C/Can

Ensuring quality data for decision-making

C/Can Responsible
Data Framework

Table of Contents

1. Background

As a data-driven organisation, City Cancer Challenge (C/Can) recognises the importance of having data of the highest quality on which to base decisions. This means having processes, definitions and systems of validation in place to assure data quality is integrated into our everyday work.

Complete and accurate data are essential to support effective decision making on the priority solutions for cancer care in cities to be addressed through the C/Can City Engagement Process including:

01 Identification of priority solutions (needs assessment)

02 Engagement of the "right" partners, stakeholders and experts

03 Activity planning

04 Design and implementation of projects to deliver priority solutions

05 Monitoring and Evaluation

06 Sustainability

2. Purpose

All data are subject to quality limitations such as missing values, bias, measurement error, and human errors in data entry and analysis. For this reason, specific efforts are needed to ensure high quality data and the reliability and validity of findings. C/Can is putting in place measures to assess, monitor and review data quality, and so that investments can be made in continuously improving the quality of data over time.

This guide lays the basis for a common understanding of data quality for C/Can as an organisation. It sets out the procedures and processes that need to be carried out to ensure good data quality management of all C/Can related data.

3. Dimensions of Data Quality

C/Can assesses data quality[1] across six core dimensions:

01 **Completeness**

Completeness is defined as the extent to which all the data required are available. Data can still be complete if optional data is missing. Measures of completeness look at the levels of data that are missing or unusable.

02 **Accuracy**

Accuracy is the degree to which the data accurately reflects the real world situation / object or event. Where possible, data accuracy can be checked by comparing the data collection with an identified and trusted second source of information. An additional measure to ensure accuracy is to have the data checked/approved before moving from data collection to data analysis.

03 **Uniqueness**

Uniqueness is a measure of the level of duplication of data within a given data set or across data sets. Unique information means that there is only one instance of it appearing in a database. Duplication should be avoided wherever possible, meaning that if there are duplicate copies of repeating data or multiple responses for the same question from the same institution or individual, then this should be cleaned and consolidated into a single data point.

04 **Integrity**

Integrity refers to maintaining information as it was inputted, and can therefore be measured by the degree of data corruption. Any unintended changes to the data as a result of processing operations (e.g. hardware / software malfunctions), storage, retrieval of data and human error is a failure of data integrity. Examples include loss of a record due to human errors in data cleaning processes.

05 **Validity**

Validity is a data quality dimension that refers to the correctness and reasonableness of the data, including whether information confirms to a specific format. For example, data on dates should follow the same format such as DD.MM.YY rather than MM/DD/YYYY. Data validation is a key process in ensuring data validity by checking that data inputted is acceptable in terms of the format as well as rational /acceptable.

06 **Timeliness**

The degree to which the data is up-to-date and available within an acceptable time frame, timeline and duration. The value and accuracy of data may decrease over time, therefore timeliness of data is critical. It can be measured as the time between when information is expected and when it is available for use.

[1] Data Quality Management, Mark Allen, Dalton Cervo, in Multi-Domain Master Data Management, 2015

03.1 Core Indicators of Data Quality

Benchmarks and sample sizes for data quality indicators should be determined on a case-by-case basis, but should cover the six core dimensions. The examples of sample indicators provided below relate to the C/Can City Needs Assessment Questionnaire.

01 **Completeness**

> Number of responses (for the Needs Assessment, this should be broken down by healthcare institutions, civil society and patients).

> Percentage of targeted institutions/individuals that have responded to all required questions.

> Percentage of incomplete data fields.

> Percentage of data fields with 'Don't know' responses.

02 **Accuracy**

> Percentage of institutions with data reviewed and submitted by institutional lead.

03 **Uniqueness**

> Percentage of duplicate responses/records.

04 **Integrity**

> Percentage of corrupted responses/records.

05 **Validity**

> Percentage of outliers (if a data value in a series of values is extreme in relation to the other values in the series).

06 **Timeliness**

> Percentage of data received within the designated time frame.

4. Needs Assessment Quality Control Methodology

C/Can has a number of interrelated processes to support high levels of data quality:

01   Setting data standards

02   Undertaking data validation

03   Checking for and acting on missing or inconsistent data

04   User access and rights arrangements for quality control

05   Reports and dashboards

06   Ethical considerations and personal data protection

The examples below apply to the City Needs Assessment Process and the newly developed questionnaire platform, but can be adapted and applied to other data collection processes throughout the C/Can city lifecycle.

01   **Setting data standards**

› Parameters are set for data entries for questions requiring numerical or answers as percentages, allowing only integers within pre-specified ranges as a response.

› Predefined answers with drop down options, multiple choice or checkboxes are used as response options wherever possible.

O2 Undertaking data validation

› Checking for the completeness of any data set and reviewing if missing information can be obtained and entered in the system.

› 'Sense-checking' any information produced and comparing to similar or previous datasets.

› Data quality checks are run monthly by the Senior MEL Manager, and followed up by the City Manager via the institutional leads, to improve data quality, completeness and validity on reportable and required fields to ensure submissions to C/Can are as comprehensive and correct as possible.

O3 Checking for and acting on missing or inconsistent data

› Institutional leads are identified for each institution and assigned responsibility to ensure that data for the institution is complete before submitting to C/Can.

› Key questions in the questionnaire are mandatory, so submission to C/Can can only occur once all mandatory data is completed.

› Errors and inconsistencies identified should be investigated and addressed by the institutional lead.

O4 User access and rights arrangements for quality control

› Each institution has a unique access code that multiple respondents can use to complete their allocated section of the questionnaire. This creates one record per institution which ensures that there is no duplication of responses for a given institution.

O5 Reports and Dashboards

› A dashboard accessible to City Managers and all relevant team members provides an overview of which sections have been submitted by institutions. City Managers can use this dashboard to check on data completeness and follow up with stakeholders to complete missing data and ensure the minimum level of completeness is met.

O6 Ethical considerations

› Patient responses to the questionnaire are anonymous; patients follow a link sent to them by a civil society or institutional representative. Consequently no individual patient data, including personal data such as names, email addresses, telephone numbers, are collected.

* Please see C/Can's data privacy policy and guidelines on the protection of personal and non-personal data for further detail and examples.

5. Governance

05.1 Roles and responsibilities

Ensuring that high quality data is collected and used at global and city levels is the responsibility of every C/Can team member. To ensure a consistent and rigorous approach, specific roles and responsibilities for data quality are however assigned across the team for key activities/ processes within C/Can. For example:

- › Oversight of the quality of all data collected as part of the city needs assessment process is primarily the responsibility of the City Manager. Support is provided by members of the Policy and Impact team.

- › Responsibility to collect and input into C/Can's CRM is distributed among each dataset owner and validation is the responsibility of their line managers.

05.2 Training

The Policy and Impact Team are responsible for the on-going programme of training and communication to staff on the content and processes set out in this guide and related documents.

Key local stakeholders including City Executive Committee members and local sustainability partners, and those involved in the City Needs Assessment process in particular should be aware of C/Can's guidelines and standards for quality data collection, and be provided with access to this document.

05.3 Routine Data Quality Review

Routine and regular data quality reviews (DQR) are an important mechanism to collect, analyse and base decisions on the highest quality data possible.

> For specific processes such as the needs assessment, data checks are run monthly by the Senior MEL Manager, and followed up with the relevant data holder, to improve data quality completeness and validity on reportable and required fields to ensure submissions to C/Can are as comprehensive and correct as possible.

> For the needs assessment, City Managers should plan to conduct at least one site visit to each institution inputting data during the needs assessment to conduct a data quality assessment. Specifically, to confirm data sources and validity on a set sample of the data being collected. The sample of the data assessed will be at the discretion of the city manager in consultation with the Senior MEL Manager.

> For Salesforce data, line managers of dataset owners are responsible to review datasets regularly for completeness, accuracy and
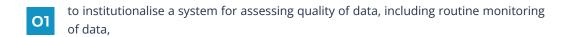
05.4 Annual Data Quality Review

C/Can will institute annual DQRs of key datasets including of at least 2.5% of the needs assessment indicators from all cities engaged in the process during that calendar year. The objective of the review will be to examine data quality across the six quality dimensions.

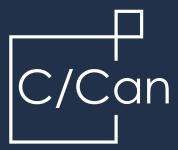The desk review has two levels of data quality assessment:

O1 an assessment of each indicator aggregated to the city level;

O2 an assessment of each indicator aggregated to the global level;

The objectives of the DQR are:

O1 to institutionalise a system for assessing quality of data, including routine monitoring of data,

O2 to identify measures to improve quality, and to develop action plans to implement such measures, and

O3 to monitor the performance of data quality over time and the capacity to produce good quality data.

Depending on the findings of the annual DQR, the Policy & Impact team may recommend that a Data Quality Improvement Plan should be prepared. The Data Quality Improvement Plan should seek to identify and address the root causes of data quality problems revealed by the DQR, including specific measures needed to strengthen the system and resolve the problem. The actions outlined in the plan should be specific, timebound and costed.

C/Can