

Statistical Best Practices to Improve Data Quality

by Cynthia Wilson and Heng Xin Yu
 Policy Planning Division,
 Singapore Department of Statistics

Introduction

Data have become increasingly important in all organisations, providing insights to facilitate planning and decision-making. Valuable insights are only possible with good quality data.

The Statistical Best Practices (SBP) 2020 handbook published by the Singapore Department of Statistics (DOS) details the statistical processes to produce quality data. The best practices described in the handbook are in line with the latest standards and recommendations by international bodies and take into account recent developments in the use of alternate data sources and methods.

While the SBP 2020 handbook was developed with the public sector agencies in mind, it is also a useful resource for private sector organisations seeking to improve data quality. This article highlights the key concepts of data quality and how the best practices are structured in the SBP 2020 handbook.

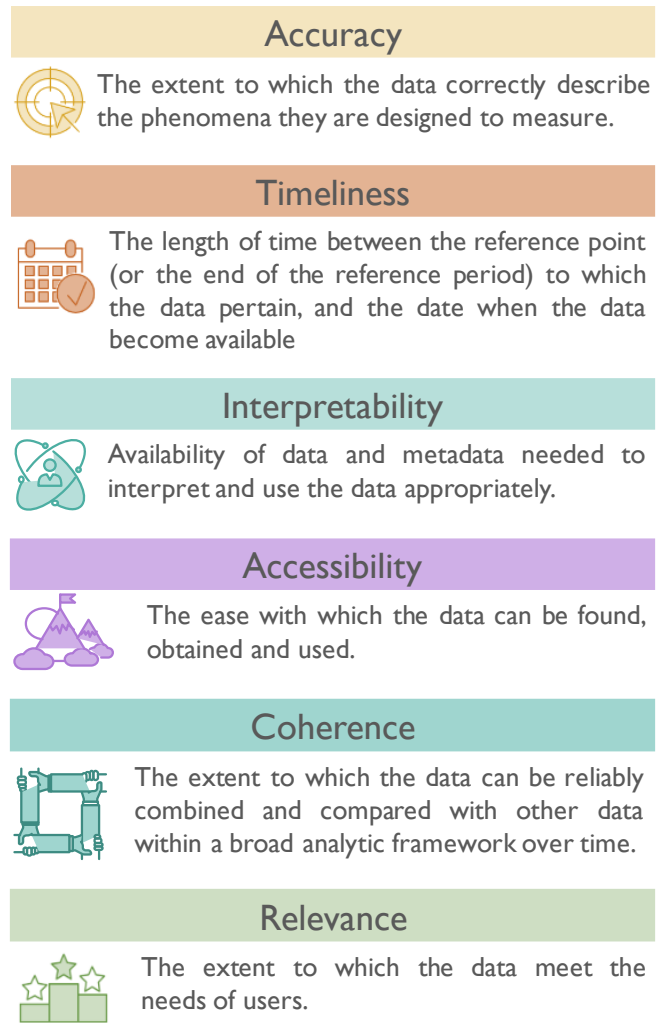
Data Quality

Quality Dimensions

To produce quality data, one must first understand the factors determining data quality. There are six quality dimensions (Chart 1) which are interrelated, with trade-offs between some of them. For example, to produce data in a timely and frequent manner, the accuracy of the data may be affected. This may arise if some sources of data are not yet available or it is not cost-effective to obtain the information within the required timeframe.

When the quality dimensions are interrelated, it implies that there are dimensions that are not mutually exclusive. For example, when data is coherent, it is comparable with other data and can be reliably combined. The new combined data may then be used to meet the needs of users, thus making it relevant.

CHART 1
 DATA QUALITY DIMENSIONS



Statistical Processes

The SBP 2020 handbook identifies five key statistical processes of a statistical activity¹, which can be further divided into sub-processes (Chart 2). While the statistical processes and sub-processes are arranged in a sequential order, some may not be applicable to every activity. For example, if the activity involves producing output from administrative data sources, Key Process 2: Survey Design will not be applicable since there is no need to conduct surveys to obtain the information.

¹ Any activity in the collection, handling and processing of information to produce statistical outputs, e.g. conducting a census or survey.

CHART 2 KEY STATISTICAL PROCESSES AND SUB-PROCESSES OF A STATISTICAL ACTIVITY

1 - Planning & Preparation 1.1 Initial Planning 1.2 Planning Proposal 1.3 Resources 1.4 Training 1.5 Confidentiality	2 - Survey Design 2.1 Coverage & Frame 2.2 Sampling Techniques 2.3 Questionnaire Design 2.4 Questionnaire Testing
3 - Data Collection 3.1 Survey Data Collection Procedures 3.2 Use of Data from Administrative Sources & Databases	4 - Data Processing 4.1 Data Capture & Coding 4.2 Editing 4.3 Imputation 4.4 Seasonal Adjustment 4.5 Data Integration
5 - Compilation, Analysis & Dissemination 5.1 Compilation & Analysis 5.2 Dissemination 5.3 Documentation	

CHART 3 SELECTED REQUIREMENTS AND RECOMMENDATIONS

5.1 Compilation & Analysis

- S5.1.1** Classifications on the compilation of data that are in broad conformity with national/international recommendations are adopted.
- R5.1.1** If more than one data source is used, the consistency and impact of data sources are considered and integrated into the analysis.

5.2 Dissemination

- S5.2.1** The confidentiality requirements governing the data being disseminated are complied with.
- R5.2.1** Prices of data outputs are disclosed clearly.

5.3 Documentation

- S5.3.1** The entire statistical activity is documented. For example, adequate information on the meaning of the data and the methodology used to collect and process them, the type of statistical analysis and techniques used, and sources and references are provided.
- R5.3.2** For regular surveys, the documentation is reviewed regularly to ensure that it meets the needs of users.

Statistical Best Processes

In the SBP 2020 handbook, the best practices are divided into two groups, namely, **requirements** (which are mandatory for DOS and Research and Statistical Units (RSUs) that collect and produce data under the Statistics Act, while other public sector agencies are encouraged to adopt) and **recommendations** (which all public sector agencies can consider adopting where appropriate). Overall, there are 113 requirements and 96 recommendations in the handbook.

For example, the fifth key process 'Compilation, Analysis and Dissemination' has three sub-processes, namely, 'Compilation and Analysis', 'Dissemination' and 'Documentation' (Chart 3). Best practices considered requirements are prefixed with the letter 'S', while those that are recommendations are prefixed with the letter 'R'.

- When compiling data, there is a requirement to adopt classifications that are aligned with national and international standards to ensure data coherence, which is one of the six data quality dimensions. One of the recommended practices is to consider the consistency and impact of different data sources.
- When disseminating data, one of the requirements is to comply with the confidentiality obligations (e.g. non-disclosure of identifiable information), while it is recommended to clearly state the prices of statistical outputs.
- When performing documentation, one of the requirements is to record the entire statistical activity, while reviewing the documentation of regular surveys is a recommended practice.

Concluding Remarks

The SBP 2020 handbook aims to improve the quality of data to facilitate planning and decision-making. It would serve as a useful resource for all organisations, including private sector organisations that are involved in producing data.



To access the full SBP 2020 handbook, visit
www.singstat.gov.sg/standards/standards-and-classifications/sbp