DATA QUALITY CHECKLIST

For data you can use and trust

OPEN NORTH

CONTENTS

METADATA

Organization or Project Name		
Document version number	Latest update (yyyy-mm-dd)	Person accountable for data quality
Contributor(s)	Priority action items	

STEP 1

Describe your data use case [1]

Description

Feedback

(1) See the Tips section below for useful tips on how to complete this step.

STEP 2

Identify potentially useful data sources [2]

Potential data sources	Relevance

(2) See the Tips section below for useful tips on how to complete this step.

STEPS 3 TO 5

Define requirements, assess data quality, and take action

Think of the following types of measures:

Technical (better tools and software, automation, features to detect duplicates, data quality metric dashboards).

Administrative (processes, protocols, policies, training).

Governance (decision-making structures).

Data quality dimensions *	Step 3 - Requirements Translate the dimensions into specific requirements for your use case. Provide a justification if you believe the dimension doesn't apply. **	Step 4 - Assessment How well does your data meet these requirements?	Step 5 - Action plan If requirements are not adequately met, what can you do to improve?
Well documented	• Define which metadata fields are mandatory to fill out. • Metadata descriptions should be provided in clear and plain language.	**Poor** **Fair** **Good**	
Timely	Percentage of timely records for a given task. Note: you can determine ranges. For example: • 75-100% = good • 50-75% = fair • below 50% = poor	**Poor** **Fair** **Good**	
Complete	• Define the expected number or percentage of complete values for a given task or analysis. You can establish ranges for completeness thresholds as well.	**Poor** **Fair** **Good**	
Reliable	• Acceptable error threshold of 1 meter when measuring distance between two objects or people (accuracy) • Zero duplicate entries (uniqueness) • 100% of dates are formatted according to the ISO 8601 standard (consistency).	**Poor** **Fair** **Good**	

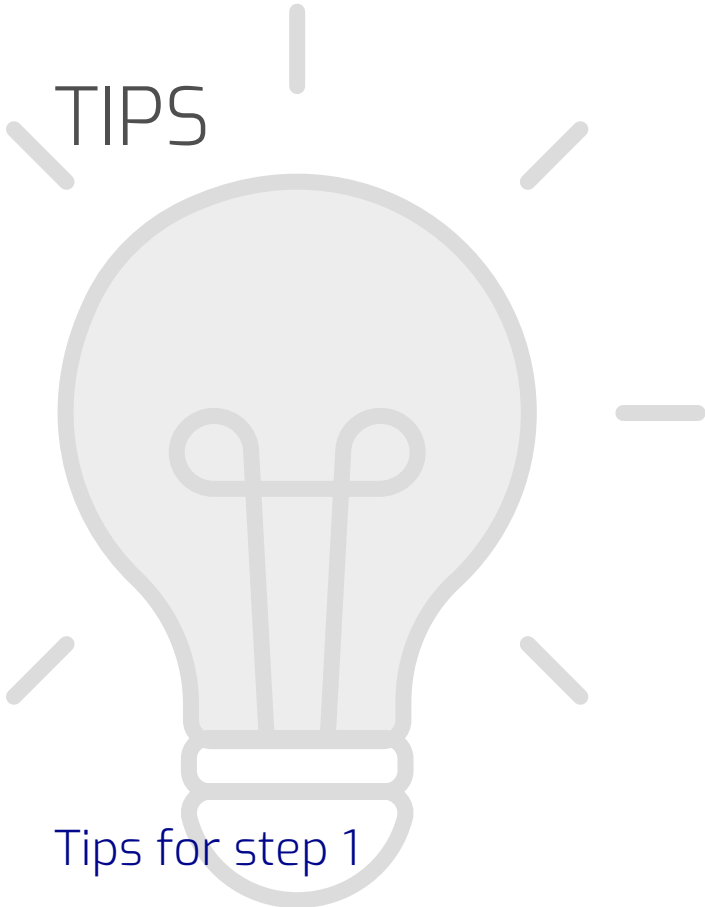* Definitions provided on the last page

** Examples are provided for each dimension

(continued)

Data quality dimensions *	Step 3 - Requirements **	Step 4 - Assessment	Step 5 - Action plan
Reusable	• All datasets are stored in a machine-readable format. • Data structures and naming conventions are standardized according to domain standards.	Poor Fair Good	
Accessible	• All staff members with the necessary training and authorizations have access to the data.	Poor Fair Good	
Secure	• All personal information has been identified and flagged for risk analysis. • Appropriate technical, administrative and physical guardrails have been planned for and implemented.	Poor Fair Good	
Relevant	• Impacted stakeholders are engaged to overcome perspective gaps. • Monitoring and evaluation strategies are implemented to measure whether the data helped reach objectives. • The socio-demographic sample data is representative of the target population.	Poor Fair Good	
(add dimension)		Poor Fair Good	

* Definitions provided on the last page

** Examples are provided for each dimension

TIPS

Tips for step 1

Be as specific as possible when describing the various components of your data use case:

- the problem statement
- the context in which the problem emerges
- the impacted stakeholders
- The desired outcome(s) and objective(s)
- your hypothesis statement on how the use of data will help reach the stated objective(s)
- any other relevant information such as conditions for success.

Tips for step 2

Here are a few general tips to help you identify fit-for-purpose data for your use case.

1. Define the data scope

Thinking about data quality begins here by asking yourself questions surrounding data relevance. In other words, finding data that is helpful and applicable for the task at hand. Here are some questions to ask yourself:

what	What do I want to observe and better understand (people, built environment, nature, services, etc.). How has this been measured in the past? What are the strengths and limits of these methods?
where	Is the data tied to specific geographical boundaries (country, region, continent) and jurisdictions (federal, state or provincial, or local)?
when	Is there a relevant time period during which the data must be collected?
topics	Is the data tied to a specific topic or domain, such as healthcare, mobility, housing, etc.? Is the topic multidisciplinary?
frequency	Does the data need to be updated at a specific frequency?
format	Does the data need to be available in a specific format?

2. Identify potentially relevant data sources

Internal data   Is your organization or team currently collecting this data or does it have the capacity to collect this data in the future?

Open data   Is the data openly published? In a push towards more open government, public organizations increasingly provide open data platforms where you can filter your searches to find open datasets. Depending on your geographic constraints, you could look at federal, state or provincial, or local level open data portals.[3] In addition, numerous data repositories routinely publish open datasets that are global in scale and scope (see examples below).

Shared data   Seek data marketplaces to purchase bespoke datasets (for example from Statistics Canada) or, alternatively, seek existing data partnerships where you can join like-minded organizations to collaborate, share, and pool data to reach common objectives. Open North's course "Fundamentals of Data Partnerships" (Code D104), available on our free online training platform, offers examples of existing data partnerships across various domains such as climate or health data.

Examples of global data repositories				
Core Certified Repositories – CoreTrustSeal	Re3data	FAIRsharing	Google Public Data Explorer	UNdata
World Bank Data	Open Knowledge Hub	Our World in Data	Global Health Observatory	OpenStreetMap

DEFINITIONS OF DATA QUALITY DIMENSIONS[4]

Well documented	Well documented data is accompanied by comprehensive and appropriate metadata.
Timely	Timely data is sufficiently current for the task at hand.
Complete	Complete data does not contain any missing values or has an acceptable number of missing values for the task or analysis at hand.
Reliable	Reliable data is the result of consistent and stable data collection processes that can be repeated with the confidence that similar results will be obtained.
Reusable	Reusable data has the potential to serve a purpose beyond its original intent.
Accessible	Accessible data can be easily used and accessed by intended users.
Secure	Secure data is safeguarded, with access limited to individuals possessing suitable training and authorization.
Relevant	Relevant data is helpful and applicable for the task at hand.

(4) For more details, see Open North's guide : Data Quality in Practice

About Open North

Open North is a Canadian nonprofit dedicated to advancing the common good. As an organization with expertise in data governance and digital strategy, we work alongside governments, nonprofits, and mission-aligned businesses to create transformative digital strategies and data governance frameworks.

Open North's team is made up of professionals with a wide range of expertise, including in government, strategic and operational planning, urban planning, community building, information technology, applied research, international development, and policy development. With our diverse backgrounds and skills, Open North's team members bring valuable perspectives and experience to all projects.

Open North is part of Montréal in Common, a project led by the City of Montréal as part of the Smart Cities Challenge, carried out with the financial support of the Government of Canada.

opennorth.ca

About the Smart Cities Challenge and Montréal in Common

Montréal in Common is an innovation community led by the City of Montréal, whose partners are experimenting with solutions regarding access to food, mobility and municipal bylaws, with a view to rethink the city. Montréal in Common projects are made possible thanks to the prize awarded to the City of Montréal by the Government of Canada as part of the Smart Cities Challenge.

Author: Samuel Kohn