Health Information and Quality Authority

An tÚdarás Um Fhaisnéis agus Cáilíocht Sláinte

Health Information and Standards

Data Quality Assessment Tool for health and social care

October 2018

Safer Better Care

Introduction

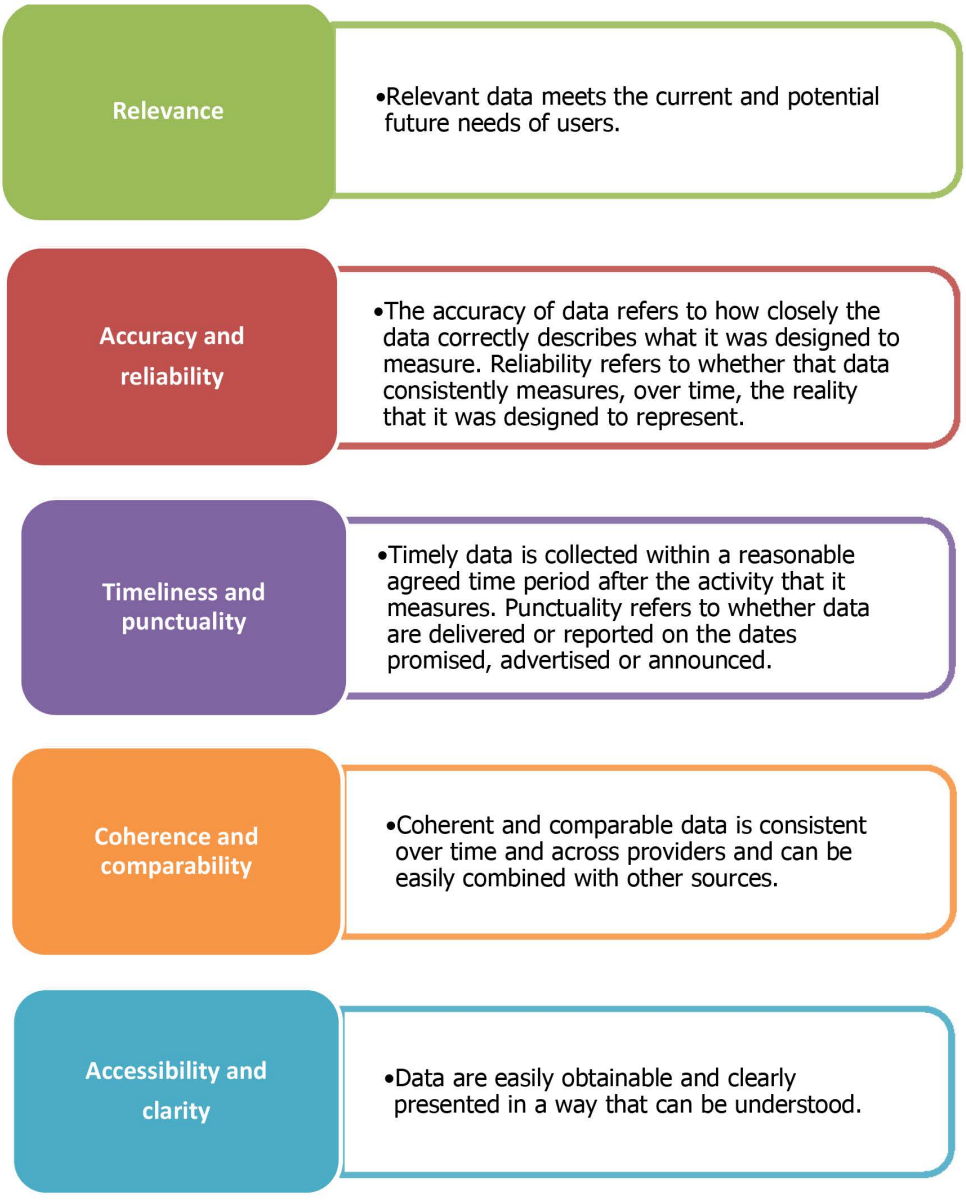
This interactive data quality assessment tool has been developed to meet the needs of a broad range of health and social care organisations. It should be read in conjunction with HIQA's *Guidance on a data quality framework for health and social care*.

This tool provides a detailed set of criteria that organisations can use to comprehensively assess its data sources across all five dimensions of quality.

The tool draws strongly from the CIHI's Data Quality Framework, CIHI's Information Quality Framework, CIHI's Data Source Assessment Tool, CIHI's Information Quality Plan, the European Statistical System Quality Assurance Framework and the United Nations Statistics Division National Quality Assurance Framework.[1,2,3,4,5,6]

The characteristics and criteria included in the data quality assessment tool serve as a guide to organisations to assist in the assessment of data quality. They are not an exhaustive list and can be amended by organisations as necessary.

Figure 1. Dimensions of data quality



1. Relevance

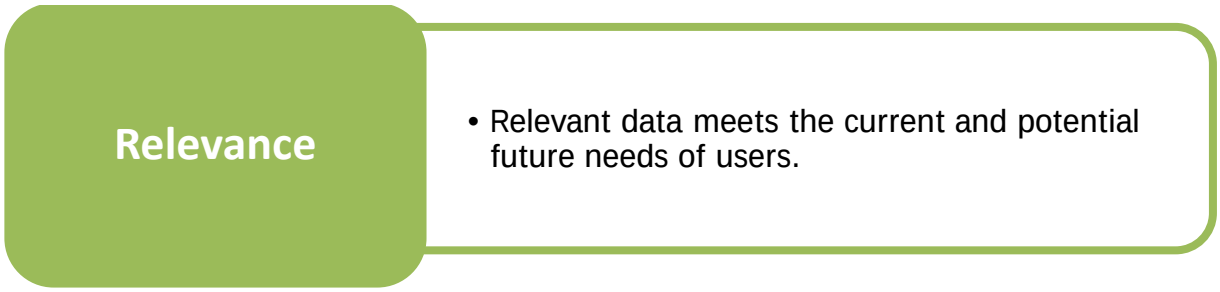


Table 1. The quality characteristics of relevance

Characteristic	Description
Release and use of the data	This refers to the way in which an organisation's data is released and used. It concerns the frequency of use of that data. It should be noted that, where data is not used regularly, it may be due to reasons other than it not being useful, for example, difficulty for the user in accessing the data.
Value of the data	This refers to the extent to which the data successfully serves the needs of data users. If the data does not meet the needs of the data user, it can become irrelevant.
Adaptability of the data source	This refers to the extent to which the data is adaptable and flexible enough to meet the current and future needs of data users. If the data is not flexible in meeting the current and future needs of data users, it can become irrelevant.

In order to assess the quality of data in relation to relevance, the assessment tool outlined in Table 2 can be used.

Table 2. Data quality assessment tool — relevance

Relevance			
Characteristic	**Criteria**	**Details**	**Assessment**
Release and use of the data	Are regular assessments carried out to determine whether all of the data that is being collected is being used?	Undertake regular reviews of data usage, for example, assessment of usage metrics such as the number of times a data set was viewed, the number of times the data was downloaded and the number of times requests were received for access to the data.	Yes ○ No ○

	Has a list of key users and their use of the data been compiled, including unmet user needs?	Compile a list of data users. Describe the known uses of the data, for example, service planning or development of funding models, to assess prevalence and incidence of diseases.	Yes ○ No ○ Partially ○
Value of the data	Are data users consulted to establish if the data available assists them in achieving their objectives?	Establish that the content of the data available is sufficient to assist the organisation in achieving its objectives, by conducting an assessment of data user needs.	Yes ○ No ○
	Are quality improvement plans in place to address required improvements in the data in order to ensure the data remains relevant to users?	A quality improvement plan is in place to ensure data is relevant to user needs. Detail actions that have been identified to improve relevance of data, for example, engagement with data users, and the impact that the implementation of these actions has had in ensuring that data remains relevant to user needs.	Yes ○ No ○ Partially ○
Adaptability of the data source	Are procedures in place to gather information on the potential future needs of data users?	Mechanisms are undertaken, for example, surveys, to gather information on the potential future needs of data users and such needs have been identified and implemented and their effectiveness measured and reported on.	Yes ○ No ○ Partially ○
	Are data user needs prioritised as a result of consultation undertaken with data users about how the data relates to their needs?	Consult with data users about their data requirements and develop prioritisation criteria to facilitate implementation of changes required.	Yes ○ No ○

2. Accuracy and reliability

Accuracy and reliability

- The accuracy of data refers to how closely the data correctly describes what it was designed to measure. Reliability refers to whether that data consistently measures, over time, the reality that it was designed to represent.

Table 3. The quality characteristics of accuracy and reliability

Characteristic	Description
Coverage	The degree to which the data available to data users covers the population or event of interest. It is critical to understand the level of coverage of a population in order to make a statement about that population. Coverage can be assessed by defining the population in question and determining and monitoring coverage rates. Any limitations to the data as a result of coverage issues should be clearly explained.
Data capture and collection	The procedures that are in place to ensure that the data is captured in a usable format and can be prepared for submission. This is assessed by identifying what measures are in place to ensure that relationships with data providers are positive and that suitable processes are in place to facilitate efficient and effective data flows.
Data processing	The transformation of data from the form in which it is received into another form that facilitates analysis. Processing can include validation and correction of the data. Assessment involves checking if processing has been documented in detail and if the processed data can be differentiated clearly from the original data received.
Completeness and validity	The degree to which individual variables are present within a data set. Incomplete and invalid data within a dataset impacts upon the fitness for use of that dataset. Assessment of completeness and validity is done by calculating and monitoring rates of valid, invalid, missing and outlier values.
Revisions to data	The extent to which the data is subject to revision or correction following new information or correction of errors in processing or estimation of data and the time frame in which revisions are produced.

In order to assess the quality of data in relation to accuracy and reliability, the assessment tool outlined in Table 4 can be used.

Table 4. Data quality assessment tool — accuracy and reliability

Accuracy and reliability			
Characteristic	**Criteria**	**Details**	**Assessment**
Coverage	Are details of the reference population explicitly stated in all information releases and is the coverage of the population quantified?	State if the reference population is documented and if coverage has been quantified. Where this has not occurred, provide details. The reference population is the population for which information should be available.	Yes ○ No ○ Partially ○
	Are significant coverage issues that may impact analysis and interpretation of data documented and made available to users?	Provide details of any significant issues in relation to coverage that have the potential to impact on data users' interpretation. Consider the impact of any identified coverage issues in terms of their impact at different levels of reporting. If coverage issues are not documented, detail why this is the case.	Yes ○ No ○ N/A ○
	Are processes in place to identify and handle duplicate and potential duplicate records within the data?	Describe the processes that are in place to identify duplicates and calculate the duplication rate to give an indication of its occurrence. Provide details of how duplicates are removed once identified. If measures are not undertaken to identify and remove duplicates, explain why this is the case and provide an estimate of the number of duplicates that remain in the data.	Yes ○ No ○ Partially ○

Data capture and collection	Are issues with the quality of data submitted that have the potential to impact significantly on analysis and interpretation of that data addressed and documented for users of the data?	Provide details of identified data quality issues, including the action undertaken to resolve the issue, the time frame to resolve the issue and any potential impact that such issues have on interpretation and analysis of the data.	Yes ◯ No ◯ N/A ◯
Data processing	Are data validation processes applied consistently and are the processes documented for data users?	Provide details of the types of validation checks undertaken, for example, checks on the structure and integrity of the data, checks for missing data, checks that the data conforms to data source specifications and checks for any unusual data, that is, outliers.	Yes ◯ No ◯ Partially ◯
Completeness and validity	Are rates of valid, invalid, missing and outlier values documented and updated routinely and reported with each data release?	Provide a data profile for data elements which include valid, invalid and missing value rates.	Yes ◯ No ◯ Partially ◯
Revisions to data	Are revisions or corrections made to the data regularly analysed to ensure effective statistical use of same?	Provide details of the reasons for, timing of, and nature of revisions. Clearly identify data as preliminary or revised.	Yes ◯ No ◯

3. Timeliness and punctuality

Timeliness and punctuality

- Timely data is collected within a reasonable agreed time period after the activity that it measures. Punctuality refers to whether data are delivered or reported on the dates promised, advertised or announced.

Table 5. The quality characteristics of timeliness and punctuality

Characteristic	Description
Submission timeliness	This refers to whether the data is submitted on time by providers. This is a critical element for achieving timely and punctual release of results.
Processing timeliness	This refers to whether the data, once received from data providers, is processed as efficiently as possible. This is critical in supporting timely and punctual release of data.
Release timeliness and punctuality	This refers to whether the data was released on time and how current the data is at the time of release. The length of time between its reference date and the date that the data is released should be short enough to ensure that the data remains relevant for the purpose it was intended. This can be assessed by determining if the data was released on time, in line with the planned release date.

In order to assess the quality of data in relation to timeliness and punctuality, the assessment tool outlined in Table 6 can be used.

Table 6. Data quality assessment tool — timeliness and punctuality

Timeliness			
Characteristic	**Criteria**	**Details**	**Assessment**
Submission timeliness	Are procedures in place to ensure the effective and timely submission of data from providers?	Ensure that data providers are aware of the submission dates, for example, through submission calendars or automatic reminders sent for data collections.	Yes ○ No ○
	Are agreements in place with data providers	Ensure formalised agreements have been put	Yes ○

	which detail planned dates for submission of data?	in place between data providers and the organisation to which data is being submitted that clearly set out the date by which data must be submitted.	No ◯
	Are follow-up procedures in place to ensure timely receipt of data, including procedures to address necessary improvements?	Identify mechanisms to improve timeliness of data submissions, for example, indicators to measure timeliness or stakeholder engagement activity to explore data submission issues. Evaluate the impact that the chosen mechanism has had on the timeliness of data submissions.	Yes ◯ No ◯
Processing timeliness	Are data processing activities regularly and systematically reviewed to improve timeliness and has an associated action plan been developed and implemented?	Describe the review process, systems used and the timeframe for reviews, including the date of the last review undertaken. Provide details of why reviews are not undertaken regularly, where applicable. Provide details of any improvements in timeliness of data processing that have occurred as a result of the review process.	Yes ◯ No ◯
Release timeliness and punctuality	Has a data release policy and procedures document, which includes targets for timeliness, been developed, published and implemented? Does the policy describe revisions for key outputs that are subject to scheduled revisions?	Confirm whether or not a publicly available policy exists which clearly identifies when data is to be released. Indicate if the policy includes details of planned revisions of data.	Yes ◯ No ◯ Partially ◯
	Do planned releases occur within a specified	Ensure that all planned data releases, as well as updates	Yes ◯

	period of time from the end of the reference period?	to data, indicate the following: ▪ the planned release date ▪ the actual date the data was released ▪ the reference period for release of the data ▪ time elapsed between end of reference period and release date.	No ◯
	In the event of delays affecting a planned release, are delays and causes documented and made available to data users?	Ensure that planned data releases indicate the following: ▪ if there was a delay in releasing the data ▪ if data users were notified of any delays ▪ the method by which data users were kept informed.	Yes ◯ No ◯ N/A ◯
	Is an up-to-date release calendar publicly available?	Confirm that a release calendar is in place which details the dates on which data is to be released. The calendar should be developed following consultation with data users. Establish that procedures are in place to regularly monitor and evaluate the punctuality of data releases, as per the release calendar.	Yes ◯ No ◯

4. Coherence and comparability

Coherence and comparability

- Coherent and comparable data is consistent over time and across providers and can be easily combined with other sources.

Table 7. The quality characteristics of coherence and comparability

Characteristic	Description
Standardisation	This refers to the degree to which data is collected using common definitions or standards. Collecting data in this way ensures that data is comparable and can be easily analysed.
Coherence	This refers to the degree to which it is possible to combine and make use of related data from different sources. Where data is coherent, data users can combine data sources, allowing them to report on something in greater detail.
Historical comparability	This refers to how the consistent use of definitions, standards and methods facilitates comparison of data over time. Data and statistics should be comparable over a reasonable period of time.
Regional comparabilty	This refers to the degree to which the data allows for reasonable comparisons within and across regions and countries. Where this level of comparability is possible, it facilitates more informed decision making and management of systems while driving improvements in the quality of the data.

In order to assess the quality of data in relation to coherence and comparability, the assessment tool outlined in Table 8 can be used.

Table 8. Data quality assessment tool — coherence and comparability

Coherence and comparability			
Characteristic	**Criteria**	**Details**	**Assessment**
Standardisation	Is data collected in line with national and international standards	Detail what standards and classifications have been applied to the collection of	Yes ◯ No ◯

	and classifications?	data and any variations used. Where standards and classifications are not applied in the collection of data, explain why.	Partially ◯
	Are metadata clearly described and made available to data users?	Employ a data dictionary, which is a descriptive list of names, definitions and attributes of data elements to be collected in an information system or database.	Yes ◯ No ◯
Coherence	Is aggregated data compared with other sources of data, for example, administrative data, that provide the same or similar information on the same phenomenon? Are divergences identified and clearly explained to data users?	Identify the type of comparison undertaken and any differences in data sources that were identified as a result of the comparison. Where aggregate data have not been compared with data from another source, explain why.	Yes ◯ No ◯
Historical comparability	Are historical changes/trends in the data documented and publicly available for data users?	Describe any changes which have occurred in the data over time, for example, data definitions, methodology, processing methods and the means by which data is captured and submitted by providers. Where changes in data over time are not formally documented and available to data users, explain why.	Yes ◯ No ◯ N/A ◯
	Are any changes in the data/trends that can potentially have a significant impact on interpretation and	Provide details of any changes to key elements of the data set, explaining the trend in changes.	Yes ◯ No ◯ N/A ◯

	analysis of data, that is, changes to key elements of the data set, documented and available for data users?	Where changes to key elements within the data set have not been documented, explain why.	
Regional comparability	Is the impact of any identified differences in data across regions documented?	Describe the limitations to the analysis and interpretation of the data that have emerged as a result of regional differences.	Yes ○ No ○ N/A ○

5 Accessibility and clarity

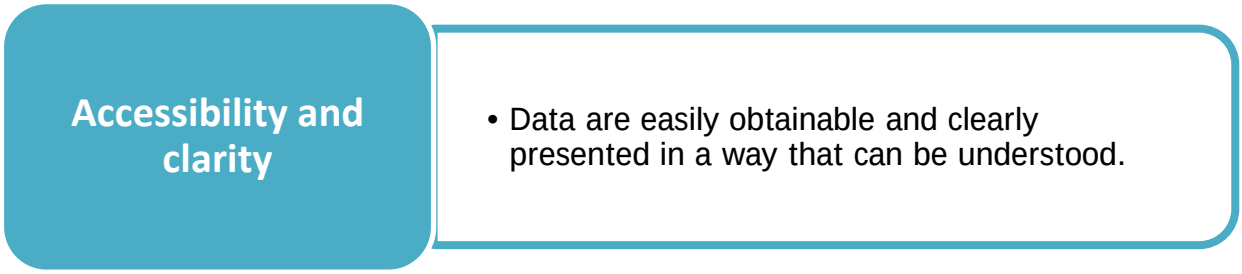


Table 9. The quality characteristics of accessibility and clarity

Characteristic	Description
Accessibility	This refers to the ease with which data can be identified, obtained and used. Awareness that the data exists is a key component of data quality. The suitability of the format of the data should be considered in terms of ease of use for data users. Accessibility of data can be assessed by establishing whether the data is available in a standardised format for each data release.
Interpretability	This refers to the degree to which users are provided with the required documentation and metadata to assist them in understanding the data. Interpretability can be assessed by establishing whether or not the necessary documentation is up to date and available to data users. This includes documentation to accompany preliminary and revised data.

In order to assess the quality of data in relation to accessibility and clarity, the assessment tool outlined in Table 10 can be used.

Table 10. Data quality assessment tool — accessibility and clarity

Accessibility and clarity			
Characteristic	**Criteria**	**Details**	**Assessment**
Accessibility	Are data available to users in a form that facilitates proper interpretation and meaningful comparisons?	Facilitate interpretation and comparison of data by making a publication catalogue or list of publications available to data users. Disclose policies and procedures detailing how data users can access and	Yes ◯ No ◯

		request data. Put in place a user support service to assist data users and other stakeholders to obtain and ask questions about data.	
	Is ICT effectively used to disseminate data and information?	Ensure that data users can navigate the website in a way that allows them to easily access data and metadata. Ensure tools and resources are available to support the needs of data users, including web-based tools to facilitate analysis and manipulation of data.	Yes ◯ No ◯
Interpretability	Are supporting documents, for example, metadata, publicly available to facilitate clarity of interpretation for data users?	Documentation to accompany data outputs should include details of the following: information on key concepts, scope, data sources, compilation, methods and statistical techniques. Publish a data quality statement for all data outputs that details the quality of the data as assessed under the five dimensions of data quality. Identify and publish any known issues in relation to data quality to facilitate the data user to assess if the data is of sufficient quality to be relevant their needs.	Yes ◯ No ◯ Partially ◯
	Does a revision policy exist which covers all data and is it available to data users?	State if a revision policy exists. Provide details of the content of the revision	Yes ◯ No ◯ Partially ◯

		policy, for example, reasons for undertaking a revision, schedule of revisions and the nature of the revisions covered. Describe the impact of the assessment, the changes that occurred and any future plans that are in place as a result, to reduce the need for a revision.	

References

1. Canadian Institute for Health Information. *The CIHI Data Quality Framework*. 2009. Available from: http://www.cihi.ca/cihiweb/en/downloads/data_quality_framework_2009_e.pdf. Accessed on: 1 June 2016.

2. Canadian Institute for Health Information. *CIHI's Data Source Assessment Tool.* 2017. Available from: https://www.cihi.ca/sites/default/files/document/cihi-data-source-assessment-tool-en.pdf. Accessed on: 06 November 2017.

3. Canadian Institute for Health Information. *CIHI's Information Quality Plan*. 2017. Available from: https://www.cihi.ca/sites/default/files/document/corporate-information-quality-plan-2016-en.pdf. Accessed on: 06 November 2017.

4. European Statistical System. *Quality Assurance Framework of the European Statistical System*. 2015. Available from: http://ec.europa.eu/eurostat/documents/64157/4392716/qaf_2012-en.pdf/8bcff303-68da-43d9-aa7d-325a5bf7fb42. Accessed on: 1 June 2016.

5. United Nations Statistics Division. *Guidelines for the template for a generic national quality assurance framework (NQAF)* 2012. Available from: http://unstats.un.org/unsd/dnss/docs-nqaf/GUIDELINES%208%20Feb%202012.pdf. Accessed on: 1 June 2017.

6. Canadian Institute for Health Information. *CIHI's Information Quality Framework.* 2017. Available from: https://secure.cihi.ca/free_products/IQF-summary-july-26-2017-en-web.pdf. Accessed on: 8 February 2018.

Health Information and Standards Directorate
Health Information and Quality Authority
Unit 1301, City Gate,
Mahon,
Cork
T12 Y2XT