between the existing regulations and instruments on data privacy and protection around the world.[7] Currently, the only UN instrument that specifically deals with data protection is the UN Guidelines on Computerized Data Files of 1990.[8] Among other non-binding frameworks are the OECD,[9] ECOWAS[10], and APEC[11]. Convention 108 of the European Council - the legally binding Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data - is another important document on data protection and privacy. However, it currently only has close to fifty signatories.[12]

It seems clear that a greater cross-border collaboration is needed, as well as a common understanding of privacy norms. However, whether such a globally accepted framework would have a positive impact – and not inhibit innovation – is still under debate. At the same time, many of the existing frameworks do not specifically address big data, and some suggest a separate framework is required to do so. It is also unclear what type of instrument would be most appropriate for such a framework, and whether it should take the form of non-binding flexible ethical standards or whether it should be a legally binding instrument.

3.1 Data Privacy and Protection in International Organisations

In addition to developing new standards for the field, self-regulation is essential. Of vital importance is building awareness among staff at international organisations on the importance of data privacy and data protection, and an understanding of the risks and harms associated with data use. The Resolution on Data Protection and International Organisations,[13] adopted by the regulatory community at the International Conference of Data Privacy and Data Protection Commissioners suggests that international organisations need to adopt appropriate standards, policies and principles, and to establish mechanisms to ensure that they are carried into effect in a manner that observes internationally recognised practices.[14]

---

[4] See complete agenda at http://unglobalpulse.org/sites/default/files/Agenda%20of%20the%20PAG%20meeting%20Hague%202015.pdf

[5] U.N. Human Rights Council, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/17/27 (May 16, 2011), available from http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Annual.aspx.

[6] K. Kittichaisaree, C. Kuner, The Growing Importance of Data Protection in Public International Law, available from: http://www.ejiltalk.org/the-growing-importance-of-data-protection-in-public-international-law/

[7] ILC, Report on the Work of its Fifty-Eighth Session, A/61/10 (1 May to 9 June to 11 August 2006), para. 257.

[8] General Assembly resolution 1990/38, Guidelines for the regulation of computerized personal data files, A/RES/45/95 (14 December 1990), available from http://www.un.org/documents/ga/res/45/a45r095.htm.

[9] OECD, Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data, C(80)58/FINAL, as amended on 11 July 2012 by C(2013)79, available from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

[10] ECOWAS, Supplementary Act A/SA.1/01/10 on Personal Data Protection Within ECOWAS (16 February 2010), available from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

[11] APEC, Privacy Framework [2005] APEC#205-SO-01.2, available from http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx.

[12] Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, CETS No. 108 (Strasbourg, 28 Jan 1981), available from http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108.

[13] 25th International Conference of Data Protection and Privacy Commissioners, Resolution on Data Protection and International Organisations, 12 September 2002, available from https://icdppc.org/wp-content/uploads/2015/02/Resolution-on-Data-Protection-and-International-Organisations.pdf.

[14] Ibid.

KEY POINTS ON FRAGMENTATION

CHALLENGES:

- Inconsistency of interpretation and application of data protection and privacy norms
- Insufficient resources and duplication of efforts
- Majority of existent laws applicable to humanitarian and development contexts do not account for the use of new data sources coming from the private sector

RECOMMENDATIONS:

- Employ a greater cross-border collaboration towards a common approach on data privacy and data protection in humanitarian and development action, without inhibiting innovation or the objectives of humanitarian and development action
- Consider a greater self-regulation, including staff training within the international development and humanitarian sector
- Implement an inclusive and open process for privacy policy innovation involving a variety of expert stakeholders
- Adapt data privacy and data protection laws to the mandates of international humanitarian and development organisations
- Develop safe and flexible approaches to public-private data collaborations to facilitate humanitarian action

There is a growing number of development and humanitarian organisations taking initial steps to develop and implement data privacy and data protection guidelines and policies including the ICRC,[15] USAID,[16] and within the UN system: IOM,[17] UNDG, UNHCR[18], UNICEF, UN OCHA,[19] WFP, WHO,[20] and others[21]. In this regard, practitioners and organisations would benefit from regular information and knowledge sharing of experiences and best practices. With the increasing use of new data sources, algorithms, and technologies by humanitarians and development practitioners, a consistent, unified, and transparent process of data handling across the ecosystem is required. For these reasons, many efforts, such as working groups within the UN and the international humanitarian and development community, are underway to develop guidelines and processes for data ethics, data protection and privacy.

Finally, as new types of data are being discovered and used, new risks and types of harm may arise. The potential and impact of data is being discovered through collaborative research [22] and experimentation requires engagement with different experts and stakeholders. Therefore, creation of such frameworks should involve consultations with a variety of expert stakeholders, including humanitarian and development decision-makers, the private sector (e.g., data controllers), researchers, regulators and representatives of affected populations.

An inclusive and adequately transparent process may help organisations in gaining public trust and may ensure accountability as well as higher impact of innovation.

3.2 Data Privacy & Protection in Development and Humanitarian Action

The data privacy landscape is rapidly developing and it must take into account the needs and nuances of the development and humanitarian sectors. Some laws cover only specific instances of humanitarian action, such as emergencies. However, referring to emergencies does not always include the use of data to prepare for and prevent emerging crises, nor does it typically consider recovery from a crisis including development responses. Although some regulations may explicitly refer to natural disasters, they omit other humanitarian contexts such as pandemics or conflicts. While there is evidence of progress that humanitarian aspects are being considered within certain privacy frameworks, use of big data for development projects has not been explicitly explored. Fast and swift action is required, and any international effort to address privacy in humanitarian programmes should also be afforded to development programmes.

During the PAG meeting, members stressed that a contextual approach to the use of big data in humanitarian and development settings is required, as it does not always pose the same risks, harms, or positive impacts as the data used for commercial purposes. Saving lives during a natural disaster or helping to create sustainable livelihoods for small-holder farmers is markedly different from commercially driven initiatives to improve a specific product or service. At the same time, humanitarian missions may require highly sensitive data, the use of which, depending on the context, may create risks otherwise not expected in commercial settings.

[15] ICRC, Professional Standards for Protection Work Carried Out By Humanitarian and Human Rights Actors in Armed Conflict and Other Situations of Violence, Publication Ref. 0999, April 4 2013, available from https://www.icrc.org/eng/resources/documents/publication/p0999.htm.
[16] USAID, ADS Chapter 508, USAID Privacy Program, available from http://pdf.usaid.gov/pdf_docs/Pdacr985.pdf.
[17] IOM, IOM Data Protection Manual, 2010, available from http://publications.iom.int/system/files/pdf/iomdataprotection_web.pdf.
[18] UNHCR, Policy on the Protection of Personal Data of Persons of Concern to UNHCR, May 2015, available from http://www.refworld.org/docid/55643c1d4.html.
[19] UN OCHA, Building data responsibility into humanitarian action, (May 2016), available at http://www.unocha.org/about-us/publications/policy-briefs.
[20] WHO, Guidance on Food Data and Record Management Practices, Working Document QAS/15.624 (September 2015), available from http://www.who.int/medicines/areas/quality_safety/quality_assurance/Guidance-on-good-data-management-practices_QAS15-624_16092015.pdf.
[21] New York University Governance Laboratory, Leiden University Centre for Innovation, Mapping and Comparing Responsible Data Approaches, (June 2016), available from http://www.thegovlab.org/static/files/publications/ocha.pdf.
[22] Examples of expert research concentrated solely on privacy and data protection include Access Now, Privacy International, Brussels Privacy Hub, and many others. Collaborative research in this area is also done by many organisations that bring together a variety of expertise– examples include International Data Responsibility Group (IDRG), Data& Society, The Engine Room, The GovLab, and others.

The Group discussion highlighted that it is critical that every use of data in humanitarian or development action should consider the context, and be viewed in terms of the proportionality of risks and harms to the positive impacts.

Additionally, safe and flexible approaches to public-private sector data collaborations, as distinct from the traditional regulatory frameworks applicable to private sector business analytics, are necessary to facilitate and support humanitarian action and not complicate it.

A dialogue raising similar concerns continued at the 37th and 38th International Conference of Data Privacy and Protection Commissioners, with the participation of some of the PAG members. Global Pulse took part in the closed session of the 37th (in 2015) and 38th (in 2016) Conference. The 37th Conference concluded with the adoption of a Resolution on Privacy and International Humanitarian Action.[23] The Resolution called for a regulatory framework that accounts for the specific conditions of humanitarian work, acknowledging the integral role data processing has in successful humanitarian action. The resolution represents an important step towards harnessing the full power of data in achieving development and humanitarian goals, and it recognises that data processing is an integral part of successful humanitarian action. The 38th Conference, which took place in 2016 in Marrakesh, Morocco followed up on the implementation of the Resolution through a working group to explore nuances of privacy in the context of International Humanitarian Action.

KEY CONSIDERATIONS FOR RISK MANAGEMENT:

CHALLENGES:

- Lack of proper guidance for addressing risks and predicting harms as well as justifying purpose of data use
- Challenge of identifying big data project purpose due to unforeseeable outcomes and outputs
- Non-use of data may be also harmful and thus should be accounted for
- Big data analytics may present a risk of bias and thus cause unpredicted harms
- Lack of awareness on the harms presented by the non-use of data

RECOMMENDATIONS:

- Employ a contextual approach, accounting for cultural, religious, social and political factors
- Justify purpose of data use through the key principles of legitimacy, necessity, purpose limitation and clarity as well as proportionality of the positive impacts to the risks and harms
- Ensure data is lawfully and ethically collected from the onset, including when data is used by non-original data collectors
- Limit collection, retention, use of, and access to data to the necessary extent only
- Account for group harms to known and unknown groups of individuals
- Increase awareness of the risks and harms presented by both use and non-use of data to provide more control and accountability as well as build trust and confidence in data-driven projects
- Employ risk management throughout the data lifecycle
- Take into account of group harms
- Base decisions regarding the risks on the level of data security and according to a data classification scheme as well as availability of privacy enhancing technologies

With the interconnected world of technology and data, a tailored framework taking into account the mandates of development and humanitarian organisations may be crucial in achieving the SDGs through the impactful use of data-driven innovations.

4. Risk Management - Assessing Risks, Harms and Benefits

Fear of the harms associated with big data use, and lack of proper guidance to address risk in development and humanitarian projects often prevent effective harnessing of data when it is most critical.

Humanitarian and development organisations are often presented with difficult choices. For example, how to seek consent in an emergency situation where consent is hard or impossible to gather? What is the legitimate basis for data use when consent is impossible to seek? How to apply the principles of legitimate interest and justified purpose in situations where there is no clear-cut answer? Most often, questions regarding big data use in development and humanitarian contexts go beyond privacy and legal compliance.

Risk management, therefore, has been a priority topic in PAG discussions as the key process for ensuring fundamental rights are respected in data-driven projects.

Risk management is generally defined as 'the identification, assessment, and prioritisation of risks followed by coordinated and economical application of resources to minimise, monitor, and control the probability and/or impact of unfortunate events.'[24]

Every data use on the project must be justified for a specific purpose, and be legitimate, necessary, and proportionate in terms of risks, harms, and positive impacts.

[23] 37th International Conference of Data Protection and Privacy Commissioners, Resolution on Privacy and International Humanitarian Action (27 October 2015), available from https://icdppc.org/wp-content/uploads/2015/02/Resolution-on-Privacy-and-International-Humanitarian-Action.pdf.
[24] Hubbard, Douglas W. (2009), The Failure of Risk Management: Why It's Broken and How to Fix It, Hoboken, New Jersey: John Wiley & Sons, Inc., at p. 46.

A contextual approach, accounting for the cultural, religious, social, and political factors that may influence how data analysis impacts certain individuals, is the key to successful decision-making when it comes to process design and execution.

4.1 Lawful, Legitimate, Fair Use and Purpose

While the reason behind risk management is primarily to identify risks, harms, and establish proper mitigation mechanisms, ascertaining the beneficial purpose of the project in a humanitarian and development context is extremely important. Assessment of not only risks and harms but also of an ultimate positive impact helps humanitarian and development practitioners make a final decision on whether to proceed and what project execution strategy to choose.

Every purpose of data use should be justified by positive impacts, and any risks or potential harms should be absent or mitigated without exceeding the identified benefits of the project. As an example, UNDP provides a useful guide for project justification, which suggests that the primary purpose of the process is to capture the project idea or concept and test it against organisational plans and overall strategies. Based on the above criteria, a decision on whether to further develop the concept of the project, or proceed to implementation can be made.[25]

The purpose of data use should be justified through the following minimum principles: legitimacy, necessity, and clarity of the identified purpose.

Legitimacy

When it comes to big data analytics and the use of new data sources (such as social media data, mobile phone data, financial transaction data), many humanitarian and development organisations do not collect the data themselves. Access to this data comes primarily through the private sector. However, even if international and humanitarian organisations do not collect such data, they should still exercise due diligence in ensuring as much as possible that the data has been legitimately collected by the private sector data provider, including in compliance with applicable privacy norms as well as the highest ethical standards.

Furthermore, organisations that receive access to data should ensure the data provider has a legitimate right to share it. Often, conducting due diligence vetting to verify legal compliance is an important early step towards successful collaboration. In addition to due diligence and legal compliance, legally binding agreements setting the key terms of responsibilities and data handling between private and public sector actors help ensure the smooth completion of a successful project.

Although such vetting and legally binding agreements are necessary, legitimacy is not always a clear-cut principle, especially when the purpose of the project can change due to various circumstances and where consent is the original ground for data processing. The question that should be addressed is whether data processing (including pseudonymised data, which may be considered sensitive and personal in certain jurisdictions) can be legitimate if consent cannot be attained? (See more on consent in Section C)

Many argue that there is a need for better conditions allowing for more responsible, efficient, and legally compliant mechanisms, given the uncertainty of conflicting jurisdictions, international mandates, and lack of regulatory certainty. Such mechanisms, if established, must account for new data sources and private-public data collaborations.

Necessity and Data Minimisation

A simplified interpretation for the principle of necessity is to ask: "What type of data do I (as an organisation) need and is (the collection/retention of) this data really necessary for the result that I have in mind?"[26]

Clearly, any collection, use of, and access to data should be limited by necessity. While any data collection, use or retention poses risks, it should be limited to the minimum amount necessary. While it is true that researchers and project managers may not predict all of the potential uses of the data at the project-design stage, they must account for and justify why the data set is required, for how long it must be retained, and in which form. In other words, any 'just-in-case' retention or data collection should not take place. Additionally, deletion of any data set should be considered wherever possible according to the retention limits envisoned at the outset of the project.

---

[25] UNDP, National Implementation by the Government of UNDP Supported Projects: Guidelines and Procedures (01 July 2011), available from: http://www.undp.org/content/dam/undp/library/corporate/Programme%20and%20Operations%20Policies%20and%20Procedures/NIM_for_Government_english.pdf.

[26] Alexander Dix, Big Data Challenge & Opportunities from a Regulatory Perspective, plenary presentation to the Global Pulse PAG Meeting, The Hague 23 October 2015. See also, UNECE Report, Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes (June 2009), Principle 6; UNHCR, Policy on the Protection of Personal Data of Persons of Concern to UNHCR, May 2015, available from http://www.refworld.org/docid/55643c1d4.html, pg. 16; The Madrid Resolution: International Standards on the Protection of Personal Data and Privacy (5 November 2009), available from http://privacyconference2011.org/htmls/adoptedResolutions/2009_Madrid/2009_M1.pdf, Principle 8.

Proportionality

The question of proportionality has great operational relevance, and once the utility, risks, and harms have been assessed, proportionality refers to the ratio between expected positive and negative impacts of the project. For example, a certain dataset may hold great potential to solve the problem at hand, but the required costs and/or implications of data processing (e.g. harms that can be caused to the beneficiaries) are considered too high given the expected impact, and so are disproportionate to the perceived benefits.

Clarity of Purpose

Any data project requires those in charge to clearly define the project's purpose.

However, identifying a specific purpose of data use at the outset of a big data project may be difficult due to the uncertain outcomes of big data analytics. While the purpose should be defined as narrowly as possible, unpredicted data re-use, or expanded use that may be needed as the project progresses in the development and humanitarian contexts, remains a dilemma. The Group found that in situations where purpose may not be clearly defined or where unpredicted data re-use may be necessary, a "Risks, Harms and Benefits Assessment" may help.

There may be trade-offs between a narrowly defined and an open-ended articulation of the purpose, particularly for more explorative data projects. Taking into consideration the fact that it is particularly difficult to predefine research outcomes, the question is to what extent the specific purpose can be appropriately described beforehand? While purpose should be defined as narrowly as possible, leaving the purpose broad or shifting it once the project started can lead to unforeseen harms. In this context, project managers should re-evaluate the project's purpose, assessing all of the risks of harms and positive impacts, their proportionality, as well as the necessity and legitimacy for changing the purpose or leaving it vague.

4.2 Risks, Harms and Benefits Assessment Tool

To address the many challenges of uncertainty presented by the unique context of humanitarian and development projects, Global Pulse developed a Risks, Harms and Benefits Assessment tool to help humanitarian and development practitioners in addressing the risks and harms associated with the use of data and technology. The Risks Harms and Benefits Assessment tool is a two phase process that consists of a) Part 1, an Initial "checklist" [27] that helps to assess whether a more comprehensive assessment of risks, harms and benefits is required and b) Part 2, a more detailed Risks, Harms and Benefits Assessment.. The tool is being developed with regular feedback from the PAG, and the meeting of PAG members was a valuable opportunity to look into specific aspects of the tool in more detail. The tool will be open to continuous feedback, enhancements, and improvements to keep pace with technological developments.

The assessment tool provides guidance to ensure that risks and harms are not disproportionate to the positive impacts of a given project. It helps to incorporate privacy considerations into a risk management practice. Having a similar concept to the widely recognised Privacy Risk Assessment, this tool combines the assessment of privacy risks with addressing the ethical dilemmas of data use, and could be an integral part of a "privacy by design" approach in development and humanitarian contexts.

The assessment should be employed as a systematic process accounting for the entire lifecycle of a project, and should act as an "early warning system" for detecting potential risks.

4.3 Harms

For effective risk management, it is critical to be aware of the actual harms that might result from the risks posed by the use of data throughout the data lifecycle. Harms can be understood as the negative impact of data applications that develop as a result of the risks presented by a project. Harms must be evaluated contextually to help determine their probability or likelihood, their magnitude and severity.

The Group also discussed how big data use on a project is not always useful or necessary, yet non-use of data in certain circumstances may also carry the risk of harm particularly in the humanitarian context where data use may help save lives by providing critical real time information. For example, during one of the sessions of the PAG meeting, it was argued that if mobile phone records had been used in the Philippines during Typhoon Haiyan, humanitarian workers could have received real-time information for faster needs assessments. The use of mobile phone data could have improved disaster response.[28]

[27] Global Pulse & UNDP, *A Guide to Data Innovation for Development from Idea to Proof-of-Concept*, at p.73; United Nations Global Pulse, Data Privacy, available from: http://www.unglobalpulse.org/privacy.
[28] Jose Ramon Albert, *Beyond the Hype: From Potential to Impact,* presentation to the Global Pulse PAG Meeting, The Hague 23 October 2015.

An important distinction for analysing harms can be made by contrasting tangible and intangible harms (examples include physical, emotional, economic, reputational harms). Additionally, it is important to identify the entity that may be harmed by a given project, i.e. an individual, a group or community, an organisation or a government. Lastly, when considering harms, it is crucial to also assess the harms that can result from not using the data.

All development and humanitarian missions are aimed at bringing public and social goods to the vulnerable. However, to properly understand how every data project will affect its intended beneficiaries, domain expertise is crucial. Additionally, the views, feelings, and considerations of the affected community or beneficiaries of the projects should be taken into consideration if possible when assessing the magnitude - but especially the severity - of the harms, and the significance of the positive impacts.

PAG members were asked to consider collective privacy and how information processing could affect communities or groups, particularly vulnerable populations, e.g., those who could be affected by political or social instability, moral or cultural attitudes, or societal norms. The Special Rapporteur for Privacy in his recent Report noted, "The impact of new technologies also means that we may have to re-visit the distinctions between individual and collective privacy as well as expectations of privacy in both public and private spaces."[29]

It is true that big data not only poses the risk of exposing information about vulnerable populations through data misuse or mishandling, it also has the potential to deepen socio-economic divides through the categorisation and segmentation of individuals into larger groups, often reflecting existing inequalities. Predictive analysis has the capacity to become a "self-fulfilling prophecy."[30]. Data typically must be representative in order to accurately inform insights. Therefore, it is important to consider that certain data sets or algorithms may contain biases. To avoid biases, data quality, accuracy and human intervention in any of the data processing activities are crucial.

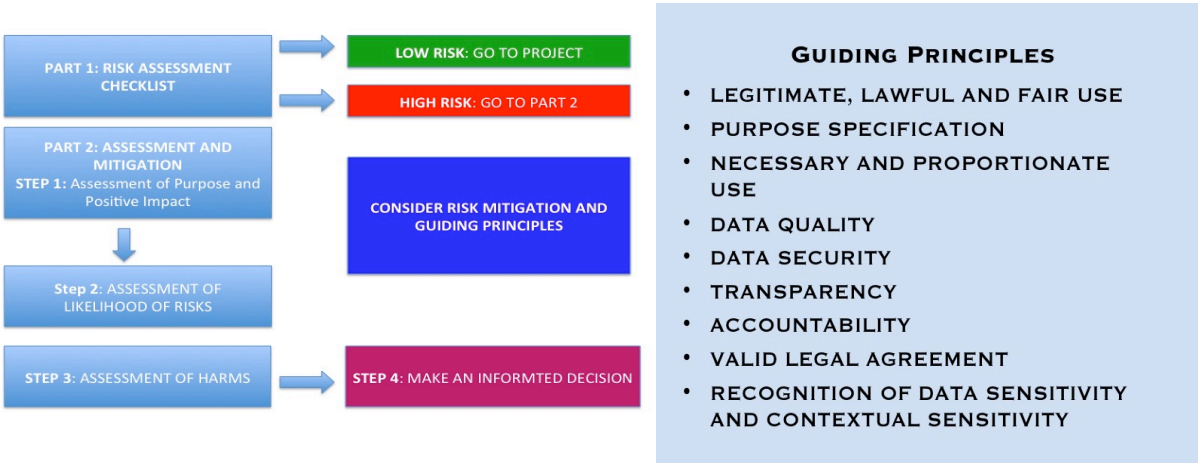
Assessment of risks and harms requires not only privacy considerations but also a variety of expert engagements (from legal, data security to social science, policy experts). For these reasons, ensuring that a multidisciplinary team is engaged in the assessment process is a key minimum requirement.

Data Sensitivity and Classification Schemes

While ensuring proportionality of the risks, harms and benefits in a given context is an important component of risk mitigation, the Group also stressed that many of the decisions should be based on the level of data sensitivity and availability of PET.

A classification scheme of data sensitivity could be a tool for reflecting the contextual nature of risk and data sensitivity. For example, social media data may be considered highly sensitive in the context of a political crisis, as compared to a natural disaster. Sensitivity can be determined by the type of data, the timeliness of data (with real-time data often being more sensitive), the degree to which re-identification is possible and who would be identified and affected by data, the means of processing and how insights will be used in the future.

Risk, Harms, Benefits Assessment Flowchart



Draft as of November, 2016

[29] U.N. Human Rights Council, *Report of the Special Rapporteur on the right to privacy* A/HRC/31/64 (8 March 2016), available from http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx.
[30] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 254 (2013).

4. Consent

KEY POINTS ON CONSENT:

CHALLENGES:

- Uncertainty in application of consent requirements in situations where consent is hard or impossible to seek
- Challenges in ensuring that consent is informed and effective
- Absence of other effective and efficient mechanisms to substitute consent when consent is impossible to obtain.

RECOMMENDATIONS:

- In situations where it is not practical or possible or where it is not clear whether it is necessary to obtain consent, assess potential harms posed from use of data, taking into consideration the principle of proportionality and legitimate interest
- Keep in mind that even if consent is obtained, it is still important to understand whether the consent was informed and freely given
- Account for expectations of privacy, cultural differences, and level of data literacy even when consent has been obtained
- Risk Assessment is still critical in sensitive situations even where consent has been obtained
- Foster public trust and data literacy through awareness building and information sharing
- Contextual self-regulation is crucial to ensure safe and responsible data use

Consent is one of the key principles of privacy regulations throughout the world, yet there is much debate over what level of consent is required for lawful collection and use of data, especially in humanitarian and development contexts. For example, some PAG members suggested that linking the requirement of consent to the potential harm posed from use of the data could be a solution. However, that would still need to be viewed through the prism of the overall proportionality of the project's risks, harms and positive impacts.

One of the main issues linked to consent is how informed, and thus, effective it is. This is particularly a concern in regions where data literacy is below average. Awareness of the risks and positive impacts posed by data use plays an important role in determining how informed consent should be. However, establishing informed consent becomes even more complicated when it is used in situations of emergency (as opposed to processing data for marketing purposes). Can a person whose life and livelihood is threatened provide informed consent where usage of their information has the potential to increase their chances of receiving critical aid?

The major problem today is the lack of awareness and education on the "good and bad" of data use in humanitarian and development contexts. Building awareness and sharing information about a project's scope and use of data with adequate transparency can help build not only higher levels of literacy but also society's trust in big data for development and humanitarian response.

Additionally, humanitarian and development practitioners should also take into account that not every piece of information shared freely and publicly on social media or radio, for example, has been shared with a proper understanding of what "public" means. Expectations of privacy can vary from one community to another. Taking into account local cultures may help determine how informed beneficiaries are when they are giving their consent.

For the development and humanitarian sector, the question is how to decide whether explicit or implicit consent is required and under what conditions consent is not a precondition for data analysis. As informed consent is not always possible, it may be worth differentiating between initiatives that require consent and those that are exempt. [31]While such schemes can help, contextual self-regulation from those collecting and using the data remains a crucial element in ensuring that data is used safely and responsibly.

5. Risks of Re-Identification

In the increasingly interconnected world of data and the Internet of Things, forming unified de-identification standards for new data sources is necessary to ensure the efficient functioning of public-private data collaborations in global development and humanitarian contexts.

Some suggest that establishing a risk threshold for re-identification would be helpful. However, the dilemma is that it may be highly dependent on the context, including the availability of other data sets, technology and skills that could make re-identification possible. Consequently, a minimum threshold approach would require regular revisions over time.

[31]Christopher Kuner, Big Data Governance in Humanitarian and Development Fields, plenary presentation to the Global Pulse PAG Group Meeting, The Hague 23 October 2015.

KEY POINTS ON RE-IDENTIFICATION:

CHALLENGES:

- Lack of de-identification standards for private sector data uses in global humanitarian and development sectors
- The risk of 'mosaic effect' and inference can lead to unforeseen data re-identification

RECOMMENDATIONS:

- Develop unified de-identification standards for new data sources by considering rapidly changing technology and data landscape
- Consider the specifics of humanitarian and development projects within existing or new standards and regulations on data de-identification
- Establish a re-identification risk-threshold may be a possible solution, which also must be subject to regular revision
- Data security is one of the key components to prevent unwanted re-identification

A threat that has become more prominent since the development of big data analytics is the 'mosaic effect,' which involves matching data-points between separate datasets in order to re-identify an individual or groups of individuals. A major issue is that it is often unknown what other datasets, identifying a certain geographical area or community, are already publicly available, and unintentional linking of datasets may cause significant unforeseeable harms to data subjects.

While many sectors already have data de-identification frameworks, especially the health industry, it will be crucial for the global community to develop common de-identification standards for each data type. It is important that such standards have guidance not only for private sector data de-identification, but also take into account humanitarian and development uses of data if not specifically tailored to those uses.[32]

In addition to developing guidelines and standards, innovative methodologies and PET to protect data are necessary. Such methodologies should ensure that data keeps its utility while being de-identified (through various methods - aggregation, k-anonymity, masking). Finally, strong data security is one of the key elements to address data protection, reduce risks of re-identification, and any other threats that are constantly posed to data in today's world, made vulnerable due to data breaches.

6. Data Security

KEY POINTS ON DATA SECURITY:

CHALLENGES:

- Inadequate data security may lead to data breaches, leaks, unauthorised access etc.

RECOMMENDATIONS:

- Employ principles of privacy by design from the start
- Apply privacy enhancing technologies and methods to protect, and where necessary, anonymise the data
- Engineer proper infrastructure for cloud computing and data storage
- Use encryption and VPN for safe data transmission where necessary
- Conduct regular audit and monitoring to identify possible system vulnerabilities
- Train personnel on the basics of data security and administrative safeguards to protect data

Since the earliest formulations of data protection principles, data security has been seen as essential to protecting privacy. Employing principles of privacy by design[33] in every data-driven project - the main concept of which is incorporating privacy protections when designing a new project - may reduce risks associated with data handling. Security is one of the essential components of privacy by design. Security can be assured with proper administrative and technical safeguards.

Additionally, many raise concerns that cloud computing and data storage may also increase risks of data leakage and breach. Having proper infrastructure for data storage and computing is critical. For example, many suggest that using an offline server in highly sensitive contexts could provide better security. Data researchers should also consider using a private cloud and ensuring that sensitive data or data that is used in sensitive contexts is stored separately from other data to avoid data linking and mixing. Encryption and use of VPN is also a recommended way to allow for safe data transmission, where necessary.

Lastly, one of the key aspects of security is to avoid and prevent human errors. For these reasons, personnel should be trained in the basics of administrative data handling, such as frequent change of passwords, use of encrypted channels, avoiding usage of publicly open Wi-Fi, etc.

[32] For example, some early efforts have been made by GSMA for mobile data anonymisation in response to Ebola Outbreak: GSMA, GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak (October 2014), available from http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/11/GSMA-Guidelines-on-protecting-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-Ebola-outbreak-_October-2014.pdf.

[33] Privacy by design is a concept developed in the 1990s by Dr. Ann Cavoukian, former Information and Privacy Commissioner for Ontario, Canada. See, Dr. Ann Cavoukian, Privacy by Design: The 7 Foundational Principles, available from https://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf.

Privacy and data protection go hand in hand with data security. The goal is to make privacy and security the default mode of operation for organisations, rather than just a means to the end of regulatory compliance.

7. Transparency

KEY POINTS ON TRANSPARENCY:

CHALLENGES:

- Inadequate openness produces a lack of public trust and accountability
- Releasing too much information may cause harm and threaten the right to privacy

RECOMMENDATIONS:

- Develop strategies for data and information release, especially in sensitive contexts
- Consider risks and harms in determining adequacy of openness

Using data for the public good necessitates transparency about data collection and use. Transparency is the key to building trust, awareness, and accountability in any data-driven project.

Impacts of data use for humanitarian and development causes is highly contextual and may be quite sensitive. Releasing too much information may be detrimental to the beneficiaries. International humanitarian and development organisations are mandated to protect and act in the best interests of their beneficiaries. This also implies a principle of "do no harm," whether this may be caused by using the data for the project's purposes or simply by disclosing information about a project. Every project manager should develop a strategy with regard to data and information release. The strategy should consider what is an adequate level of transparency. A risks, harms, and benefits assessment must play a key role in determining the level of transparency.

8. Private – Public Data Collaborations

There are a number of successful examples of data collaborations between the private sector and public organisations seeking to leverage data.[34] However, data sharing is not a standard practice in data-driven analysis, and companies tend to shift from data sharing to real-time data analytics and sharing of insights only. Many obstacles impede private-public collaborations. These include risks to privacy that may result in reputation damage and liability costs, the lack of unified standards for anonymisation and a fragmented regulatory landscape. A lack of appreciation for the incentives (monetary, operational or otherwise) for buy-in from the corporate sector, as well as a lack of evidence of successful case studies are also challenges for establishing a common way for private-public data collaborations. All of the existing uncertainties in the data privacy and ethics field lead to uncertainty over the success and efficient implementation of private-public data collaborations.

Additionally, so long as there is no recognised and accepted framework for private-public data collaborations for development and humanitarian causes that is recognised by the private sector and regulatory community, private-public sector collaborations will face slow progress.[35]

Some suggest that the international humanitarian and development community can play a neutral role as facilitator in identifying solutions that are acceptable to all parties, by creating a safe space framework limited to public–private data collaborations in the context of humanitarian and development response. Developing model clauses for data handling that cannot be changed once accepted by the key stakeholders would be one of the elements of such a framework. Capacity building within countries is another important process in ensuring that data collaborations are done responsibly and do not take unfair advantage of the gaps in regulation.

[34] Orange Telecom hosted the "Data for Development" Challenge in which datasets of de-identified and aggregated data were made available for research teams to utilize for conducting research projects related to development issues (ranging from transportation, to health, to agriculture) in the countries of interest (Ivory Coast 2013; Senegal 2014). In this release the data curators note the value that both longitudinal data and data with a high spatial accuracy offer REF. So as not to preclude possible lines of scientific enquiry, the data release included a sample of 50,000 users with high spatial resolution over a 2-week period and a sample of 50,000 users with a spatial resolution limited to the sub-prefecture level. The second round of the challenge also offered analogous datasets (see http://www.d4d.orange.com/en/Accueil)
In 2013, BBVA hosted "Innova Challenge Big Data", a contest to which BBVA invited developers from all over the world to create applications, services and content based on anonymous card transaction data (anonymous and aggregate data from transactions performed with cards or at BBVA point-of-sale terminals, with timestamps). For more information, see https://info.bbva.com/en/information/bbva/.
In 2013, Telecom Italia hosted a "Big Data Challenge" in which the company made a dataset of mobile phone data (millions of de-identified and geo-referenced records of calls from Milan and Trento, covering the period of November to December 2013) available for researchers and app developers to work with. Complementary datasets were also made available including utility usage and social media messages. This data has since been made completely open (see, http://www.telecomitalia.com/tit/en/bigdatachallenge.html).

[35] See, World Economic Forum, Data-Driven Development - Pathways for Progress, (January 2015), available from http://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf.

KEY POINTS ON PRIVATE-PUBLIC DATA COLLABORATIONS:

CHALLENGES:

- Private-public data sharing is not standard practice and there may be alternative methods for achieving positive impacts through the use of data, without data sharing
- Lack of understanding of the incentives and sustainable mechanisms and therefore, lack of successful case studies for private-public data collaborations to drive demand
- Many obstacles to successful data collaborations are due to the gaps in data privacy and data protection regulations and standards (many of which were discussed in this report)
- Majority of existent laws and legal mechanisms do not consider private-public data collaborations in the context of humanitarian and development action

RECOMMENDATIONS:

- Private-public data collaborations in humanitarian and development contexts cannot happen without an inclusive process accounting for the interests of various key stakeholders (private companies, regulators and humanitarian and development decision-makers)
- Work towards a framework for public-private data collaboration specific to international development and humanitarian contexts
- Develop model clauses for data access and use in humanitarian and development contexts acknowledged and approved by the key stakeholders
- Work towards closing gaps in data privacy and data protection
- Build better understanding between public and private sector stakeholders on the incentives and develop sustainable mechanisms for such collaborations
- Build capacity within countries on responsible data collaborations; leverage practitioner expertise and support knowledge-sharing
- Localise analytics to reduce the burden of managing cross-border flows

Conclusion

In April 2015, the UN Secretary-General announced, "The number of people in need of humanitarian assistance around the world has doubled in just ten years."[36] As the challenges facing the development and humanitarian sectors increase, it is crucial that practitioners embrace the use of technology and big data to improve decision-making. At the same time, the risks that accompany data use must be accounted for and addressed; and central to this effort is an on-going and inclusive dialogue between stakeholders.

The Group discussed possible ways forward including a new framework for the development and humanitarian contexts, a framework and terms and conditions for public-private data initiatives, comprehensive data aggregation guidelines, a big data classification scheme, and the Risks, Harms and Benefits Assessment tool, which will help mitigate risks and harms while maximising the benefits of a project. Global Pulse, with the assistance from the PAG, will continue to contribute to these efforts. The above-mentioned tools will be crucial in tackling issues of privacy, data security, transparency and effective humanitarian and development response.

However, as with every international multi-stakeholder initiative, consensus can only be achieved through cooperative information-sharing and an active exchange between privacy professionals, technology experts, researchers and professionals in the development and humanitarian sectors about what works and what methods should not be repeated. Such efforts would consolidate the community of practice and expand the network of practitioners. Development and humanitarian organisations can fulfill the role of catalyst in supporting the work of a broader network of responsible data professionals across the public and private sectors. The dialogue fostered within the multi-expert PAG aims to encourage international harmonisation across the humanitarian and development data ecosystems for the successful and responsible achievement of the SDGs.

For more information on the Privacy Advisory Group, please visit:

http://www.unglobalpulse.org/data-privacy-advisory-group

For more information on UN Global Pulse, please visit:

http://www.unglobalpulse.org/

[36] UN OCHA, Global Humanitarian Overview 2016: A Consolidated Appeal to Support People Affected by Disaster and Conflict, available from https://docs.unocha.org/sites/dms/Documents/GHO-2016.pdf, noting that "[a]t the beginning of 2015 we were aiming to reach 57.5 million people in 22 countries needing assistance."