

# From Principles to Practice: a consultation on joined-up data standards

July 2016

## About this paper

This consultation paper aims to start a discussion on what technical and political solutions to joining up data standards could look like at a global level.

The consultation builds on the research that we have undertaken within the Joined-up Data Standards (JUDES) project over the past year. It works on the assumption that data standards are relevant wherever data exists and is used and recognises the opportunity that the 2030 Agenda for Sustainable Development presents to help develop lasting solutions for currently entrenched problems. Our thematic focus has been on how global data standards impact on the needs of developing countries in both meeting and monitoring these goals. We recommend what policy coordination mechanisms are needed at the international level to ensure that suitable technical solutions are found and implemented.

There are many different types of standards,<sup>1</sup> and as many interpretations of how they should be defined and classified. We focus on ‘measurement standards’ – not basic units of measure, but those that define, classify and group data – that are used in more than one country, often across multiple subject matter domains, and are governed by global or representative institutions that play credible and authoritative roles in upholding standards.

The consultation paper is divided into five chapters. Chapter 1 explains the data challenges that exist, the rationale for this project and the methodology we have employed in our research. Chapter 2 sets out our research findings in joining-up data standards; Chapter 3 describes the international policy landscape; and Chapters 4 and 5 present our preliminary conclusions and recommendations. Consultation questions can be found at the end of each chapter, and also all collated at the end of the paper.

We hope to receive a broad range of responses to this consultation from international standard-setting institutions, open data standard initiatives, data producing organisations and the users of development data themselves. We invite responses to this consultation to be submitted by 31<sup>st</sup> April 2017, using this [feedback form](#) and will engage in a series of international events to consult on the paper in the coming months. More details on the consultation process will be published on our website, [juds.joinedupdata.org](http://juds.joinedupdata.org) in the near future.

## Acknowledgements

This paper was prepared by Beata Lisowska and Tom Orrell and edited by Bill Anderson. We would like to thank the many people who have helped in putting this consultation paper together: it is the culmination of our collective work to date. We would like to particularly thank all of our colleagues at Development Initiatives, Publish What You Fund, and Jen Claydon, our copy-editor. We would also like to thank the Omidyar Network for its support.

We would also like to express our collective gratitude to all the statisticians, data and policy experts from governments, multilaterals, think tanks and civil society who have taken the time to engage with us and have helped shape our thinking over the past year. The opinions expressed here are, however, our own.

## Acronyms

API: Application program interface  
COFOG: UN Classification of Functions of Government  
CRS: Creditor Reporting System  
DAC: Development Assistance Committee  
DHS: Demographic and Health Survey  
EUROSTAT: Statistical Office of the European Union  
FAO: Food and Agriculture Organization (UN)  
FS: Fragile state  
HIPC: Heavily indebted poor country  
HLG: High-Level Group  
IAEG: Inter-Agency and Expert Group  
IATI: International Aid Transparency Initiative  
ICT: Information communications technology  
IDB: Inter-American Development Bank  
IMF: International Monetary Fund  
JUDS: Joined-up Data Standards  
LDC: Least developed country  
LHD: Low human development country  
LIC: Low income country  
LSMS: Living Standards Measurement Study  
MDG: Millennium Development Goal  
MICS: Multiple Indicator Cluster Survey  
NSO: National statistical office  
ODA: official development assistance  
OECD: Organisation for Economic Co-operation and Development  
SDG: Sustainable Development Goal  
SIDS: Small island developing state  
SKOS: Simple Knowledge Organisation System  
UN: United Nations  
UN-OHRLLS: UN Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States  
UNCTAD: UN Conference on Trade and Development  
UNDP: UN Development Programme  
UNECA: UN Economic Commission for Africa  
UNESCO: UN Educational, Scientific and Cultural Organization  
UNICEF: UN Children's Emergency Fund  
UNSC: UN Statistical Commission  
USAID: US Agency for International Development  
WHO: World Health Organization

# Contents

Executive summary .....	4
1. Introduction.....	5
1.1 Data production and use .....	5
1.2 The complexity of competing data systems.....	6
1.3 The Joined-up Data Standards (JUDS) project .....	8
1.4 Research approach.....	9
1.5 Discussion papers .....	11
The SDGs: joining-up new standards in a disconnected world .....	11
SDG 2: joining-up new standards for ending hunger .....	11
Small island developing states: a case study of standards in defining supranational regions and groupings.....	12
Household surveys: do competing standards serve country needs? .....	12
2. Research findings.....	13
2.1 The world is full of competing standards.....	13
2.2 Joining up existing standards involves numerous challenges ....	16
2.3 New standards are not yet following good practices .....	17
2.4 A checklist for new data standards .....	18
3. Principles, institutions and policies .....	19
3.1 The principle of interoperability is established.....	19
3.2 The standard setting landscape is complex.....	20
3.3 The digital divide is far from being bridged .....	21
3.4 Open data and multi-stakeholder initiatives offer breadth and depth	22
3.5 Agendas, revolutions and transformations.....	23
4. Conclusions .....	26
4.1 The policy environment is conducive to joined-up data standards initiatives	26
4.2 Turning new principles into practice is a challenge .....	26
4.3 Solutions are demonstrably achievable .....	26
5. Recommendations.....	27
5.1 New standards must be ‘joined-up’ .....	27
5.2 We need joined-up leadership.....	27
5.3 We urgently need translation services .....	27
6. Questions for consultation .....	28
Notes .....	28

## Executive summary

Data has immense potential to help drive poverty eradication and international development, but it remains incredibly difficult to join up data on resources, people and results because it is published in different formats and to different standards. Overcoming this challenge requires both technical solutions and political will. The challenge facing governments, international institutions, civil society, academics and the private sector alike is how to make sense of the vast quantities of data now being generated in order to create a coherent, holistic picture.

The aims of the Joined-up Data Standards project are to: explore the problems caused by incompatible data in international development; work with partners to find common solutions to these problems; and to build international consensus that all data should be joined up. We have to date produced four discussion papers that cumulatively explore: the way in which global institutions define and classify geographic, sectoral and results data; the overlaps that occur between competing standards; and the policy landscape that governs international data standard setting.

This paper aims to start a discussion on what practical solutions to joining up data standards could look like.

We have reached the following preliminary conclusions.

- **The policy environment is conducive to joining up data**  
Global and regional institutions are recognising the value of joined-up data and interoperability is now an internationally accepted principle. Official statistics bodies increasingly recognise the importance of embracing all producers and users of data as partners in their work.
- **Turning new principles into practice is a challenge**  
While international commitments on interoperability are now a given, standard-setting work still takes place in highly specialised forums and data silos persist, limiting the comparability of data.
- **Solutions are demonstrably achievable**  
Technologies are allowing machines to speak more easily to each other, to understand different languages and to translate between them.

We suggest three preliminary recommendations that form the starting point of our consultation.

- **New standards must be joined up**  
There is no longer any reason why new standards that duplicate existing standards should be developed. All standard-setting bodies must commit to making new standards and their components fully compatible with existing standards and build interoperability into their architecture from the outset.
- **We need joined-up leadership**  
We need united and integrated leadership, structures and mechanisms to drive the Data Revolution and the Transformative Agenda for Official Statistics forward at speeds commensurate with both the aspirations and urgency of current global ambitions.
- **Translation services are urgently needed**  
In the immediate future, the many disparate standards that relate to the Sustainable Development Goal (SDG) goals, targets and indicators need to be mapped and compared to enable data from different standards to be 'cross-walked' through a translating machine. This is the responsibility of all standard-setting bodies.

We are inviting responses to a number of questions outlined below and will also be holding a series of events to consult on the paper in the coming months. For more details on the consultation process see [juds.joinedupdata.org](https://juds.joinedupdata.org)

# 1. Introduction

## 1.1 Data production and use

More data is being produced than ever before and more of it is publicly available. Improving connectivity and data processing technology allows for increasing numbers of users with increasingly powerful machines to access increasing numbers of datasets.

Most of this expansion has taken place in the developed global North – the United States (US) government portal, for instance, contains over 180,000 datasets from over 70 separate departments or agencies.<sup>2</sup> However, it is becoming a global phenomenon. On a different scale, but no less important, 700 unique primary sources of development-related data have been identified in Uganda with 87% of these accessible online<sup>3</sup>.

One outcome of this digital revolution is the increasing use of data to formulate evidence of needs and performance that is fed into policy and decision-making processes: a process that requires joining up data on inputs with that on outcomes.

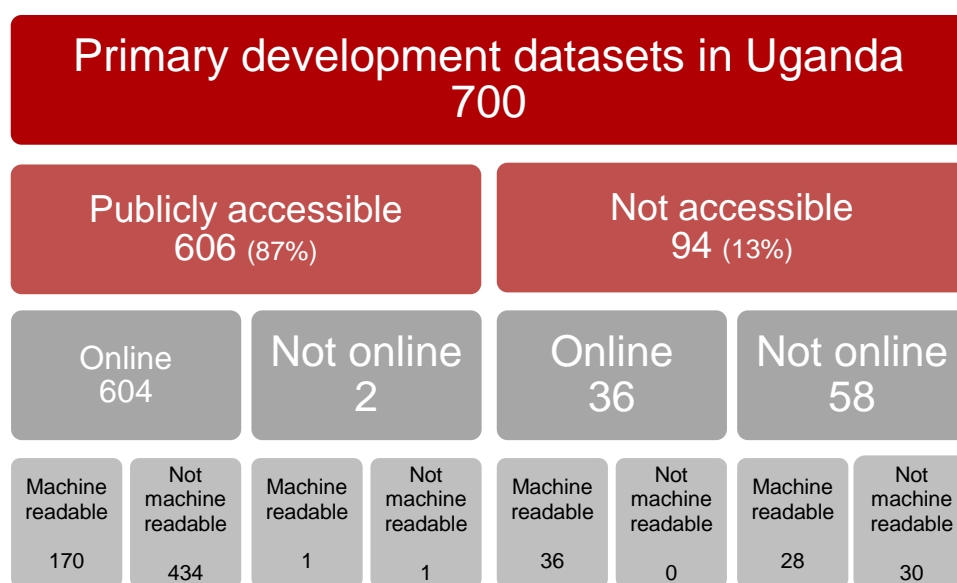


Figure 1: Availability of primary development data in Uganda

Another outcome of the digital revolution is the increasing use of data by civil society to hold government and the private sector to account. The emerging professions of data journalism and data science are indicative of this trend. They fulfil the function of translating and contextualising complex data from different sources into meaningful information. They join up data.

The production of vast quantities of data does not automatically guarantee its usefulness. One-third of the datasets found on the United Kingdom (UK) government's portal have never been used.<sup>4</sup> In the period March to May 2016 one-third of the datasets on the US government portal were accessed, but only 4% were downloaded more than 10 times<sup>5</sup> (Figure 2).

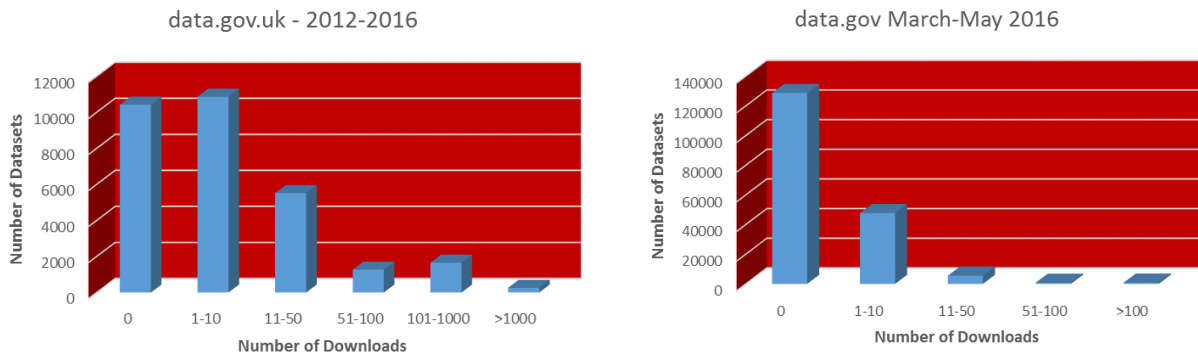


Figure 1: Dataset usage from UK and US government portals<sup>6</sup>

There are a number of reasons why data usage is not keeping up with production: the data may not be particularly useful; users may be unaware of its existence; the volume of data is overwhelming. Most importantly, data users can no longer rely on manual processes and the machines that produce the data, one dataset at a time, are not very good at reading and combining other machine's datasets.

In most cases this is not the fault of the machine, but of the person who designed the dataset in the first place. Machines are good at technical jobs: reading data from different operating systems on different platforms in different file formats and through different application programme interface syntaxes. There is a relatively finite set of mostly standardised rules that a machine needs to be programmed with to navigate these hurdles.

## 1.2 The complexity of competing data systems

When it comes to the data itself the complexity mounts. Figure 3 shows a complex data ecosystem: a mapping of how international and domestic resource flows relate to one another and contribute to (or undermine in the case of illicit flows) development. The challenge facing governments, international institutions, researchers and campaigners alike is how to make sense of the data generated by the institutions involved to create a coherent understanding of the whole. Most of the flows involve financial transactions, so amounts of money, currencies and dates are a common starting point. That is where the simplicity ends. For example:

- Data from the Organisation for Economic Co-operation and Development (OECD) uses country codes that are incompatible with the International Organization for Standardization and the UN
- The UN and OECD define supranational regions differently
- There is no global standard for identifying institutions
- There is no global system for classifying financial flow types
- Functional sectors (such as health, education) are classified in many different ways by, among others, the OECD (Creditor Reporting System [CRS] purpose codes), UN (Classifications of Functions of Government), World Bank (themes and sectors) and the IRS (National Taxonomy of Exempt Entities, used by US foundations)
- There is no standard for defining data labels (the column header in a spreadsheet)

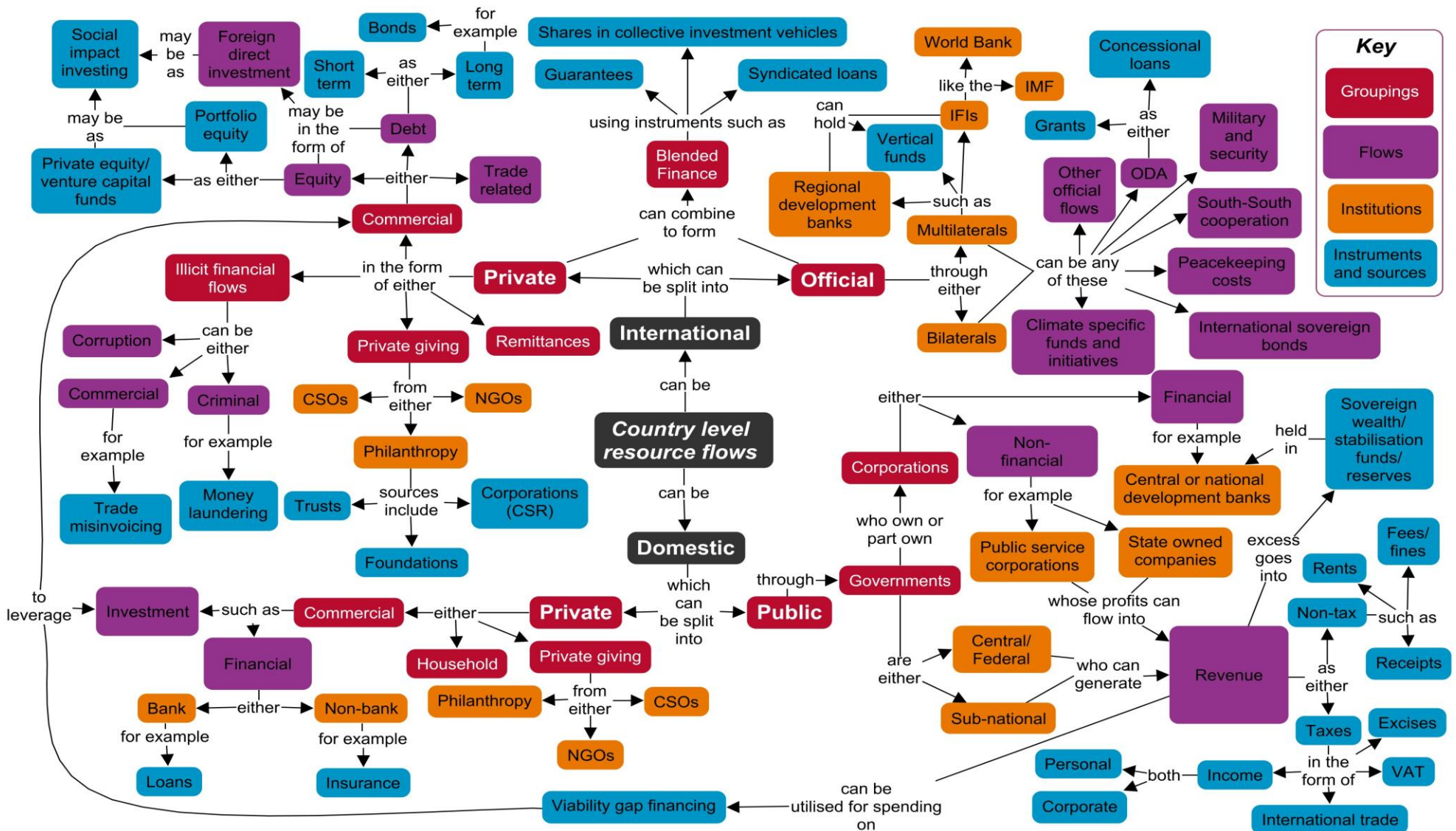


Figure 2: Mapping of international and domestic resource flows<sup>7</sup>

Notes: CSO: civil society organisation; CSR: corporate social responsibility; IFI: international financial institutions; IMF: International Monetary Fund; NGO: non-governmental organisation; ODA: official development assistance; VAT: value added tax

### 1.3 The Joined-up Data Standards (JUDS) project

In November 2013, Development Initiatives and Open Knowledge presented a paper at the Open Government Partnership London Summit that scoped out the data intersections between five transparency initiatives.<sup>8</sup> It argued that although the construction, extractives, aid and contracting standards operated in different spheres, they shared many of the same building blocks – such as geo-referencing, organisation identifiers and sector definitions – and that it was in the interests of data producers and users alike for these standards to therefore share, where appropriate, the same methodologies and coding systems.

Table 1: Common building blocks between open data standards

Data type	CoST	EITI	IATI	Open contracting
Geodata	•	•	•	•
Organisation identifier	•	•	•	•
Contracts	•	•		•
Project identifier	•	•	•	•
Money values	•	•	•	•
Related documents	•	•	•	•
Sector	•		•	•
Transaction/flow type		•	•	
Contract information	•		•	•
Goods (production/export)	•	•		•

Notes: CoST: Construction Sector Transparency Initiative; EITI: Extractives Industries Transparency Initiative; IATI: International Aid Transparency Initiative

In taking the work further it was recognised that most of these building blocks were developed by global institutions, the traditional curators of data standards, and that there are many instances where, as described in the previous section, it is very difficult to compare data from different sources as a result of the competing standards that underpin them. Supported by the Omidyar Network, Development Initiatives and Publish What You Fund – who have collaborated on the International Aid Transparency Initiative (IATI) standard from its beginnings in 2009 – joined forces to work on Joined-up Data Standards in April 2015. The aims of the project are to:

- explore the challenges of joining up standards
- work with partners to find common solutions to these challenges
- build international consensus that all data should be joined up.

Chapter 3 describes the policy gains that have been made in the last couple of years in establishing the principles of data interoperability and comparability. Unlike in 2013, these are now a given in the joined-up data debate. We no longer have to persuade anyone that joining up data is a good idea. We only have to figure out how to do it. The challenge facing all concerned is how to turn these principles into practice.

The aim of this paper is to stimulate this debate. It attempts to distil the findings of our research and engagements over the past year into questions and challenges. It aims to generate discussions with:

- the official statistics community, both at global and national levels
- all those involved in the 2030 Agenda for Sustainable Development (2030 Agenda) who realise that data is a critical element for both meeting and monitoring the Sustainable Development Goals (SDGs)
- the data revolution and open data communities including all those working with innovative technologies and new data sources.



## 1.4 Research approach

In preparing for the 2013 Open Government Partnership summit, two interoperability problems were explored. The first concerned intersections between datasets. Put simply, if two datasets share the same concept (such as 'country') you can only compare the data if both sets use the same definition of 'country'. Moreover, the 'country' in both sets must not only look the same to the human eye, but must be identical so that machines recognise that they are the same.

The second concerned identifying entities, such as institutions and activities. It is almost impossible, for example, to create a single finite list of all institutions or contracts in the world. It is, however, possible to agree a standard methodology that recognises identifiers issued by national agencies that maintain authorised registers of organisations.

Over the past year our scoping research has explored these and other barriers to joined-up data, seeking out contradictions, attempting to understand the timelines and circumstances that have led to unnecessary disconnects, highlighting the challenges and, occasionally, proposing possible solutions.

- We have explored the different ways in which global institutions subdivide the world into a variety of geopolitical and socioeconomic groupings.
- We have explored the ways in which sectoral classifications used for resource flows are difficult to match with the indicator classifications used to describe development impacts.
- We have monitored the progress, and sometimes lack of, in developing the SDG indicators, and their connections with existing indicator databases.
- We have investigated the data overlaps that occur between competing household survey programmes.
- We have also ventured unsuccessfully into a number of issues, seeking narratives that have not added up: how data is being shared in tracking the Zika virus and how data from satellite imagery is digitally connected to subnational administrative districts are two such forays that have not produced any meaningful insights.

The research methodology for this project is rooted firmly in semantic technologies using linked data that enables computers to discover and share data. Standards are grouped into projects and mapped as concept schemes in the [JUDS Thesaurus Manager](#).<sup>9</sup> The [Simple Knowledge Organisation System](#)<sup>10</sup> (SKOS) is used to define relationships between concepts both within and across concept schemes. The Thesaurus Manager also hosts a project containing [government identifiers](#)<sup>11</sup> maintained by the [Natural Resource Governance Institute](#).<sup>12</sup>

Table 2: Current contents of the JUDS Online Thesaurus

Project	Concept scheme
Indicators	Millennium Development Goals (MDG)
	Sustainable Development Goals
	World Development Indicators
Sectors	OECD Development Assistance Committee (DAC) Creditor Reporting System
	UN Classification of Functions of Government
	UN International Standard Industrial Classification of All Economic Activities
	National Center for Charitable Statistics National Taxonomy of Exempt Entities
	World Bank themes and sectors
	Development Initiatives' <i>Investments to End Poverty</i>
Supranational regions and groupings	Countries
	Economic groupings: DAC income groups, IMF low income developing countries, UN least developed countries, World Health Organization (WHO) income region, World Bank heavily indebted poor countries, World Bank income groups and World Bank lending groups
	Geographical groupings: IMF regions, MDG regional groupings, OECD DAC geographic regions, Pan-African regions (African Development Bank regions, African Union regions, UN Economic Commission for Africa/UNECA regions), UN geographic regions, World Bank geographic regions
	Health groupings: WHO Global Burden of Disease regions, WHO regions
	Other groupings: Food and Agriculture Organization (FAO) groups, Human Development Index Groups (UNDP), International Telecommunication Union regions, Landlocked Developing Countries (UN), Small Island Developing States (UN), Small States (World Bank), UNESCO regions
	Security groupings: World Bank fragile and conflict-affected situations
Household surveys	Multiple Indicator Cluster Surveys (UNICEF)
	Demographic and Health Surveys (USAID)
	Living Standards Measurement Study (World Bank)

SKOS provides the facility to accurately define relationships between concepts in different standards; mappings can be exact, close, broader, narrower or related. The Thesaurus Manager stores these relationships in both human and machine readable formats (Figure 4). Machines are thus able to automatically navigate their way from one standard to another. A user friendly interface will be in place soon to provide manual and automatic 'translation services' between standards.

Joined-up metadata from the Thesaurus Manager can also be exported in formats to drive visualisations such as [Chord](#)<sup>13</sup> and [Sankey](#)<sup>14</sup> diagrams that simplify the narratives needed to explain relationships between standards.

```

</rdf:Description>
<rdf:Description rdf:about="http://joinedupdata.org/Surveys/how-many-sons-are-alive-but-do-not-live-with-you_women_dhsq7">
  <skos:prefLabel xml:lang="en">
    how many sons are alive but do not live with you_women_dhsq7
  </skos:prefLabel>
  <skos:broader rdf:resource="http://joinedupdata.org/Surveys/reproduction_women_dhsq7"/>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-05-09T09:59:10Z</dcterms:created>
  <skos:broadMatch rdf:resource="http://joinedupdata.org/Surveys/how-many-sons-are-alive-but-do-not-live-with-you-or-how-many-daughter-are-alive-but-do-not-live-with-you_women_mics5"/>
  <skos:narrower rdf:resource="http://joinedupdata.org/Surveys/sons-elsewhere_women_dhsq7"/>
</rdf:Description>
</rdf:RDF>

```

Figure 4: Example from the JUDS Thesaurus of a machine readable 'broad Match' relationship between the Demographic and Health Survey (DHS) and Multiple Indicator Cluster Survey (MICS) household surveys

## 1.5 Discussion papers

Four discussion papers have been published to date and are available on our [website](#).<sup>15</sup> Their purpose, similar to this paper, has been to stimulate discussion on practical challenges and solutions.

### *The SDGs: joining-up new standards in a disconnected world*

The goals, targets and indicators of the SDGs are viewed as three interlinked standards. What are the challenges inherent in creating new standards in the context of their predecessors (the MDGs) and, in the case of the indicators, other existing global statistical databases?

Both the choice of indicator and its associated methodology pose a number of dilemmas to the data expert.

- Do you design a 'pure' indicator that is theoretically precise in its accurate interpretation of the target?
- Do you make a pragmatic choice based on the availability of data?
- Do you make a choice based on the potential interoperability of the data?

In the ideal world all three conditions should be met, but this is rarely possible. What should at least be recognised is that it is critical for the designers of indicators to be fully aware of the importance of getting this balance as 'right' as possible.

### *SDG 2: joining-up new standards for ending hunger*

Digging deeper and pragmatically into one of the 17 goals, SDG 2, this paper highlights the huge amount of work that still needs to be done to build functional indicators that have a credible methodology and have access to both current and historical data.

The biggest lesson learned is that the starting point in the design of standards must involve a review of existing standards. In many instances relevant data sources and methodologies do already exist and need to be more effectively harnessed. An effective monitoring framework can be compiled through conducting a global analysis of all existing indicators, assessing their universality and quality, and co-opting the best and most appropriate into the SDGs.

### ***Small island developing states: a case study of standards in defining supranational regions and groupings***

The differences in how international organisations classify small island developing states are examples of a wider problem facing the interoperability of geopolitical data standards. While a joined-up data standards ‘cross-walk’ approach can be used to link many similar terms between the standards, it is the unmappable gaps that make comparing data challenging and can lead to real economic impact on the countries affected. While this plethora of classifications may be navigable for large global institutions, developing countries may be paying the price for this complexity.

### ***Household surveys: do competing standards serve country needs?***

There are three major international household survey programmes in use around the world. These have become increasingly similar – two-thirds of the questions in the two most widely used surveys are either identical or similar enough to be practically comparable – but each contains unique, useful modules.

This presents developing countries with a dilemma. Do they, at great expense, commission multiple surveys, or do they accept that they cannot afford to collect all the data they require?

There are two ways to solve the problem of competing standards: combine them into one, or establish functional links between them. We argue that the interests of developing countries would be best served by the integration of the three programmes, and that until this is possible it is critical that data from different surveys is capable of being joined up.

#### **Questions for consultation**

- **What, if any, are the main interoperability challenges that you experience in your sector/work?**
- **What do you think is the most effective technical approach to enabling interoperability between international data standards?**

## 2. Research findings

### 2.1 The world is full of competing standards

The existence of competing standards has emerged in different ways within the various pieces of research we have undertaken to date. Three examples highlight this issue.

The Alliance of Small Island States has 44 members and observers. Of these the UN Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States (UN-OHRLLS) recognises 43 of them. UNESCO's list of small island developing states (SIDS) includes 39 of them, the UN Conference on Trade and Development (UNCTAD)'s list only 28. The World Bank maintains a category called 'small states' that contains 31; and the IMF has a similar list that contains 29.

Table 3: Selected members of the Alliance of Small Island States and their inclusion in global institution groupings<sup>16</sup>

Country	Alliance of Small Island States <sup>17</sup>	UN-OHRLLS <sup>18</sup> Small island developing states	UNESCO <sup>19</sup> Small island developing states	UNCTAD <sup>20</sup> Small island developing states	World Bank <sup>21</sup> Small states	IMF <sup>22</sup> Small states
American Samoa	•	•				
Antigua and Barbuda	•	•	•		•	•
Belize	•	•	•		•	•
Cuba	•	•	•			
Dominican Republic	•	•	•			
Guam	•	•				
Guinea-Bissau	•	•	•		•	
Guyana	•	•	•		•	•
Haiti	•	•	•			
Jamaica	•	•	•	•	•	
Nauru	•	•	•	•		
Netherlands Antilles	•					
Papua New Guinea	•	•	•	•		
Puerto Rico	•	•				
Singapore	•	•	•			
Suriname	•	•	•		•	•

The UN, World Bank (WB), IMF and OECD all maintain different systems of classifying the poorest developing countries.

Table 4: Matrix of the number of countries in selected developing countries' categories<sup>23</sup>

Categorisation	Owner	LDC	LIC	LHD	HIPC	FS
Least developed country (LDC)	UN	49	30	39	31	24
Low income country (LIC)	WB		35	28	26	16
Low human development country (LHD)	UNDP			46	32	22
Heavily indebted poor country (HIPC)	WB + IMF				39	18
Fragile state (FS)	WB + OECD					36

Notes: UNDP: UN Development Programme

The geographic region into which Afghanistan falls is also open to interpretation.

Table 5: Geographical classification of Afghanistan<sup>24</sup>

Institution	Geographic region encompassing Afghanistan
IMF	Middle East, North Africa, Afghanistan, and Pakistan
OECD	South & Central Asia
WB	South Asia
UN	Southern Asia
WHO	Eastern Mediterranean

The institutions responsible for these categorisations have perfectly reasonable explanations as to why they need to divide up the world in different ways in order to fulfil their specific policy and operational objectives. This argument would hold if these listings were only used for internal administration. It does not. Firstly, global institutions are keen to project the intellectual frameworks that underpin their policies. Secondly, developing countries need to negotiate the overlapping interests of these institutions. Thirdly an increasing number of data users are attempting to make sense of the world through regional as well as country-level analysis.

In the early 1970s two new systems of classification were introduced to describe activities related to resource flows. The OECD CRS created a list of purpose codes to describe aid flows. The UN Statistics Division created its Classification of Functions of Government (COFOG) to describe government expenditure.

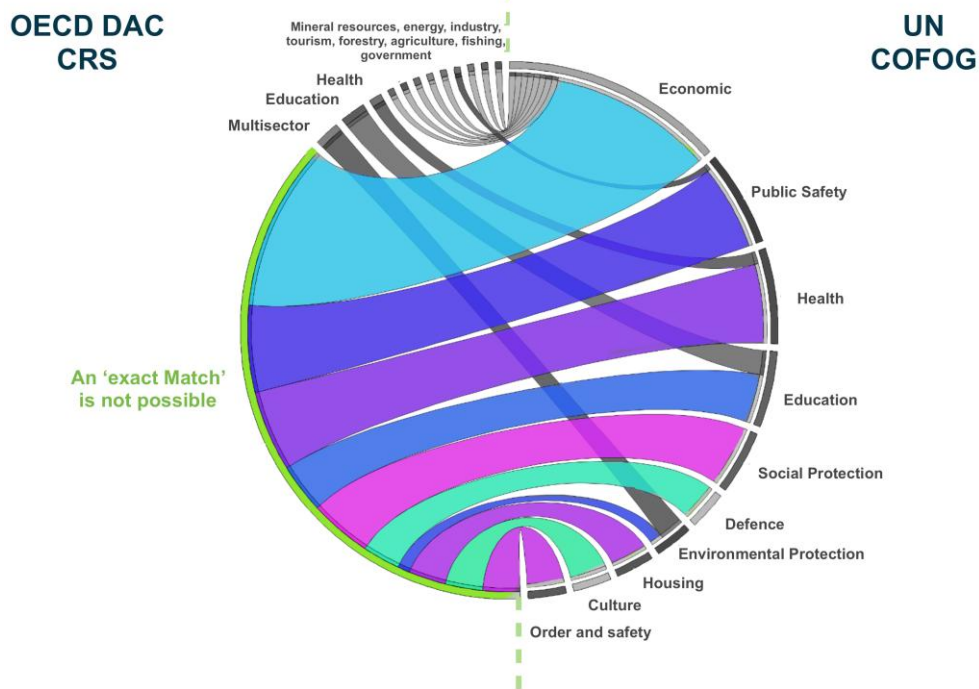


Figure 5: Semantic mapping between UN COFOG and CRS purpose codes<sup>25</sup>

They are very different. The OECD also uses a coding system for countries and regions that is unique to itself. Forty years ago when data was collected on paper and the connections between aid and national development were weak these anomalies were understandable. That it is still virtually impossible to access a single country dataset that contains both international and domestic resource flows illustrates the difficulties getting in the way of joined-up data.

The United States Agency for International Development (USAID), UNICEF and the World Bank run household survey programmes around the world that, in order to meet country needs and global statistical databases, are steadily converging in their content. As Table 6 shows, 77% of all questions found in the UNICEF Multiple Indicator Cluster Survey (MICS) are the same or sufficiently similar to questions in the USAID Demographic and Health Survey (DHS) survey.

Table 6: The extent of question overlap between the latest operational version of MICS and DHS household surveys<sup>26</sup>

Questions	Questionnaire						Total (common questionnaires)	
	Women		Men		Household		MICS	DHS
	MICS	DHS	MICS	DHS	MICS	DHS		
Total number of questions	283	327	111	164	69	53	463	544
Exact or close matches with counterpart	189	178	78	78	49	43	316	299
All matches with counterpart	225	217	82	96	51	48	358	361
Percentage of all matches with counterpart	80%	66%	74%	59%	74%	91%	77%	66%

Data from the DHS and MICS surveys is available for 98 countries. However, not only do the two programmes label their data differently so it can only be merged manually, but different versions of the same survey are coded differently. As Table 7 illustrates this global dataset is, from a data user's point of view, effectively siloed across six incompatible standards. The argument for the pooling of resources is strong.<sup>27</sup> Yet, despite having signed a public statement of collaboration an integrated solution does not appear part of the discussion.

Table 7: Most recent survey data currently available for 98 countries from DHS or MICS household surveys<sup>28</sup>

Demographic and Health Survey (DHS)			Multiple Indicator Cluster Survey (MICS)		
V	VI	VII	3	4	5
<b>2005</b> Moldova <b>2006</b> Azerbaijan, India <b>2008</b> Albania, Bolivia, Sao Tome and Principe <b>2009</b> Guyana, Maldives	<b>2009</b> Lesotho, Timor-Leste <b>2010</b> Armenia, Burundi, Colombia <b>2011</b> Angola, Cameroon, Congo, Ethiopia, Mozambique <b>2012</b> Benin, Cote D'Ivoire, Comoros, Gabon, Guinea, Haiti, Honduras, Indonesia, Jordan, Mali, Niger, Pakistan, Peru, Tajikistan, Tanzania <b>2013</b> DR Congo, Dominican Republic, Gambia, Liberia, Madagascar, Namibia, Nigeria, Philippines, Sierra Leone, Yemen, Zambia <b>2014</b> Egypt, Togo	<b>2014</b> Bangladesh, Burkina Faso, Cambodia, Ghana, Malawi, Senegal, Uganda <b>2015</b> Kenya, Rwanda	<b>2006</b> Guinea- Bissau, Somalia, Syria, Trinidad and Tobago, Turkmenistan, Uzbekistan <b>2007</b> Vanuatu	<b>2010</b> Bhutan, Central African Republic, Chad, Kazakhstan, South Sudan, Suriname, Swaziland <b>2011</b> Afghanistan, Belize, Bosnia and Herzegovina, Costa Rica, Iraq, Jamaica, Laos, Macedonia, Mauritania <b>2012</b> Algeria, Barbados, Belarus, Moldova, Saint Lucia, Thailand, Tunisia, Ukraine <b>2013</b> Uruguay	<b>2013</b> Mongolia, Montenegro <b>2014</b> Kosovo, Kyrgyzstan, Nepal, Palestine, Serbia, Sudan, Viet Nam, Zimbabwe
8	39	9	7	25	10

## 2.2 Joining up existing standards involves numerous challenges

We have used the metaphor of language as a simple way of framing the joined-up data debate. If two people who speak different languages want to understand each other they have two options: either to agree to speak the same language or to use an interpreter.

Data standards must accommodate the needs of different stakeholders. The interests of stakeholders in different constituencies are not necessarily the same. The way in which data is organised, classified and described may therefore need to be different.

While we argue that the time is ripe for an integrated household survey, the competition between the three programmes (Living Standards Measurement Study/LSMS, DHS and MICS) has been a key driver in improving overall content, methodologies and quality.

If the content of competing standards allows for accurate one-to-one or one-to-many mappings between all their elements, then it is totally feasible for a machine translation to join them up. Even if the standards contain some unmappable elements this may still be a pragmatic approach that covers most use cases.

The full range of SKOS mappings adopted by this project are useful for providing an accurate portrayal of the relationships between standards, but do not necessarily allow for a machine to decide how to translate a particular concept (for example where a concept in one standard is similar to two concepts in another). For this reason, we are exploring the use of a new relationship – ‘Best Match’ – which instructs a machine which one of multiple options to use.



There are, however, a set of conditions which, if all met, make a compelling case for the integration of existing standards into one:

- It improves the ability of data users to extract maximum benefit
- It allows for a wider range of data to be compared
- It is cost effective
- The disruption to existing use of a standard being made redundant is not excessive
- Translation tools are available to handle legacy data
- There is sufficient political consensus and will to drive the integration.

Table 8 applies these conditions to the case of household surveys.

Table 8: The case for an integrated household survey

Condition	Case
<b>Improves the ability of data users to extract maximum benefit</b>	Country statisticians and data users generally will be able to combine survey data across programmes and years in a single dataset.
<b>Allows for a wider range of data to be compared</b>	Survey data needs to be included and compared with a variety of other data such as resource flows, statistics from administrative data. A single standard would increase the incentive for such mappings by reducing the complexity of doing multiple mappings.
<b>Cost effective</b>	Countries will be able to build more sustainable infrastructure and capacity by investing in one rather than multiple programmes.
<b>Disruption to existing use of a standard being made redundant is not excessive</b>	Data is only currently available to statisticians who have the experience and capacity to convert and adjust it.
<b>Translation tools are available to handle legacy data</b>	There is currently a strong argument for the three programmes to perform a joint mapping exercise. This has already been done by the JUDS project for DHS-VII and MICS-5.
<b>There is sufficient political consensus and will to drive the integration.</b>	The three programmes have agreed to increase collaboration. The UN Statistical Commission, through the Transformative Agenda for Official Statistics, is calling for the mainstreaming of an integrated household survey.

The increasing prevalence of accessible data, machine readable interfaces (application program interfaces/APIs), and technologies such as Linked Open Data in general and SKOS in particular, provide new opportunities for connecting standards. The work that the JUDS project is doing to build an online repository of mapped standards as a shared public good provides proof-of-concept as to how easy it now is for standard-setting institutions to engage in such collaborative exercises.

This work requires three core commitments:

- the collaboration of subject matter experts
- investment in sound technical infrastructures
- the political will to make it happen.

### 2.3 New standards are not yet following good practices

The goals of the 2030 Agenda have been adopted and embraced with almost unanimous consensus and support. Constructing the associated indicators has proved to be far more challenging. The meeting of the UN Statistical Commission in March 2016 recognised that work has only reached “a practical starting point” and that “the development of a robust and high-quality indicator framework is a technical process that will need to continue over time”.<sup>29</sup> The Inter-Agency and Expert Group (IAEG) tasked with developing the indicators wisely devised a tier system to assess the status of the indicators. The preliminary work done on these tiers in the lead up to the IAEG’s meeting at the end of March 2016 produced the following startling

picture (Table 9) with little more than one third of the indicators assessed to have both a credible methodology and data available from most countries.

Table 9: Definition of tiers used to define status of SDG Indicators<sup>30</sup>

Tier	Definition	No. Indicators as of March 2016
I	Indicator conceptually clear, established methodology and standards available and data regularly produced by countries.	99
II	Indicator conceptually clear, established methodology and standards available but data are not regularly produced by countries.	54
III	Indicator for which there are no established methodology and standards or methodology/standards are being developed/tested.	85

What makes this more surprising is that the architects were not starting from a blank page. They inherited the 15 years of data and experience from the MDGs and they could draw on the existing corpus of global statistics. Moreover, a pragmatic approach to SDG 3 (Health) could be to rely more heavily on the World Health Organization’s (WHO) *Indicator and Measurement Registry*.<sup>31</sup> A pragmatic approach to filling the SDG 2 (Hunger) indicators that do not have methodologies could be to adopt a similar indicator already maintained by the Food and Agriculture Organization (FAO) or others.

These assertions are made purely from a standards point of view. The subject matter experts who have proposed Tier II and III indicators have no doubt excellent reasons for challenging statisticians and development agencies to produce the data that will provide the correct answers to their most pressing questions.

There is, however, a danger that the ‘purist’ approach will lead to elegantly defined indicators for which no data exists, or, even worse, for which elaborate algorithms will be created to forge the appearance of data when in fact none exists.<sup>32</sup>

## 2.4 A checklist for new data standards

The lessons learned from our research in general, but particularly from monitoring the progress of the technical and political work behind the work of the IAEG-SDGs leads us to suggest a work-in-progress checklist for consideration when embarking on a new data standard:

- Is there a clear need and demand?
- Does it duplicate the efforts of or compete directly with standards that already exist?
- Is the design of the architecture and individual elements intellectually, logically and methodologically sound?
- Do components (building blocks) within the standard adopt other existing standards wherever possible?
- Is it designed to ensure comparability and interoperability with other standards?
- Will the data be available through open, sustainable and easily accessible channels?
- Is there political buy-in from the institutions that will drive the standard until it gains acceptance?
- Is there political buy-in from the institutions that need to produce the data?
- Are timelines for development, implementation and adoption realistic?
- Does the data already exist that can feed the standard?
- Is it realistic to expect that new data can be produced to feed the standard?
- Does any historical data exist that can act as a ‘rear view mirror’ for the standard?

#### Questions for consultation

- Should similar international standards be integrated? What would be the benefits/drawbacks of this?
- Is there a need for a globally recognised checklist for new data standards to ensure interoperability is built into new standards from the outset? If so, what should it include?
- What do you think is the most effective way of joining up the standards that underpin the SDG framework?

### 3. Principles, institutions and policies

#### 3.1 The principle of interoperability is established

A number of major documents published in the past two years dealing with data and development have recognised the importance of joined-up data standards (Box 1). The call has moved beyond the need for open data alone, as typified in the [G8 Open Data Charter](#) of June 2013, to the recognition that standards are the essential foundations for comparable and interoperable data.

#### Box 1: Recognition of the importance of interoperable standards

Comparability and standardisation are crucial, as they allow data from different sources or time periods to be combined, and the more data can be combined, the more useful they are.

*[A World that Counts](#)* – November 2014

I recommend that, under the auspices of the Statistical Commission of the United Nations, a comprehensive programme of action on data be established. This includes the building of a global consensus, applicable principles and standards for data, a web of data innovation networks to advance innovation and analysis, a new innovative financing stream to support national data capacities and a global data partnership to promote leadership and governance.

*[Synthesis report of the UN Secretary-General on the post 2015 sustainable development agenda](#)*  
– December 2014

Innovations also have to be introduced in promoting metadata driven production, data and metadata exchange and IT processes, which will allow further standardization of statistical processes, analytics and applications for which close and enhanced national, regional and international collaboration will be necessary in building and implementing common statistical products and services.

*[Transformative Agenda for Official Statistics](#)* – January 2015

We will support efforts to make data standards interoperable, allowing data from different sources to be more easily compared and used.

*[Addis Ababa Action Agenda](#)* – July 2015

Engage with domestic and international standards bodies and other standard setting initiatives to encourage increased interoperability between existing international standards, support the creation of common, global data standards where they do not already exist, and ensure that any new data standards we create are, to the greatest extent possible, interoperable with existing standards.

*[International Open Data Charter](#)* – September 2015

The IAEG-SDGs second work-stream, on identification of inter-linkages across goals and targets, is currently developing a methodology to encourage a cross-sectoral monitoring of the goals and targets. In doing so, the group hopes to move from the current indicator long-list towards a reduced, more integrated set.

*[Interlinkages working group of IAEG-SDGs](#)* – March 2016

While there is a clear global *recognition* of the need to move beyond data silos, there is no clear consensus yet on what a global *solution* looks like.

### 3.2 The standard setting landscape is complex

The UN Statistical Commission (UNSC), representing the interests of the national statistics offices of all UN members, sits at the centre of a constellation of global and regional standard-setting institutions. It hosts an array of technical initiatives. Their names reflect the complexities of UN governance arrangements. Much of the work of these groups, supported by a relatively small secretariat, the UN Statistics Division (UNSD), is concerned with maintaining and coordinating standards in their specific areas of expertise. They report back to the annual UNSC meetings but do not, on the face of it, appear to collaborate *between* themselves.

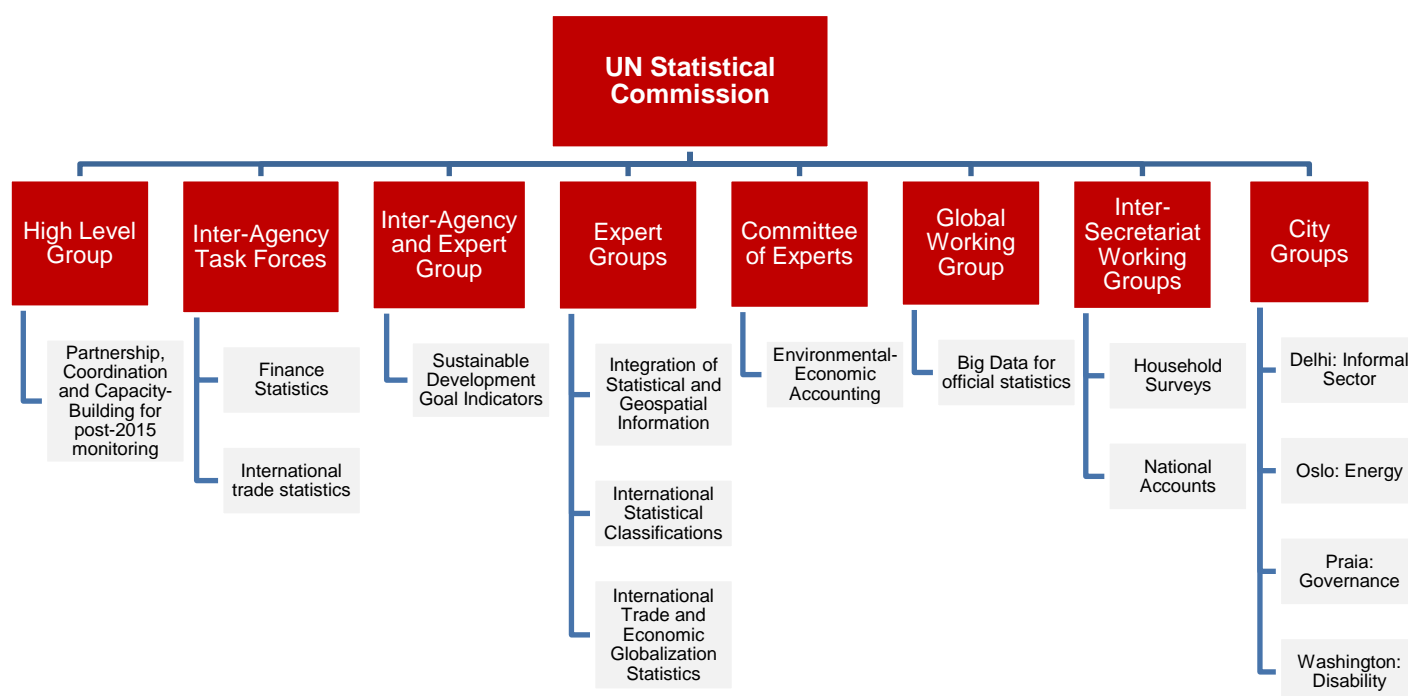


Figure 6: Selected technical groups, committees and task forces that report to the UN Statistical Commission

The Intersecretariat Working Group on National Accounts, for example, is comprised of five members: the European Commission, IMF, OECD, UN and World Bank. For over 30 years it has overseen the UN System of National Accounts, the internationally agreed standard set of recommendations on how to compile measures of economic activity.

Over the past couple of decades, the UNSD has facilitated the establishment of informal and ad hoc **City Groups**, tasked with addressing “selected problems in statistical methods” in an agile and timely manner. According to the UNSD, “It is important to note also that technical expertise and certainly practical experience resides mainly with national statistical offices. It was, therefore, recognised that these informal consultation groups are an innovative way to use country resources to improve and speed up the international standards development process.”<sup>33</sup>

The **Washington Group on Disability Statistics**, for example, has been tasked with coordinating international cooperation “in the area of health statistics by focusing on disability measures suitable for censuses and national surveys. The aim is to provide basic necessary information on disability which is comparable throughout the world.”<sup>34</sup> The Group’s work in developing a discrete set of internationally comparable indicators has been very successful, with new internationally-comparable survey modules containing disability indicators being agreed within a relatively short space of time.

Other UN departments and agencies (notably WHO and FAO), the World Bank, the IMF, the OECD and Eurostat also play key roles in global standards setting. These international and supranational agencies are part of the Coordination of Statistical Activities (CCSA) hosted by the UNSC.

The African official statistical landscape is a good example of where the need for cooperation to enable harmonisation between standards is explicitly recognised. The three pan-African institutions – the African Union Commission, the UN Economic Commission for Africa (UNECA) and the African Development Bank are tasked with coordinating statistical activities across the continent.

For example the Committee of the Directors-General of African National Statistical Offices, established by the African Union Commission, now shares its annual conference with the Statistical Commission for Africa established by UNECA.<sup>35</sup> Together they represent the highest decision-making body for statistical coordination and supervise the work of 14 technical working groups.<sup>36</sup> They also oversee the work of the AU Institute for Statistics<sup>37</sup> and the Economic Commission for Africa-led African Group on Statistical Training and Human Resources.<sup>38</sup> An annual Africa Symposium on Statistical Development is also held, which focuses on specific challenges facing African statisticians.<sup>39</sup>

This political commitment to an integrated approach is further reflected at a policy level. The African Charter on Statistics<sup>40</sup> and the Strategy for the Harmonisation of Statistics in Africa<sup>41</sup> are forward-looking documents that rise above the capacity challenges faced by many African countries and lay out road maps towards data quality and compatibility.

“The development of comparable statistical data, across time and space, on the continent calls for the adoption of harmonised and standardised definitions and concepts; the adaptation of international norms to African realities and specificities, and the utilisation of common methodologies for statistical production and dissemination by all African countries.”

*Strategy for the Harmonisation of Statistics in Africa*

### 3.3 The digital divide is far from being bridged

Much is made of the role of technology in creating a new level playing field for the whole world. Much has been written about the benefits that internet connectivity, innovative ICT and big data can bring to developing countries. This is all true, but it is still work in progress. In Uganda, as outlined in Table 10, radio remains the main source of information and while 75% of adults in Africa own mobile phones, less than 20% of them have smartphones with internet connectivity (see Figure 7). The SDGs call for internet access for all by 2020 while a report by the [Alliance for Affordable Internet](#) estimates that at current growth rates, this target will only be reached by 2042.<sup>42</sup>

Main Source of Information	2002	2014
Radio	49.2	55.2
Word of mouth	47.8	19.7
Internet	0.6	7.3
Television	0.6	7.2
Other	0.9	10.6

Table 1: Main source of information in Uganda<sup>43</sup>

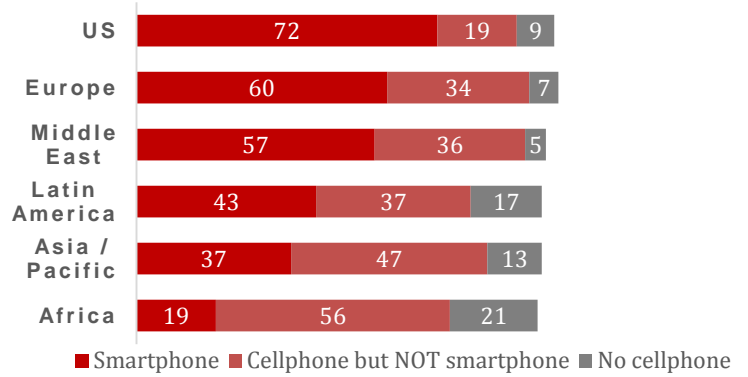


Figure 7: Regional medians of adults owning cell phones

This digital divide applies to data standards as well. The leading standard setters in the world are global institutions – the UN; the World Bank and IMF; the OECD and Eurostat – based in the developed North and engaged in collecting data to fill global datasets. National statistics offices in developing countries spend much of their limited resources servicing global databases rather than the needs of their own governments.

Developing countries need socioeconomic data disaggregated down to the lowest level of geographic administration in order to plan and monitor the delivery of services. Global databases of comparative statistics only require national aggregates. Household surveys have been a pragmatic and relatively cheap solution in creating a corpus of global statistics. Their results are also of use to national policy makers, but have little benefit for subnational planning and service delivery. Uganda is governed through 112 district administrations. Surveys provide data disaggregated by 10 statistical domains. The arithmetic of survey pricing would require a tenfold increase in cost to produce district data of use to local government.

Standards do not produce money, but the acceptance by standards bodies of a status quo that prioritises globally compatible national statistics over subnational data – global monitoring over national planning – is, to a large extent, responsible for the underinvestment in sustainable data infrastructures in developing countries. Developing countries need the same systems as their more developed counterparts: civil registration systems that use real numbers to count births and the causes of death; health and education management information systems that are used to make health facilities and schools function properly. Data has become political, and standard setters are, whether they choose to be or not, involved in the debate.

### 3.4 Open data and multi-stakeholder initiatives offer breadth and depth

Multi-stakeholder transparency initiatives have emerged over the past three decades – the International Budget Partnership in 1997, the Extractive Industries Transparency Initiative in 2002, the International Aid Transparency Initiative (IATI) in 2009, OpenCorporates in 2010, the Open Government Partnership in 2011, and the Open Contracting Partnership in 2012. This has led to new voluntary open data standards that borrow ‘building-block’ components from, and require interfaces to, official statistics.

Not only do these initiatives borrow from the official world. They see the benefit of engaging with traditional standard setters in order to improve their content. The relationship between IATI and the OECD is a case in point. When the IATI publishing standard was drafted it borrowed heavily from the reporting guidelines of the OECD CRS. One key building block inherited from the CRS is the list of purpose codes used to classify functional sectors. An IATI working group has spent the last five years establishing a better method of linking aid flows to recipient country budgets and, in April 2016, the OECD’s Working Party on Statistics agreed to include over 50 new codes in the CRS standard – an unprecedented breakthrough in the relationship between a voluntary

multi-stakeholder standard and a mandatory reporting system used by the world's wealthiest countries.

At a national level the dividing line between official and 'non-official' statistics is also becoming increasingly blurred. National statistics offices are beginning to reach out to other data producers and users while civil society, the private sector and academia are looking to more inclusive relationships with government.

### 3.5 Agendas, revolutions and transformations

In September 2015 the UN General Assembly adopted the 17 SDGs together with their accompanying 169 targets "to stimulate action over the next 15 years in areas of critical importance for humanity and the planet". Some 230 indicators (still not finalised at the time of press) are being designed to ensure that the goals can be met and monitored.

The adopted document recognised the critical role of data, and, in particular, the need to overcome considerable challenges:

"We will support developing countries, particularly African countries, least developed countries, small island developing states and landlocked developing countries, in strengthening the capacity of national statistical offices and data systems to ensure access to high-quality, timely, reliable and disaggregated data. We will promote transparent and accountable scaling-up of appropriate public-private cooperation to exploit the contribution to be made by a wide range of data, including earth observation and geospatial information, while ensuring national ownership in supporting and tracking progress."<sup>44</sup>

2030 is an important political target to focus the minds and activities of the world. It would, however, be naïve to assume that the data challenges facing all developing countries – for example, fully functioning civil registration and vital statistics systems producing credible data on the causes of death<sup>45</sup> – can be solved within 15 years. There is a danger that to satisfy short-term data needs to meet short-term political goals, longer-term investments in sustainable data systems may be overlooked.<sup>46</sup> There is, however, increasing evidence that this will not be the case.

In July 2012 the UN Secretary-General created a High Level Panel of Eminent Persons to make recommendations that led to the 2030 Agenda. Its report, published in May 2013<sup>47</sup> called for "a data revolution for sustainable development, with a new international initiative to improve the quality of statistics and information available to citizens."

In response to this, in August 2014 the UN Secretary-General established an Independent Expert Advisory Group on the Data Revolution for Sustainable Development to provide him with inputs to shape "an ambitious and achievable vision" for a future development agenda.

Within three months the group had published its report, *A World That Counts*, which recommended establishing a UN-led "Global Partnership for Sustainable Development Data".<sup>48</sup>

## **Box 2: Excerpts from key recommendations, *A World that Counts***

- 1. Develop a global consensus on principles and standards:**  
The disparate worlds of public, private and civil society data and statistics providers need to be urgently brought together to build trust and confidence among data users....
- 2. Share technology and innovations for the common good:**  
... mechanisms through which technology and innovation can be shared and used for the common good ...
- 3. New resources for capacity development:**  
... A new funding stream to support the data revolution for sustainable development should be endorsed at the Third International Conference on Financing for Development, in Addis Ababa in July 2015...
- 4. Leadership for coordination and mobilisation:**  
A UN-led 'Global Partnership for Sustainable Development Data' is proposed, to mobilise and coordinate the actions and institutions required to make the data revolution serve sustainable development...

The IEAG worked on a tight deadline in order to feed in to the UN Secretary-General's synthesis report, *The road to dignity by 2030: ending poverty, transforming all lives and protecting the planet*<sup>49</sup> published in December 2014. The report recommended that "under the auspices of the Statistical Commission of the United Nations, a comprehensive programme of action on data be established. This includes the building of a global consensus, applicable principles and standards for data, a web of data innovation networks to advance innovation and analysis, a new innovative financing stream to support national data capacities and a global data partnership to promote leadership and governance."<sup>50</sup>

Well before the publication of these two reports, the UNSD was busy considering its response to the high level panel's recommendations. At the end of February 2014 a seminar was held on 'Managing the Data Revolution', The objective was "to clearly emphasise the strategic necessity of modernising the national statistical systems in order to respond not only to the regular requests for sound official statistics, but also to emerging needs."<sup>51</sup> A few days later the UNSC itself adopted a resolution<sup>52</sup> that led to the convening of the [Global Conference on a Transformative Agenda for Official Statistics](#) in January 2015.

A dominant theme throughout the conference, and notably in the outcome document, was integration and coordination as the approach through which official statistics can make the data revolution happen.<sup>53</sup>



### **Box 3: Proceedings of the Global Conference on a Transformative Agenda for Official Statistics**

The Conference suggested that there is a need to formulate a truly coordinated global statistical programme around the overall objective of modernising statistics and strengthening global, regional and national statistical systems in order to respond more effectively and efficiently to the new policy requirements, such as the post-2015 development agenda.

Moreover, it was agreed to structure the global statistical programme around the specific objectives of:

1. developing and implementing standards and principles
2. integrating statistics with national planning and development
3. supporting transparency and openness
4. promoting the use of new methods, technology and data.

Furthermore, the broad programmatic areas of such a global statistical programme could be further elaborated to include:

1. coordination
2. statistical standards and methods
3. data collection, production and dissemination
4. innovative methods, techniques and IT infrastructure, including standards-based business architecture
5. communication/advocacy including resource mobilisation and technical cooperation assistance and training.

The conference also recognised the importance of stakeholders beyond the world of official statistics. A working group identified the challenge of:

“how to move from a silos system to an integrated system and who drives the process. The head of the [national statistical office] NSO was considered an important and leading figure in this process: the head of the NSO should drive the process and engage his staff within the institution and stakeholders outside its institution.”<sup>54</sup>

In March 2015, the UN Statistical Commission supported the formation of:

“a new high-level group (HLG) to provide strategic leadership for the SDG implementation process; such a group should consist of national statistical offices, and regional and international organisations as observers operating under the auspices of the Statistical Commission. The high-level group is tasked with promoting national ownership of the post-2015 monitoring system and foster capacity-building, partnership and coordination for post-2015 monitoring.”<sup>55</sup>

As the above brief history shows, what brings the official Transformative Agenda, Data Revolution and 2030 Agendas together is that they all recognise the need for clearer coordination between stakeholder groups at the international level. In addition to the efforts made by these UN-led processes to develop coordination mechanisms such as the HLG, the Global Partnership for Sustainable Development Data (GPSDD) was launched in September 2015; having grown out of a broad base of stakeholders responding to the “World That Counts” report.

The HLG and GPSDD represent two sides of the same coin and following the announcements made in March 2016 by the HLG for a Global Action Plan for Data and a World Data Forum, there is now a real opportunity for the two processes to develop closer ties.

### Questions for consultation

- What are the main changes you would like to see at the international level when it comes to the governance of standard-setting processes?
- Should there be a global strategy for the harmonisation for statistics, based on the African example?

## 4. Conclusions

### 4.1 The policy environment is conducive to joined-up data standards initiatives

The principles of comparability and interoperability are now a given. Since the end of 2014 every major global initiative relating to data has included a commitment to the principle of interoperability.

There is an increasing awareness that open data is not an end in and of itself: unless it can be converted into contextualised, usable information it has no benefit. Context invariably involves the joining up of different data, and if that data is not compatible it does not work.

Global institutions are beginning to recognise the value of opening up. National statistics offices as well as regional and global statistics bodies increasingly recognise the importance of embracing all producers and users of data. Parts of the private sector are willing to explore how their data can contribute to the public good. Academia and civil society are beginning to seek partnerships with government. Multi-stakeholder initiatives with representative leaderships can unify divergent interests.

There is currently a lack of mutual understanding and appreciation between many of the advocates of the Data Revolution and the Transformative Agenda for Official Statistics. They are in fact two sides of the same coin and a great opportunity exists for their language and focus to converge and merge.

Most importantly the aspirations of the 2030 Agenda have, for the moment, captured the imagination of the world. This optimism will not last forever. The moment needs to be seized.

### 4.2 Turning new principles into practice is a challenge

At the top of the evolving data ecosystem is a coherent suite of joined-up thinking ready to be turned into effective programmes. At the bottom are a number of highly proficient technical working groups in which highly professional experts seek solutions to the nuts and bolts, but often working on their specialisations in isolation from one another. There appears to be a disconnect between the top and the bottom.

Operationalising new policy consensus takes longer than it needs to. Standards bodies are, correctly, cautious and slow-moving by nature, and not naturally suited to revolution.

Principles and policies

Strategies and plans

Tactics and working groups

### 4.3 Solutions are demonstrably achievable

Notwithstanding the evidence presented in the first two chapters of this paper, there is much to be hopeful about. The number of technical working groups involved in the coordination of standards is testament to the energy that is being thrown into this field. Technologies are

allowing machines to speak more easily to one another, to understand their different languages and to translate between them. Our project has demonstrated the utility of semantic mappings and there are many similar initiatives taking root. Discrete pieces of work being done by some of the UNSD City Groups show how focused approaches can have far-reaching successes.

#### **Questions for consultation**

- **How do you think that specialised working and expert groups working on specific standards or thematic areas could coordinate their work more closely?**
- **How can formal UN and other official statistics bodies better share knowledge and experience with multi-stakeholder initiatives; at the global, regional and/or national levels?**

## **5. Recommendations**

### **5.1 New standards must be ‘joined-up’**

There is no longer any reason for new standards to be created that unnecessarily duplicate existing standards, or that do not build where possible on existing components and methodologies.

All global and regional data standards bodies, official or otherwise, should commit to a simple undertaking: new standards must be joined up.

### **5.2 We need joined-up leadership**

We need an integrated leadership body to drive the Data Revolution and transformative agenda forward at speeds commensurate with both the aspirations and urgency of current global ambitions. As recognised by both the *World that Counts* and synthesis reports, this should be led by the UN. As recognised by all the policy instruments discussed in this paper, it should be a multi-stakeholder partnership.

A solution involving the High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda and the Global Partnership for Sustainable Development Data would appear fitting.

This solution should include a specific commitment and programme to actively promote joined-up data standards throughout the ecosystem.

### **5.3 We urgently need translation services**

Solving the problems caused by existing incompatibilities between data standards will involve journeys with two destinations. The one in the distance involves thorough rationalisations between standards; possibly causing temporary disruptions in data flows; changing the culture of producers and users; and treading on the toes of those resistant to change. In many cases this process may well be deemed to be permanently unfeasible.

The closer destination can, however, always be reached, if only the travellers are prepared to make the effort. We urgently need all standard setters to recognise the benefits available to data users when data from different standards is ‘cross-walked’ through a translation machine.

This is particularly urgent in the immediate future for the many disparate standards that govern data that now need to be mapped and compared with SDG goals, targets and indicators.

### Questions for consultation

- **What do you consider to be the ‘next steps’ in transforming international commitments on the need for interoperability into workable technical solutions?**

## 6. Questions for consultation

- What, if any, are the main interoperability challenges that you experience in your sector/work?
- What do you think is the most effective technical approach to enabling interoperability between international data standards?
- Should similar international standards be integrated? What would be the benefits/drawbacks of this?
- Is there a need for a globally recognised checklist for new data standards to ensure interoperability is built into new standards from the outset? If so, what should it include?
- What are the main changes you would like to see at the international level when it comes to the governance of standard-setting processes?
- Should there be a global strategy for the harmonisation for statistics, based on the African example?
- How do you think that specialised working and expert groups working on specific standards or thematic areas could coordinate their work more closely?
- How can formal UN and other official statistics bodies better share knowledge and experience with multi-stakeholder initiatives; at the global, regional and/or national levels?
- What do you consider to be the ‘next steps’ in transforming international commitments on the need for interoperability into workable technical solutions?
- What do you think is the most effective way of joining up the standards that underpin the SDG framework?

## Notes

---

- 1 Russell A (2014) "Open Standards and the Digital Age". Cambridge University Press. [www.amazon.com/Open-Standards-Digital-Age-Enterprise/dp/1107612047](http://www.amazon.com/Open-Standards-Digital-Age-Enterprise/dp/1107612047).
- 2 USA.GOV (2016) "Federal Agency Participation". Available at: [www.data.gov/metrics](http://www.data.gov/metrics) , accessed on: 4 July 2016
- 3 Development Initiatives (Anderson B and Ntawiha W) (2016) "Uganda's data ecosystem". Available at: <http://devinit.org/#!/post/ugandas-data-ecosystem>
- 4 Data.Gov.UK (2016) "Analysis of site usage statistics". Available at: <https://data.gov.uk/data/site-usage/dataset> ,accessed on: 22 June 2016
- 5 Data.Gov (2016) "Analysis of visitor metrics". Available at: <https://catalog.data.gov/dataset/data-gov-visitor-metrics>, accessed on: 22 June 2016
- 6 Author's analysis based on Data.Gov, UK 2016 and Data.Gov, 2016. Op cit
- 7 From unpublished research by Development Initiatives on international and domestic resource flows
- 8 [Construction Sector Transparency Initiative](#) (CoST), [Extractive Industries Transparency Initiative](#) (EITI), [Global Initiative for Fiscal Transparency](#) (GIFT), [International Aid Transparency Initiative](#) (IATI) and [Open Contracting Partnership](#)
- 9 Joined-Up Data Standards (2015) Online Thesaurus Manager. Available at: <http://joinedupdata.org/>
- 10 W3C Semantic Web (2012) "Introduction to SKOS". Available at: [www.w3.org/2004/02/skos/intro](http://www.w3.org/2004/02/skos/intro)
- 11 Natural Resource Governance Institute (2015) Government entities project in the JUDS Thesaurus Manager. Available at: <http://joinedupdata.org/NRGI.html>
- 12 Natural Resource Governance Institute (2016) Available at: [www.resourcegovernance.org/](http://www.resourcegovernance.org/)
- 13 Wikipedia, 'Chord Diagram'. Available at: [https://en.wikipedia.org/wiki/Chord\\_diagram](https://en.wikipedia.org/wiki/Chord_diagram)
- 14 Wikipedia, 'Sankey Diagram'. Available at: [https://en.wikipedia.org/wiki/Sankey\\_diagram](https://en.wikipedia.org/wiki/Sankey_diagram)
- 15 Joined-up Data Standards website. <http://juds.joinedupdata.org/discussion-papers/>
- 16 Joined-Up Data Standards (2016) "Small island developing states: a case-study of standards in defining supranational regions and groupings". Accessed at: <http://juds.joinedupdata.org/discussion-papers/paper-3-sids/>
- 17 Alliance of Small Island States 2016 membership list. Accessed on 5 July 2016 at: <http://aosis.org/about/members/>
- 18 UNOHRLLS (2016) membership list. Accessed on: 5 July 2016 at: <http://unohrlls.org/about-sids/country-profiles/>
- 19 UNESCO (2016) SIDS list, accessed on: 5 of July 2016 at: [www.unesco.org/new/en/natural-sciences/priority-areas/sids/about-unesco-and-sids/sids-list/](http://www.unesco.org/new/en/natural-sciences/priority-areas/sids/about-unesco-and-sids/sids-list/)
- 20 UNCTAD (2016) SIDS list. Accessed on: 5 July 2016 at: [http://unctad.org/en/Pages/ALDC/Small\\_Island\\_Developing\\_States/SIDS-map.aspx](http://unctad.org/en/Pages/ALDC/Small_Island_Developing_States/SIDS-map.aspx)
- 21 World Bank (2016) Small states list. Accessed on: 5 July 2016 at: [www.worldbank.org/en/country/smallstates](http://www.worldbank.org/en/country/smallstates)
- 22 IMF (2016) Small states list. Accessed on: 5 July 2016 at: [www.imf.org/external/np/pp/eng/2013/022013.pdf](http://www.imf.org/external/np/pp/eng/2013/022013.pdf)
- 23 Fialho D and Bergeijk AG (2014) "Noodles and Spaghetti: why is the developing country differentiation landscape so complex?" Accessed at: [wp.peio.me/wp-content/uploads/PEIO8/Fialho,%20van%20Bergeijk%2004.09.2014.pdf](http://wp.peio.me/wp-content/uploads/PEIO8/Fialho,%20van%20Bergeijk%2004.09.2014.pdf)
- 24 Joined-Up Data Standards (2015) Afghanistan entry in the Supranational Regions and Groupings Project in the JUDS Thesaurus Manager. Available at: <http://joinedupdata.org/geo-pol/af.html>
- 25 Joined-Up Data Standards (2015) Chord diagram created from mappings in the JUDS Thesaurus Manager. <http://joinedupdata.org/Sectors/cofog.html>
- 26 Joined-Up Data Standards (2016) "Household surveys: do competing standards serve country needs?" Available at: <http://juds.joinedupdata.org/household-surveys-do-competing-standards-serve-country-needs/>

- 
- 27 UN Statistical Division (2015) "The concept note on The Transformative Agenda for Official Statistics repeatedly calls for the mainstreaming of an integrated household survey programme". Available at: <http://unstats.un.org/unsd/nationalaccount/workshops/2015/NewYork/Outcome.pdf>
- 28 From research by data scientists at Development Initiatives working on global poverty data
- 29 UN Statistical Commission (2016) "Report on the forty-seventh session". Available at: <http://unstats.un.org/unsd/statcom/47th-session/documents/Report-on-the-47th-session-of-the-statistical-commission-E.pdf>
- 30 IAEG-SDG (2016) "Provisional Proposed Tiers Global SDG Indicators". Available at: <http://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-03/Provisional-Proposed-Tiers-for-SDG-Indicators-24-03-16.pdf>
- 31 WHO (2016) WHO indicator registry. Available at: [www.who.int/gho/indicator\\_registry/en/](http://www.who.int/gho/indicator_registry/en/)
- 32 The way in which maternal mortality is calculated in many countries is a case in point. An algorithm that employs GDP and the fertility rate is used in the absence of real data.
- 33 UN Statistical Division (2016) City Groups. Available at: <http://unstats.un.org/unsd/methods/citygroup/index.htm>
- 34 UN Statistical Division (2016) Washington Group on Disability Statistics. Available at: <http://unstats.un.org/unsd/methods/citygroup/washington.htm>
- 35 UN Economic Commission for Africa (2014) Joint ECA-AUC Statistical Commission for Africa and Committee of Directors General of National Statistics Offices StatCom-Africa/CoDG. [www.uneca.org/statcomdodggim-africa](http://www.uneca.org/statcomdodggim-africa)
- 36 UN Economic Commission for Africa (2014) "First Joint Session of the Committee of Directors General of National Statistics Offices and the Statistical Commission for Africa: Draft annotated agenda". Available at: [www.uneca.org/sites/default/files/uploaded-documents/Statistics/statcom2014/annotated\\_agenda\\_statcom\\_final\\_en.pdf](http://www.uneca.org/sites/default/files/uploaded-documents/Statistics/statcom2014/annotated_agenda_statcom_final_en.pdf)
- 37 African Union (2016) Specialised Agencies, Technical Offices, Treaty Bodies and other Institutions. Available at: [www.au.int/en/organs/agencies](http://www.au.int/en/organs/agencies)
- 38 UN Economic Commission for Africa (2014) "African Group on Statistical Training and Human Resources". Available at: [http://www.uneca.org/sites/default/files/uploaded-documents/Statistics/statcom2014/agrost\\_report\\_-\\_statcom\\_iv\\_-\\_2014\\_en.pdf](http://www.uneca.org/sites/default/files/uploaded-documents/Statistics/statcom2014/agrost_report_-_statcom_iv_-_2014_en.pdf)
- 39 Statistics South Africa (2016) Africa Symposia on Statistical Development. Available at: [www.statssa.gov.za/?page\\_id=2085](http://www.statssa.gov.za/?page_id=2085)
- 40 African Union (2015) African Charter Statistics. Available at: [www.au.int/en/treaties/african-charter-statistics](http://www.au.int/en/treaties/african-charter-statistics)
- 41 UN Economic Commission for Africa, 2010, Strategy for Harmonisation of Statistics in Africa (SHaSA). Available at: [www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/AfDB,%20SHaSA\\_web.pdf](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/AfDB,%20SHaSA_web.pdf)
- 42 Alliance for Affordable Internet (2016) "Affordability Report 2015/16". Available at: <http://a4ai.org/affordability-report/report/2015/>
- 43 Uganda Census (2014) "Percentage distribution of Main Source of Information in the Household, 2002–2014". Available at: [www.ubos.org/onlinefiles/uploads/ubos/NPHC/NPHC%202014%20FINAL%20RESULTS%20REPORT.pdf](http://www.ubos.org/onlinefiles/uploads/ubos/NPHC/NPHC%202014%20FINAL%20RESULTS%20REPORT.pdf)
- 44 UN General Assembly (2015) "Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015". Available at: [www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/70/1&Lang=E](http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E)
- 45 Development Initiatives (2016) Visualisation of current state of registrations of death in Africa. Available at: <http://bit.ly/29Irfd7>
- 46 Anderson, B (2015) "What to do with one and a half billion dollars". Available at: <http://devinit.org/#!/post/what-to-do-with-one-and-a-half-billion-dollars>
- 47 UN (2015) "The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda". Available at: [www.un.org/sg/management/pdf/HLP\\_P2015\\_Report.pdf](http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf)
- 48 UN Data Revolution Group (2014) "A world that counts, mobilizing the data revolution for sustainable development". Available at: [www.undatarevolution.org/report/](http://www.undatarevolution.org/report/)

---

49 UN General Assembly (2014) “Synthesis report of the Secretary-General on the post-2015 sustainable development agenda”. Available at: [www.un.org/ga/search/view\\_doc.asp?symbol=A/69/700&Lang=E](http://www.un.org/ga/search/view_doc.asp?symbol=A/69/700&Lang=E)

50 UN General Assembly (2014) “Synthesis report of the Secretary-General on the post-2015 sustainable development agenda”. Available at: [www.un.org/ga/search/view\\_doc.asp?symbol=A/69/700&Lang=E](http://www.un.org/ga/search/view_doc.asp?symbol=A/69/700&Lang=E)

51 UN Statistical Commission (2015) Global Conference on a Transformative Agenda for Official Statistics: “Towards a Strategic Framework for Statistics in Support of the Post-2015 Development Agenda”. Available at: [http://unstats.un.org/unsd/nationalaccount/workshops/2015/NewYork/NY\\_D2.pdf](http://unstats.un.org/unsd/nationalaccount/workshops/2015/NewYork/NY_D2.pdf)

52 UN Statistical Commission (2014) “Decision 45/103 – Programme review: broader measures of progress. Report on the forty-fifth session of the UN Statistical Commission”. Available at: <http://unstats.un.org/unsd/statcom/45th-session/documents/statcom-2014-45th-report-E.pdf>

53 Joined-Up Data (2016) Analysis on the outcome document for the word ‘integrate’ (107 times) and ‘coordinate’ (114 times). Source of document: <http://unstats.un.org/unsd/nationalaccount/workshops/2015/NewYork/Outcome.pdf>

54 UN Statistical Division and the Statistical Office of the EU (Eurostat) (2015) “Proceedings of the Global Conference on a Transformative Agenda for Official Statistics: Outcomes and Summaries of Sessions” Session 3, Group 2. Available at: <http://unstats.un.org/unsd/nationalaccount/workshops/2015/NewYork/Outcome.pdf>

55 UN Statistical Commission (2015) “Decision 46/101. Report on the 46th session of the UN Statistical Commission”. Available at: <http://unstats.un.org/unsd/statcom/46th-session/documents/statcom-2015-46th-report-E.pdf>