

Data Governance and Data Quality



Last updated: 14 Apr 2022

What's inside?

i)	How to Use This Guide	2
ii)	Introduction	3
iii)	Data Governance	4
	a) Data Governance Roles and Responsibilities	4
	b) Data Policies	5
	STB's Data Trust Charter	5
	Our Commitment:	6
	Commitment from our Data Partners:	7
iv)	Data Sensitivity in Stan	9
v)	Challenges and considerations during data sharing	9
vi)	Data Standardization	11
	a) Tourism data reference model	11
vii)	Data Quality Management	13
	a) Data Profiling	13
	b) Data Quality Processes	16
viii)	Data Quality Mindset Awareness	18

How to Use This Guide

The Data Governance and Data Quality Playbook aims to help tourism industry stakeholders understand the enablers for managing and exploiting data in Stan. The policies and frameworks adopted in Stan seek to enable an environment for (1) safe and secure storage and usage of data and (2) to promote clean and quality data.

Key to this playbook is the data governance and data quality management framework. The framework can be examined closely from a few perspectives, namely, (1) data governance from a centrally controlled data management of a robust data integration platform; (2) compliance with tourism industry data standards; (3) building the right processes to facilitate and sustain good data quality; and lastly, (4) instilling the data quality mindset in our people. Each chapter discusses specific aspects of data governance and data quality.

As business needs evolve, there may be a requirement to look at enabling data collaboration among stakeholders and partners to support decision making, while ensuring that the data provided is of good quality. Product features such as data marketplace and sandbox of (Stan) serve to maximise opportunities for data sharing and collaboration.

Introduction

“You can have data without information, but you cannot have information without data.”

– Daniel Keys Moran, computer programmer and science fiction writer

Most organisations are clear with the consistent theme that good data sharing is key to driving business needs: organisations are increasingly using data to create meaningful and content-safe interactions with their consumers. Such organisations are also the ones noticing positive business results through the use of data analytics to provide real insights and solutions to problems in today’s competitive business environment.

A collaborative data study by Boston Consulting Group and Google in October 2018 shows that companies in the top quartile of maturity that engaged consumers with personalized experiences were able to achieve an average of 11% annual incremental revenue and 18% cost efficiency.



of Asia companies believe they are below average or average at using data to draw insightful meanings to drive sales or improve business outcomes



of Asia companies consider data to be critical to their success in marketing or sales efforts

This playbook contains some best practices on data governance, data quality and data standards for our tourism industry stakeholders’ reference.

Data Governance

As more data is being used to drive business decisions, the inherent risks of mis-using inaccurate or irrelevant data are ever growing. The risks are both internal and external. Data governance is therefore, more than ever before, becoming a major concern in most organisations and recognised as an indispensable measure to mitigate these risks.

a) Data Governance Roles and Responsibilities

Organisations can develop a data governance framework consisting of pivotal data governance roles including data owners and data custodians. Together with the senior management with decision making roles, the team will be responsible for the data governance and data stewardship success of the organisation.

With the right data governance policy in place, formalised accountability of the roles can be appointed to take the organisation forward for the company's best interests in ensuring good data governance.

Data owners can help to provide direction in complying with data governance policies and frameworks as implemented. As data stewardship issues are likely to surface over time, data owners can also help to appoint business domain users as data custodians and provide the necessary training and communication to these users.

In Stan, users are the owners of the datasets they contributed. The data owners in your organisation will be responsible to:

- approve the data sensitivity level for the data contributed
- approve appropriate data access for the data contributed
- ensure the quality of the data owned, including monitoring and analysing data quality metrics and ways for improvement to quality of source data
- work with STB to resolve data related issues

With the responsibility that comes with data ownership, STB has also drawn up a data charter to describe what that responsibility entails.

b) Data Policies

STB's Data Trust Charter

STB has drawn up the Data Trust Charter to support data sharing between STB and the industry stakeholders.

It represents STB's public commitment to secure and responsible data sharing and usage of the data contributed by the industry stakeholders (Data Partners).

It mirrors the relevant Singapore data-related laws, guidelines and directions that govern STB's conduct.

Equally, it describes the commitment expected from the Data Partners.

Our Commitment:

We are committed to ensure that Stan remains a safe environment for our Data Partners to contribute data via the following commitments:

1. Data Usage & Transparency

Data is collected to fulfil STB's legislative functions such as to develop and promote Singapore as a travel and tourist destination. For any other use, STB will seek prior consent from the industry through Terms of Use in Stan or collaboration agreement.

2. Data Confidentiality & Sharing

Any data shared by a Data Partner will not be disclosed to any other Industry Stakeholder without prior consent. Aggregated benchmarks and insights that are derived from fewer than five individual tourism stakeholders will not be shared;

No benchmark or insight will be shared where the contribution of any one tourism stakeholder exceeds 50% of the aggregated data

3. Data Protection

Reasonable measures will be taken to ensure the security of the data shared with us, such as the administration of access control on all systems holding data shared.

4. Data Accuracy & Relevancy

Reasonable measures will be taken to ensure the data collected, compiled and shared is complete and accurate.

5. Personal Data Privacy

Reasonable measures will be taken to ensure privacy of individuals will not be compromised through sharing data with us. STB has in place a Personal Data Protection Policy for this.

6. Modification & Withdrawal

Data Partners have the right to request for modification and withdrawal of data shared, subject to reasonable justifications and mutual agreement, to maintain the integrity of the data sharing platform.

Commitment from our Data Partners:

STB welcomes organisations who are coming on board with us as Data Partners. The responsibility remains with the Data Partners to ensure that the data contributed via Stan is of good quality.

With the data charter in place, STB would like the Data Partners to have the same commitment to only contribute data that can be trusted and used by stakeholders in the tourism industry:

1. Data Quality & Format

Reasonable measures shall be taken to ensure the data shared is complete and accurate and in a form that STB may reasonably require.

2. Frequency of Data Sharing

Data shall be shared in a timely manner in line with a mutually agreed frequency.

3. Variety of Data

The different types of data collected shall be considered and the types of data to be shared shall be mutually agreed upon. Data Partners should provide as great a variety of data as possible to be achieve greater insights and value for the industry.

4. Volume of Data Shared

The volume of data to be shared shall be mutually agreed upon. Data Partners should provide as much data as possible to enhance the insights that may be generated for mutual benefit.

5. Data Granularity

The granularity of data shall be mutually agreed upon. Data Partners should provide data that is as granular as possible to enhance the usefulness of insights generated.

The full document may be found at:

<https://www.stb.gov.sg/content/dam/stb/documents/corporate-governance/Singapore%20Tourism%20Board%20Data%20Trust%20Charter.pdf>

Data Sensitivity in Stan

STB respects the commercial sensitivity of industry data and will take reasonable measures to ensure that stakeholders' specific data remains confidential. Any industry benchmarks and insights shared by STB will be done at an aggregated and anonymised level to ensure that the data provided will be neither identifiable nor directly attributable. Data marked by STB as "Confidential" will be subjected to Confidentiality clause in Stan's Terms of Use.

Private Space

Industry stakeholders should not upload data that is personal identifiable data or commercially sensitive data that could cause financial loss if it is disclosed without authorisation (e.g. Intellectual Property yet to be patented/protected).

For more information on Stan Terms of Use, please refer to <https://stan.stb.gov.sg/portal/terms-of-use.html>

Challenges and considerations during data sharing

In today's data environment, many organisations face the difficulty of integrating data due to the silos data residing in legacy systems or inter-systems incompatibility arising from data being stored in different formats in the different databases and information models. Hence, we will develop the Tourism Data Reference Model (TDRM) progressively that specifies the data standards (such as the meaning, format etc.) for data elements that are commonly used across tourism industry so that data users are able to merge consistent datasets for analysis between STB and industry data and enable inter-operability of systems.

With the rolling out of a common data reference model, organisations can look forward to adopting the new data standards by converting and harmonising current data to the required format before data contribution to Stan.

As data sources become more diverse, the need for data standardisation has grown exponentially. It would therefore be more beneficial if the data is collected in a standardised format right from the start. The following are some of the benefits of data standardization:

- Organisations are able to work with the data more effectively with less preparation time. This in turn will support STB's efforts to develop a data sharing repository for all submitted data.
- The use of these data standards is also expected to benefit the industry by streamlining the flow of data from collection through submission, and also facilitating data interchange between partners and providers.

Data sharing is a multi-disciplinary process which involves technology and legal considerations. It is common to hear of trust and security concerns being raised despite the huge benefits that large amount of data can be leveraged on for data analytics, machine learning and artificial intelligence uses.

The next few chapters will address these trust and security concerns for data sharing and how data owners and data custodians in your organisations could submit datasets with good data quality at the point of data ingestion, whether it is via Stan online portal webforms or via direct uploading of datasets.

Data Standardization

Data from different sources depicting the same data element, for example, 'date' be defined in a different format, data type, data size. A date can be:

- April 2, 2020
- 04-02-20
- 04/02/2020
- 2020-04-02

This will potentially create problem when one tries to consolidate the same data elements from the different sources. Without data standardisation, it is unlikely that data from different sources would be able to merge to form a single source of data for sharing. If all data sources use the same data standard for dates, then there will not be a need to do this extra step of data conversion.

With this key issue in mind, STB is working towards having a data reference model for the tourism industry comprising of common data standards. Organisations are recommended to adopt the data standards early during their systems implementation so as to enjoy the benefits of the large data pool that is available at STB for their analytics modelling.

a) Tourism data reference model

The TDRM is an evolving model that STB is rolling out progressively with inputs and advice from industry stakeholders.

The goal is to help organisations understand and recognise the usefulness of tourism data reference standard as a means to derive real business value from the sharing of data within the tourism industry. With the

adoption of data standards, organisations can look forward to sharing of data not just with STB but also with other collaborators through the use of STB's data analytics workspace such as sandbox and marketplace.

The principles for developing TDRM are:

First, STB data standards to be used where applicable.

Second, if data element is not included in STB data standards, relevant International standards such as ISO standards would be referenced.

Lastly, if there are no existing reference standards in both STB data standards and international standards, industry standards would be referenced.

With the implementation of the TDRM, industry stakeholders can look forward to having a common set of data standards to work with. STB will continuously seek inputs from the tourism key stakeholders to identify business needs and how the TDRM can be further improved upon with each new revision.

The datasets that stakeholders contribute to Stan will be subjected to the data quality checks at the point of collection. Therefore, the data elements in the dataset to be provided to Stan must align with our data reference standards.

The latest version of the TDRM can be accessed via the Data College page:

<https://www.stb.gov.sg/content/stb/en/trade-events-and-resources/data-college/learn-about-data.html>

Data Quality Management

To ensure sustainability of good data quality, it is important to have the right tools to surface the problem areas. One of the ways to do so is to implement data profiling to ascertain the baseline and to start working on improving the quality from the baseline profile results. The data profiling process, which is described below, is the first step towards good data quality.

a) Data Profiling

In the earlier chapters, the playbook explains some of the key areas that organisation can look at to create and sustain data quality. In this chapter, guidelines are given on how data quality can be assessed and monitored using 1) data profiling and 2) data quality processes compliance.

Data quality is the assessment of the condition that the data is in according to pre-defined dimensions. By quantifying the various aspects of data quality, it allows us to know at a glance what are the potential shortfalls and issues with the ingested data sets.

In Stan environment, a single agreed-to set of data quality dimensions has been defined. The data quality generally focuses on 7 different dimensions: whether there is enough data (completeness); whether it is correct (accuracy, validity); how well it fits together (consistency, uniqueness); and whether it is available (relevance), up to date (timeliness), accessible, usable and secure.

Data Quality Metric	Definition
Completeness	Refers to the degree to which all required data are available in the data set
Validity	Refers to the measure of conformity to defined domain of value. Checks are done against the TDRM.
Uniqueness	Refers to the discrete measure of duplication of identified unique data record within a data set
Consistency	Refers to the absence of differences (i.e. pattern, frequency, values) between the data items representing the same objects across different data sets
Timeliness	Refers to the degree to which the data is up to date and available within acceptable time frame, timeline and duration
Accuracy	Refers to the degree to which data correctly describes the real-world object or event being described.
Relevancy	Refers to the closeness between the data set required and the data provided

Data quality assessment involves creating metrics and thresholds around the quality of the data that is captured. The data quality assessment is based on the findings of data profiling. Data profiling is therefore, an indispensable step in evaluating the quality of the data elements that were identified in the business glossary.

As an organisation begins to understand where their largest data quality issues lie, they can start setting targets for improving the data quality.

Organisations can perform data profiling either with tools that are provided by technical vendor (for example, a report that shows which fields are blank most frequently) or through manual review processes such as looking at dashboards from the previous day. The end goal of data profiling is to understand where issues lie with the data and identify areas of improvement in the quality of data being recorded.

Data profiling can provide the following benefits:

- Identification of critical quality issues;
- Discovery of duplicate records;
- Supporting a more thorough data quality assessment at a later date and;
- Discovery of where existing policies and best practices are not being followed and identification of new policies, procedures, and best practices.

Using data profiling as a tool to drive data quality initiatives would ensure that consolidated datasets can support decision making that directly affects organisations' business outcomes.

b) Data Quality Processes

Organisations can begin by establishing data governance policy and defining data cleansing requirements that identify what critical data needs to be cleansed. To further enhance the process, procedures that should be used to improve the capture of data at the point of origin should be clearly specified.

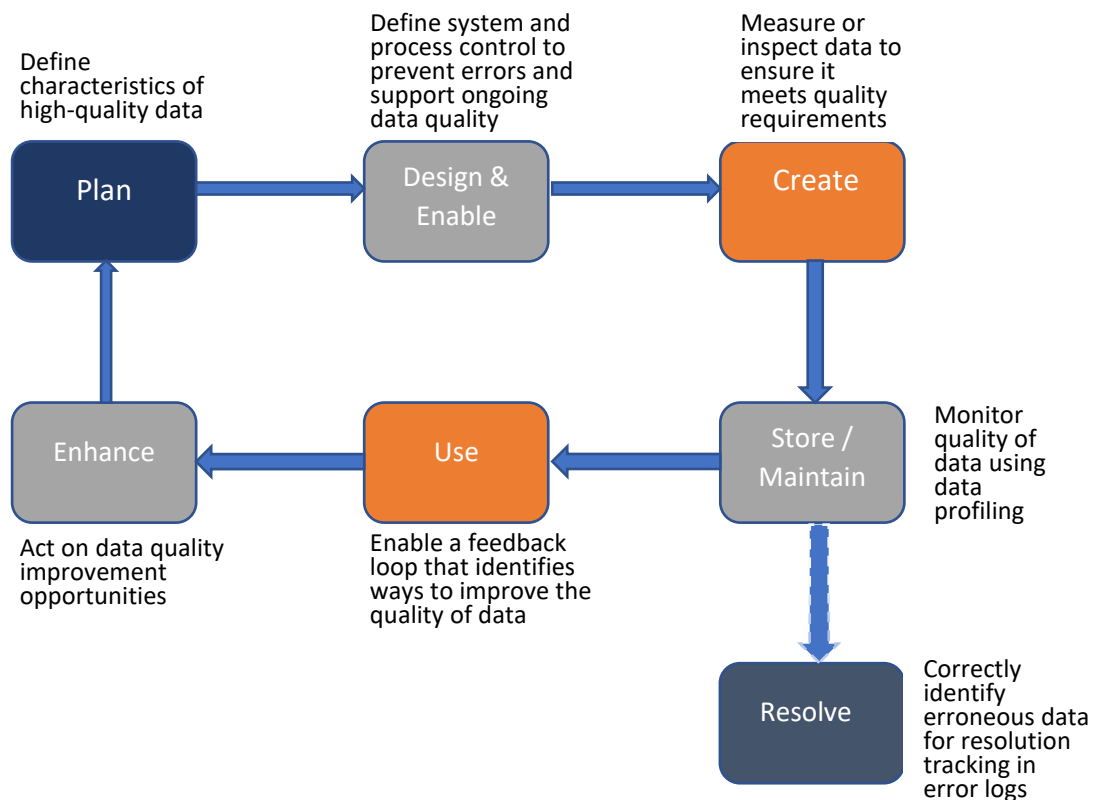


Figure 1. Adapted from DAMA Data Management Book of Knowledge (DMBOK2)

Once the business rules are determined, validation checks can be built into systems or a controlled data management platform to automate data checks during data collection. The point-in-time checks will help to reduce the errors during data ingestion to the database and ensure that there is always only a single view of any data available for consumption at all times.

Similarly, if a data is consistently captured in the wrong field, data owners can work with their data governance team to have field checks that ensure numerical data is captured for

fields like phone number, while alphanumeric data is captured for fields like address.

With the use of data validation checks before data ingestion, organisations can identify where the root problems are and teams can work together to specify areas of improvement. For example, if it is identified that a specific data element is not captured, data owners can work with their data governance team to make the field mandatory.

Lastly, the life cycle of a data quality process should enable a feedback loop to continuously monitor data quality scorecards populated with metrics. Data owners can act on these improvement opportunities and work on the feedbacks received to minimise the percentage of errors found in their datasets.

Data Quality Mindset Awareness

It is imperative to reiterate that data quality is part of an organization's culture. When employees are made to realize the importance of accurate data, they will be extra careful in handling data whether it is during collection or when it is being processed. When heads of different departments are aligned under a common goal of being data-driven, data quality will automatically be on the agenda.

It is no longer enough to want to be data-driven. Organisations must implement a data-conscious culture, which is only possible when data quality is prioritized, and employees are aware and trained in Data Management best practices.

In retrospect, ask yourself and your team — do you have data you can trust?

Therefore, basic data quality training should be part of an organisation's data quality plan. Everyone in the organisation needs to receive basic data quality training. Employees should be able to understand their role in collecting, recording, managing, and handling data and what constitutes bad data as the lack of understanding of the effects of poor quality data can be detrimental to the success of the organisation. Having a data quality mindset will raise the awareness of what practices employees can implement to reduce the occurrence of bad data.

The responsibility of maintaining good data quality should not be solely that of an individual data consumer or a functional team. Rather it should be the responsibilities of all employees within the organisation to ensure good quality enterprise-wide data is generated. In other words, every employee in the organisation should adopt a data-driven mindset and enable the culture to change throughout the organisation.

How to begin?

For a start, it is recommended that the organisation carries out an assessment of the data governance and data quality planning. The goal of the assessment is to help the organisation understand, identify and document the gaps in the processes that could cause poor data quality.

A data quality plan generally includes data quality goals and objectives, data quality principles, issues that the organisation will address, an identification of roles that are responsible for each portion of the plan, and any toolsets that will be used. It is possible that organisation may not have a data quality plan early on, since they may not know where their data quality issues exist. However, as organisations become more experienced with managing their data quality, they may find it helpful to create a plan that will help them profile, assess, and cleanse their data on an ongoing basis.

The development of a good data quality plan aids in propelling organisations forward to receive the following benefits:

- Move from ad hoc efforts to a planned and staffed quality process, which may decrease cost and save time in the long term;
- Have approved prioritization of efforts, which may create a “quality culture;” and
- Create reusable methods, procedures, techniques, etc., that can be tested and improved over time.

To encourage organisations to start developing their transformation plan, organisations can tap into STB’s available tools in Stan by performing a self-assessment using STB’s Tourism Transformation Index (TXI). The TXI assessment helps organisations to understand their current

state of digital transformation and benchmark their score against their industry's and sector's performance. The TXI score helps organisations to assess their strengths and identify areas of opportunities in their digital transformation roadmap.

Along with the TXI score assessment, organisations can also look forward to having a head start in their digitalisation transformation by joining STB's DASH (Data Analytics SHift) programme to build their capability in data. Through the DASH programme, nominated individuals from tourism organisations will acquire the necessary data skills to drive data usage and improve business performance within their organisations.

Online knowledge and educational materials will also be disseminated and updated from time to time in STB's Data College designed to help organisation in their journey to achieve good data governance and quality.