Connecting Data Communities: Introducing the Joined-Up Data Maturity Assessment

Introduction

Background and context to the Joined-Up Data Maturity Assessment

Interoperability is the ability to join up data streams in ways that allow both machines and humans to understand and contextualize the data they contain. It's thanks to interoperability that you can use Microsoft Office tools on an Apple iPad, send and receive emails from a range of email providers, or collaborate remotely with colleagues on the same spreadsheet using a multitude of devices. Within the data revolution for sustainable development, interoperability enables the overlaying of earth observation, administrative, or mobile-derived data with statistical data, helping to both achieve and monitor progress towards the Sustainable Development Goals (SDGs). Ultimately, interoperability allows decision-makers to join up and contextualize a variety of data in ways that are most useful to them.

Since 2017, the Global Partnership for Sustainable Development Data (the Global Partnership) and the United Nations Statistics Division have jointly managed and maintained a Collaborative on SDG Data Interoperability (the Collaborative). In 2018, the Collaborative published two documents, "Using Data to Join Up Development Efforts" (the Brief) and "Data Interoperability: A Practitioner's Guide to Joining Up Data in the Development Sector" (the Guide). Taken together, these two documents provide an overview of what interoperability means for statisticians, development professionals who manage data, and IT specialists working in sustainable development. Between them, the Brief and the Guide define and explore five pillars of interoperability identified by the Collaborative:

- Data management, governance, and interoperability.
- Canonical data and metadata models.
- Classifications and vocabularies.
- Standardized interfaces.
- Implementation of linked-data approaches.

This document builds on the foundations laid by the Guide and Brief and introduces the Joined-Up Data Maturity Assessment (the Maturity Assessment) developed by the Collaborative in 2019–2020. (See Annex A for the full Maturity Assessment.) The Guide and the Brief were

endorsed by the UN Statistical Commission at its 50th session in 2019 and are now curated by the Commission's Working Group on Open Data.

How to use the Joined-Up Data Maturity Assessment

The Maturity Assessment introduced in this document is designed to be used by official statisticians and professionals who regularly produce, collate, or use diverse data sets in the sustainable development sector. It builds on the concepts defined and established in the Brief and Guide and is designed for use in strategic planning and review.

The Maturity Assessment is the product of an analysis of several commonly used **maturity models**, including among others:

- IBM's "Information Governance Model" (IBM 2007).
- Gartner's "Maturity Model for Data and Analytics" (Gartner 2018).
- CARE USA's "Responsible Data Maturity Model for Development and Humanitarian Organizations" (CARE USA 2019).
- Open Data Institute's "Open Data Maturity Model" (ODI 2015).

The Maturity Assessment draws on elements from all of these models and tries to create an adaptable and flexible tool that can be used by national statistical offices (NSOs) and other entities that control or process data in the development sector.

The Maturity Assessment (see Annex A) has three components: layers of interoperability, dimensions, and levels of maturity. It starts with the four layers of interoperability elaborated in the Brief and Guide, as originally conceptualized by John Palfrey and Urs Gasser in *Interop: The Promise and Perils of Highly Interconnected Systems* (Palfrey and Gasser 2012). These layers are:

- Organizational interoperability.
- Human interoperability.
- Data interoperability.
- Technological interoperability.

These four layers have a total of 19 dimensions, or data management functions that correspond to each layer. The Maturity Assessment then identifies characteristics that are found in each dimension that indicate an organization's level of maturity, starting with *undefined,* and increasing in maturity to the *emerging, learning, building,* or *consolidating* levels. These levels of maturity are not mutually exclusive; it is entirely plausible for a user to be *consolidating* progress in one dimension, but only *emerging* in others.

The Maturity Assessment is designed to be used either holistically or in part. Some of the circumstances in which this tool could be used include the following:

- While producing a new data management framework or strategy in an NSO, international organization, or government department that produces, collates, or uses diverse data sets.
- During a strategic review or update of an existing data management framework or operational business process, from the development of a national statistical development strategy to the adoption of the generic statistical business process model.
- As part of a broader data governance maturity assessment being undertaken by an international organization that produces and manages data to strengthen its data management capacity.

Users who wish to find out more about the value of interoperability and its relevance to sustainable development processes should consult the Brief and Guide before using this document.

The remainder of this document is divided into four chapters, each exploring dimensions in the Maturity Assessment under one of the four layers of interoperability. Each dimension is explained and linked to relevant sections of this document and parts of the Guide. Other resources and examples are provided as required. The full Maturity Assessment can be found in Annex A and a Glossary is located in Annex B.

Chapter 1: Organizational interoperability

Organizational interoperability can be interpreted in two ways. On the one hand, it can refer to how **data governance** and **data management** functions that touch upon interoperability are distributed across an organization.[1] Alternatively, it can be viewed as the ways in which whole organizations, or individual departments within them, engage with the broader data ecosystem or national statistical system to decide the degree of interoperability they would like to achieve between their collective data assets. The Maturity Assessment identifies seven dimensions of organizational interoperability. These dimensions and their associated characteristics for each level of maturity are described in Tables 1.1 to 1.7.

1.1 Strategic objectives

Table 1.1: Organizational interoperability layer: Strategic objectives.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
Interoperability is not recognized as a strategic objective	The ability to join up data is recognized but it is not explicitly identified as a strategic objective	Interoperability is identified as a strategic objective in an organization's technical units, but not outside of them	The need to join up data across systems is recognized as a strategic objective in an organizational data strategy The value of standards and robust data governance is recognized	The strategic value that joined-up data can bring to decision-making is recognized in organizational strategies Interoperability forms part of an organization's external engagement strategy with other data producers and users

Data management can loosely be defined as the development, implementation, and monitoring of strategies and plans that allow the value of data to be safely unleashed. Therefore, it is a central function for organizations that manage data to develop a data management strategy that sets out choices and decisions that chart a course of action (DAMA International 2017, pp. 17 and 31).

For organizations that manage, collate, or process diverse data sets, interoperability should be a key tenet in their data management strategy. Strategic choices about what data sets should be made interoperable, the degree to which they should be interoperable[2], the desired degree of interoperability in relation to the rest of the data ecosystem, and similar considerations need to be captured in a data management strategy.

[1] For the purposes of this document, we have followed the definitions for data governance and management in the "Data Management Body of Knowledge" (DAMA International 2017). Definitions for the two terms are in Annex B, and an overview of the concept of data governance is set out in the Guide (Gonzalez and Orrell 2018), p. 19.

[2] See p. 11 of the Guide (Gonzalez and Orrell 2018) for a discussion on optimal interoperability.

In recent years, the value of interoperability as a strategic objective in and of itself has been increasingly recognized in the field of official statistics. For instance, since 2018, Mexico's NSO, the National Institute of Statistics and Geography (INEGI), has been working on an information governance and architecture strategy. Interoperability is recognized as one of the strategy's three pillars, alongside data quality and security and confidentiality.

When setting strategic objectives relating to interoperability, it is important to allocate budgetary and human resources for them in accompanying implementation plans.

1.2 Leadership and management

Table 1.2: Organizational interoperability layer: Leadership and management dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no defined leadership over interoperability issues	Ad hoc leadership on interoperability issues emerges organically but is not coordinated	Leadership around interoperability emerges across various technical units but remains fragmented Silos persist	There is a coordinated hierarchy of leadership over interoperability issues Clear functions relating to interoperability are established across an organization	There is a data governance committee or council and it has an explicit mandate to lead on interoperability issues The value of joined-up data is understood by organizational leaders and managers, and is clearly identified as a function in relevant job descriptions

The Leader's Data Manifesto (Data Leaders 2017), produced by a coalition of thought leaders in the data management field, recognizes that

> "data offers enormous untapped potential to create competitive advantage, new wealth and jobs; improve healthcare; keep us all safer; and otherwise improve the human condition [...] still we find no examples of fundamental lasting change without committed leadership and involvement of everyone at all levels of the organization." (Data Leaders 2017)

Leadership is therefore crucial to data management, and by extension is central to both the role of NSOs as data stewards (section 2.1) and ensuring that interoperability is treated as a strategic objective (section 1.1). In order for organizations to realize their data assets' potential, senior management must be engaged in strategic processes that relate to data interoperability and be aware of, and understand, the value that interoperability brings to their organizations. In larger organizations, leadership over interoperability issues is frequently structured at various management levels and manifests in the form of data governance committees or councils, which are also central to oversight and accountability (DAMA International 2017) (section 1.3).

1.3 Oversight and accountability

Table 1.3: Organizational interoperability layer: Oversight and accountability dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no oversight or accountability over interoperability issues	An organization is aware of the need to create accountability chains to oversee how data is being joined up, but is not yet taking steps to create them	Oversight structures emerge across different technical units but are not coordinated or aligned Accountability over how interoperability efforts are undertaken is fragmented and unclear	Oversight and accountability functions are embedded in an organization's strategy and reflected in the leadership structure Fragmented chains start to join up and common standard operating procedures emerge	A clear chain of oversight and accountability flows from an organization's data governance committee or council, down to operational staff Organizational units are clear about their functions relating to interoperability and who they are accountable to

Oversight and accountability structures are central tenets of data governance, and to management functions relating to the setting of strategic objectives (section 1.1), leadership and management (section 1.2), and data stewardship (section 2.1). Ensuring that there is effective oversight and accountability built into organizational processes helps to ensure that the processes of making data interoperable — and therefore integratable, sharable, and accessible to others — are properly managed and maintained. There also needs to be oversight of data sets that may contain data that could reveal an individual's identity or other sensitive attributes about individuals or vulnerable groups (sections 2.2 and 3.5).

Oversight and accountability can be internal or external. Internal oversight and accountability can be structured around data governance councils and committees[3], and can be built into individual staff members' job descriptions. External accountability includes ensuring that an organization's actions comply with any applicable legal regulations, binding ethical guidelines, and partners' expectations.

[3] See pp. 18 and 19 of the Guide (Gonzalez and Orrell 2018) for a discussion on data governance councils and committees.

1.4 Legal compliance

Table 1.4: Organizational interoperability layer: Legal compliance dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no awareness around any applicable legal obligations relating to joining up interoperable data	There is a general understanding that the actions interoperability facilitates — data transmission, sharing, and use — might be regulated, but it is unclear how	Active steps are taken to better understand legal compliance requirements around data retention, transmission, sharing, and use, and make them available to data users	Compliance with applicable laws on data sharing, transmission, and use is embedded in oversight and accountability functions, and is reflected in an organization's data strategy, which is published online	An organization's data transmission, sharing, and use activities fully comply with applicable laws and sometimes exceed legal standards

Legal and regulatory challenges can arise for interoperability when it comes to the sharing and integration of data assets between organizations and across national borders. Laws set enforceable boundaries of what is acceptable conduct and what is not. In some instances, they govern *how* data can be shared (for instance, laws that regulate and set standards for data reporting, security, and protection) and in others, they govern *what* data can, or more often *cannot*, be shared and integrated (for example, data protection and privacy laws).[4] Legal and regulatory issues are therefore closely related both to people's rights over data (section 2.2) and to organizations' duties to protect data (section 3.5).

International data transmission can be especially complex, as the European Court of Justice case *Data Protection Commissioner v Facebook Ireland and Maximillian Schrems* (ECJ 2020) highlights. In this case, the court found that the primary mechanism that enabled the sharing of personal data between the European Union and United States of America was deficient because it failed to guarantee European Union citizens' data privacy and protection. This has a clear knock-on effect on the ability of entities in these two jurisdictions to integrate data sets containing personal data. This case may require that these entities limit the ability of certain types of data in their control to be interoperable with other systems.

Legal and compliance issues pertaining to data interoperability are complex and multifaceted. They are likely to require expert engagement from lawyers on a case-by-case basis to establish the possible effects of particular laws in specific jurisdictions on an organization's desire to enable interoperability of data sets containing personal or sensitive data. If a project involves cross-border transmission of personal or sensitive data, legal advice will likely be needed at the design phase to ensure that activities comply with any applicable laws.

[4] See p. 19 and Annex B of the Guide (Gonzalez and Orrell 2018) for further information on legal and regulatory frameworks.

1.5 Data ethics

Table 1.5: Organizational interoperability layer: Data ethics dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no awareness of the ethical questions that interoperable data might give rise to	There is a general understanding that joining up data may sometimes give rise to ethical questions, but it is unclear how	Active steps are taken to better understand the ethical impacts that joining up data might have and to understand how they might unintentionally cause harm Rudimentary ethical impact assessments are undertaken on an ad hoc basis	The types of ethical questions that joined-up data might give rise to are understood and appropriately categorized Appropriate steps are taken to minimize harm caused by a breach of ethical standards Ethical impact assessments are routinely undertaken	The risks of harm posed by joined-up data are well understood and ethical reviews are undertaken across the data life cycle to monitor issues and course correct as needed Ethical assessments are published transparently online An organization joins up data only once it has undertaken, and documented, a review of the potential risks of harm it might give rise to, and has taken appropriate steps to mitigate those harms

Interoperability can give rise to ethical dilemmas. These differ from legal issues in that they are likely to relate to questions of equity or what is considered right or wrong in a particular context, as opposed to lawful or unlawful. An ethical interoperability issue might be whether international data processing firms that obtain access to national demographic microdata as part of a development program should have a right to reuse and profit from that data in the future by integrating it with other data sets. Depending on the national laws of a particular jurisdiction, this might be legal, but it still raises ethical questions about the equitable distribution of the value extracted from national data assets.

In some countries, data ethics frameworks are being adopted to establish operational principles and actions that can be taken by public servants to work through ethical dilemmas. The United Kingdom's "Data Ethics Framework" (U.K. Government 2020) sets out concise and clear guidance on data ethics and serves as a useful example to others. Civil society groups such as the Open Data Institute have also produced easy-to-use tools, in this case the "Data Ethics Canvas" (ODI 2019a), to help organizations think through the ethical implications of their work.

Sometimes, legal and ethical issues can merge and the route out of an ethical dilemma might be legal. Keeping the example above in mind, robust licensing regimes that limit how data can be reused and reintegrated might offer a solution to that specific ethical challenge (section 1.7).

It is a good practice to regularly assess the ethical issues relating to specific data sets (including issues that touch on interoperability), document them, and consider steps to mitigate or resolve them.

1.6 Procurement

Table 1.6: Organizational interoperability layer: Procurement dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
An organization is not aware of the impact of the procurement of technical and data solutions on interoperability	There is an emerging understanding of the need to join up data across procurement activities, but there is no coherent approach There is a heavy reliance on outside contractors to fill gaps but no coordination between units on how this is done, often resulting in incompatible data solutions being procured	Coordination across organizational units around procurement of technical solutions begins to materialize but is still not formalized Occasionally, units informally coordinate on the hiring of external contractors to ensure that new data systems are compatible with existing data infrastructure	The procurement of compatible and interoperable data systems across an organization is formalized and coordinated There is a common procurement policy across the organization that requires staff to consider interoperability issues when procuring new systems Reliance on external contractors is strategic and coordinated	An organization integrates the procurement of new interoperable software and data processing services into its data strategy and includes forward looking plans Units strategically plan and think through their common procurement needs and ensure that any new data system or service that is procured is both backwards compatible with existing infrastructure and meets likely future needs

The procurement of IT services, data systems, or other solutions that relate to the processing, storage, or sharing of data all raise important interoperability issues for organizations.

Numerous NSOs and government ministries, departments, and agencies around the world have anecdotally reported that **vendor lock-in** is a serious obstacle to interoperability for them. Vendor lock-in occurs when organizations separately procure different proprietary digital and data solutions as part of different programs of work or donor-sponsored projects without considering existing data architectures or digital infrastructure. This can create siloed information systems that produce systems or data sets that do not interoperate or whose data outputs cannot be integrated with each other.

To prevent vendor lock-in, organizations need to use a coordinated approach and set of common guidelines across departments that consider interoperability when procuring new digital or data systems. Departments across the Government of the United Kingdom, for instance, have adopted a set of Open Standards principles that state: “Government frameworks for IT procurement must: specify the use of open standards for software interoperability, data and document formats, unless there is a clear reason why this is not possible” (U.K. Government 2018). Wherever possible, open-source standards and software packages should be used to minimize the risks associated with vendor lock-in.

1.7 Links to broader data ecosystems

Table 1.7: Organizational interoperability layer: Links to broader data ecosystems dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no awareness of how data is used across a data ecosystem and the role of data interoperability in that	There is an emerging understanding that joined-up data across entities in a data ecosystem can give rise to both opportunities and challenges An organization engages with other parts of the data ecosystem informally and in an ad hoc manner	An organization starts to attach pro forma licensing terms with provisions on data integration to data that it releases, transmits, or shares but does not monitor or engage with data users An organization starts to document the data that it receives from other organizations An organization starts to document and coordinate its engagements with other parts of the data ecosystem	An organization effectively categorizes its data and licenses it for use appropriately An organization documents all data that is shared with it and has a general understanding of what it can and cannot do with it An organization engages with other parts of the data ecosystem in a coordinated way, pursuant to its data strategy	There is a well-established and bespoke set of licenses that set out clear parameters for use, including integration depending on the category of data involved An organization documents all data that is shared with it and has clear guidance and procedures in place that govern whether and how that data can be joined up with other data sets in its control An organization makes engagement with other parts of the data ecosystem a strategic priority and has a well-coordinated approach with clear processes for joining up its data with external data

Modern NSOs — and in fact, many other government departments, international organizations (including SDG target custodian agencies), private sector actors, and non-governmental organizations — can all be viewed as part of a common data ecosystem, a network of interconnected organizations, data systems, digital infrastructure, and applications.[5]

How organizations choose to engage with other entities within their own, or overlapping, data ecosystems is a strategic decision with important interoperability considerations. Section 3.3 of the Maturity Assessment considers several technical issues at play. The agreements organizations put in place with other parts of the data ecosystem are especially important in ensuring that any planned interoperability is strategically thought through and documented. Data licensing agreements — legal documents that specify what a user is allowed to do and not do with data — are one type of tool for doing this.[6]

[5] Both Development Gateway and Open Data Watch have produced useful guidance on the structure and value of data ecosystems. See, for instance, *Understanding National Data Ecosystems* (Development Gateway 2019) and *The State of Development Data Funding* (ODW 2016, p. 6).

[6] In advance of the UN Statistical Commission in 2020, the Commission's Working Group on Open Data produced a background report entitled "Guidance on the Implementation of Open Data in National Statistical Offices" (UNSD 2020b), which includes guidance on the role of open data licenses at Part II.

Chapter 2: Human interoperability

Human interoperability can refer to two things. On the one hand, it refers to the need to ensure that as data is made interoperable across data systems, it remains readable and usable by human users. On the other hand, it refers to the ability of individuals, groups, teams, and departments of individuals within organizations to be able to communicate and work together in ways that foster interoperability across their workstreams and data assets. The Maturity Assessment identifies five dimensions of human interoperability These dimensions and their associated characteristics for each level of maturity are described in Tables 2.1 to 2.5.

2.1 Data stewardship

Table 2.1: Human interoperability layer: Data stewardship dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
No staff are assigned as data stewards	Joining up data forms part of certain staff members' function but it is not reflected in their job descriptions and is ad hoc	Staff in different units have recognized functions relating to data interoperability, but there is little or no central coordination	There is a coordinated staffing plan that reflects the various dimensions and roles relating to data stewardship, including interoperability, across the organization Units are coordinated and communicate with each other about what data they are joining up, how, and why	A strategically thought-through plan for data management is overseen by an organization's data governance council or committee and includes a clear plan for stewardship of data, including data interoperability functions Staff across the organization are aware of how data is used, joined up, and shared with other entities

In the field of data management, **data stewardship** is the "practice of managing data assets on behalf of others and in the best interests of the organization" (DAMA International 2017, p. 76). Effective and accountable data stewardship is key to interoperability. It is often data stewards who are charged with implementing procedures and practices that enable greater interoperability.

Data stewards are often individual staff members, strategically appointed throughout an organization, to oversee key data management functions in which they have substantive expertise. For instance, chief data stewards might chair data governance councils or committees (sections 1.2 and 1.3), while technical data stewards might be IT professionals who oversee data integration and such in specific departments. It is not uncommon for organizations to operate hierarchies of stewardship, with a high-level data governance council delegating issues to departmental heads, who in turn appoint data stewards to implement the decisions of the data governance council.

Within the data revolution for sustainable development, the concept of data stewardship is being examined in broader terms. For instance, The GovLab, based at New York University, has undertaken extensive research into the role of data stewardship in data collaboratives and it now maintains a Data Stewards Network (The GovLab 2020). Moreover, in the field of official statistics in particular, data stewardship was the focus of the 2020 High Level Forum on Official Statistics that precedes the annual UN Statistical Commission session. In the official statistics community, the concept of data stewardship is also being expanded from relating to individual data management functions to exploring the role of an NSO as a national data steward, coordinating across the national statistical system and entities across broader national data ecosystems (section 1.7) (UNSD 2020a). The definition and scope of data stewardship are continually evolving, with different fields taking approaches that most effectively meet their needs.

2.2 Privacy and confidentiality preservation

Table 2.2: Human interoperability layer: Privacy and confidentiality preservation dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is little to no understanding of the risks to privacy and the need to preserve confidentiality in interoperable data sets	There is emerging understanding of the risks posed to individuals or vulnerable groups if data is combined	There is awareness of applicable privacy and confidentiality related (international) law, normative principles, best practices, and guidance but they are not routinely considered or followed when an organization's data assets are integrated with other data or otherwise used	There is routine consideration of applicable law, principles, best practices, and guidance An organization undertakes privacy impact assessments before and during data-related projects and those assessments include considerations pertaining to the risks associated with data interoperability	The preservation of individual privacy and data confidentiality form part of an organization's legal and ethical review and are integrated across the data life cycle An organization adheres to the highest applicable standards of privacy and confidentiality preservation An organization integrates privacy and confidentiality preservation as part of its data strategy and explicitly provides guidance surrounding the risks of interoperable data, such as the **mosaic effect** An organization is forward-looking and cognizant of the potential privacy risks inherent to interoperable data posed by emerging technologies such as the Internet of Things, biometric ID verification, or general automated processes

Establishing interoperability between systems requires a careful balance between determining what data can be opened up, shared, or integrated with other data sets and what data should not be. There are numerous reasons why an organization might not want to make its data

interoperable with other data sets, but one key reason is the desire to maintain confidentiality and protect the right to privacy that data subjects hold over their personal and sensitive data.[7]

The need to preserve confidentiality and respect individuals' right to privacy is well established. The 6th Fundamental Principle of Official Statistics stipulates that "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes" (ECOSOC 2013). The Universal Declaration of Human Rights of 1948 and the International Covenant on Civil and Political Rights of 1966 both establish a right to privacy for all people, including over information relating to them.

In the case of statistical organizations, individual records collected from surveys, censuses, administrative sources, telecommunication providers, and so on are highly sensitive. Sound data privacy and security protocols are particularly important when linking records from different data sources through common identifiers of individuals, households, businesses, or geographies in the process of producing official statistics. This is also crucial to avoid any reputational damage and retain the trust of all stakeholders.

Within the data revolution for sustainable development, and considering the role of NSOs as key players in the national data ecosystems (section 1.7) and as data stewards (section 2.1), research is ongoing into where the boundaries lie between NSOs' ability to open data up for public use and the need to preserve confidentiality. In 2019, the UN Sustainable Development Solutions Network's Thematic Research Network on Data and Statistics published "Maximizing Access to Public Data: Striking the Balance Between "Open by Default" and Targeted Data Sharing" (TReNDS 2019), which explores these issues in more detail. Moreover, UN Global Pulse's "Risks, Harms and Benefits Assessment" (UN Global Pulse 2020) provides a useful resource for entities seeking to map the risks associated with their data sets.

A further issue to consider is the concept of **sensitive group data**. Sensitive group data is data points that might not pose a risk to individuals if revealed, but may pose a risk to vulnerable groups. An example might be the potential identification of forest dwelling Indigenous peoples' homes in earth observation data sets. Although individual members of a group might not be identifiable, the data may identify what group a member belongs to. If this data were to fall into the hands of criminal loggers, for instance, it may expose the group to harm. The UN's Principles on Personal Data Protection and Privacy (UN 2020) recognize that confidentiality preservation may apply to non-personal data in certain circumstances, particularly "in a sensitive context that may put certain individuals or groups of individuals at risk of harms" (UN 2020).

A final dimension of privacy and confidentiality preservation that touches upon interoperability is the mosaic effect. The Centre for Humanitarian Data, managed by the UN Office for the Coordination of Humanitarian Affairs (OCHA), has defined the mosaic effect as occurring "when

[7] See, for instance, Helen Nissenbaum's book (available online), *Privacy in Context: Technology, Policy and the Integrity of Social Life* (2010) for further information.

multiple datasets are linked to reveal significant new information. While such information could be used to gain insight, it could be used by bad actors to do harm" (OCHA 2020). Risks associated with the mosaic effect might include the risk of re-identifying individuals or vulnerable groups in highly disaggregated anonymized data sets when they are integrated with one another. While each data point might be anonymous, when several are linked together, the resulting insights may be able to identify individuals within the data sets. Keeping the risks associated with the mosaic effect in mind is an important part of strategic planning (section 1.1). More detailed information and guidance on the phenomenon can be found in Care USA's "Responsible Data Maturity Model for Development and Humanitarian Organizations" (CARE USA 2019).

2.3 Staff knowledge and skills

Table 2.3: Human interoperability layer: Staff knowledge and skills dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
Staff do not have the necessary knowledge or skills to join up data	Some staff have the knowledge and skills to join up data, but this is not reflected in their job descriptions and is tangential to their main functions	Knowledge and skills relevant to interoperability start to be recognized as part of job descriptions in some units, but the approach is fragmented	There is a coordinated approach to knowledge and skill strengthening across an organization that explicitly recognizes and addresses interoperability needs	The value of interoperability is recognized by numerous parts of the organization, including non-technical units Training courses relating to data governance issues, including interoperability, are available to all staff

Staff skills and knowledge are central to human interoperability. For organizations that manage diverse data sets and want to improve their interoperability, it is essential that as they develop data management strategies and set strategic objectives (section 1.1), consider oversight and accountability issues (section 1.3), and consider data stewardship functions (section 2.1), they also consider what skills and knowledge staff will need to effectively perform their duties.

Organizations' ability to efficiently integrate multiple data inputs to generate valuable knowledge products increasingly depends on the technical skills of team members. For instance, staff in many organizations need to develop foundational skills to enable them to design, build, and maintain data integration, processing, and dissemination pipelines.

In the field of official statistics, the Partnership in Statistics for Development in the 21st Century (PARIS21) released "Guidelines for Developing Statistical Capacity: A Roadmap for Capacity Development 4.0" (PARIS21 2020) in 2020 that elaborates on the ways in which NSOs can support continuous staff training and knowledge building.

2.4 Internal and external communication

Table 2.4: Human interoperability layer: Internal and external communication dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is no internally or externally coordinated communication reflecting the value of joined-up data	Examples of good practice and value generated as a result of interoperability emerge in an organization but are not communicated internally or externally	Staff and units start to share examples of good practice with each other, but this is not coordinated The value of interoperability starts to be understood by non-technical staff but is not yet communicated externally	Mechanisms to facilitate internal communication and sharing of best practices around interoperability form part of an organization's data strategy Cross-unit communication helps to translate best practices and examples of value generation for external audiences	An organization has a variety of coordinated internal communication channels open between units and staff, enabling the sharing of best practices and examples of value generation An organization is a champion of the value of joined-up data to data ecosystems and actively communicates its experiences and examples with others in compelling and effective ways, including through engagement with data journalists and storytellers

Another tenet of human interoperability is ensuring that there is effective internal and external communication around key data management issues that affect users' ability to use and integrate data.

Internally in an organization, formal communications channels need to be established between individuals, teams, and departments when setting strategic objectives in data management strategies (section 1.1), leadership and management functions (section 1.2), oversight and accountability (section 1.3), and data stewardship functions (section 2.1).

Externally, it is essential to establish user feedback loops that enable communication of data quality, gaps, usability, and other dimensions of human interoperability to be raised and resolved. Similarly, for non-technical audiences, it is important that organizations working in official statistics or the data revolution for official statistics can effectively communicate with policymaking audiences who are the ultimate users of their data. Further guidance on user-centered design in the context of official statistics can be found in "Guidance on the Implementation of Open Data in National Statistics Offices" (UNSD 2020b), produced by the UN Statistical Commission's Working Group on Open Data as a background document for the Commission's 51st session, held in March 2020.

The importance of communicating the meaning and value of data effectively to non-technical audiences is elaborated on in PARIS21's "Guidelines for Developing Statistical Capacity: A Roadmap for Capacity Development 4.0" (PARIS21 2020). Moreover, the Global Partnership has published numerous data stories, blogs, and opinion pieces on its homepage that convey

the importance of data issues to non-technical audiences. For example, “5 Useful Things I Learned About Data Storytelling: Lessons from #Visualize2030” (Jacobson 2019) and “Joining-up Data for Universal Healthcare in Kenya: The View from the Ministry of Health” (Orrell 2020) are two resources published on the Global Partnership’s homepage that respectively provide guidance on how to tell stories with data, and give an example of data storytelling focused on interoperability.

2.5 Adaptability

Table 2.5: Human interoperability layer: Adaptability dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
Processes relating to staff functions and oversight of data interoperability are rigid and hard to change	There is emerging understanding of the value of adaptability in functions and oversight to data management generally, but no specific approach	Disparate units across an organization start to formally recognize the need to ensure that staff’s functions and oversight of data systems are adaptable so as to ensure that value continues to be generated from their data assets	The value of empowering staff to be adaptable in how they use data, including in how they join it up with other data, is recognized by an organization and is reflected in its data strategy Staff have the authority to adapt their working processes and oversight of organizational data assets in ways that enhance its value, including by joining them up	An organization becomes a leader in adaptive management, and staff feel empowered and are confident in their ability to adapt their oversight of data systems as needed, including how they join up data, to maximize value

An organization’s ability to ensure that its internal data management processes are adaptable is an important dimension of human interoperability. As digital and data innovations flourish, organizations and their staff need to ensure that their data management strategies keep apace of developments. Adaptability requires communication and empowerment of team members, as well as flexibility to change course at any point in order to meet evolving stakeholders’ needs.

Thinking through ways in which decision-making functions are distributed across organizations is a key part of strategic planning (section 1.1), links to broader data ecosystems (section 1.7), data stewardship (section 2.1), staff knowledge and skills (section 2.3), and internal organizational communication (section 2.4). Chapter 17 of the “Data Management Body of Knowledge 2nd Edition” (DAMA International 2017) provides further guidance on data management and organizational change management.

A good example of the value that being adaptive can add relates to the need for individual and organizational awareness around the emergence of new data infrastructure components. For instance, organizations seeking to improve the interoperability, accessibility, and use of their data assets should be aware of trends pertaining to the accessibility and application of digital

infrastructure, and be able to adapt to them or adapt them to their needs. Broadly speaking, there are two dimensions to this.

First, digital and data innovations occur very rapidly, with new techniques, tools, and uses for data emerging at a high frequency. This means that organizations seeking to improve the accessibility, interoperability, and use of their data assets need to ensure that they are regularly monitoring new technologies and incorporating them where possible. This requires planning at a strategic level (section 1.1), maintaining proactive links with the broader data ecosystem (section 1.7), and ensuring that data stewards are given responsibility to monitor innovations (section 2.1).

Second, it means that organizations also need to be aware of the potential opportunity costs of missing out on new digital technologies. While it may seem counterintuitive, evidence suggests that as more digital innovations are rolled out around the world, this harms existing digital divides — in many instances making them worse. The UN Conference on Trade and Development's "Digital Economy Report 2019" (UNCTAD 2019), for instance, finds that as digital divides worsen, so too do inequalities as those with access to the latest technologies edge ever further ahead of those without. Understanding the dynamics of these issues is an important part of the strategic planning process. It enables organizations to make the most of what they have and plan ways in which they can incrementally improve their access to advanced digital infrastructure, including technologies that enable or enhance interoperability.

Chapter 3: Data interoperability

Data interoperability is the need to ensure that data systems and data sets are designed to enable interoperability. This means ensuring that they can collate, store, and process data in **machine-readable formats** and that interoperability needs are considered when data are modeled and classified. The maturity assessment identifies five dimensions of data interoperability. These dimensions and their associated characteristics for each level of maturity are described in Tables 3.1 to 3.5.

3.1 Data and metadata modeling capacity

Table 3.1: Data interoperability layer: Data and metadata modeling capacity dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is little or no ability to model data or metadata	There is an emerging understanding of the value that data and metadata modeling can confer to data assets, but data modeling is not a priority for technical units	Disparate units across an organization recognize the value of data and metadata modeling, including its importance to data interoperability, and take steps to align their modeling techniques and start to coordinate their efforts	Technical units coordinate their approach to both data and metadata modeling and align efforts to consistently model data based on their organizational needs Internal needs are still prioritized over external groups, but data is modeled consistently	An organization routinely utilizes canonical data and metadata models that follow standardized patterns, making them reusable and conducive to data sharing The selection and application of canonical models is done through careful planning, including through engagement with data users and other entities in the data ecosystem

Data and metadata modeling are the foundations of data interoperability. They are part of the design process in which the internal structure and inter-relations across different data sets are defined, optimized, and described to capture all relevant business metrics and dimensions (Chang 2018). The process of modeling data and metadata includes a number of steps, ranging from the discovery, analysis, and scoping of data requirements through to communicating them in the form of a model (DAMA International 2017).

Data modeling needs to be fit for purpose and remain flexible because the patterns used in specific applications or for analytical purposes often involve trade-offs that need to be frequently revisited, such as accepting **data redundancy** to facilitate usability. However, a case-by-case approach to data modeling harms the ability of data to interoperate across models and flow across a data ecosystem.

Common data and metadata modeling patterns are needed in order to avoid organizational, or even departmental, silos. Chapter 2 of the Guide (Gonzalez and Orrell 2018) elaborates on the relationship between interoperability and data modeling and highlights the value of **canonical**

data models and industry standards to enable interoperability. Similarly, Chapter 5 of the "Data Management Body of Knowledge" (DAMA International 2017) provides robust technical guidance on how organizations can improve their data modeling processes.

3.2 Data organization and classification capacity

Table 3.2: Data interoperability layer: Data organization and classification capacity dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
An organization is unaware of the importance of data classification to interoperability and does not have a clear idea of its data assets	Units start to inventory their data Units are aware of the need for standardized data classification, but only use them on an ad hoc case-by-case basis	There are informal attempts between units to use common classifications, but these are not formalized or coordinated across all relevant units There is some, but not consistent, use of common classifications across the organization	There is a coordinated approach to the use of data classifications across the organization Units work together to identify the most appropriate classifications for their data and ensure that the data under their control is appropriately classified	The organization not only routinely and appropriately uses data classifications but also produces its own classifications to fill gaps and ensure consistency The organization engages actively with other entities in the data ecosystem to improve commonly used classification systems and establish new ones as needed The organization effectively communicates the value of consistent data classification for interoperability

Coordination within and across organizations, and whole national statistical systems or data ecosystems for that matter (section 1.7), to decide on common classifications and identifiers is key to enabling data interoperability. Standard vocabularies, classifications, and unique identifiers are part of the basic data infrastructure of a country or organization. They help improve consistency and avoid ambiguity in the description of a data set, while enabling users to more easily locate and link together related data elements.

Data classification and identifier systems "shape the way that data is collected, processed, analyzed and shared with users. [...] The use of common classifications and vocabularies allows data to be shared efficiently and for users to more easily find related information across numerous data platforms" (Orrell 2018). Standard classifications enable the integration of multiple data sets, as well as their consistent analysis and interpretation. They provide the taxonomic basis for managing and describing statistical data and are fundamental components of key frameworks for the compilation of official statistics (e.g., the System of National Accounts).

Both classification and identifier systems need to be interoperable themselves, and require adequate governance, maintenance, and oversight. This is particularly important in the case of unique identifiers, in order to ensure that they are issued to all relevant entities by a recognized central authority, and that they are never reissued and remain valid over the entire lifetime of the entities to which they relate.

Chapter 3 of the Guide (Gonzalez and Orrell 2018) explores the value of standard classifications and vocabularies to interoperability in more detail and provides examples of their use in the official statistics and data revolution for sustainable development communities.

3.3 Data access, openness, and sharing

Table 3.3: Data interoperability layer: Data access, openness, and sharing dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
An organization has little or no knowledge of interoperability considerations when planning to responsibly manage data access, share data, or open it up for use	Disparate units across an organization are aware of interoperability considerations when planning, responsibly manage data access, share or publish data as open data, but this knowledge is not uniform or universally applied	There are coordinated efforts in technical units to ensure that data is accessible and shared responsibly, including relevant licensing permissions or limitations for future data integration and use Some data is made open on an organizational platform, but data sets are incomplete, not timely, or have not been quality assured	Data is shared responsibly in ways that protect any rights that third parties may have over it Data that is published openly is done so in machine readable formats under a clear open data license with terms of use, and has been stripped of attributes that may result in the re-identification of individuals or vulnerable groups Open data portals are accompanied by relevant contextual information and are visualized in ways that promote use by numerous audiences	An organization operates an effective data sharing policy that provides guidance on the various ways in which data sharing should take place, from publication under an open license, through to the use of data sharing or processing agreements Legal advice is available to staff wanting to share data that will be integrated with other data sets by third parties Open data is not just published in machine and human readable formats but is also made available as linked data through the semantic web There are feedback loops with key audience groups and the organization is responsive to user needs

While Chapters 1 and 2 of this document touch upon dimensions of data access, sharing, and openness, there are also technical considerations at the data interoperability layer.

At the data interoperability layer, considerations around how to strategically plan for the publication of open data (section 1.1) also touch upon how data sets are modeled (section 3.1) and data is classified (section 3.2). Where possible, open-source software and data standards should be used for open data and metadata publication because these will allow data users to more easily integrate new data sets into their systems.

There is a growing interest in tools and technologies that allow for the publication of data in such a way that machines can easily identify and integrate related information sources directly through the world wide web. This “linked data” approach facilitates interoperability and data integration by connecting both structured and unstructured data sets with special metadata elements designed to be referenced over the web. This in turn enables machines to retrieve meaningful information from across a potentially enormous range of sources. Google’s Dataset Search (Google 2020) tool operates a linked data approach using standards set by the Schema.org consortium.

Linked data approaches are particularly important for development data, because the “indivisible” nature of the Sustainable Development Goals makes it more urgent than ever to join up the information resources and data assets owned and managed by different sectors and communities. Chapter 5 of the Guide (Gonzalez and Orrell 2018) covers linked open data in more detail.

3.4 Data analytics and automation

Table 3.4: Data interoperability layer: Data analytics and automation dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is little to no awareness of how to enable interoperability between data sets to undertake data analytics or how to join up data to train algorithms (machine learning)	There is disparate understanding of the role of interoperability in undertaking automated data analytics across organizational units There is limited understanding of how interoperable data should be used to train algorithms	A coordinated approach between organizational units starts to emerge and some units start to produce scrubbed, quality assured, and consistent data sets that are available for integration and automated processing There is a coordinated effort to understand how data sets can be combined to train algorithms	Data analytics and machine learning functions are reflected in an organization’s data strategy The relative benefits and risks of running automated analytics over interoperable data, or using it to train algorithms, are generally understood but there is not yet a consistent approach across an organization	An organization’s data strategy includes forward looking plans for how data analytics tools can be responsibly applied to multiple, interoperable data sets in future There is a nuanced and well-established understanding of the relative benefits and risks of running automated analytics over interoperable data or using the data to train algorithms and appropriate risk and cost-benefit assessments are applied as needed An organization proactively engages with other entities in a data ecosystem to share its learnings and uses open-source analytics tools whenever possible to enable transparent scrutiny

As digital and data revolutions have exploded around the world, so too has the ability of a broader range of entities and individuals to collect, organize, structure, and analyze more diverse types of data. As more data, and types of data, are joined together, interoperability

becomes more important to ensuring that they are modeled (section 3.1) in ways that enable processing and analysis, including by automated means.

In the official statistics community, in recent years there has been increased international cooperation around how analytics, including automated analytics, can be run over statistical data sets. The UN Global Working Group on Big Data for Official Statistics (UNSD 2020c), for instance, explores the role that insights gleaned from big data can play in supporting the attainment of the SDGs. Interoperability between data derived from multiple sources is often needed in order to run automated processing techniques over data sets.

As NSOs and other entities in the data revolution for sustainable development experiment and innovate with new data sources, including by using automated processing techniques to analyze them, it is important that the risks associated with these practices are well understood and managed. Organizations should be especially aware of the risks associated with the mosaic effect, discussed in more detail in section 2.2. CARE USA's "Responsible Data Maturity Model for Development and Humanitarian Organizations" (CARE USA 2019) also contains further information on the risks associated with automated processing, including the mosaic effect.

3.5 Data protection

Table 3.5: Data interoperability layer: Data protection dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is little to no understanding or awareness of the links between data interoperability and data protection techniques, including anonymization, pseudonymization, and encryption	There is some knowledge and understanding of the need to protect data that will be combined with other data, including through the use of appropriate pseudonymization, anonymization, and encryption techniques as needed, but this knowledge is not uniformly understood, and data protection techniques are not consistently applied	Disparate units routinely apply appropriate data protection techniques to their data sets before data integration, but there is little to no consistency in how those techniques are applied There is some, but limited, understanding of the risks of re-identification inherent to interoperable data	Personal, sensitive, and sensitive group data is subject to appropriate protections before being integrated, shared, or processed through automated analytics tools Risks of re-identification inherent to interoperable data are understood and are applied, but not routinely	All data is protected using the appropriate techniques and either responsibly archived or permanently deleted at the end of its intended life cycle Access to sensitive data sets is monitored and documented to ensure accountability over data protection Prior to integration, sharing, or processing through automated analytics, all data is assessed for risks of re-identification or other harms and is only used when there is a high degree of certainty that the data will remain safe following reuse An organization helps to set standards for data protection within the broader data ecosystem and champions responsible data use

At its most basic level, data protection is the regulation of how access and use of data stored on computers, digital devices, and paper records is controlled.

Data protection can be viewed through three lenses. First, it is a series of legal obligations in jurisdictions where data protection laws have been passed that establish rules around how data should be protected. The European Union's General Data Protection Regulation (EU 2016) is a good example of this. Second, data protection can relate to the protection of individuals' personal data and is closely related to privacy and confidentiality preservation (section 2.2). Finally, it can relate to a mix of data-related and technical considerations. Technically, it is related to cybersecurity and incident response (section 4.2), while at the data layer it can relate to data security, including issues pertaining to "the planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access and auditing of data and information assets" (DAMA International 2017).

In terms of its links to interoperability, there are several considerations to keep in mind, many pertaining to the potential identification of individuals if data is used or integrated with other data in ways that reveal individuals' or (vulnerable) groups' identities (sections 1.4, 1.5, and 2.2). As organizations consider their approach to data protection, they therefore need to be aware of their jurisdiction's legal obligations, any applicable normative standards (such as the Fundamental Principles of Official Statistics) or ethical codes, and customary practices in a particular sector or domain. Moreover, steps need to be taken within organizations to anonymize and pseudonymize data sets that may contain data points that may result in the identification of individuals. The Open Data Institute's "Anonymisation and Open Data: An Introduction to Managing the Risk of Re-identification" (ODI 2019b) provides a helpful overview of key issues to consider when seeking to anonymize data sets.

Data sets that may contain sensitive or personally identifiable information should have access to them restricted in order to uphold individuals' data rights and any applicable duty of confidentiality (section 2.2). This data should be encrypted and securely stored and should establish a hierarchy of permissions for access. Maintaining permission lists and monitoring their implementation and integrity should be the function of data stewards within technical departments of a data producing or processing organization (section 2.1).

Finally, links to data access, openness, and sharing (section 3.3) must also be considered at the strategic planning stage (section 1.1) and layers of encryption must be added to restricted data sets.

Chapter 4: Technological interoperability

Technological interoperability requires appropriate data infrastructure to enable data, human, and organizational interoperability in meaningful ways. The maturity assessment identifies two primary dimensions of technological data interoperability. These dimensions and their associated characteristics for each level of maturity are described in Tables 4.1 and 4.2.

4.1 Data infrastructure

Table 4.1: Technological interoperability layer: Data infrastructure dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
An organization faces shortages of key infrastructure to store, manage, exchange, and process data, such as hardware and software components, a reliable electricity supply, or Internet connectivity	There is adequate access to key infrastructure components but there is a shortage of organization-specific data storage and content management solutions, resulting in non-standardized and non-aligned data management systems	All appropriate staff members have access to adequate hardware and software tools, as well as network connectivity There are secure servers and data repositories, but they are used inconsistently by staff and organizational units; there is little oversight of digital infrastructure	All appropriate staff members are aware of, and trained in, how to use an organization's data management and processing systems Secure servers and data repositories are routinely used by staff members and oversight of digital infrastructure is part of an organization's data strategy	An organization's data strategy includes provisions for the maintenance, regular review, and upgrading of its digital infrastructure, and budget lines are set aside for this purpose Data policies and standards on procurement, data sharing, and infrastructure oversight are aligned An organization is forward thinking in its approach to digital infrastructure and actively strategizes and plans on how it can make best use of emerging technology to improve the interoperability of its data systems

Although it might be self-evident, it is worth explicitly stating that all entities and organizations hoping to engage with the broader data ecosystem (section 1.7) and working towards improving the accessibility and use of their data assets need to have basic digital infrastructure, such as laptops, database capacity (server or cloud), and Internet connectivity.

However, the concept of what constitutes digital infrastructure is vastly broader than what it was even just a few years ago. There are now numerous categories of infrastructure that all rely, and impact, upon interoperability. A few examples of interconnected infrastructural components could include the following:

- Data storage infrastructure (whether local servers or cloud-based).
- Data management infrastructure, such as bespoke information management systems.
- Data processing infrastructure (increasingly cloud-based for large data sets).
- Data dissemination infrastructure composed of data platforms, community hubs, etc.

Choosing, acquiring, and maintaining digital infrastructure is a key function in data-driven organizations. Factors such as the types and volume of data assets that will need to be stored, accessed, and analyzed need to be considered when deciding what solutions to adopt. Whether data needs to be accessed and processed in batches or in real time or whether open-source technologies need to be supported are further relevant considerations.

All of these components require and improve interoperability at the data, organizational, and human layers. As more data storage, processing, and dissemination functions become cloud-based, for instance, this raises the need for considerations around cybersecurity (section 4.2); data access, openness, and sharing (section 3.3); ethical issues pertaining to the rights granted to third-party data processors (section 1.5); and many more. Ultimately, what is important is that infrastructure is reflected clearly in organizations' strategic objectives (section 1.1) and that there is effective management (section 1.2) and oversight (section 1.3) of how these components are applied.

4.2 Cybersecurity and incident response

Table 4.2: Technological interoperability layer: Cybersecurity and incident response dimension.

Characteristics of each increasing level of maturity →				
Undefined	***Emerging***	***Learning***	***Building***	***Consolidating***
There is little to no awareness of the risks of cyberattacks or other breaches to an organization's data systems, including the specific risks associated with potentially reusable, interoperable data No data breach protocol or policy is in place	Disparate staff and units across an organization have awareness or show concern about the risks posed to their reusable data by a cyberattack or other data breach Champions emerge who push for a data breach protocol or policy	A data breach protocol is drafted, but risks associated with the potential reuse of stolen interoperable data remain vague and there is inconsistent understanding and application of the policy	A clear data breach policy setting out sequential steps and responsibilities is established Staff receive training on what they should do in the event of a data breach and are taught about the risks associated with the reuse of interoperable stolen data	An organization is able to deal with data breaches swiftly and effectively, and takes active steps to ensure that its technological infrastructure is as secure as possible The data breach policy is regularly reviewed and updated, and explicitly covers risks associated with interoperable data reuse Appropriate staff are routinely trained on how to respond to a data breach

As more key data management functions move online and remote work arrangements become the rule rather than the exception, cybersecurity and protocols for incident response become more important. Take the example of cloud computing. Due to its reliance on hardware independent virtualization technology, cloud computing enables organizations to quickly back up data, applications, and even operating systems to a remote data center, and to deploy them to multiple users in many different locations. However, all this data transmission over the Internet exposes it to cybersecurity threats.

Data security threats usually succeed at the weakest points of organizations' data systems. As web-enabled approaches to interoperability begin to flourish (section 3.3), cybersecurity and incident response issues need to be considered in organizations' data management strategies (section 1.1). Equally, staff need to be assisted to assume responsibility over the data security of their own devices (laptops, mobile phones, etc.), encouraging "a reflective, analytic approach" to cybersecurity, instead of simply providing rules-based training that may result in overly rigid or narrow decision-making (HBR 2020). This area is also closely aligned to issues pertaining to data protection (section 3.5); data stewardship (section 2.1); and data access, openness, and sharing (section 3.3).

The "Data Management Body of Knowledge 2nd Edition" (DAMA International 2017) identifies five key activities relating to the data security strategy: identifying relevant data security requirements, defining a policy, defining security standards, assessing risks, and implementing controls and procedures. Moreover, CARE USA's "Responsible Data Maturity Model for Development and Humanitarian Organizations" (CARE USA 2019) provides additional guidance relating to the development of approaches to manage incident response situations.

Annex A: The Joined-Up Data Maturity Assessment

Insert

Annex B: Glossary

Canonical data models are data models that follow specific standardized patterns that make them highly reusable and conducive to data sharing (Orrell 2018).

Data governance is how control over data is exercised across its life cycle. Data governance guides all other data management functions and ensures that data is managed properly, according to policies and best practices (DAMA International 2017).

Data management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their life cycles (DAMA International 2017).

Data redundancy occurs when the same piece of data exists in multiple places at the same time. Data redundancy can cause confusion and inconsistency in how data is categorized, reducing its usefulness and value.

Data stewardship is the "practice of managing data assets on behalf of others and in the best interests of the organization" (DAMA International 2017).

Machine readable formats are data formats that can be "read" and processed by a computer. Generally speaking, machine readable formats require data to be formally structured.

Maturity models are tools that set out criteria and steps that help organizations measure their ability and continuous improvement in particular fields or disciplines.

Mosaic effects occur when multiple data sets are linked to reveal significant new information. While such information could be used to gain insight, it could be used by bad actors to do harm (OCHA 2020).

Sensitive group data is data points that might not pose a risk to individuals if revealed, but may pose a risk to vulnerable groups.

Vendor lock-in occurs when different proprietary digital and data solutions are procured separately as part of different programs of work or donor-sponsored projects without considering existing data architectures or digital infrastructure. When this occurs, siloed information systems can be created that produce systems or data sets that do not interoperate or whose data outputs cannot be integrated with each other.

References

Bower, T. 2020. Boost Your Resistance to Phishing Attacks. *Harvard Business Review*. hbr.org/2020/09/boost-your-resistance-to-phishing-attacks

CARE USA. 2019. Responsible Data Maturity Model for Development and Humanitarian Organizations. ictworks.org/wp-content/uploads/2019/11/Responsible_Data_Maturity_Model_10-16-19.pdf

Chang, R. 2018. A Beginner's Guide to Data Engineering — Part II. medium.com/@rchang/a-beginners-guide-to-data-engineering-part-ii-47c4e7cbda71

DAMA International. (2017). Data Management Body of Knowledge (2nd Edition). New Jersey: Technics Publications.

Data Leaders. 2017. The Leader's Data Manifesto. dataleaders.org/manifesto/

Development Gateway. 2019. Understanding National Data Ecosystems. developmentgateway.org/blog/understanding-national-data-ecosystems

European Court of Justice (ECJ). 2020. The Court of Justice invalidates Decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield (press release No 91/20). curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf

European Union (EU). 2016. General Data Protection Regulation (2016/679). eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32016R0679

Gartner. 2015. Take Your Analytics Maturity to the Next Level. gartner.com/smarterwithgartner/take-your-analytics-maturity-to-the-next-level/

Gonzalez, L. and Orrell, T. 2018. Data Interoperability: A Practitioner's Guide to Joining up Data in the Development Sector. data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner's%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf

Google. 2020. Dataset Search. datasetsearch.research.google.com

IBM. 2007. IBM Data Governance Maturity Model. lightsondata.com/data-governance-maturity-models-ibm/

Jacobson, A. 2019. 5 Useful Things I Learned About Data Storytelling: Lessons from #Visualize2030. data4sdgs.org/news/5-useful-things-i-learned-about-data-storytelling-lessons-visualize2030

Nissenbaum, H. 2010. Privacy in Context: Technology, Policy, and the Integrity of Social Life. crypto.stanford.edu/portia/papers/privacy_in_context.pdf

Open Data Institute (ODI). 2015. A Guide to the Open Data Maturity Model: Assessing Your Open Data Publishing and Use. scribd.com/document/260481608/ODI-Maturity-Model-Guide-Assessing-your-open-data-publishing-and-use#download

Open Data Institute (ODI). 2019a. Data Ethics Canvas. theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf

Open Data Institute (ODI). 2019b. Anonymisation and Open Data: An Introduction to Managing the Risk of Re-identification. docs.google.com/document/d/1CoXniaTnQL_4ZyQuji9_MA_YCEElQjx4z1SEdB08c2M/edit

Open Data Watch (ODW). 2016. The State of Development Data Funding. opendatawatch.com/wp-content/uploads/2016/09/development-data-funding-2016.pdf

Orrell, T. 2018. Using Data to Join Up Development Efforts. data4sdgs.org/sites/default/files/services_files/Data%20Interop%20Brief_0.pdf

Orrell, T. 2020. Joining-up Data for Universal Healthcare in Kenya: The View from the Ministry of Health. data4sdgs.org/resources/joining-data-universal-healthcare-kenya-view-ministry-health

Palfrey, J. & Gasser, U. (2012). *Interop: The promise and perils of highly interconnected systems*. New York: Basic Books.

Partnership in Statistics for Development in the 21st Century (PARIS21). 2020. Guidelines for Developing Statistical Capacity: A Roadmap for Capacity Development 4.0. paris21.org/sites/default/files/inline-files/UNV003_Guidelines%20for%20Capacity%20Development%20WEB_0.pdf

The GovLab. 2020. Data Stewards. medium.com/data-stewards-network

Thematic Research Network on Data and Statistics (TReNDS). 2019. Striking the Balance Between “Open by Default” and Targeted Data Sharing. sdsntrends.org/research/2019/3/4/maximizing-access-public-data

United Kingdom (U.K.) Government. 2018. Open Standards Principles. gov.uk/government/publications/open-standards-principles/open-standards-principles

United Kingdom (U.K.) Government. 2020. Data Ethics Framework. gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020#overarching-principles

United Nations (UN). 2020. Principles on Personal Data Protection and Privacy. unsceb.org/principles-personal-data-protection-and-privacy

UN Conference on Trade and Development (UNCTAD). 2019. Digital Economy Report 2019. Value Creation and Capture: Implications for Developing Countries. unctad.org/en/PublicationsLibrary/der2019_en.pdf

UN Economic and Social Council (ECOSOC). 2013. Fundamental Principles of Official Statistics (E/RES/2013/21). unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf

UN Global Pulse. 2020. Risks, Harms and Benefits Assessment. unglobalpulse.org/policy/risk-assessment/

UN Office for the Coordination of Humanitarian Affairs (OCHA). Centre for Humanitarian Data. 2020. Exploring the Mosaic Effect on HDX Datasets. centre.humdata.org/exploring-the-mosaic-effect-on-hdx-datasets/

UN Statistics Division (UNSD). 2020a. Data Stewardship: A Solution for Official Statistics' Predicament? unstats.un.org/unsd/statcom/51st-session/side-events/20200302-3A-high-level-forum-on-official-statistics/

UN Statistics Division (UNSD). 2020b. Guidance on the Implementation of Open Data in National Statistical Offices. unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3v-Guidance_OD-E.pdf

UN Statistics Division (UNSD). 2020c. UN Global Working Group on Big Data for Official Statistics. unstats.un.org/bigdata/

Acknowledgements

The Joined-Up Data Maturity Assessment and its introductory guide were authored by Tom Orrell, Managing Director of DataReady Limited, on behalf of the Global Partnership for Sustainable Development Data, and Luis Gerardo González Morales, Statistician at the United Nations Statistics Division.

The Guide was supported by all members of the Collaborative on SDG Data Interoperability, to whom the authors extend their gratitude and thanks. The authors would like to thank in particular members of the Advisory Group from within the Collaborative who supported the production of the Maturity Assessment, including Eric Swanson at Open Data Watch; Josh Powell and Paige Kirby at Development Gateway; Karen Carter at the United Nations Children's Fund; Malarvizhi Veerappan at the World Bank; Mahsa Hedayati at the UN Office of Information and Communications Technology; and Neil Jackson at the U.K.'s Foreign, Commonwealth & Development Office.

The authors would also like to extend a special thanks to Bill Anderson at Development Initiatives, whose foresight helped put interoperability on the data revolution policy map; Shaida Badiee and the whole Open Data Watch team, who have continuously supported this work; Manuel Cuéllar and Enrique Jesus Ordaz López at Mexico's National Institute of Statistics and Geography for their continued support and feedback; Fredy Rodriguez at CEPEI (Centro de Pensamiento Estratégico Internacional); Rachael Beavan, Edafe Onerhime, and their colleagues at the Foreign, Commonwealth & Development Office; the UN Statistical Commission's Working Group on Open Data; and Francesca Perucci at the UN Statistics Division and Claire Melamed at the Global Partnership and their teams, for their support for the Collaborative on SDG Data Interoperability over the last four years.