LEAVING NO ONE BEHIND IN THE DATA REVOLUTION

An analysis of four years of High-Level Political Forum reports

Barbara Rosen Jacobson

Summary

• Based on summary documents of the High-Level Political Forum (2014–2017), data is becoming an increasingly central element in discussions around sustainable development.

• The sustainable development goals (SDGs) and their aim of ‘leaving no one behind’ has generated a growing demand for disaggregated data, leading to increased pressure on small and developing countries to report a comprehensive amount of data.

• Digital technologies have given rise to new data forms and sizes that can be used for monitoring the SDGs. Nevertheless, capacities to effectively collect and analyse this data remain unevenly distributed, leading to increased inequality between countries to gather the data that is needed to monitor the SDGs. Capacity building initiatives are of vital importance to bridge this data capability gap.

• Given the large variety of data sources, as well as the growing inequality between countries in their ability to ‘mine’ and use this data, the HLPF discussions signal a need for harmonisation and standardisation of data for the SDGs, to be developed by the international community.

• It remains to be seen whether the increased centrality of data mirrors an in-depth recognition of its importance or is the result of the pursuit of a hype. To adequately respond to the ‘data revolution’, a down-to-earth understanding of the potential, limitations, and challenges of data for development is necessary.

I. Introduction

Since 2015, global development efforts have been guided by the sustainable development goals, 17 global goals that frame the global development agenda. The SDGs have become a central element in guiding discussions on development policy. References to them have become abundant in global forums. And with the SDGs, another term has taken main stage: data. This emphasis on data comes as a lesson learned from the millennium development goals (MDGs). For example, the 2015 MDG Report states that ‘better data are needed for the post-2015 development agenda’ and that we need to ‘measure what we treasure’.¹

From 10 to 19 July 2017, governments, international organisations, scientists, business leaders, and civil society organisations gathered in New York, for the yearly High-Level Political Forum (HLPF). The HLPF is the central
venue for discussion, follow-up, and review of the 2030 Agenda for Sustainable Development and the Sustainable Development Goals. It was conceptualised in 2012, when governments agreed on *The Future We Want*, which put the process of creating the SDGs in motion. The first (one-day) HLPF was held in 2013.

Reading reports on this year’s HLPF and following discussions around the SDGs throughout the rest of the year, one cannot help but notice the growing centrality of data. The MDGs, which preceded the SDGs and determined the development agenda during the period from 2000 to 2015, lacked proper monitoring, evaluation, and accountability. Most worrisome, there were important data gaps that made vulnerable groups invisible in general averages. Coupled with fast-paced innovation, data arose in SDG discussions, and its importance was recorded in the UN Document *A World That Counts: Mobilising The Data Revolution for Sustainable Development*, published in 2014.²

The present study looks into the changes in discussions about data at the HLPF. Has the prominence of data changed over the last four editions? In which context was data mentioned?

II. Methodology

To find out how data has been discussed at the four HLPFs, this study consults two different reporting services: IIID Reporting Services and UN Meeting Coverage files. Both services provide publicly available daily summaries of the discussions. A full list of reports consulted for this study is available in the resources section.

Summary reports provide a great resource for in-depth analysis, as they include the most relevant content for the day in a version broadcasted to a much wider audience. Yet, they cannot provide a complete image and are dependent on the interpretation of the rapporteur. However objective the reporting aims to be, individual rapporteurs select the words that make most sense to them to describe the discussions, and the reports that they produce are therefore a secondary resource.

We analysed the reports using word count techniques (looking for the number of times the term ‘data’ was mentioned) and further categorising the use of the word. Every mention of data was then given a relevant tag, denoting the context in which it had been mentioned. By analysing the tags over time, we can identify how data has been discussed, and whether the focus on discussions about data has changed throughout the last four years.

Each mention of data was given one of the following tags:

1. Monitoring
   The need to collect and analyse data for the SDGs is mainly in relation to monitoring progress towards achieving the goals. The tag ‘monitoring’ was given whenever data was explicitly mentioned in relation to monitoring.

2. Access to data
   The prerequisite for monitoring the SDGs is to have access to data. Data can be obtained through data collection, as well as by consulting open or shared data. Yet, there are important gaps in the availability of data. The following tags were given whenever data was mentioned in reference to access.
   a. Data collection
   b. Availability and data gaps
   c. Data sharing and open data

3. Data type
   With rapid innovation in the data-field, data was mentioned in references to specific forms that are thought of as particularly useful in monitoring the SDGs. These include both innovative forms of data that should be explored, as well as more traditional forms of data that should not be forgotten.
   a. Big data
   b. Geo-data
   c. Qualitative data
   d. Statistics
   e. Crowdsourced data
4. Data quality
While access to data is important, the data itself needs to be of sufficient quality in order to accurately measure progress towards the achievement of the SDGs. Discussions around data often refer to the need for ‘proper’, ‘timely’, ‘reliable’, and ‘high-quality’ data, which are grouped under the heading ‘general quality indicators’. A separate heading is given to disaggregated data, as one of the most prominent quality indicators in data discussions.
   a. General quality indicators
   b. Disaggregated data

5. Resources
Collecting and analysing all of the data necessary to measure the 169 SDG targets is a demanding task that requires both human and financial resources, leading to concerns related to capacity building and funding.
   a. Capacity building
   b. Financial resources

6. Examples
Occasionally, data is referred to when discussing a specific example of an analysis. These occurrences are relatively scarce, bearing in mind that this study is based on summary reports where examples are only occasionally mentioned.

7. Other
Sometimes the term ‘data’ did not directly refer to any of the above-mentioned categories. Despite the generic term, this label might be most interesting to analyse, as it captures the specific topics that were focused on during the different editions of the Forums that cannot be captured by the ways in which discussions about data are usually addressed.

III. Analysis

1. The growing centrality of data in HLPF debates
Table 1 indicates the number of times the term ‘data’ was mentioned in the IISD and UN reporting services. The first Forum, in 2013, was not taken into consideration as it was a one-day event, and can consequently not be compared to the forums from 2014 onwards.

The 2014 forum was one day shorter than the forums in 2015–2017, which means that the cumulative number of data-mentions slightly misrepresents the prominence of the topic in the HLPF discussions. To correct this, we took the average number per day for 2014 and multiplied it by the number of days of the later conferences (8). These scores are indicated with an asterisk.

<table>
<thead>
<tr>
<th></th>
<th>IISD</th>
<th>United Nations</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>19*</td>
<td>58*</td>
<td>77*</td>
</tr>
<tr>
<td>2015</td>
<td>13</td>
<td>54</td>
<td>67</td>
</tr>
<tr>
<td>2016</td>
<td>38</td>
<td>73</td>
<td>111</td>
</tr>
<tr>
<td>2017</td>
<td>52</td>
<td>87</td>
<td>139</td>
</tr>
</tbody>
</table>

Table 1. Total number of data-mentions in IISD and UN reports of 2014–2017 HLPFs
2. The context in which data was discussed (2014–2017)

We are not just interested in knowing whether data was discussed, but also in what context it was discussed, and whether this has changed throughout the years of the HLPF.

2014: Lessons from the MDGs and preparing for ‘the data revolution’

With the HLPF inaugurated on 24 September 2013, the second forum – and first full conference – took place from 30 June to 9 July 2014, under the overarching theme ‘Achieving the Millennium Development Goals and charting the way for an ambitious post-2015 development agenda including the SDGs.’ There was reportedly ‘a great sense of opportunity,’ yet at the same time, participants were in a wait-and-see mode, recognising that most of the work would start after the adoption of the 2030 Agenda. The meeting concluded with the adoption of a Ministerial Declaration (United Nations Economic and Social Council, 2014a).

The number of data-mentions in the reports of IISD and the UN are displayed in Figure 1. We can observe a growing number of mentions of ‘data’ in both reports. Whereas 2014 and 2015 showed similar results, the growth in prominence of data in HLPF discussions really took off in 2016 and continued its rise in 2017.


Figure 1. Number of data-mentions in HLPF reports 2014–2017.

The number of data-mentions in the reports of IISD and the UN are displayed in Figure 1. We can observe a growing number of mentions of ‘data’ in both reports. Whereas 2014 and 2015 showed similar results, the growth in prominence of data in HLPF discussions really took off in 2016 and continued its rise in 2017.

2. The context in which data was discussed (2014–2017)

We are not just interested in knowing whether data was discussed, but also in what context it was discussed, and whether this has changed throughout the years of the HLPF.

2014: Lessons from the MDGs and preparing for ‘the data revolution’

With the HLPF inaugurated on 24 September 2013, the second forum – and first full conference – took place from 30 June to 9 July 2014, under the overarching theme ‘Achieving the Millennium Development Goals and charting the way for an ambitious post-2015 development agenda including the SDGs.’ There was reportedly ‘a great sense of opportunity,’ yet at the same time, participants were in a wait-and-see mode, recognising that most of the work would start after the adoption of the 2030 Agenda. The meeting concluded with the adoption of a Ministerial Declaration (United Nations Economic and Social Council, 2014a).

The context of data-mentions in HLPF 2014, by main field.

Figure 2. The context of data-mentions at HLPF 2014, by main field.
Figure 2 shows the context in which the term ‘data’ was mentioned at the HLPF, as retrieved from the IIISD and UN reports, according to the main categories. Figure 3 depicts this in greater detail, as the data is broken down into subcategories. The figures show that in 2014, data was predominantly mentioned in relation to data quality. Many mentions were rather general, with speakers stressing the need for ‘high-quality’ data. Some discussants were more specific, and clarified the importance of ‘objective’, ‘unbiased’, and ‘reliable’ data. Nevertheless, almost all mentions defined what was needed in the future, without clarifying the current situation in much detail. There was one exception, where a speaker spoke about the ‘sketchy’ data that misrepresented a region.

The need for quality data indicated a lack thereof in the present, making access to data an important topic, including data sharing and open data. The need for data collection was also coupled with the experience of the MDGs, as some mentioned the failure to adequately gather data to monitor progress.

New technologies and data sources could complement data gaps, and participants pointed out the potential of mobile phones and satellite images in particular. Monitoring was sometimes mentioned in combination with data, as well as the lack of capacity and the need for resources. Interestingly, while the reports do not extensively mention the need for capacity building, the only reference to data in the Ministerial Declaration emphasises ‘strengthening of ongoing capacity-building for data collection and analysis in developing countries’.

Other mentions of data included other uses of data (e.g. for ‘fuelling innovation and economic growth’ and as an early warning system), as well as references to the ‘data revolution’ (‘contributing to a global data revolution’; ‘a need for a data revolution to track progress’; and ‘the data revolution should be central to the post-2015 development agenda’). Although never specified in the reports what the speakers actually referred to, the term was most likely inspired by the Expert Advisory Group on a Data Revolution for Sustainable Development that had been appointed by the UN Secretary-General, and that would produce A World That Counts later that year.

2015: ‘Leaving no one behind’: Disaggregation and data collection
The third Forum was convened from 26 June to 8 July 2015, with the theme ‘Strengthening integration, implementation and review – the HLPF after 2015’. The meeting took place as the negotiations about the post-2015 Development Agenda were being finalised, focusing predominantly on its future implementation and review. Participants agreed on the urgency of implementing the Agenda after its upcoming adoption, and agreed that reviews should be ‘voluntary and state-led’. The Forum concluded with the adoption of a Ministerial Declaration.
The context of the data-mentions at the 2015 HLPF are shown in Figures 4 and 5, broken down by main field and subcategory, respectively. While the 2014 Forum focused on the need for quality data in a relatively general sense, the need for disaggregated data in particular was an overriding concern in the 2015 Forum. Although data disaggregation was most often related to gender, other dimensions were also mentioned, such as ‘age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts’.

The need for such a high level of disaggregation puts pressure on obtaining this data, which added more depth to the way in which data collection was discussed. A predominant concern related to capacity building, especially for countries like Small Island Developing States (SIDS) to set up appropriate data collection systems.

Thinking about ways to improve data collection, several speakers mentioned the role of citizens and the opportunity for crowdsourced data. One discussant pointed at the role of civil society in data collection. Compared to 2014,
geospatial data and big data were no longer in focus compared to statistics and crowdsourced data.

Other mentions of data focused on the way in which data is used. One speaker stressed it should be used ‘more effectively’, while another emphasised that data should not just be generated for the sake of data collection, but it should ‘help countries identify progress, gaps and lessons learned’.

**2016: Filling data and capacity gaps to ensure no one is left behind**

The 2016 Forum took place from 11 to 20 July 2016, and was the first since the adoption of the 2030 Agenda. It discussed a wide range of topics under the theme of ‘Ensuring that no one is left behind’, including implementation, mainstreaming the SDGs, and ownership at the national level. In addition, the first round of voluntary national reviews were presented.

**Figure 6. The context of data-mentions at HLPF 2016, by main field.**

**Figure 7. The context of data-mentions at HLPF 2016, by subcategory.**
Figures 6 and 7 present the number of data-mentions by main issue field and issue subcategory, respectively. The 2015 Forum’s emphasis on disaggregation remained in 2016, rather unsurprisingly given its theme of ‘Ensuring that no one is left behind’, as ‘missing data implies missing people’. There was a particular focus on the need for gender disaggregated data as participants noted an ‘incomplete picture of how women and girls were being impacted by the SDGs’. Overall data quality remained a concern as well, with speakers expressing the need for ‘high quality’, ‘reliable’, ‘timely’, and ‘robust’ data.

With an ambitious number of 230 SDG indicators to be monitored, many countries talked about the challenges of collecting data. For developing countries, and especially SIDS, this posed a significant obstacle, which could be mitigated with capacity building efforts. In addition, some called for the international community to analyse the situation and to pool its knowledge to fill data gaps. International agencies were also considered important in making sure that data was comparable between countries. Finally, an additional source of data could come from citizens, and crowdsourced data was mentioned as an important element to be integrated into the monitoring process.

Other data-mentions related to concerns of transforming data into meaningful information for policymakers. Several speakers also noted that data needs to be gathered responsibly with ‘ethical standards and methodologies’ and in a more transparent way. Finally, some participants stressed that while data gathering is important, how it is used is what matters, and a lack of data ‘should not be an excuse for inaction’.

The increased focus on data was mirrored by the Ministerial Declaration that was adopted at the end of the Forum. Paragraph 16 stresses that the availability and use of accessible, timely, reliable and high-quality disaggregated data underpins our efforts to leave no one behind by, inter alia, identifying inequalities. The same paragraph emphasises the need for capacity building in ‘data collection, disaggregation, dissemination and analysis at all levels, taking into account that the global review of the 2030 Agenda will be based primarily on national official data sources’.

2017: Towards harmonisation to fill data gaps and inequalities

The 2017 Forum took place from 10 to 19 July, with the overarching theme ‘Eradicating poverty and promoting prosperity in a changing world’. It was the first time that a number of goals (1, 2, 3, 5, 9, and 14) were reviewed in detail. It also included 43 Voluntary National Reviews.
The need for disaggregated data remained the major hurdle for the 2030 Agenda in discussions at the HLPF (Figures 8 and 9). Many civil society organisations highlighted the need for disaggregated data in their particular issue areas, be it gender, age, disability, or ethnicity. At the same time, countries continue to struggle to obtain even basic data, let alone disaggregated data, leading to calls for the international community to share best practices in collecting disaggregated data. One participant called for a ‘minimum’ set of disaggregated data to mitigate this challenge.

Although data collection poses a challenge for all countries, the 2017 Forum brought into focus the large inequalities in data collection and data capacities around the world. These gaps were only highlighted as the Forum started to look into progress on specific SDGs, and it became clear that there were ‘gaps and distortions in the indicator framework due to limitations in data collection’. Addressing the considerable variety in data collection not only needs capacity building, but also investment, as ‘investment in data is as important as investment in infrastructure.’

Yet, while many emphasised the lack of data, some highlighted the ‘fantastic amount’ and ‘wealth’ of data that has been emerging. In particular, participants urged the exploitation of the potential of big data and machine learning, although others emphasised the continued importance of complementary sources of data, such as statistics and qualitative data. Some participants mentioned that the potential of big data needs to be accompanied with measures related to privacy regulations and transparency.

To address the global differences in data collection, participants turned their attention to the need for coordination and harmonisation, both in the collection of data and between data sources and their analysis (technical interoperability and data comparability). The harmonisation of data could also be used to improve the understanding of the relationship between the SDGs and their targets.

While some of the calls for harmonisation are captured by the other category, this label also includes the increased mentions of data without further explanation, such as ‘the importance of data,’ ‘data remained a challenge,’ or someone who had ‘data’ in his job title ‘Director of Data, Research and Policy’ – apparently data has become so mainstream that it does not require additional clarification or elaboration. Finally, some participants cautioned against the ‘trendiness’ of the topic of data: it should be considered seriously and coherently, while fetishising it should be avoided.

Data was also included in the Ministerial Declaration that was adopted at the close of the meeting. Paragraph 27 stresses ‘the need for improved and coordinated collection, analysis, dissemination and use of statistics and high-quality, accessible, timely and reliable data disaggregated by income, sex, age, race, ethnicity, migration status, disability, geographical location and other characteristics relevant in national contexts.’ Recognising the challenge coupled with the need for disaggregation, the declaration urges further strengthening of collaboration for capacity building and best-practice sharing. The declaration furthermore notes the potential of ‘data-driven decision-making and innovation’ and the importance of building capacity to produce, analyse, and use ‘various forms of data’.

![Figure 9. The context of data-mentions at HLPF 2017, by subcategory.](image)
2. The evolving discussion of data at the HLPF

![Discussion on data issues at HLPF](image1)

*Figure 10. The evolving context of data-mentions at HLPF 2014-2017*

![Context of data-mentions 2014-2017](image2)

*Figure 11. The evolving context of data-mentions at HLPF 2014-2017*
Throughout the past four years, discussions around data at the HLPF seem to have become more central, as well as more in depth. The growth of data as an important element of HLPF discussions can be seen in Figures 10 and 11, broken down by issue category. Whereas the 2014 Forum often referred to the need for ‘better data’ and ‘data collection’, these topics are now being discussed in more detail, while their challenges are becoming increasingly evident. The 2015 Forum was an interesting outlier in our analysis: the number of data-mentions dropped slightly, even though the theme of the meeting – strengthening integration, implementation and review – would intuitively be most closely related to data. Yet it was the only Forum during which the term data was not mentioned in the resulting Ministerial Declaration. The distribution of topics shows general trends: an overwhelming concern related to data quality (including disaggregated data) and access to data. We might be able to distil some trends.

I. The focus on disaggregated data is most likely coupled with the SDGs determination to leave no one behind; an uncompromisable goal that ensures the global reach of the SDGs. At the same time, this emphasis has brought into motion a major need to gather disaggregated data. States are being put under increasing pressure to produce data about all segments of society.

II. The availability of data remains a concern, but the focus seems to be shifting away from the overall quantity of data, towards in their distribution and the inequality of their accessibility. There seems to be a growing realisation that there is sufficient data 'out there', generated by digital technology, but countries greatly differ in their ability to gather it and analyse it.

III. Connected to this concern, there seems to be a growing realisation about the great capacity gap and the need for training and resources in this area, especially in small and developing countries.

IV. The international community seems to be increasingly expected to harmonise data collection efforts, fill in data gaps, and provide standards for data comparison.

V. With (big) data being a growingly ‘trendy’ topic, concerns are slowly generated regarding the need for an informed discussion away from the hype, as well as for additional measures to ensure that data collection and analysis is being done properly, ethically, and transparently.

Overall, this analysis shows that data has become a central element at HLPF debates. The assumptions behind the analysed reports is that what cannot be measured will not be achieved. This in line with the phrases used, for example in the 2015 MDG Report, such as ‘measure what we treasure’ and ‘only by counting the uncounted can we reach the unreached’. In fact, the MDGs’ failure to adequately measure progress brought in motion the recognition of the need to gather data. Coupled with the increased importance of data and data-driven processes in debates and sectors far beyond of the realm of international development, data has grown into an important part of the SDG debates.

The way in which data has been debated has become increasingly detailed. With the operationalisation of the SDGs in 2015, the challenges related to data collection became tangible and immediate, and added an additional dimension of urgency to data-related topics. Despite the advent of new technologies and big data resulting in an enormous amount of data that is being generated daily, the key challenge is now to properly access, combine, and analyse this data. With the pressure of providing increasingly disaggregated data in line with the aim of leaving no one behind, there are growing obstacles of capacity and resource constraints, leading to increased inequality among countries in being able to properly manage this data.

Monitoring is surprisingly invisible in this analysis, even though the main purpose of data discussions is to explore how to monitor the SDGs. The explanation might simply be that ‘monitoring’ is implied in discussions around data, and ‘data’ is implied in discussions around monitoring, making the two almost synonymous.

Whereas there is a clear trend towards more talk about data, the next step will be to translate all these discussions into concrete action. The data hype is most likely here to stay, which begs for a more down-to-earth understanding of what is needed to capture the potential of new types of data, while recognising their limitations and mitigating their challenges. The key will most likely lie in providing resources, developing capacity building programmes, and designing other mechanisms for small and developing countries to better collect and analyse data. Ensuring no one is left behind includes addressing the unequal capabilities to measure inequalities.

### III. Conclusion

- States are being put under increasing pressure to produce data about all segments of society.
- The availability of data remains a concern, but the focus seems to be shifting away from the overall quantity of data, towards in their distribution and the inequality of their accessibility. There seems to be a growing realisation that there is sufficient data 'out there', generated by digital technology, but countries greatly differ in their ability to gather it and analyse it.
- Connected to this concern, there seems to be a growing realisation about the great capacity gap and the need for training and resources in this area, especially in small and developing countries.
- The international community seems to be increasingly expected to harmonise data collection efforts, fill in data gaps, and provide standards for data comparison.
- With (big) data being a growingly 'trendy' topic, concerns are slowly generated regarding the need for an informed discussion away from the hype, as well as for additional measures to ensure that data collection and analysis is being done properly, ethically, and transparently.
Endnotes


Appendix I: List of reports used for the analysis - IISD (in chronological order)


Appendix II: Reports used for the analysis - UN Press Centre (in chronological order)


Ms Barbara Rosen Jacobson is Programme Manager at DiploFoundation (Diplo) and the Geneva Internet Platform (GIP), where she manages several diplomacy and Internet governance-related projects, including Diplo’s activities in the area of Data Diplomacy. In addition, she is part of UN SDG Lab’s Innovation Sprint, a cohort of individuals from diverse sectors and disciplines who share a common passion for advancing progress on the SDGs. Before joining Diplo in 2014, she worked with the Applied Research Institute Jerusalem in the Palestinian Territories and RACI Project in Indonesia. Ms Rosen Jacobson holds a BA in Social Sciences from Amsterdam University College and a MA in International Affairs from the Graduate Institute of International and Development Studies in Geneva.

We look forward to your comments – please e-mail them to barbarar@diplomacy.edu

Diplo’s policy papers and briefs can be downloaded from www.diplomacy.edu/policybriefs

If you are interested in publishing a policy paper or brief with us, please get in touch with Katharina Höne, at katharina@diplomacy.edu

Please cite as: Rosen Jacobson B (2017) Leaving No One Behind in the Data Revolution: An analysis of four years of High-Level Political Forum Reports. DiploFoundation Policy Papers and Briefs, No. 5. Available at https://www.diplomacy.edu/sites/default/files/Policy_papers_briefs_05_BRJ.pdf

Disclaimer: The content of the policy brief represents the personal opinion of the author and should not be attributed to any organisation with which the author is affiliated.