

**Expert Meeting on Statistics for SDGs**

31 March - 1 April 2022, Geneva

Session 2: Data from non-traditional sources (including non-official statistics and their quality)

**Assessment of non-official statistical sources for SDG reporting**  
Prepared by Atanaska Nikolova, Office for National Statistics (ONS)

**I. INTRODUCTION**

1. The 17 Sustainable Development Goals (SDGs) cover a very wide range of topics - from monitoring poached wildlife, to growth rate of GDP, and everything in-between. As the United Kingdom's national statistical institute, the Office for National Statistics (ONS) is responsible for sourcing, reporting and monitoring the SDG indicators for the UK. This is done through the UK's SDG data site<sup>1</sup>, which is managed and populated by a dedicated SDG team.

2. Sourcing appropriate data for the indicators is a very challenging task. In order to provide an accurate picture of the UK, we usually use official sources from official statistics producers. For example, sources from government departments or United Nations' agencies. But for some indicators it is very difficult to find official sources that report exactly what the United Nations (UN) metadata requires, and we then turn to non-official alternatives. For example, we use data from UK's largest cash machine network LINK to report on number of cash machines, which forms part of SDG indicator 8.10.1b: Number of automated teller machines (ATMs) per 100,000 adults.

**II. QUALITY ASSESSMENT**

3. One of the benefits of using non-official sources, such as charities, academia and businesses, is that it promotes collaboration and inclusiveness for the purposes of public good and prosperity. But it also comes with risks. We need to ensure that the statistics and data we use are of good quality and are trustworthy, regardless of whether they are official or non-official. To achieve this, the SDG team at ONS has developed a protocol for assessing non-official sources<sup>2</sup>. It is a tool that allows the UK SDG team at ONS to score a statistical source based on criteria that is key in the production of official statistics, and the specific SDG requirements set by the UN. It is based on a model initially developed by Statistics Netherlands, but we have modified it so it considers general principles from the Code of Practice for Statistics<sup>3</sup>. Although the assessment tool is SDG-specific, it can be adapted to other contexts, and we welcome others to do so. The Code is designed for official statistics in the UK, so we do not expect that non-official statistical sources would satisfy all the principles. We aimed to identify only the most relevant and critical dimensions.

---

<sup>1</sup> <https://sdgdata.gov.uk/>

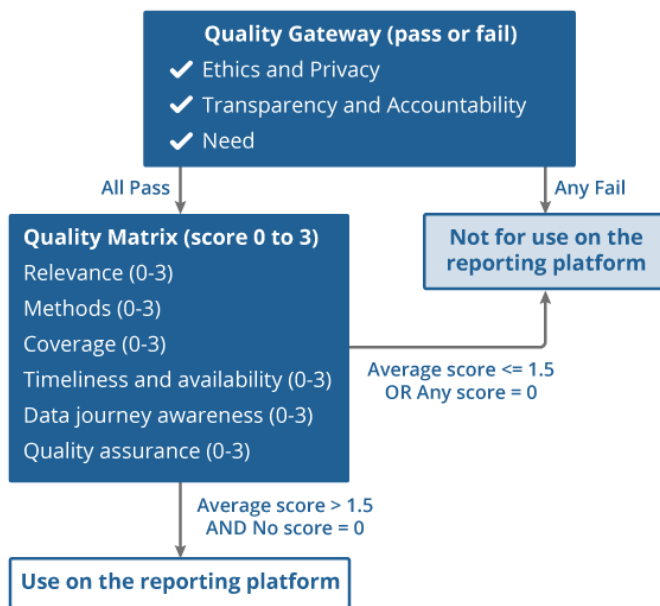
<sup>2</sup>

<https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/uksustainabledevelopmentgoalsuseofnonofficialsources>

<sup>3</sup> <https://code.statisticsauthority.gov.uk/the-code/>

4. The assessment protocol has two stages – an initial pass or fail gateway, followed by a scoring matrix. Both stages are based on a list of quality principles that need to be satisfied to a various degree (see Figure 1). To pass the Gateway stage, the source must comply with the quality dimensions of ethics and privacy, transparency and accountability, and need. Any fail on these criteria suggests that the source is not suitable for use towards SDG reporting. If the source passes all criteria on the gateway stage, it gets scored from 0 (not acceptable) to 3 (ideal score) along the quality dimensions of relevance, methods, coverage, timeliness and availability, data journey awareness, and quality assurance. An average score above 1.5 across these dimensions and no score of 0 suggests the source is suitable for SDG reporting. The spreadsheet template we use to aid with the assessment contains free-text fields where the assessor will enter a justification for each score. The template also contains detailed information to help with the decision-making process, such as descriptions of what it takes to score a ‘pass’ or any of the numerical values of the matrix stage. For a detailed description of the quality criteria, please refer to the detailed source assessment process section of the ONS publication<sup>4</sup>. The final score is also checked by another member of staff before finalising the decision.

Figure 1  
**Non-official sources protocol assessment flow schematic**



5. It is worth noting that in most cases for this assessment we refer to non-official statistical sources that are already publicly available by the supplier, rather than raw data. For example, these could be online reports and statistics produced by charities or think tanks. One of our key requirements is that all information we publish on our data website is publicly available. However, in some cases a non-official producer may hold microdata that are useful for the calculations of an SDG indicator, but these data are not publicly available, or may be underlying a published statistic. In these cases, we would contact the supplier and discuss if they could provide us with the raw data, which we would store securely on our internal systems. We have a dedicated Data Acquisitions and Operations team that assists us with this process,

<sup>4</sup>  
<https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/uksustainabledevelopmentgoalsuseofnonofficialsources#detailed-source-assessment-process>

so that any legal aspects are properly addressed (for example, a data sharing agreement). After obtaining the raw data, we would produce the necessary statistics for the indicator and publish the result as an ad-hoc dataset publication on the ONS website. This final publication would still be classed as non-official and would have to satisfy all the requirements of our protocol. Therefore, the assessment tool is not designed for raw data, but rather for the final statistical output based on the raw data. The quality of the latter is assessed using existing tools, such as the UK Office for Statistics Regulation's Quality Assurance of Administrative Data tool<sup>5</sup> and the Government Data Quality Framework<sup>6</sup>.

6. One key benefit of using this tool is that even if a source fails the assessment, the reason for that can be clearly identified and remedied in some cases. For example, we recently identified a suitable non-official proxy source for indicator 5.5.1: Proportion of seats held by women in national parliaments and local governments. The source originates from Oscar Research - a company collecting and analysing data on the UK public services. Although the information scored well on most dimensions, it was an overall 'fail' because the data published on the provider's website gets overwritten at each update, so no historic time series were publicly available. Therefore, it could not pass the 'timeliness and availability' dimension of the scoring matrix. Through communication and engagement with the supplier, we explained our standards and needs to use their data. Shortly after, Oscar Research published a historic trace of the figures we needed, so the source could then pass the protocol with a satisfactory score.

### III. NEXT STEPS

7. Our protocol for assessment of non-official sources is a dynamic document. Through increased usage, we will refine the quality dimensions so that they are better suited to assessment of potential SDG sources. The current criteria were first published in July 2021. When we assess whether these should be updated we will consider experiences of using the assessment and feedback received. The outcome of this might lead to changes to some of the scoring detail. One example is potentially introducing more clarity to the 'timeliness and availability' dimension, to accommodate statistics for environmental indicators where timely data cannot be made available on an annual basis and is likely to take longer periods of time. At present, the highest score of 3 is only given to sources that are updated annually or more regularly, putting possible environmental statistics sources at a disadvantage. We may also seek to clarify or merge the 'Data journey awareness' dimension into the 'Methods' one. Refining of the 'Quality assurance' aspect may also be necessary, as we find that many non-official sources may lack published detailed information on this. It is of key importance to be flexible with our approach, but also transparent. Of course, the tool does not come without limitations and risks, but we have addressed those in the original publication<sup>7</sup> and through our review process, we are hoping to improve our scoring system

8. We hope that other National Statistical Institutes (NSIs) could adapt our tool for their own use. As it currently stands, it is heavily reliant on the UK Code of Practice for official statistics, and most of the description for the dimensions scoring refers to specific clauses from the three pillars of the Code: trustworthiness, quality, and value. Essentially, the assessment

---

<sup>5</sup> <https://osr.statisticsauthority.gov.uk/guidance/administrative-data-and-official-statistics/>

<sup>6</sup> <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>

<sup>7</sup>

<https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/uksustainabledevelopmentgoalsuseofnonofficialsources#limitations-and-considerations>

should be done in parallel with reference to the Code of Practice. Although the principles are relatively high level and comparable to similar international standards, such as the United Nations National Quality Assurance Framework and the Fundamental Principles of Official Statistics, other NSIs are welcome to adjust the specific dimensions to fit with their national context.

\*\*\*