

STAT 441/2/177A

Data stewardship and National Statistical Offices: defining a stock-taking exercise

(Amended following 18th HLG-PCCB meeting, as of 6 July 2020)

Background

1. Official statistics' role as the dominant source of information is being challenged in the new data ecosystem as new data actors emerge while at the same time demand for data and statistics ever increases. New technologies allow for an unprecedented level of data collection, integration and analysis on human behaviour and overall societal trends, which also require attention to safeguard the use of these data. Data is now the basis of many business models and this has created some of the most valuable companies on the planet, who profit based on building vast and ever-increasing databases of individual data. To discuss official statistics' position in this new data ecosystem, the topic of this year's High-level Forum on Official Statistics (HLF-OS) at the 51st session of the UN Statistical Commission (UNSC) asked the question "Data stewardship – a solution for official statistics' predicament?"
2. The summary of the HLF-OS was presented in a statement to the UNSC. This summary concluded the idea of data stewardship had different interpretations and connotations depending on the national context, expectations and institutional set-up. However, there was a consensus that the adoption of a data stewardship approach enables the statistical office to evolve from being a statistics producer to becoming a service provider that facilitates a joined-up approach to data and statistics across different data and statistics communities. As a result of the HLF-OS discussion and summary, the UNSC at its 51st session requested the HLG-PCCB via decision 51/120:

to establish a **working mechanism** open to other stakeholders and **linked to existing initiatives** at global and regional levels to address the issue of **data stewardship and the role of statistical offices in the new data ecosystem**; the work should consist of a **stock-taking exercise** and **specific recommendations** on the **way forward**, to be provided to the fifty-second session of the Commission.
3. This paper is updated following the first meeting subgroup for this work comprising of Argentina, Finland, Mexico, Saudi Arabia, State of Palestine, the United Kingdom and UNESCAP. This paper plans a response to decision 51/120 in terms of three parts covering; the scope of the stock-taking exercise, the working mechanism and

the timetable. This paper is informed by previous work¹ but also considers some new approaches.

Scope: defining the stock-take

4. This paper proposes a *theory of change* methodology² is applied to the stock-take to help identify the key drivers of change needed to adopt or support the role of a data steward in a National Data Ecosystem (NDE). This method's starting point is to define a desired impact statement or goal to be attained and then map backwards the necessary pre-conditions and foundations such as legal, technical and institutional mandates. Under this methodology we define the broad goal of a data steward is to:

Improve the use of data in society

5. Following this method, a set of outcomes are defined which will lead to this goal. Since improved use is determined by fitness for purpose, the desired outcomes are informed by the common quality dimensions³ used in official statistics, which are derived from the Fundamental Principles of Official Statistics (FPOS)⁴. Four outcomes are proposed as follows:

Better governance More oversight and better governance by establishing a strong independent role of the data and statistical authority to provide an expert opinion of data produced within Government and across society to ensure trust, protect data rights and privacy, and counter misinformation and data misuse.

Better collaboration More collaboration between producers and users within Government and across other communities (i.e. private sector, academia etc.) and improved coordination by brokering new partnerships to ensure data of public interest⁵ that is relevant, interoperable, comparable, and meets needs.

¹ This includes the work of UNECE at its 67th plenary session on “Emerging roles of national statistical offices as offices for statistics and data” (ECE/CES/2019/22), the Sustainable Development Solutions Network Thematic Research Network on Data and Statistics (SDSN TReNDS) “Counting on the World to Act: A Roadmap for Governments to Achieve Modern Data Systems for Sustainable Development” and the High Level Forum on Official Statistics at the 51st session of the UN Statistical Commission (and subsequent UNWDF webinar) “Data stewardship – a solution for official statistics’ predicament?”

² See https://en.wikipedia.org/wiki/Theory_of_change for further references on the methodology.

³ See <https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAFManual-WEB.pdf>

⁴ See <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>.

⁵ Public interest is a common concern among citizens in the management and affairs of local, state, and national government. It does not mean mere curiosity but is a broad term that refers to the body politic and the public weal.

Better methods	More transparency to ensure openness to a range of new data methods and sources from different parts of the statistical and data systems, public and private, and mitigating latency in information systems
Better access	More consultation to bring in the viewpoint of the user and enable users to work with data more easily and target population groups who maybe have more challenges to using data fully in their lives

6. These outcomes need to be supported by outputs which bring about these changes which in turn are supported by activities to achieve each output and the inputs needed to accomplish the activity. Since there is no common view on what data stewardship requires, this paper proposes some to initiate a discussion about what outputs, activities and inputs are needed to bring about these outcomes (see Annex). Over the course of the stock-take, it is hoped a better understanding emerges of how NSOs can position themselves within the National Data Ecosystem and the kind of changes a data stewardship approach could usher in. The stock-take will acknowledge and incorporate different starting points, institutional environments and national/ regional set-ups to be relevant to most settings and situations.

Working mechanism: development of case studies

7. Case studies on aspects of a data stewardship model (as defined in Table) will be the main source of information to inform the stock-take. The type of “data” referred to in the impact statement will be informed by the case studies collected, but “data of public interest” (see earlier definition) are a suggested starting point.
8. Different forms of data stewardship being followed in National Data Ecosystems will be used. Key scenarios and national examples already identified are as follows:

Scenario	Examples
The NSO takes the formal data steward role	Colombia, Estonia and New Zealand, Philippines,
NSO supports the role and activities of a data steward who is located elsewhere in Government or outside it.	Australia, Finland
The data stewardship role is independent from the NSO and the NSO plays no oversight role.	Chief Data Officers at city level ⁶
There is no role or activities of data	

⁶ A few examples of these are provided in document: <https://countingontheworld.sdsntrends.org/>

stewardship in the country (counterfactual) – to describe situations where data stewardship approach is needed	
---	--

9. In addition, it would be useful to look at examples by different types of organisational and data system circumstance like level of capacity⁷, institutional set-up (i.e. centralised or decentralised statistical system at the national level or even localised or city statistical systems at the sub-national level) and those NSOs within countries with special circumstances (i.e. small islands, least developed, landlocked etc.)
10. The case studies will provide a better understanding of the roles and activities of a data steward and how they fit into different contexts (i.e. capacity, institutional, organisation). This will then give the basis for an initial set of options or drivers of change to be considered for further discussion by the global statistical and data community. The findings as well as any gaps in knowledge will be presented at the 52nd session of the UN Statistical Commission as part of the official report of the HLG-PCCB as well as presented in a more detailed background report and will provide the basis for further work by the commission and the broader data community

Timetable for the stock-table

1 June	1 st subgroup meeting
11 June	HLG-PCCB meeting to further review way forward
w/c 13 July	2 nd subgroup meeting (with wider stakeholder group)
23 July	Review of progress and direction at a special meeting of the HLG-PCBC
July - September	Further desk research, stakeholder engagement, collection of case studies and other evidence to identify options on the drivers for change to a data steward approach
30 Sept	1 st draft of case study compendium with executive summary
18 – 22 Oct	Session on data stewardship at UNWDF
30 Oct	2 nd draft of case study compendium with executive summary
November	Secretary-General's report on the HLG-PCCB for the 52 nd session
Jan – Feb	More detailed background report for the 52 nd session of the UNSC

⁷ <http://datatopics.worldbank.org/statisticalcapacity/>

Annex: Scope of the “data stewardship” stock-take

Outcome	Outputs	Activities	Inputs / needs
Better governance	Role of Chief data officer / national data steward / another institutional role	Coordination within national context and institutional setup to understand the best way to establish and maintain leadership and direction	<ul style="list-style-type: none"> - What is the scope of the role? - Implications for the NSOs changing functions (i.e. capacity and mandate)? - How is the role complementary to the chief statistician? - How does it fit in an existing NSO’s organisational structure (as new position or part of existing one)? - What is the role in setting ethical standards and safeguarding privacy of data to ensure trust in data? - What are the key elements to assess the NSO baseline related to data stewardship implementation approach? - What are the key elements to consider a shift from statistics producers to service providers?
	National data strategy		<ul style="list-style-type: none"> - What is the purpose of these strategies and how do they differ from code of practice and other documents released by the NSO? - How does professional independence and accountability of NSOs impact on data stewardship role?
	UN Statistical Commission and GGIM to discuss broader data issues		To be developed
	Fundamental Principles for data of public interest		To be developed

Outcome	Outputs	Activities	Inputs / needs
	Quality assurance: metadata and data documentation / open data	Build out skills within the organisation(s) such as data officers (skills investment)	<ul style="list-style-type: none"> - How can professionals within the NSO be used to enable data stewardship? - Can existing frameworks be adapted easily to indicate quality of non-official statistics and non-Government data? - How do NSO policies and practices on these areas change with a data stewardship approach?
Better collaboration	Data sharing agreements	Co-ordination role of the NSS on data sharing	<ul style="list-style-type: none"> - What are the challenges of data sharing agreements with different partners (i.e. Govt, private sector, legal capacity challenges, academia etc.)? - What is best practice in agreeing them? - What role does the NSS have to facilitate sharing within and outside the NSS? - How can countries improve their legislation to adopt a data stewardship approach (i.e. access to administrative public or private data, and big data)
	Data security and privacy standards	Technology investments	<ul style="list-style-type: none"> - Does data stewardship require investment in any new technology? - What opportunities are there for public-private partnerships? - How does policies and techniques of security and privacy under official statistics apply to the wider data system?
		Data science investments	<ul style="list-style-type: none"> - How does incorporating data science within the NSO assist with collaboration and partnerships? - Should a methodology unit for data be established?

Outcome	Outputs	Activities	Inputs / needs
Better methods	More integrated data strategy	Non-official statistics	<ul style="list-style-type: none"> - How can these sources be organised and handled with a data stewardship approach? - How can these sources capitalize from linkages with traditional sources of data? - How do NSO engagement on these outputs change with a data stewardship approach and encourage greater responsibility for innovation?
		Mobile phone, social media, web-scraping and scanner data	
		Citizen generated data and citizen data science	
		Earth observations / geospatial imagery	
		Data lab, artificial intelligence	
Better access	More integrated data literacy strategy	Learning (including data science)	<ul style="list-style-type: none"> - How does data stewardship expand the NSO role for informing users and training its own staff, especially on subjects like data science? - How do programmes reach youth and other groups who need more support to use data fully?
	Open data	Adoption of open data practices and expansion of open data policies to more sets of data	- How best to use existing assessment tools (i.e. ODIN)?
	Geospatial information	Integration of geospatial systems and information with the statistical production and dissemination systems	To be developed

Outcome	Outputs	Activities	Inputs / needs
	Microdata	Expansion of microdata data policies to more sets of data	Use of the work being undertaken by the Inter-secretariat Working Group on Household Surveys
	Interoperability / data platforms	Expansion of interoperability and data platforms to more sets of data	To be developed