Non-traditional data quality assessment tool

Atanaska Nikolova

Office for National Statistics (ONS), UK

SustainableDevelopment@ons.gov.uk
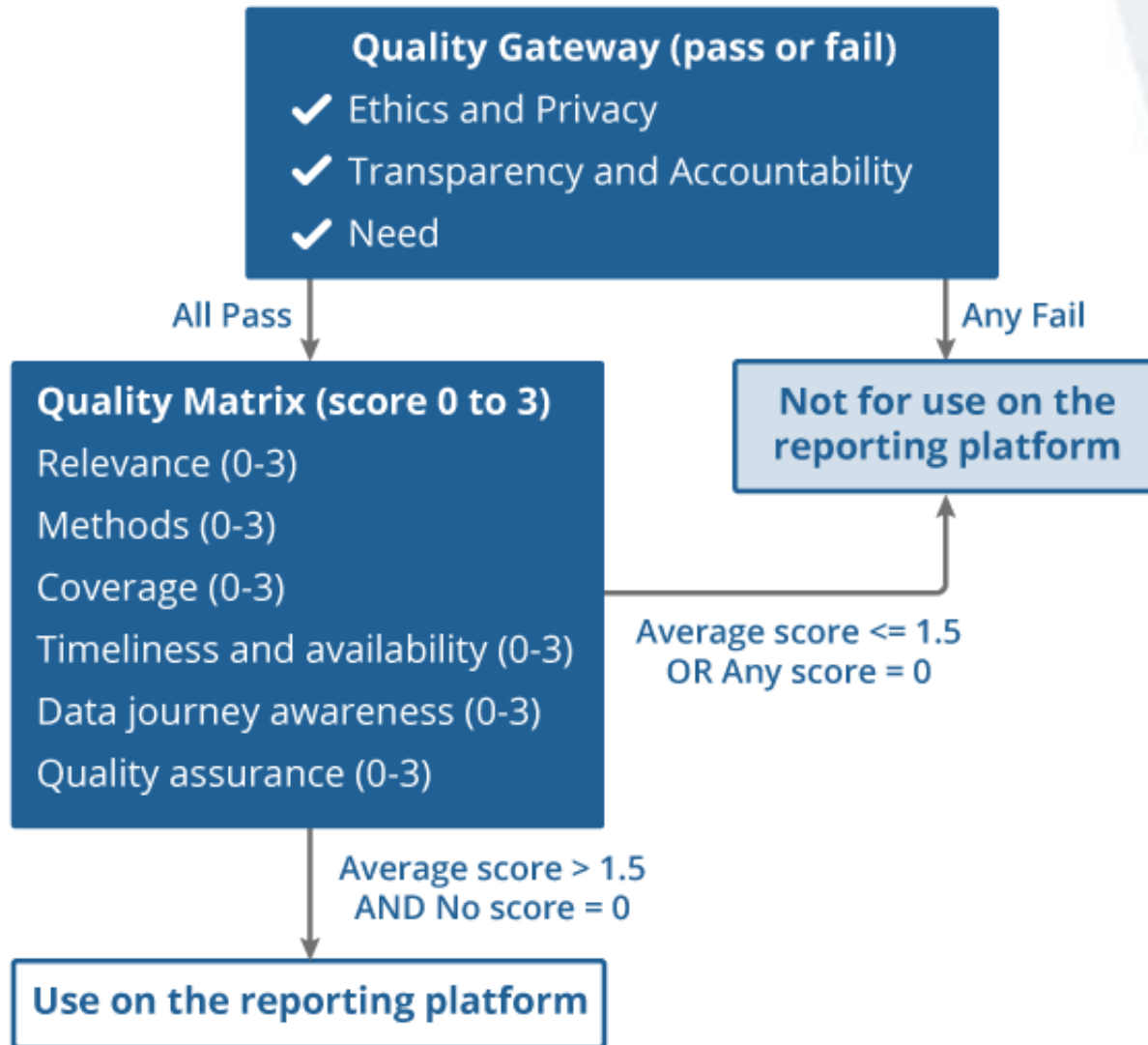
What are non-traditional data?

- Data or statistics (sources) produced by non-official organisations, not covered by government or other formal standards.

- Non-official organisation: not a UK government department or government-related body, local or devolved authority, or an official international reporting body (e.g. Eurostat or UNSD).

- Examples:
  - charities
  - businesses
  - academia

Protocol for Assessment of Non-official Sources

- Based on a model from Statistics Netherlands, and aligned with UKSA Code of Practice and Ethics Advisory Committee guidelines

- Covers assessment of statistical data sources

- SDG-specific, but could be adapted to non-SDG context for wider use

Non-official Protocol flow

Quality Gateway (pass or fail)
- ✔ Ethics and Privacy
- ✔ Transparency and Accountability
- ✔ Need

All Pass

Any Fail

Quality Matrix (score 0 to 3)

Relevance (0-3)

Methods (0-3)

Coverage (0-3)

Timeliness and availability (0-3)

Data journey awareness (0-3)

Quality assurance (0-3)

Not for use on the reporting platform

Average score <= 1.5
OR Any score = 0

Average score > 1.5
AND No score = 0

Use on the reporting platform

Trustworthiness
Confidence in the people and organisations that produce statistics and data

Quality
Data and methods that produce assured statistics

Value
Statistics that support society's needs for information

Assessment tool – quality gateway

Criterion	Criteria		Assessment	
	Pass	**Fail**	**Source score**	**Explanation**
Ethics and Privacy	No ethical concerns, or any concerns are fully documented and actions are in place to minimise identified risks. Fully compliant with all parts of the Code principle T1 (honesty and integrity) and T6 (data governance). Use is in line with the terms and conditions of the source. Privacy policy is compliant with the General Data Protection Regulation (GDPR) for the UK and the Data Protection Act 2018.	Significant ethical concerns that may damage ONS' reputation. Not compliant with all parts of principle T1 and T6 of the Code. The source's terms and conditions prevent use of the data as required. Not compliant with GDPR and/or the Data Protection Act 2018.	Pass	No personal data involved and the dataset is regularly updated and publicly available
Transparency and Accountability	Source meets principles T4.1 and T4.5 of the Code; ensuring processes for all parts of the data journey are transparent. If metadata information is not already in the public domain, permission must be granted to place this in the public domain.	Source does not meet T4.1 and T4.5 of the Code. Source may not be fully transparent about any data quality issues and there is no metadata available.	Pass	The LINK website has detailed documentation on Governance and data privacy and policy statements. The abide by the code of conduct of Bank of England
Need	There is a clear identified need for the source, either due to a headline data gap or a priority disaggregation gap (see above). In exceptional circumstances, it is possible that a non-official source could replace an official source, where the official source provides a proxy which scores lower on the Quality matrix than the non-official source.	The proposed source does not fill a headline or priority disaggregation data gap. The source does not improve on information relative to another source of non-official data currently reported on the Platform, or any existing proxy data.	Pass	Fulfills part of the indicator to UN specifications

8.10.1

ATM Statistics and Trends (LINK)

Assessment tool – matrix scoring

Criterion	Criteria				Assessment		Quality Matrix (score 0 to 3)
	3 - high	2 - medium	1 - low	0 - not acceptable	Source Score	Explanation	
Timeliness and ongoing availability	Source is sufficiently current to be informative, with a time series from at least 2015 and no time lag greater than 15 months for annual data, or 6 months for monthly data. No gaps (missing data) in the time series. The source is reasonably expected to be regularly updated and available in the	Source is sufficiently up to date to be informative, with a time lag no greater than 2 years. There are no gaps in the time series. New timely sources without previous data points that are expected to be updated and available in the future would be included.	Source is older than 2 years, but is still meaningful in the social, environmental, or economic context of the indicator. The time series may have gaps, or only one data point is available.	Source is too old to be meaningful, with the latest data point(s) before 2015, or has too great a time lag, and/or no reasonable expectation of future updates.	3	Time lag is less than a year	Relevance (0-3) Methods (0-3) Coverage (0-3) Timeliness and availability (0-3) Data journey awareness (0-3) Quality assurance (0-3)
Relevance	Close match with UN SDG metadata, or gives more detail than the metadata requires. Fully compliant with the Code principles Q1.1 and Q1.5 on suitable data sources. Disaggregations specified in the SDG indicator title are reported as a minimum, potentially supported by additional Inclusive Data Charter	Partial match with UN metadata and/or disaggregation in indicator title. Fully compliant with the Code principles Q1.1 and Q1. May enable reporting of additional IDC disaggregations.	Does not fully report the indicator but is an appropriate proxy relevant to the UK national context. Fully compliant with the Code principles Q1.1 and Q1.5. May enable reporting of relevant disaggregations.	Does not align well with the metadata for the indicator and/or provides no appropriate proxy for headline or relevant disaggregation gaps.	3	Matches UN requirement, there are no required disaggregations	

Average score > 1.5 AND No score = 0

Use on the reporting platform

6

Assessment tool – matrix scoring

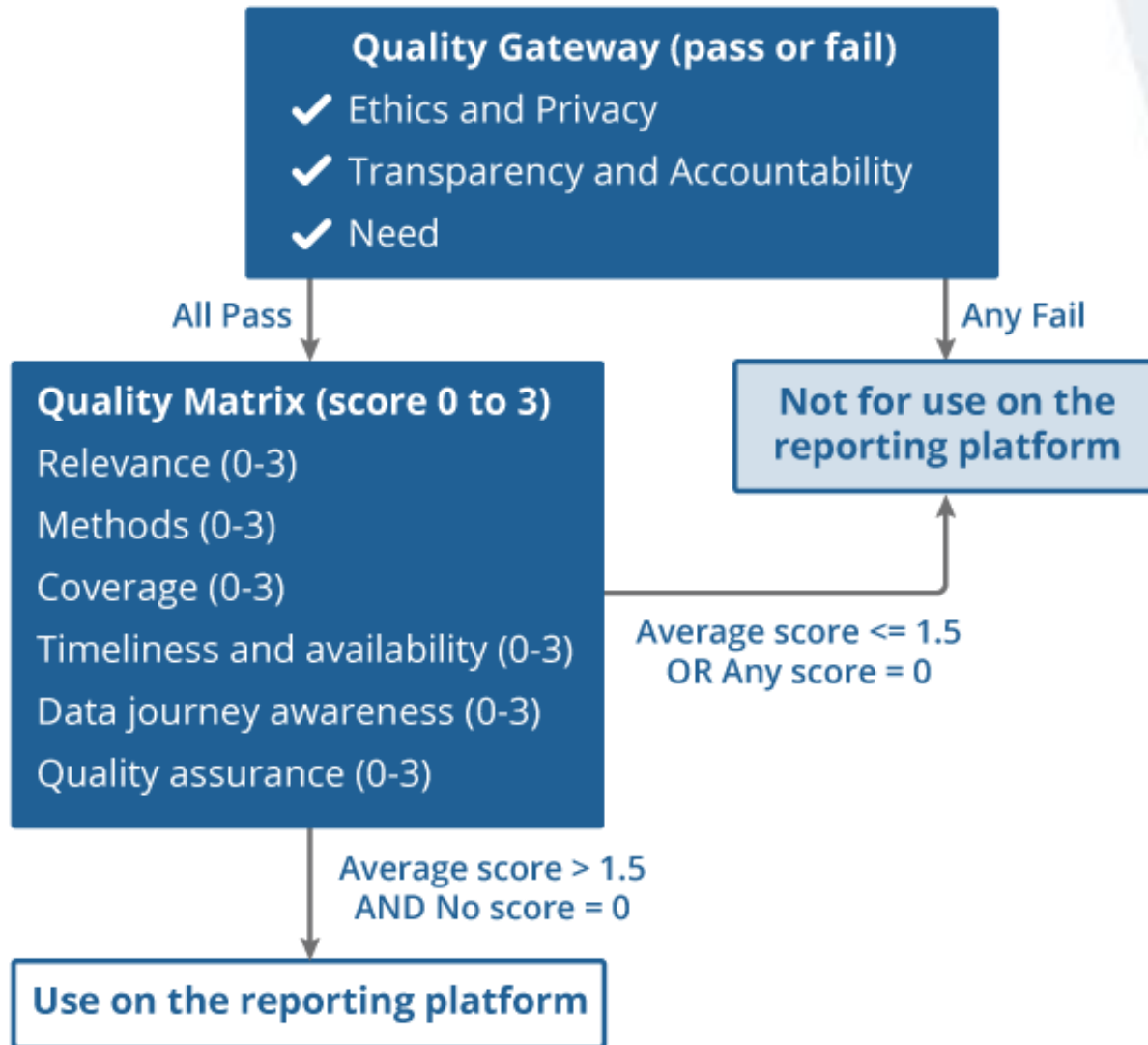|							
Data journey awareness	There is a comprehensive, fully documented data journey from collection to publication phases. Data validation procedures are outlined. It is clear how the source was collected and pre-processed (if relevant).	The data journey is outlined from collection to publication phases.	Parts of the data journey be may be described with limited detail, but the data journey path can be followed.	No data journey available, or parts of the data journey are missing or too poorly described to assess the stages it has been through.	1	No comprehensive metadata on this; the numbers are publicly available and supplied directly by LINK	**Quality Matrix (score 0 to 3)** Relevance (0-3) Methods (0-3) Coverage (0-3) Timeliness and availability (0-3) Data journey awareness (0-3) Quality assurance (0-3)
Quality Assurance	Fully documented and published QA process, fully meeting the quality standards in the Code principle Q3 (assured quality).	Largely compliant with standards in the Code principle Q3, specifically with Q3.3 (covering accuracy and reliability). Some particular or minor details of the process may be missing from the documentation.	Some basic checks have been conducted, such that accuracy and reliability of the source can be established. Further work may be needed to explicitly meet standards in principle Q3 of the Code.	No QA process has been conducted, is not presented, or the process is too poor to present. No evidence of data accuracy and reliability.	2	Regular self-assessment conducted and published in line with the principles of the Technical Committee of the International Organization of Securities	Average score > 1.5 AND No score = 0 **Use on the reporting platform**
Overall comments and average score					2.166666667	Pass	

Non-official Protocol flow

Quality Gateway (pass or fail)
- ✔ Ethics and Privacy
- ✔ Transparency and Accountability
- ✔ Need

All Pass

Any Fail

Quality Matrix (score 0 to 3)

Relevance (0-3)

Methods (0-3)

Coverage (0-3)

Timeliness and availability (0-3)

Data journey awareness (0-3)

Quality assurance (0-3)

Not for use on the reporting platform

Average score <= 1.5 OR Any score = 0

Average score > 1.5 AND No score = 0

Use on the reporting platform

Trustworthiness
Confidence in the people and organisations that produce statistics and data

Quality
Data and methods that produce assured statistics

Value
Statistics that support society's needs for information

SUSTAINABLE DEVELOPMENT G⚙ALS

Current non-official sources

- Indicator 8.10.1 – ATM numbers from LINK

- Indicator 6.6.1 – Ground water levels from British Geological survey and river flow data from Centre for Ecology and Hydrology (in progress)

- Indicator 3.1.1 – Maternal mortality from The National Perinatal Epidemiology Unit and MBRRACE charity

- Indicator 12.3.1 – Food waste per capita from WRAP charity

- Indicator 1.4.2 – Rights to land and tenure security from Prindex

Thank you for listening!

Contact:
Atanaska.Nikolova@ons.gov.uk
SustainableDevelopment@ons.gov.uk

UK Sustainable Development Goals:
use of non-official sources