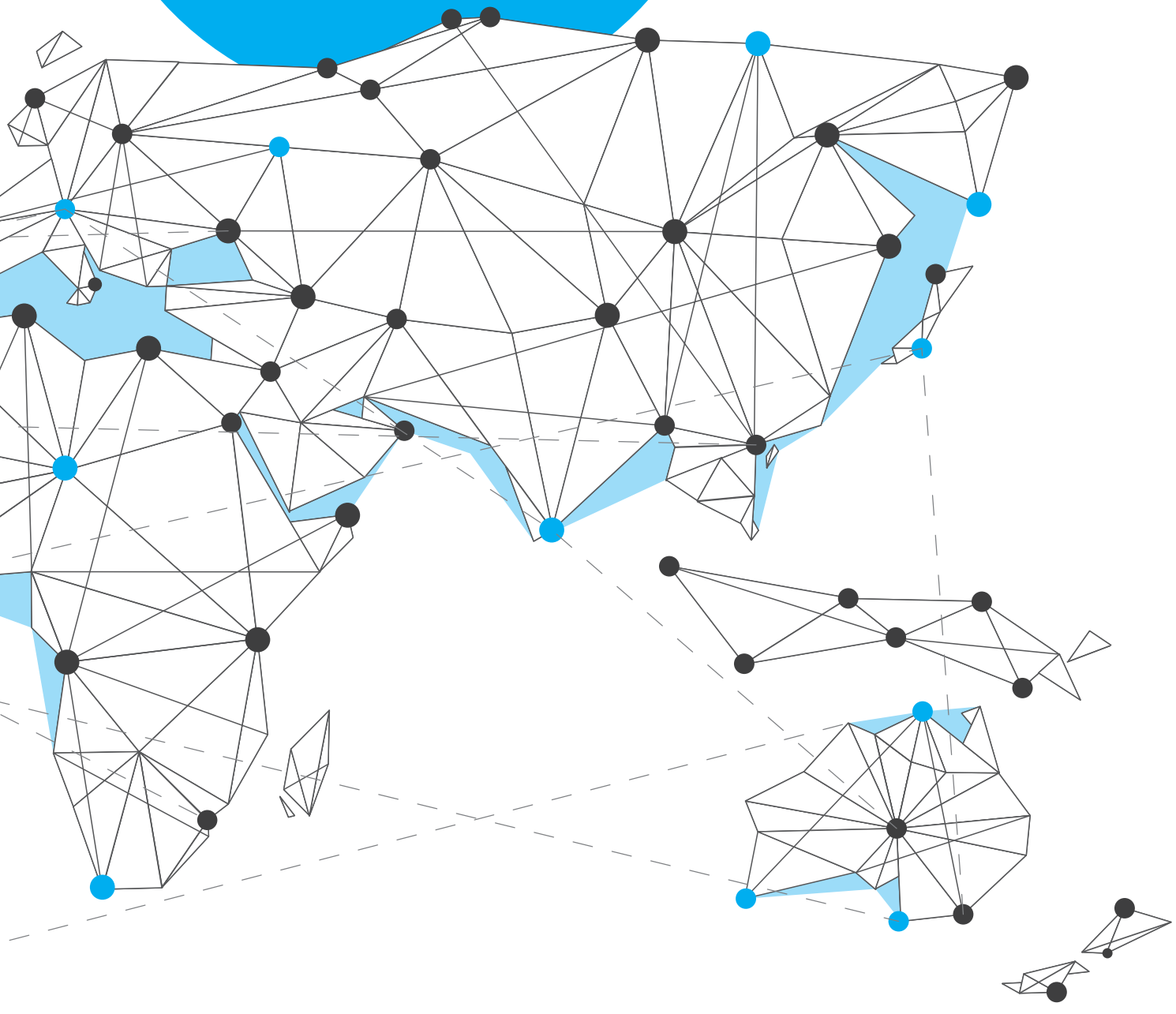
DATA AND INTERNATIONAL ORGANISATIONS:

Navigating cross-sectoral data challenges

Geneva Internet Platform



February 2018

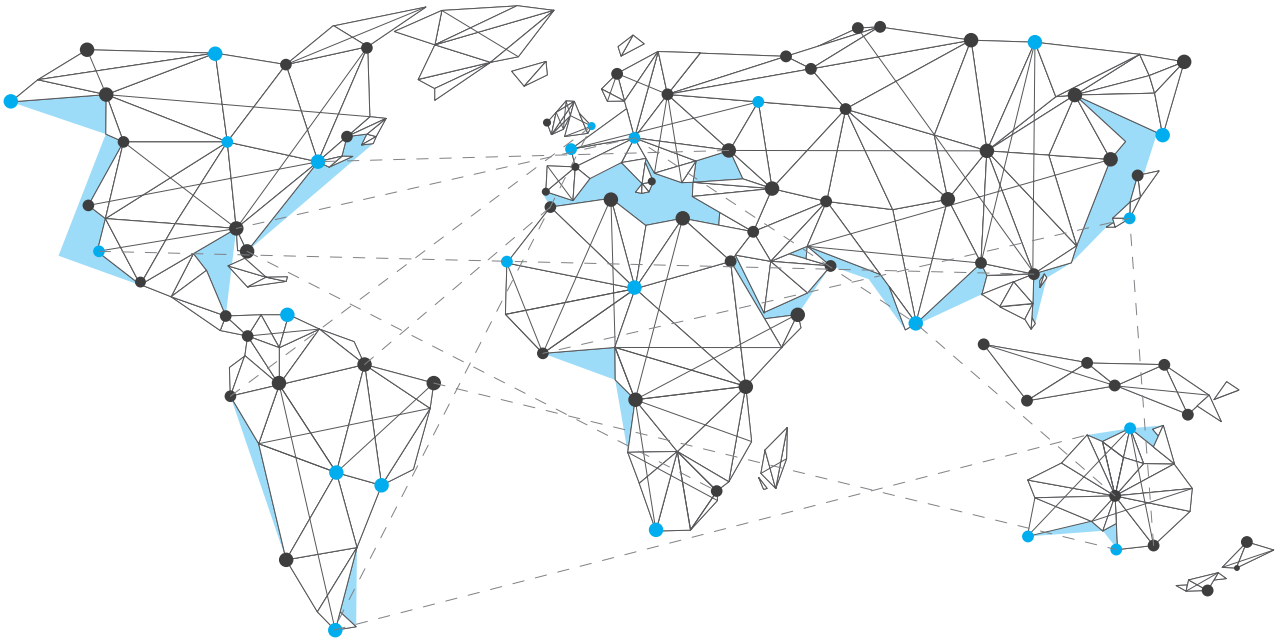
Introduction

Modern society is characterised by a spike in the amount of data generated, which is available in a growing number of forms and formats. From big data to open data and from crowdsourced data to social media, many sectors have taken advantage of the opportunity to better inform their work, guided by this rising tide of the big data sea. The potential of this 'data revolution' also reaches the shores of international organisations (IOs), which face the challenge of continuously adapting to this ever-changing data environment, to be able to integrate the potential of new forms of data in their operations and mandates.

The Sustainable Development Goals (SDGs) have increased calls to make better use of new forms of data, especially bearing in mind the pressing needs to monitor the 232 SDG indicators at disaggregated levels. In the humanitarian sector, there is a growing realisation that new forms of data - such as mobile, GPS, or crowdsourced data - can enhance needs assessments and early warning systems, in order to deliver better target humanitarian aid. Climate and environmental developments can be better monitored through the numerous new sensors that have appeared on digital devices, and health emergencies and epidemics can be closely scrutinised by identifying the movement of populations.

Although there is a growing awareness of the potential of the data revolution, this document focuses on tackling some of the key challenges that arise in practice within this new, digital environment, characterised by new forms and sizes of data, new demands for analysis and data storage, new skills and changing legal contexts. As observed from Geneva, each IO faces particular data-driven challenges and opportunities related to their scope of work. Yet, all IOs ultimately face similar tests in their digitization efforts – thus sharing approaches and issues on which they could learn from each other is beneficial. As the operational centre of the UN, Geneva hosts a total of 33 international organisations, 23 of which with headquarter agreements here. The institutions hosted here range from specialised bodies of the UN to the European Organization for Nuclear Research (CERN) and the World Trade Organization.

The reflections in this paper have been informed by the Data Talks, which converge representatives from IOs in Geneva with the aim of sharing best practices, experiences, and lessons learned about data across silos and institutions. In addition, the document has been enriched with insights from data-related sessions at the 12th Internet Governance Forum, which took place from 18-21 December in Geneva.



Acknowledgments: *This paper has been made possible by the valuable contribution of the participants at the Data Talks and the Internet Governance Forum, in particular the representatives from CERN, European Broadcasting Union, International Labour Organization, International Committee of the Red Cross, International Federation of the Red Cross and Red Crescent Movements, International Institute for Sustainable Development, International Organization for Migration, UN Economic Commission for Europe, UN Environment, UN International Computing Centre, UN Joint Inspection Unit, UN Office for the Coordination of Humanitarian Affairs, and the World Meteorological Organization.*

Key issues

Engagement with new forms of data can give rise to a number of challenges. Below, we list six issues to keep in mind when using new forms of data to advance the activities of IOs: data analysis, data storage, data sharing, data protection, data regulation, and data capacity. It is important to keep in mind that these issues are not clearly delineated, and the real challenges arise in their interaction.

I. Data analysis: Data harmonisation in an increasingly unstructured world

While the generation of new forms of data is impressive and provides great opportunities, the data itself is of little value if it is not combined with sound analysis. Yet analysis can be incredibly complex, as it demands both new skills and new methodologies. Looking at the rising demands for data collection for the SDGs, the question that arises is: how to harmonise a dataset if the data derives from many different sources and in many different formats?

New and diverse data sources will require increased standardisation. At the same time, standards need to remain simple, easy to use, unambiguous, and relatively flexible, not to stifle the innovative potential of data. In this context, there are many obstacles to standardisation: some are technical, others political. For example, there are interests in having 'your' practice standardised across a sector. In addition, it can be difficult to ensure compliance with standards, especially if they are voluntary. Yet, these obstacles might be worth engaging with, as standards can provide many benefits: they ensure the quality of data, they promote the exchange of data, and they allow for the comparability of data.

In the context of the SDGs, IOs are tasked to take care of the harmonisation and international comparability of the 232 SDG indicators. The initial data is first collected at the national level, and then reported to UN agencies. Yet, there are sometimes different measures for one indicator, and there is often a lack of transparency in how countries calculate indicators, which makes it difficult for IOs to carry out their task. Standards and definitions could assist, although they are often plagued with the afore-mentioned political and technical obstacles; and countries might want to stick to preferred methodologies and definitions that work in their favour. In addition, the harmonisation of SDG data is not only an issue of incomparability, the problem often lies in multiple areas, such as different access to technology and available resources, as well as limited data infrastructure and oversight. Among other structural issues, these factors hinder the capacity of national governments to fulfill their reporting commitments. This challenge might be mitigated through providing assistance and capacity building while integrating local and national needs.

Key insights:

- The current pressure on IOs to better monitor global developments, from the SDGs to emergencies and climate change, coupled with the rise of unstructured data, has generated challenges to better harmonise and standardise data
- It is important to start from an agreed-upon definition and scrutinise methodologies and data collection processes in advance in order to ensure the compatibility of data
- Standards are important to streamline data collection analyses, but they should be developed bearing in mind the political sensitivities that could arise. Without such considerations, standards might not be implemented by all parties, and quickly become ineffective
- In addition, IOs in charge of harmonising data should pay attention to the capacity and the resources of the entities from which the data is collected, and capacity building and awareness-raising among these entities could further enhance the compatibility of the data

Take a look...

- UNECE develops standards on data harmonisation in the area of trade facilitation through the UN/CEFAT unit
- WMO develops standards for the collection of meteorological and hydrological data
- The IFRC has experience in harmonising data through their Federation-wide databank

II. Data storage: Finding leakproof spaces for data lakes

The collection of a growing amount of data raises the question of how this data can safely and effectively be stored. With increased storage needs, some have started to look into the possibilities of the cloud. Introducing or further elaborating policies on cloud computing in the UN system has the potential to make the system's operations

more flexible and scalable. Some internal cloud solutions are being developed for the UN system, particularly by the UN International Computing Center, yet there have been calls for the UN to increasingly connect to the public cloud. There are many opportunities related to connecting to the public cloud, which range from a quicker provision of services to standardisation, innovation and cost-effectiveness.

Yet cloud computing engenders questions related to the control over data, risk management, and the protection of privacy and confidentiality, especially when connecting to the public cloud. There are risks that the cloud provider could misuse information, and that state actors will use legal means to access public cloud services. In fact, IOs might feel concerned about relinquishing control over their data when connecting to external cloud providers. Legal issues may also arise, as there is a complex interplay between data protection laws in the host country, the immunities of the IOs, and the jurisdiction over the IO's server.

Key insights:

- The increased collection of digital data in IOs requires updated and reliable solutions for their storage, which could include cloud computing
- IOs will only choose cloud computing if there is sufficient trust between them and the cloud provider, especially on the protection and integrity of the data
- Trust can be enhanced through partnerships that include clear commitments for data security and is clear on data ownership and risks
- Legal uncertainties might arise, especially with the changing legal framework (see section V). This changing legal framework should be assessed from the outset, to better understand what is and what is not possible for cloud computing and partnerships

Take a look...

- UN JIU is conducting research on cloud computing in the UN system
- UNICC is working on cloud integration and support in the UN system, and they assess, deploy, manage and assist with private and public cloud environments
- Several IOs, including the IOM, are looking into the legal framework of cloud computing

III. Data sharing: The promises and perils of open and shared data

There is an extensive amount of data 'out there', ready to be analysed in the private or public sector or out of cross-sectoral partnerships. Yet this data can often only be captured and utilised by entering into data-sharing agreements with other organisations. To avoid duplication and best make use of what is available, data sharing will be an important element of an IO engagement with data. At the same time, it might be difficult to obtain data that is held by other organisations and other sectors. Difficulties arise due to data protection concerns, an uncertain legal framework, a lack of trust, and the reluctance to share strategic assets.

Besides data sharing across international organisations, data is increasingly collected, stored and deciphered in partnership with industry players, recognising the wealth of data that is collected by companies. These partnerships can be extremely valuable to gain new insights into populations, society, and the environment, for example based on social media patterns, GPS signals, or sensors. Yet these partnerships are challenging to negotiate, requiring extensive trust between two or more parties to properly manage the data, and an awareness of the interest of the parties to engage in such a partnerships.

In addition to bilateral data sharing, making datasets publicly available raises many opportunities, as organisations across the sector will be able to enhance their understanding and improve their activities with the available data in mind. Yet, while sharing between individual organisations can be a challenge, publishing datasets openly can further increase privacy risks and raises questions of ownership, responsibility and data management once the data is made available. Organisations can be incentivised to share data by providing services as data visualisation or standardisation. Ultimately, joint efforts in sharing data following privacy-enhancing guidelines are essential to explore the full potential of the data revolution.

Key insights:

- Data sharing is essential to capture the benefits of the data revolution and avoid the duplication of efforts
- Data sharing will only be effective with sufficient trust between the parties and clarity on interests and incentives
- Any data-sharing agreement needs to bear in mind data protection concerns, and some data might be too sensitive to share (see section IV)
- Open data can further enhance the potential of the data revolution, but only if the data is easily organised, findable, and user-friendly. Open data platforms should be clear from the outset on the ownership and responsibility of the data that is shared

Take a look...

- UN OCHA's Humanitarian Data Exchange and Centre for Humanitarian Data for best practices on open data and data sharing in the humanitarian sector
- Humanitarian OpenStreetMap for an example of capturing the potential of crowdsourced data to map vulnerable areas
- CERN's Open Data portal and how the organisation established standards in data preservation and publicly available data

IV. Data protection: Upholding the 'do no harm' principle in the data age

One of the most pressing issues identified during the Data Talks discussions was data protection. While data provides ample opportunities to better identify needs and target interventions, this information can also be misused if it falls into the wrong hands. Potential consequences include privacy breaches, discrimination, and even risks to human safety, especially in conflict-prone environments. Ensuring data protection is therefore paramount, especially for humanitarian actors, and their duty to extend the 'do no harm' principle to the consequences of mishandling of personal data. In addition, data breaches in IOs would be detrimental to the public trust they enjoy from the populations they serve. Any collection and use of personal data therefore needs to be carried out in accordance with international data protection standards and respect for people's privacy. An additional challenge is that there is no possibility for testing and failing, as personal data explicitly deals with living individuals, rather than applications or environmental factors. For this reason, it is important for organisations to conduct privacy impact assessments.

Not all data is personal data; yet, even data that does not contain personal identifiers can obstruct collective privacy. With the proliferation of data, even data that is anonymised could sometimes be re-identified through the combination of different datasets. In addition, even data that is not collected at the individual level (such as satellite images) could put a community at risk. 'Community identifiable information' can be used to identify or monitor a geographic, ethnic, religious, economic, or political group. While this is important for needs assessment and the collection of disaggregated data, it can be misused by actors that are on a mission to repress and discriminate. In addition, there are important challenges related to consent, which is needed for any use of personal data. Yet, what constitutes informed consent? What if providing consent is a prerequisite for the provision of services, such as the distribution of food or medical services; do individuals really have a choice?

Not all data protection concerns relate to personal or humanitarian data. For example, organisations can be concerned with data integrity. The manipulation of data by third actors, or the distribution of false data could potentially cause serious harm. In addition, even though the term 'data protection' seems to be rather well-established by now, there are still uncertainties and grey zones, especially with the continuous evolution of digital technology. For example, are IP addresses personal data in every circumstance? How to distinguish between data generated about employees when it becomes increasingly difficult to distinguish between their public and private use of digital tools? How does data protection relate to machine-to-machine communication, self-learning algorithms and the handling of metadata?

Data protection dilemmas

The most pressing challenges often arise with conflicting interests and unclear priorities. This is also the case with data protection concerns. They are often in conflict with other ideals, such as the timely sharing of data in case of emergencies, or the preservation of data to maintain institutional memory.

Dilemma #1: Data protection vs data sharing

With on the one hand encouragements to share data across IOs, and on the other hand calls to take no risks on data protection, it can be difficult to strike the right balance between the benefits and risks of data. Generally, data should be shared according to the sensitivity of the data provided. UN OCHA recently established the Centre for Humanitarian Data, which analyses this challenge and is developing a framework for data sharing in relation to the sensitivity of the data.

Whether and how data is shared depends on both the sensitivity of the data and the actor the data is shared with. It is important to keep in mind the incentives of the counterpart, the mechanisms of data protection of the counterpart, the need to inform data subjects of a potential transfer of their data to third parties, and the necessity of the data transfer.

Dilemma #2: Data protection vs data preservation

Data preservation will be one of the most pressing issues in the times to come, yet it is one of the most overlooked challenges in digitisation. While traditional information has usually been stored in physical archives, today, almost all information is stored digitally, in different formats – that may or may not be compatible with newer technology - and not always neatly organised. If institutional memory is at risk of not being available one day, how do we preserve it in the long-term? How valuable is digitally-stored data if it cannot be easily retrieved?

At the same time, proper data protection guidelines require that data is not stored for longer than necessary. How to mitigate these two requirements? Again, there are no 'either-or' solutions. Rather, they can be found along a spectrum of sensitivity: personal identifiable information should always have a limited storage time, determined at the outset, while aggregated and less sensitive data might be stored for a longer timeframe. In addition, the sensitivity of the data should be matched with the rights storing solution. Sensitive data should be kept as confidential and secure as possible, even within an organisation, while other types of data might better be stored in places that are more publicly accessible, to enhance the findability and usability of the data.

Key insights:

- Breaches of digital data can put populations at physical risk. Not properly using and protecting data might do more harm than good
- Any use of personal data needs to be accompanied by privacy impact assessments and matched against the legal framework
- Assessments should include the risks related to the re-identification of individuals and the protection of community identifiable information
- Principles of data integrity and risks to manipulation should be incorporated in data security strategies

Take a look...

- ICRC has published the *Handbook on Data Protection in Humanitarian Action*, which is an excellent guide on how data protection measures can be implemented by humanitarian organisations
- UN Global Pulse provides data privacy and data protection principles in dealing with big data
- The UN Development Group has published *Data privacy, ethics and protection*, which provides guidance on big data for the SDGs

V. Data regulation: Navigating the GDPR disruption

The shifting role of data in society has been noticed by policy-makers. In particular, the privacy risks that arise with the increased generation of personal data has caused politicians to think about whether and how data could be better protected through regulation. In this context, the European Union has created the General Data Protection Regulation (GDPR), to enter into force in May 2018, which is likely to revolutionise the way in which organisations (both in the public and private sector) are able to manage data. The regulation applies to the data of EU citizens, even if the data resides on servers outside of the EU.

Many IOs, especially those that are collecting an extensive amount of personal data, will also be affected by this changing legal framework. Responding to the GDPR will require the collaboration among both legal and technical departments within IOs, especially those related to technology or innovation and legal departments. Generally, IOs have privileges and immunities, which means that the data that they are processing is not subject to the jurisdiction of the country in which the IO is based. Yet sometimes, these privileges and immunities are questioned, especially if those who are affected by poor protection of data are not provided with an effective remedy when their data is breached.

One of the main areas that is affected by the regulation is data sharing. Under the GDPR, transfers of personal data of EU citizens to IOs can only take place after a so-called adequacy decision (a decision by the European commission that the organisation ensures an adequate level of protection, which no IO has passed so far). Without the adequacy decision, individual transfers can still take place if they are subject to appropriate safeguards or if they are conducted for important reasons of public interest. As these concepts are currently undefined, it remains to be seen how practice develops after 25 May 2018.

Of course, there is still a large number of grey zones. What happens when data is outsourced for processing outside of an IO? Will the immunities of the IO be attached to the data, or will the data be subject to the domestic laws of the processor? What happens when data is stored in one jurisdiction and processed in another? While these unclarities are likely to be resolved by policy-makers, it is important for IOs to stay up to date of this changing legal framework.

Key insights:

- Legal departments at IOs, especially those dealing with large amounts of personal data, will be under increased pressure to keep up with the changing legal environment
- Legal experts across IOs are encouraged to work together to share insights on how the legal requirements apply to their sector. In this way, challenges common to a number of IOs can be faced more effectively.
- IOs should closely follow developments related to the GDPR, especially after it enters into force, to better understand the evolution of the regulation in practice and to be able to clarify some of its grey areas

Take a look...

- The topic of data protection within international organizations is discussed during regular workshops organised by the European Data Protection Supervisor. The latest one was co-organised by the IOM and included the role of the GDPR
- Several organisations, including the ICRC, UPU and IOM, are developing specific legal expertise on the transition towards the GDPR.

VI. Data capacity: Updating the IO to new data needs

What are the data skills that should be acquired by the staff of IOs? What are the capacity building needs to best capture the potential of data, and how can these needs be met?

A first step is to raise a general awareness on the growing possibilities of data and technologies, and a practical understanding of what is and is not possible. This will also help those who are working with data analysts to ask the right questions based on the data that is available. An awareness of the biases of data is also important in order to avoid data misrepresentation. Problems will arise when no one is able to question the validity and objectivity of data.

Ultimately, benefiting from data goes beyond technical skills of collecting and analysing data, to include an understanding of the legal framework, the policy context, and the ability to communicate the outcomes of data analyses. Different staff members will have different expertise, and it is in their smooth collaboration that the potential of data can be captured.

Besides know-how on analysing data, there is also a need for awareness on how to manage data responsibly to minimise data protection risks. Some organisations choose to implement compulsory data training to be aware of the risks of data security, as some sort of 'first aid kit' or safety check for all employees that engage with data. While everyone working with data needs to be aware, this holds especially for middle management, which is usually held accountable for the internal control of data.

Ultimately adapting the organisation to a new environment, issues such as digitisation, cloud computing and data management could best be addressed in a comprehensive manner across an organisation, rather than in a separate, additional division. This way, efficiency within existing bodies can be enhanced, instead of creating new units and adding organisational layers. Another model is to create a small unit tasked with innovation, that has contact points, or 'data champions' across the wide spectrum of the organisation's units and divisions.

Adopting an organisation-wide perspective on data management could enhance learning and awareness across the organisation's departments. For example, legal and technical departments in IOs are often relatively detached, leading to a lack of understanding of issues such as data security, privacy and preservation among legal experts, and a lack of the legal environment and requirements among those working on the technical aspects of data.

Ultimately, it could be fruitful to build a 'data culture' across the organisation, using the knowledge and skills that people often already possess in different departments, and sharing this expertise across departments by identifying leaders with know-how.

Key insights:

- Data-related skills and capacity building needs come in many forms, from data analysis to protection, and from the legal to the policy context
- Basic data protection awareness, as well as awareness on the possibilities, constraints and biases of data, should be promoted for everyone who engages with data
- More in-depth skills should be available in a way that it can be used across units. As these skills are diverse, they will likely be found in a team, rather than in one perfect data scientist
- Learning across the organisation, and knowledge exchange between departments, should be encouraged to create a data-conducive culture.

Conclusion

IOs are well on their way to adapt to the ever-changing data environment. Expertise can be found across the sector, yet is often still confined to niches within organisations. Regular encounters across IOs on these issues can help advance the understanding of these new and complex issues, and promote an awareness of where expertise can be found.

Ultimately, tackling these challenges effectively can be accelerated through discussions across IOs, as well as with the engagement of other sectors, most prominently the private sector, which traditionally has more experience with similar data challenges, and could provide valuable insights. Understanding how challenges are tackled in other industries could help provide a solution.