GUIDE TO DATA VALUATION FOR DATA SHARING

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

pdpc PERSONAL DATA PROTECTION COMMISSION SINGAPORE

IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

In support of:

SG:D EMPOWERING POSSIBILITIES

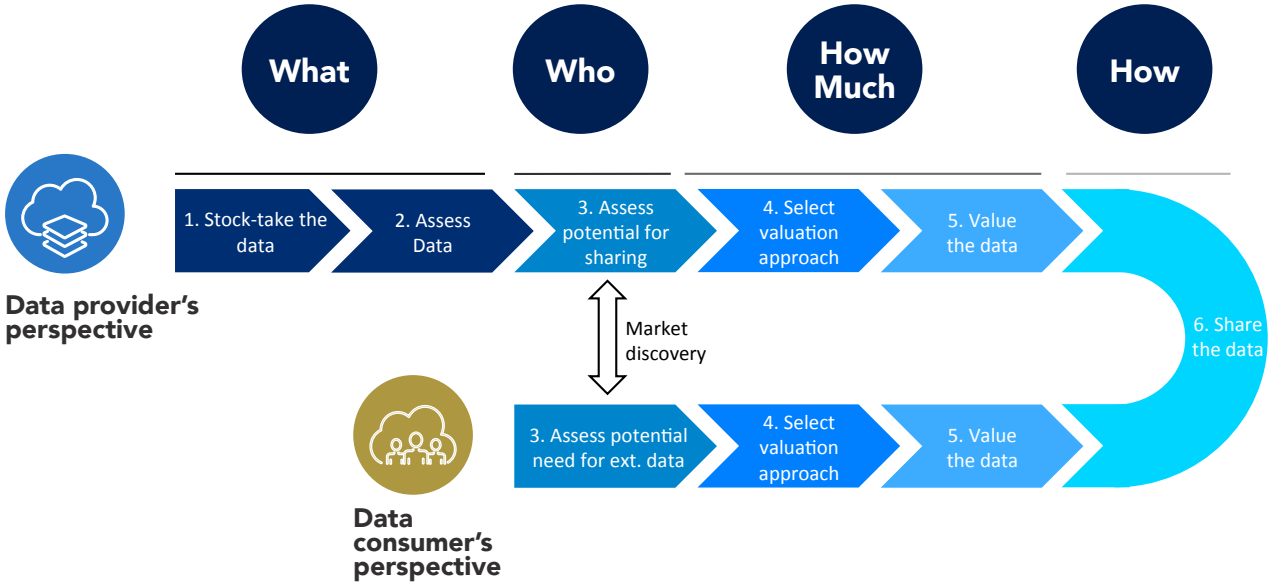
CONTENTS

INTRODUCTION

In today's digital economy, as more organisations begin to harness the power of data analytics and big data, data sharing is becoming essential for organisations to maximise the value of the data and gain access to new insights.

Organisations and individuals are increasingly receptive to data sharing, especially where they are able to see the benefits. A key obstacle to data sharing is the difficulty in assessing the value of the data assets. While it is widely accepted that intangible assets like data are valuable, the challenges of measuring that value is an issue.

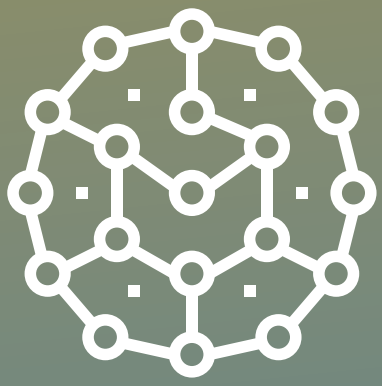
This Guide has been written to help organisations assess and value their data to enable greater sharing of data for competitive advantage. It has been designed for use by data providers[1] and data consumers[2] to aid discussions relating to potential data sharing arrangements, and not intended to be used to derive a retail price for data assets. An overview is provided in the diagram below.



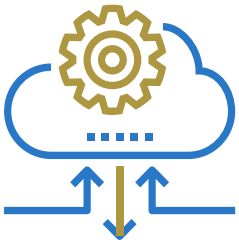
[1] *Data providers can be considered creators or owners of data, for example, government agencies and businesses.*

[2] *Data consumers collect or buy external data to generate additional insights and supplement internal functions.*

While appearing to be a linear process, the process of valuing and sharing the data is usually iterative, and the value of the data asset is unlikely to remain static. For instance, the value of the data asset may increase as organisations discover new use cases (e.g. through discussions between data providers and consumers) and gain a better understanding of the market. The value would also take into consideration the efforts by organisations in creating and managing useful, high-quality data for trusted data sharing purposes.



DATA VALUATION



STOCK-TAKING THE DATA

To value data, organisations have to first identify the various data assets available. Many organisations generate large amounts of data in their day-to-day business, making it challenging to conduct a comprehensive stock-taking exercise. A useful starting point would be to understand what constitutes a data asset, how it is generated, and what types of data are present.

What are Data Assets?

Data assets are intangible, and generally:

- **Are identifiable and definable** – Data assets may be made up of specific files, specific tables, or records within a database.
- **Promise probable future economic benefits** – To have value, data assets need to have a useful application. Identifying productive uses for data is often necessary to assign value to the asset.
- **Are under the organisation's control** – The organisation must also have rights to use the data in a way that is consistent with its rights under applicable law and any contractual licensing arrangements, while protecting the data and restricting access to it by others.

Data Categorisation

There is no 'one-size-fits-all' solution when it comes to categorising data types. Depending on the industry and context, organisations have to define their own data taxonomy. While this is generally a lengthy and challenging exercise, organisations could set up their own unique taxonomy to share the data internally across business units and overseas, and externally.

As a useful starting point, organisations could consider the following dimensions:

(i) data source;
(ii) data domain; and
(iii) data category.

A data domain, in particular, provides an additional lens to consider potential use cases for data sharing. For example, there could be a lot of value for the data in other domains (e.g. telco data used by credit industry or transportation data used by advertisers).

Other considerations include the geographical source of the data, how strategic the data is to the organisation, and rights to use or share the data.

The following examples of data taxonomy provide a reference to guide organisations' stock-taking and categorisation.

Data Taxonomy Example 1:

Data Source	Examples
Authored Data Typically created through some kind of creative process	• Architectural drawings • Photographs • Music soundtracks
User-entered Data Data purposefully entered by users into a system without any expectations	• Social media posting • Reviews on third-party rating sites
Captured Data Recorded from events occurring in the real world or digitally	• Financial transactions • Web browsing logs • CCTV recordings
Derived Data Typically generated by combining, aggregating and otherwise processing other data	• Credit scores • Aggregated transactions

Data Taxonomy Example 2:

Data Category	Examples
Master Data Describes people, places and things that are critical to a firm's operations	• Customer data (name, address) • Supplier data (contact details) • Product data (product features) • Employee data (name, position)
Transactional Data Describes an internal or external event or transaction that takes place as part of the organisation's business	• Sales data (purchase history, credit card payments, sales order) • Payment data (payment date) • Geospatial data (current location)
Reference Data Information that is used solely for the purpose of categorising data	• Jurisdictions (area code) • Currencies (currency code) • Industry standard data (country code) • Demographic fields
Metadata Characterises other data, making it easier to retrieve, interpret or use the data	• Date of creation tag • File author identity tag • Audit trail data (accesses, changes) • Descriptive data (author, abstract)
Unstructured data Data lacking a consistent format or syntax to describe objects and attributes	• Social media posting • Car movements • Weather data • Photography

ASSESSING THE DATA

Once organisations have a stock-take of their data assets, the next step is to assess the data's value. The value of data lies in its usefulness in achieving organisational or business goals. Data value drivers are the key factors that drive the usefulness of the data, which will in turn, impact the value of the data on businesses. The diagram below provides an overview of these drivers.

Overview of Key Data Value Drivers

Usefulness

- The value of data lies in its **usefulness.**
- Data without concrete or potential **application** has no inherent value.
- **Data value drivers** are the key underlying factors that drive the usefulness of the data.

Completeness

- Generally, the more complete a dataset, the more valuable it is, as it reduces bias.
- Completeness of the data has to be determined subject to the use case.

Consistency

- Data is consistent if it conforms to the syntax of its definition.

Accuracy

- The degree to which data correctly describes the "real world".
- Provenance is a critical aspect of accuracy, as it lets the user understand the history of the data and account for errors.

Timeliness

- The more timely and up-to-date the data is, the more valuable it is.
- However, timeliness is a relative measure, which is dependent on the use case intended.

Exclusivity

- The more unique the dataset, the more valuable the data is.
- The extent of the impact of exclusivity on value is dependent on the use case.

Usage Restrictions

- The less restricted the use of the shared data is, the higher its value.

Liability and Risks

- Potential liability and risks associated with the data shared could reduce its value.

Interoperability/ Accessibility

- All things being equal, a consumer will choose the most accessible dataset.
- Interoperability or joinability of the data is often critical for the consumer to use the data.

The importance of each driver in determining the value of the data is dependent on the use case of the data. For example, timeliness may be more critical for an organisation seeking to use data to analyse emerging stock market trends, compared to an organisation seeking to use data to derive insights from examination results over the past decade. An elaboration on data value drivers, including how each may affect a dataset's value, can be found in Annex A.

ASSESSING POTENTIAL FOR SHARING

When assessing potential use cases for the data, either as a data provider or a data consumer, an organisation should consider all potential stakeholders in the whole value chain or ecosystem that it operates in. This is in view that potential use cases could be generated from new insights derived from combining data with stakeholders across the value chain, beyond the immediate key suppliers and consumers. For example, credit card providers could share data on purchasing patterns with goods manufacturers and not just retailers.

It is also useful to consider the motivations for the sharing of data. Common motivations may be broadly categorised as:

- **Income generation** – Organisations can leverage data to generate new products and services, or even new streams of business;

- **Cost reduction** – Data can be leveraged to reduce costs through better planning and optimisation of operations, as well as minimising and managing risks; or

- **Public or sectoral good** – Other than financial gain, data can be shared to improve public welfare or sector efficiency and productivity (e.g. sharing of data with public sector or industry organisations that are able to improve the operating environment for the business or general public).

Illustrations of Common Data Sharing Motivations

One way sharing

Transaction

Data Service Providers

Motivations

Data providers

Generate income	Strengthen existing product	Social good
Raw data may be licensed directly to a user or through an exchange or broker.	Data may be used to increase or better demonstrate product utility.	Organisations may share data to further public good objectives. This objective may or may not benefit the provider directly.

Data consumers

Enable or improve a product	Optimise investments	Improve operations
Third party data may enable a new product or improve an existing product.	Use third-party data to identify potential customers and focus advertising and sales resources.	Integrating data from other value chain participants, companies are able to improve their operations.

Examples

Credit card transaction data sold by banks to data brokers to understand consumer purchasing and target advertising.	Business offers customer analytics on their sales data to charge higher merchant fees.	Transport business sharing real-time traffic data with the government to help identify and address any traffic issues.	Market data platforms purchase pricing and other data from exchanges to improve their products.	Online retailer use customer profiles generated by data brokers to identify potential customers and target marketing expenditure.	Supermarket sharing shelf data with its suppliers for free in return for efficient stocktaking of Inventory.

Mutual sharing

Transaction

Data Service Providers

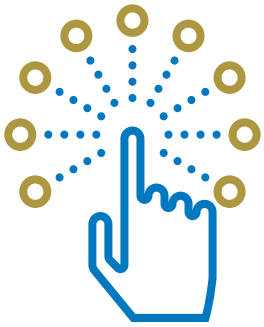
Motivations

Both parties

Industry development	Control fraud and abuse	Mutualise data management costs
Organisations may choose to strategically share data with a long-term view to benefit from industry development.	Firms may choose to share data to reduce fraud losses - this typically benefits all participants and society.	Some non-differentiating data can be managed centrally to reduce costs.

Examples

Logistic business sharing data as an industry sharing.	A group of organisations sharing information on known fraudsters to allow members to reduce fraud losses.	A group of organisations having a central management of customer information so that each member need not maintain a duplicate copy.



SELECTING VALUATION APPROACH

With potential data sharing use cases in mind, organisations, as data providers or data consumers, can start to value the data for these uses.

General valuation approaches and principles that are typically used to determine market value[3] of assets can be applied to valuation of data assets:

a. **Market Approach** – The Market Approach considers the market value for available identical or similar data assets, and would typically provide the best evidence of the market value of the data.

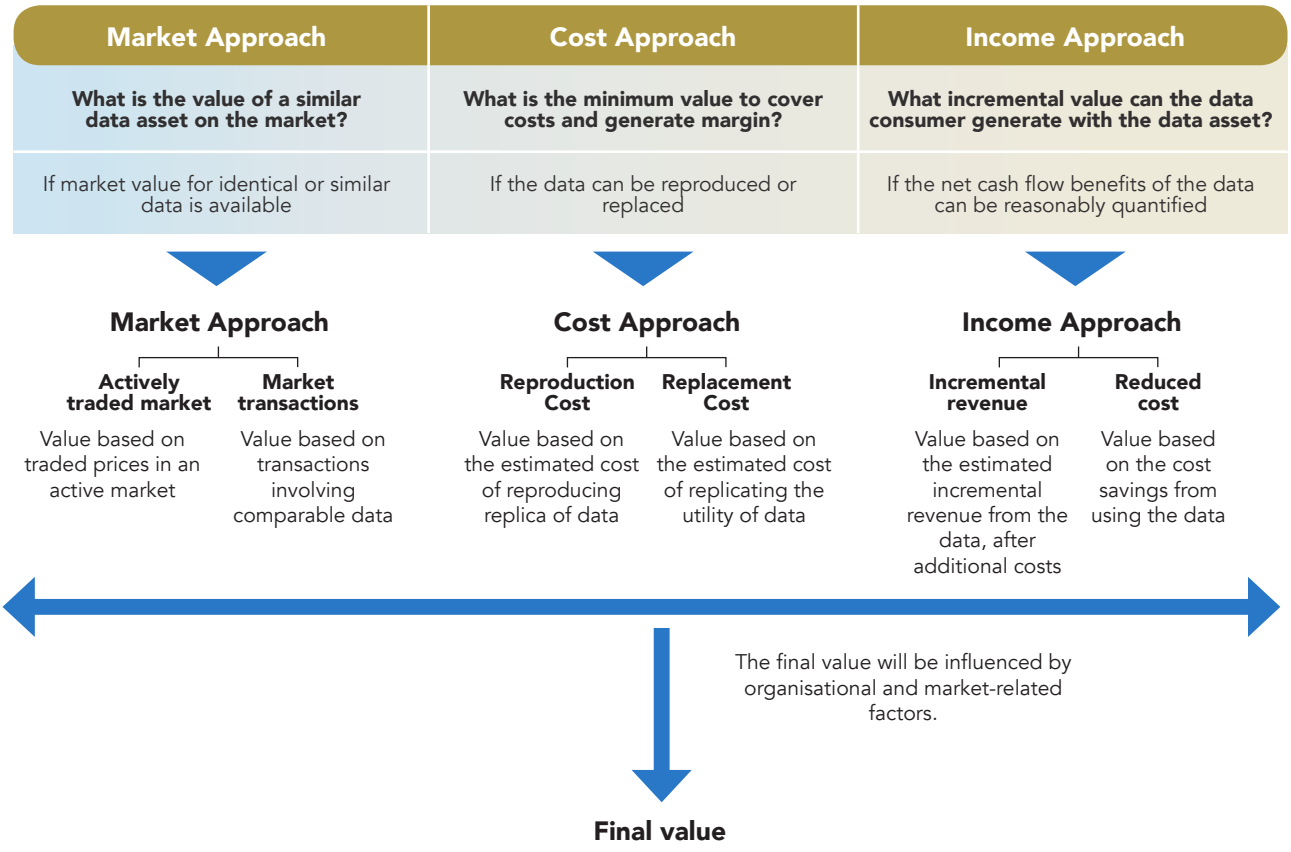
b. **Cost Approach** – From the perspective of the data provider, the Cost Approach involves the costs incurred to create the data assets. From the perspective of the data consumer, the Cost Approach considers whether it would be feasible to reproduce or replace the data, which can be characterised as a 'make-or-buy-decision'. The Cost Approach is typically calculated either as the cost to produce the data or to reproduce the data using similar input and methods ("**Reproduction Cost Approach**"), or the cost of replicating the utility of the data without creating an exact copy of the data ("**Replacement Cost Approach**"). The Cost Approach is typically used to provide a base value of the data, and does not capture future economy returns which the data may be capable of generating.

[3] *For the purpose of this guide, 'market value' is taken to be "the estimated amount for which an asset or liability should exchange…in an arm's length transaction" See International Valuation Standards Council [IVSC] 2017, International Valuation Standards 2017 ("IVS 2017"), paragraph 30.1, page 18*

c. **Income Approach** – The Income Approach considers that the value of an asset should depend on its ability to generate future cash flows for its owner or rights holder, and tries to take into account the incremental cash flow benefits to the data consumer from purchasing and using the data. It is the purest and most robust approach from a theoretical standpoint, but requires technical valuation knowledge to apply.

In determining which valuation approach to use, organisations may take into consideration the characteristics of the data asset, circumstances of the transaction and availability of information. While there are three main approaches to valuation, organisations are encouraged to apply more than one valuation approach in valuing the data so as to help the organisation arrive at a more robust view of the value of the data.

Overview of the Market, Cost and Income Valuation Approaches



Typically, the value range derived by the data provider and data consumer would overlap and the final value[4] of the data would be a result of other factors. While this guide focusses on the market value of data, the final value ultimately negotiated may be influenced by other subjective factors, such as the organisations' bargaining power, financial circumstances, asymmetry of information on the potential of the data[5] and ability to market the data among other market-related factors.[6]

[4] *The final value of the data can only be determined after the data is made available and is influenced by other market-related factors, even if the starting point is the market "value" of the data.*

[5] *For instance, the data provider may not be aware of how a data consumer is able to use its data to generate value. This may mean the provider accepts a lower value than the true value of the data.*

[6] *Number of consumers and providers, market mechanics, market size, and market maturity are examples of other market-related factors that can potentially influence the value of a data asset. In a matured market where there are many providers of the same type of data, a data provider may have to offer its data at a price lower than its value.*

Considerations for Approaches

The following table sets out broad data sharing circumstances and highlights key considerations when deliberating whether to share or acquire the data, through the perspectives of the three data valuation approaches:

	Data Provider		
Scenario	**Data sharing venture**	**Opportunistic data sharing**	**Data sharing optimisation**
Background of Data	Data produced specifically for the venture	Data is a by-product of core activity	Data already in existence and being shared
Valuation Question	Would the value that my organisation gets through the data sharing venture justify the investment?	What value could my organisation get by sharing the data?	Could my organisation generate more value by sharing an enhanced version of the dataset?
	Application		
Market Approach	Check the value of comparable data on the market.	Check the value of comparable data on the market.	Assess the gap between the market value of comparable products to the enhanced dataset versus the value of the existing data.
Income Approach	Estimate the potential extra cash flow that data consumers could get with the data.	Estimate the potential extra cash flow that data consumers could get with the data.	Compare the potential cash flow data consumers could generate with the enhanced datasets to the value of the current data sharing.
Cost Approach	Estimate the cost to produce the data plus an acceptable margin to validate that the value derived by the Market Approach and Income Approach will yield a higher amount to the Cost Approach.	Ensure that the costs to package and distribute the data for data sharing are low enough to have a Cost Approach, with the target margins below the market and income approaches.	Validate that the extra costs required to improve the data for data sharing is not more than the extra value assessed through the Market Approach and Income Approach, to enable similar or superior margin of return.

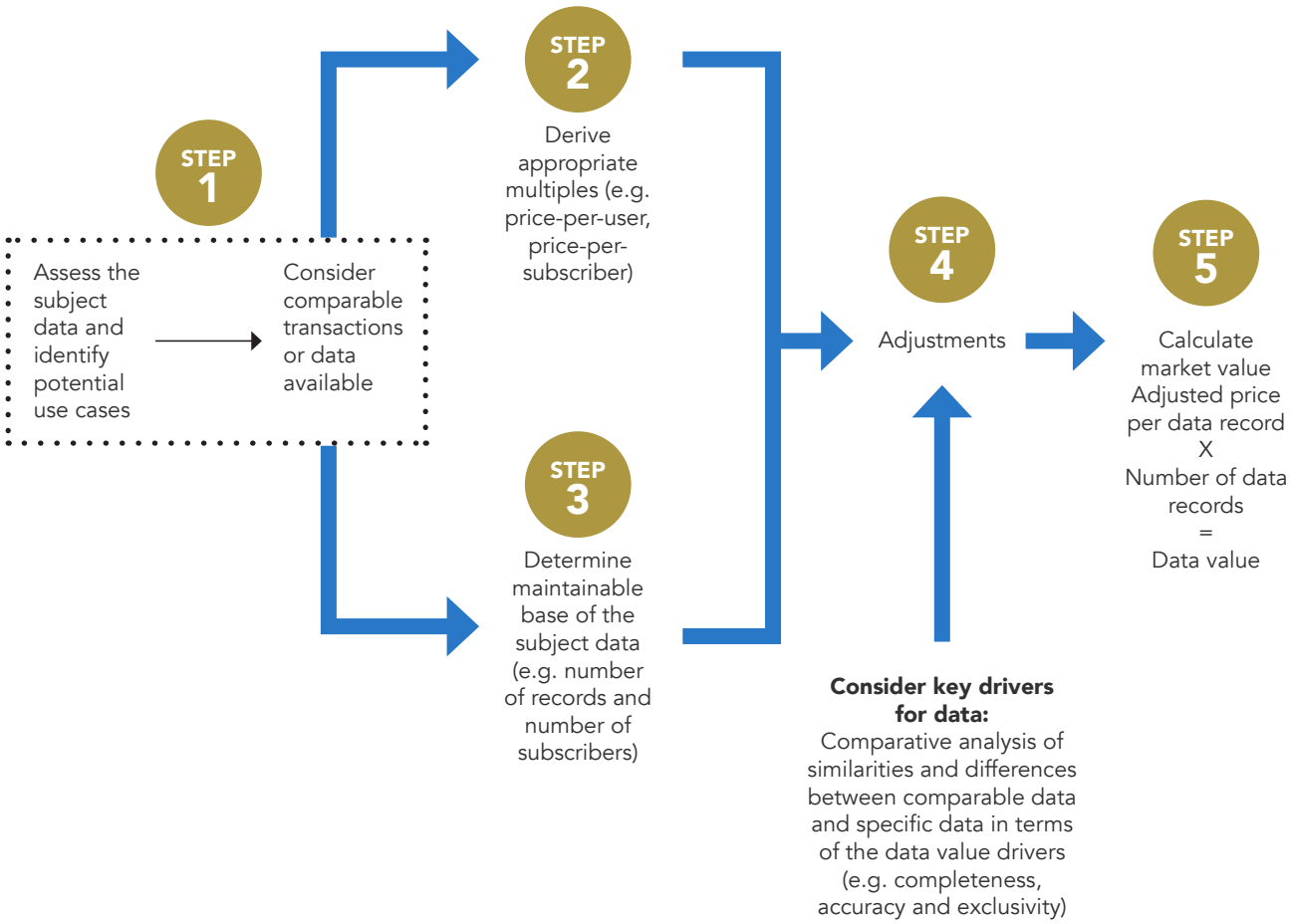
	Data Consumer		
Scenario	**Acquiring the data**	**Making/obtaining the data**	**Aggregating the data**
Background of Data	No similar data to the one in question	Similar dataset to the one shared	Complementary dataset to the one shared
Valuation Question	What is the value of the data for my organisation?	Is the additional dataset shared of a higher value than the one currently acquired / produced?	What extra value could my organisation generate by combining the shared dataset with the existing one?
	Application		
Market Approach	Check for comparable data on the market and its market value.	Check for comparable data on the market and its value.	Assess the gap between the market value of the two datasets combined (complementary + shared data) against the value of the existing data.
Income Approach	Estimate the additional income that may be generated with the data.	Estimate the additional income to the organisation if the data is not fully identical to the existing data the organisation has or is able to produce.	Estimate the additional income that may be generated with a combined dataset (complementary + shared data).
Cost Approach	Estimate how much it would cost to produce similar data (if possible). Typically, the data sharing would be ideal to the organisation if the amount derived from the Cost Approach is higher (i.e. costlier) than the Income and Market Approaches.	Assess the cost of producing a similar dataset. If the amount derived from the Cost Approach value is higher than the income and Market Approach, it would be economical to obtain the data from the other organisation instead of creating the data internally.	Estimate cost to produce a similar dataset if possible.

VALUING THE DATA

This section provides further details on each of the valuation approaches and how it applies to valuing data for sharing.

Market Approach

Where there is an actively traded market[7] for the data in question, organisations (whether from a data consumer or data provider's perspective) can look up the market price of the data. Where there is no active market, but examples of similar data assets are available, or valuation evidence from transactions involving similar data assets, or even existing arrangements where an organisation is providing similar data assets to others, the organisation can still review these evidences or transactions to derive the value of the data. The key steps, including the considerations of the data value drivers, are shown in the diagram below:



[7] *For instance, where there is evidence of regular trading frequency and volume on a data marketplace to provide pricing information on an ongoing basis.*

A case study showing how an organisation uses the Market Approach can be found in Annex B.

Cost Approach

If the costs to reproduce or replace the data can be estimated, organisations could use the Cost Approach to benchmark the value of the data. In general, the value of the data may be derived from estimating the costs of building and maintaining a database (e.g. input costs, software and storage costs, human capital costs, selling and marketing costs, and overhead cost), and the opportunity costs of investing in the data asset over other assets. The Cost Approach is typically used to provide a base value as it does not capture future economic returns which the data may be capable of generating. There are two methods for the Cost Approach:

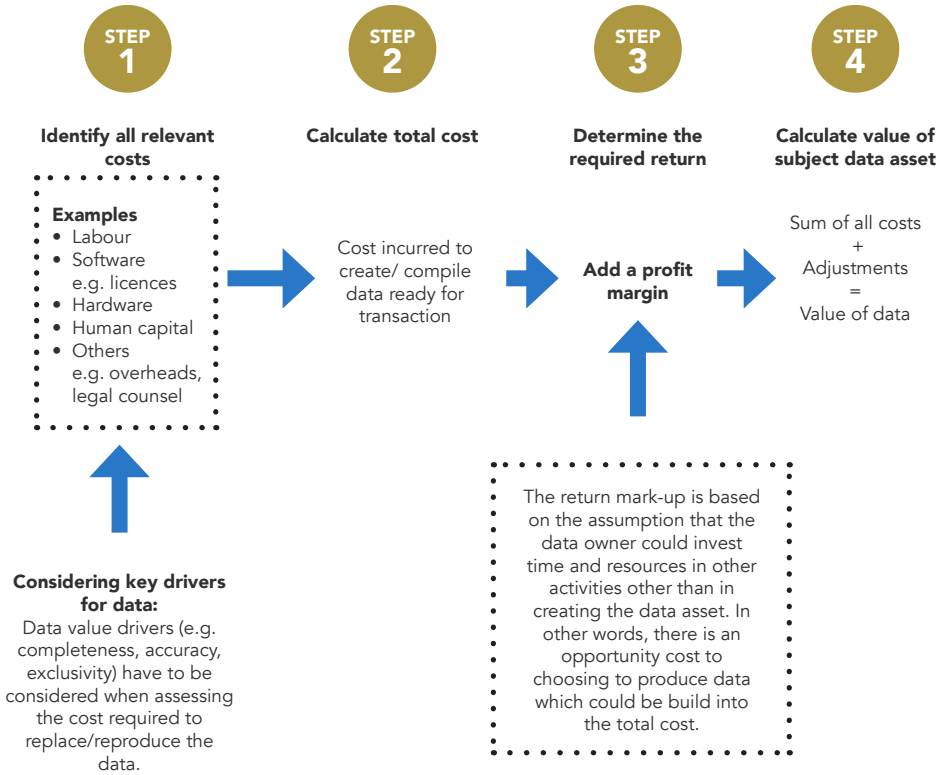
a. The Replacement Cost method represents the current cost of replacing a similar data asset with equivalent utility to the data asset being valued. Essentially similar to the Market Approach, the value derived from the Replacement Cost method would be equivalent to the cost of replicating the utility of the data, and not creating an exact copy of the data.

   To do so, organisations can observe transactions involving similar data assets currently available. Assuming these data assets are closely comparable to the data asset being valued, these transaction values can be considered to represent the Replacement Cost of the data asset.

   In particular, from a data provider's perspective, the organisation should also consider if there are alternatives to its data which would give the same utility and value to the user. This is because if there are alternative sources of data which give the same utility and value, then the Replacement Cost of the data may represent the maximum amount that the data consumer is prepared to pay.

b. The Reproduction Cost method represents the current cost of producing or reproducing the data in-house using similar inputs and methods.

Data providers should be able to establish the cost of building and maintaining a database and the opportunity costs of investing in the data asset over other assets. Data providers should also consider whether other organisations (such as potential data consumers) are able to recreate the data, or able to recreate it at a lower cost.[8] The Reproduction Cost would also be a useful indicator of the minimum value that the organisation should accept for the data.[9] The Reproduction Cost method, including the considerations of the data value drivers, can be applied using the following steps:



A case study showing how an organisation uses the Cost Approach, as well as an elaboration on the different cost components, can be found in Annex B.

[8] *For example, if the core business of the data consumer does not lie in recreating the data, then its costs in doing so would be higher than the data provider which is a specialist in creating the data.*

[9] *Data providers should also factor in a fair return for efforts to prepare the data for the data consumer.*

Income Approach

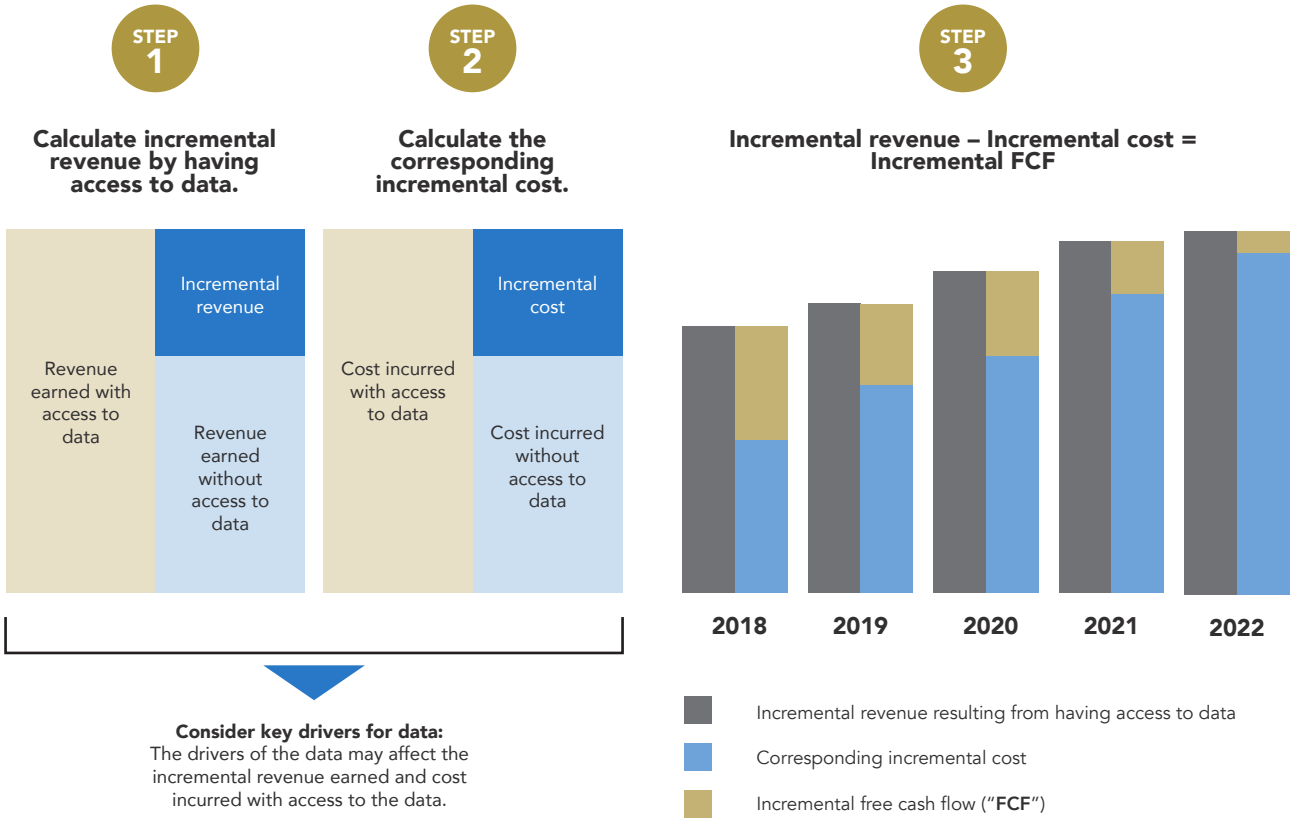
The premise of the Income Approach is that the value of the data asset is equal to the value of future cash flow it can be expected to generate over a specified time period (e.g. the remaining useful lifespan of the data). The value of the data would be derived by looking at the difference between the revenue generated by the organisation with access to or without access to the data ("**incremental cash flow**"). If incremental cash flow generated by the data can be reasonably forecasted and quantified, the Income Approach could be used to estimate the value of the data.

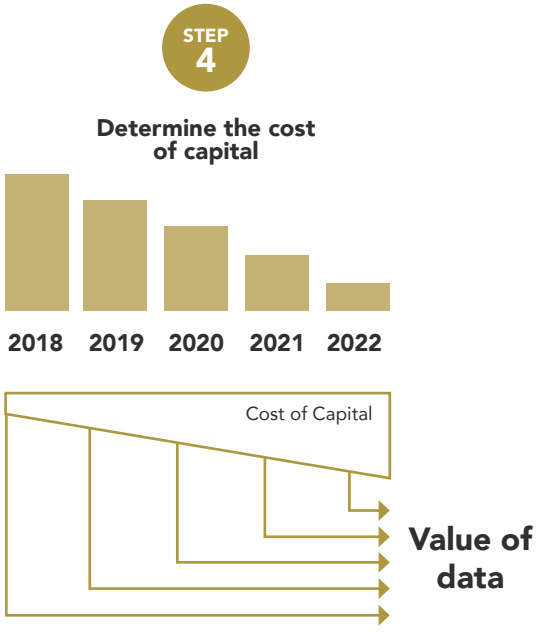
Incremental cash flow could result from either an increase in revenue (incremental revenues) or a decrease in cost (reduced costs), for example by:

- driving the creation of innovative services and products;
- generating insights allowing entry to new markets;
- improving existing operations (e.g. minimise idle machine time, better working capital management and preventing fraud); and
- reducing marketing costs (e.g. targeted customer advertising).

After identifying the incremental cash flow generated as a result of the data, organisations would need to factor in risks associated with achieving those earnings or savings and the time value of money. They would also have to consider if the incremental cash flow benefits might be impacted by legal rights to use the data, or by the useful economic life of the data from a purely commercial perspective.

The Income Approach, including the considerations of the data value drivers, can be applied using the following steps:





Annex B provides illustrations of worked examples of deriving value using the Income Approach.

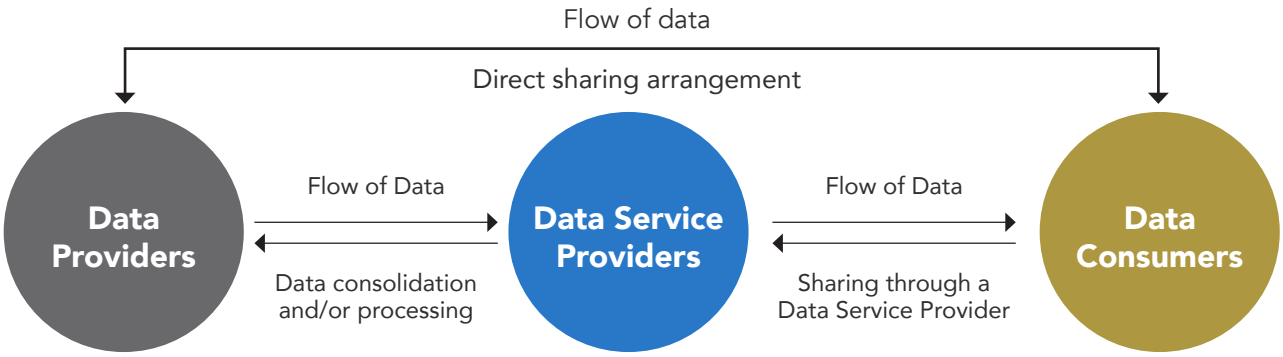
SHARING THE DATA

Organisations that wish to share data and leverage insights from the data should consider the ecosystem for the data sharing activities (e.g. the emergence of third parties such as data service providers to connect data providers and consumers). Organisations should also consider the regulatory obligations associated with the collection, use and disclosure of data, in addition to the data value chain.

Considerations when Sharing Data

Traditionally, data has been exchanged between a data provider and a data consumer. However, data service providers have emerged as facilitators of data sharing as illustrated below.

Overview of Data Service Providers in a data sharing context



	Data Providers	Data Service Providers			Data Customers
Role	are the creators or owners of data. This data may be transferred to consumers directly or indirectly through data vendors.	**Data vendors** are individuals or organisations that aggregate data from providers and supply data to consumers or marketplaces. They can process the data to enhance its value.	**Data marketplaces** are platforms in which data exchanges occur, connecting consumers with data providers or vendors. They can be industry- or region-specific.	**Other data service providers** could include third-party providers of data management services, for example, data encryption and data cleansing.	are individuals or organisations that collect or buy external data. This data may be processed with data analytics to generate insights and supplement internal functions.

Identifying Restrictions on Collecting, Using or Disclosing Data

Given the potential for misuse and unethical use of data, data sharing may be regulated to prevent the unauthorised collection, use or disclosure of data. For example, when it comes to personal data, organisations should ensure compliance with their obligations under the Personal Data Protection Act ("**PDPA**"). Organisations looking to share data may also consider some of the regulatory considerations below:

Theme	Considerations
Data sovereignty	• Where the data resides, and the jurisdiction in which the data falls under. • Data could be subject to the regulations of the country in which it is collected, stored or used. • Whether the data is being transferred internationally or handled by an overseas third party. • Data sovereignty challenges arise when there is a cross-border element in data sharing.
Ownership/ Intellectual Property	• Who the owner(s) of the data is/are. • If there is a credible claim in copyright over the data. Although data itself cannot be copyrighted, compilations of data which display sufficient creativity in the arrangement, annotation or selection can be protected. Mere data aggregating and input is insufficient. If protected by copyright, such data cannot be used without the permission of the copyright owner. Compilations which have been protected in Singapore include telephone directories, street directories, railway tables, examination papers, trade catalogues, a racing information service, football fixtures, betting lists, and listings of television broadcast programmes. • Outright acquisition of the ownership of the data allows for broadly unfettered usage of the data, while licensing may place limitations on the use of the data, depending on the scope and terms of the licence. • Organisations should understand licensing terms before engaging in data sharing to avoid copyright infringement or a breach of contract.

Theme	Considerations
Broad and/or sector-specific regulations	• Where the data is personal data, organisations are required to comply with the data protection provisions of the Personal Data Protection Act. Depending on the data, organisations may be required to comply with other sector-specific regulations. • Consider the impact of compliance costs where there are enhanced user rights in legislation (e.g. protection of minors' rights and data portability rights).

ANNEX A: DATA VALUE DRIVERS

Completeness

Completeness refers to how much of a known universe the dataset covers. In general, the more complete a dataset, the more valuable it is due to the increased accuracy of insights derived.

Enhancing the completeness of data increases the value of the data. To enhance the completeness of data, an organisation needs to define boundaries of the dataset. There are two approaches to enhancing the completeness of data:

a. Ensuring data is captured in a comprehensive manner, through the improvement of data management infrastructure and processes
b. Acquiring and aggregating datasets from external sources (e.g. open source data) to complement the existing dataset. This requires some analytic processing to match and merge data from various sources.

Consistency

Data is consistent if it conforms to the syntax of its definition. For example, structured data such as storekeeping inventory and business transactions are data that conform to a pre-defined syntax and format. These data have a high degree of organisation, making analysis and processing easier.

On the other hand, unstructured data such as images and sounds may require some degree of processing to conform to defined rules and syntaxes so as to enhance their consistency. In such cases, the more processing is done, the more valuable the data is.

Accuracy

Accuracy describes the degree to which data correctly describes an object or event. Reliability of the data significantly impacts the value of the data. Inaccurate data produces unreliable insights, which makes it ineffective for any organisation seeking to utilise the data. In addition, knowing the data's provenance (or source) is a critical aspect of determining its accuracy, as it informs the data consumer of the history of the data and account for errors, if any.

Timeliness
Timeliness refers to the degree to which the data is up-to-date at the required point in time of use. In general, the more up-to-date the data, the more valuable it is. However, timeliness is a relative measure, which is dependent on the intended use case for the data.

Exclusivity
Exclusivity refers to the uniqueness of the data. In general, the fewer existing alternatives for the data, the more valuable the data is. The key driver of value for exclusivity lies in the competitive advantages and revenue opportunities afforded by the data. Exclusivity can be enhanced by:

a. creating unique datasets by integrating and enriching existing data with data from other sources;
b. controlling access to the data through technical and procedural means; and
c. identifying new sources of data or creating new means of capturing data.

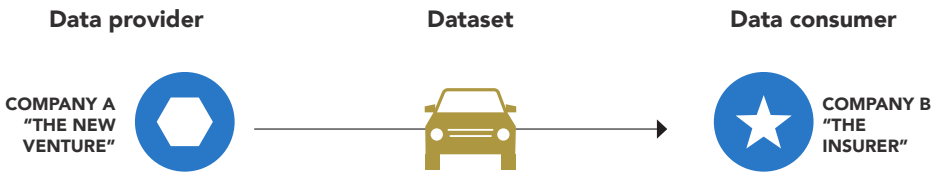
Interoperability/Accessibility
Interoperability/Accessibility is critical to the value of the data. This is because in many cases, the value of data lies in its potential to be combined with an internal dataset. Without the ability to combine and enrich the data, the data is generally of little value to potential consumers.

Restriction, Liability and Risk
Usage restrictions have to be compatible with the sharing use case for the data to have value. In general, the less restricted the use of data, the higher its value.

In addition, potential liability and risks associated with the data could reduce the value of the data, and are very often the main deterrents to data sharing. Sharing arrangements between organisations would require organisations to design contractual obligations to satisfy internal risk management policies.

ANNEX B: CASE STUDIES AND WORKED EXAMPLES

Case Study #1: Single Data Provider and Consumer



Company A intends to expand into a new business.
It installs monitoring devices inside passenger vehicles to record driving habits, and make the data available to insurance companies and other interested parties.

The dataset to be shared contains the driving behaviour of 20,000 drivers over the past year.

Company B is an insurance company that is considering ways to improve its ratio of claims to premiums collected.
One option is to obtain data on driving habits to understand risk profiles.

Company A thinks its data can help Company B and considers how to value its data using the different approaches.

MARKET APPROACH

In the case where there is an actively traded market for the data asset, an organisation can look up the market price of its data. Where there is no active market, but there are examples of similar data assets available, or valuation evidence from transactions involving similar data, the Market Approach may still be applied. The key steps are shown below:

Key Steps	Explanation	Worked example with illustrated values
Step 1 Assess data and consider comparable transactions or available data.	Identify relevant comparable transactions and calculate the valuation metrics for those transactions.	While there is no actively traded market for similar data, three months ago, a US based company sold a dataset of 5,000 US driver's records for S$1,800,000 (a record is defined as being the driving history data of one driver for a one-year period).
Step 2 Identify appropriate multiples.	To apply the pricing from the comparable transaction to the current valuation, a common price multiple has to be identified (e.g. price-per-record).	The price-per-driver record in the comparable transaction was $1,800,000 / 5,000 = $360 per record.
Step 3 Choose the valuation base for the subject data.	The value base of the price multiple identified should be identified for the data being valued (e.g. number of records).	For the purpose of this transaction, Company A estimates that it will have approximately 20,000 unique driver records based on the current number of drivers.
Step 4 Estimate appropriate adjustments in consideration of the data value drivers.	Identify the differences between the comparable data and the subject data and account for the differences by making adjustments, if any, to the valuation metrics (e.g. accuracy, timeliness and consistency).	When comparing the subject data asset to the transacted data asset, Company A notes that its data is more accurate and up-to-date. It therefore judges that a 10% premium to the benchmark price-per-record of S$360 is appropriate and estimates that its data is worth S$396 per record.
Step 5 Calculate market value.	Having made the suitable adjustments, the price multiple can be applied to the identified value base for the subject data.	**S$** **Adjusted price/record** 396 **Number of records** 20,000 **Value of Company A's data** 7,920,000

INCOME APPROACH

The premise of this approach is that the value of the data is equal to the value of the Cash Flow ("**CF**") an owner/licensee of the data is expected to generate over the remaining useful lifespan of the data. The present value of the additional income earned or costs saved through the use of the data represents the value of the data.

Key Steps	Explanation	Worked example with illustrated values
Step 1 Calculate the incremental revenue or cost saving by having access to the data, including the data value drivers.	Data can generate income through, for example: - driving better customer targeted advertising campaigns spending habits; - generating insights allowing entry to new markets; and - allowing more accurate or timely pricing. Examples of how data can help companies save costs include: - improving existing operations (e.g. minimise idle machine time; preventing fraud); and - reducing marketing costs via more customer targeted advertising In addition, the relevant income and cost savings may also be affected by the data value drivers.	By better identifying high-risk driver categories and adjusting its customer premiums accordingly, Company B is able to reduce the value of claims It benefits from these cost savings over the period of an insurance policy, which is three years on average for Company B. By adjusting premiums according to driver habits and behaviours, Company B loses some revenue as some high-risk drivers choose to switch insurers. There is a net overall loss of revenues in this case. Below is the expected annual cost of claims for the insurance company over the next three years with and without the use of Company A's data.

S$	Forecast Year 1	Forecast Year 1	Forecast Year 1
Cost of claims pre- data use	20,000,000	22,000,000	23,000,000
Cost of claims post data use	(17,000,000)	(16,000,000)	(15,000,000)
Net loss of revenue	(1,000,000)	(1,100,000)	(1,200,000)
Cost savings	**2,000,000**	**4,900,000**	**6,800,000**

Key Steps	Explanation	Worked example with illustrated values
Step 2 Calculate the relevant incremental cost.	The next step is to estimate the total incremental costs incurred by the use of the data asset. In addition, the relevant incremental cost may also be affected by the data value drivers.	Company B is also expected to incur additional implementation and analysis costs to make use of the data (e.g. costs to analyse and incorporate the data into its premium setting process). Company B estimates the additional costs as follows:

S$	Forecast Year 1	Forecast Year 1	Forecast Year 1
Additional data analysis cost	150,000	150,000	150,000
Additional marketing cost	50,000	50,000	50,000
Total additional costs	**200,000**	**200,000**	**200,000**

Key Steps	Explanation	Worked example with illustrated values
Step 3 Calculate the incremental CF.	With both the estimated incremental revenue as well as the incremental costs, the incremental impact on cash flow can be calculated	Below is Company B's expected incremental CF from having access to and using the data. Singapore's statutory tax rate of 17% is applied to calculate the after-tax incremental CF.

S$	Forecast Year 1	Forecast Year 1	Forecast Year 1
Cost Savings n et of revenue reduction	2,000,000	4,900,000	6,800,000
Less: Additional Implementation Cost	(200,000)	(200,000)	(200,000)
Incremental CF - Tax	1,800,000 (306,000)	4,700,000 (799,000)	6,600,000 (1,122,000)
After-tax Incremental CF	**1,494,000**	**3,901,000**	**5,478,000**

Key Steps	Explanation	Worked example with illustrated values
Step 4 Determine the cost of capital.	The cost of the capital needs to be determined as it will be used to derive the net present value of the incremental cash flows.	Company A estimates the cost of capital for an insurance company like Company B to be approximately 10%.
Step 5 Calculate the present value of the incremental CF.	Discounting the future cash flows will provide an estimate of the value of the subject data.	

S$	**Forecast Year 1**	**Forecast Year 1**	**Forecast Year 1**
After-tax Incremental CF	1,494,000	3,901,000	5,478,000
Discounted factor	0.91	0.83	0.75
Present value of Incremental CF	1,358,182	3,223,967	4,115,702
Total present value	**8,697,851**		

COST APPROACH

The key rationale for utilising the Cost Approach is that the minimum amount a data asset provider will be prepared to accept for its data is the cost it incurred in creating and marketing the data, plus a return on that cost.

On the other hand, a potential data consumer applying the Cost Approach has to consider whether to create or acquire the data.

There are two Cost Approach methods:

- The **Replacement Cost** method represents the current cost of a similar data with equivalent utility to the data being valued. Value is equivalent to the cost of replicating the utility of the data, and not creating an exact copy of the data.
- The **Reproduction Cost** method represents the current cost of reproducing the data using similar inputs and methods.

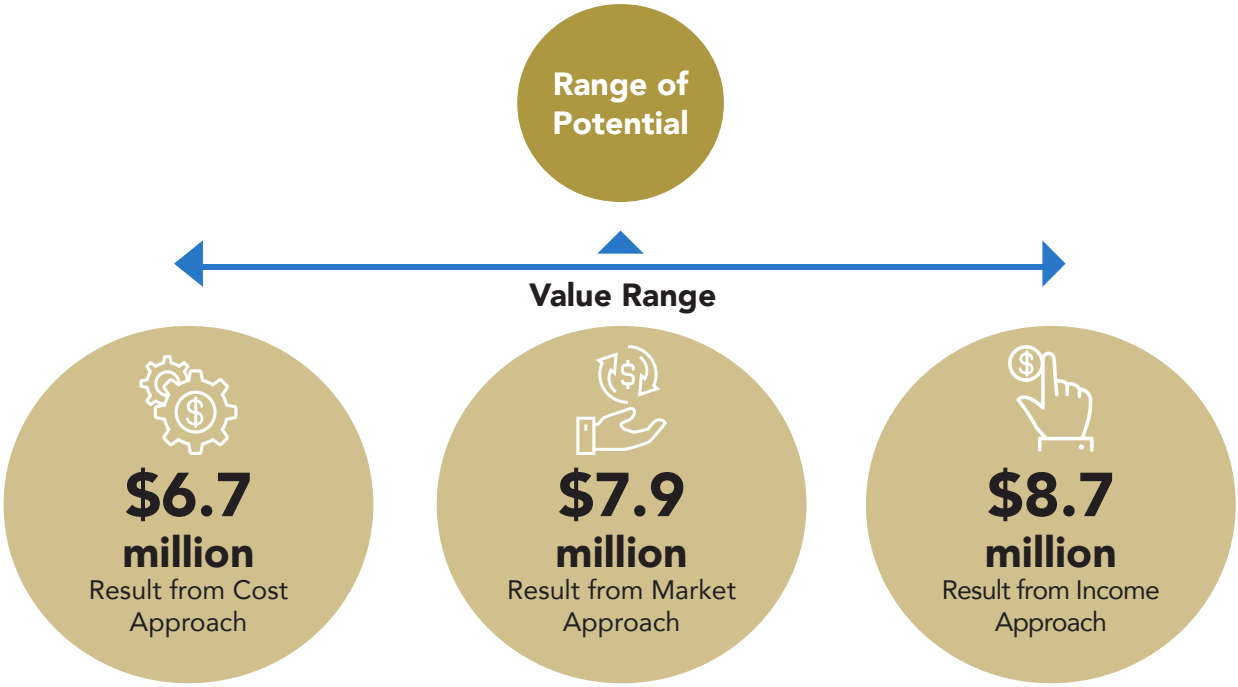
Given that the Replacement Cost method is similar to the Market Approach, the following table will detail the Reproduction Cost method from the data provider's perspective.

Key Steps	Explanation	Worked example with illustrated values
Step 1 Identify all relevant costs, including the consideration of the data value drivers.	There are typically five key cost components to building and maintaining any database: • Input cost • Software and storage cost • Human capital cost • Selling and marketing • Other overhead costs In addition, the relevant costs may also be affected by the data value drivers. For example, there may be additional costs to ensure that the data is accurate and up-to-date.	In this case, Company A identifies the following costs: • Input cost – cost to purchase, install and maintain the monitoring equipment that collects the data • Software and storage cost – cost of tools required to transmit, process, manage and store the data • Human capital cost – cost to hire people to build, develop and manage the data system and architecture as well as to collect and process the data asset • Selling and marketing costs – cost required to market and promote the data to potential consumers • Other overhead costs – e.g. rent, legal and administrative costs and finance expenses

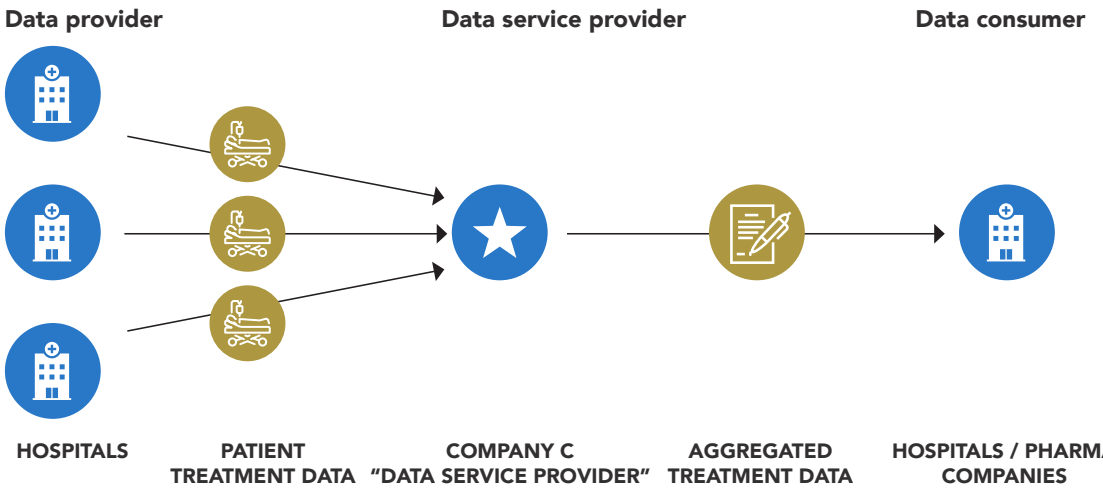
Key Steps	Explanation	Worked example with illustrated values
Step 2 Estimate the total cost.	Once all the cost components relevant to the creation and maintenance of the data are identified, the total costs can be calculated.	**S$** **Input cost** 4,400,000 **Software and hardware cost** 1,000,000 **Human capital cost** 226,000 **Selling & marketing cost** 100,000 **Other overheads cost** 100,000 **Total cost** **5,826,000**
Step 3 Determine the profit margin.	A key element of the Cost Approach is the profit or return element, A data asset owner could have invested in other assets other than the subject data asset. These other investments would have provided a certain return for the data asset owner. Accordingly, a data owner should be reasonably expecting a return on its investment in the data asset. A common proxy for the required return might be the company's return on equity. This return could be adjusted depending on the risk of the investment.	Company A could deploy the resources it uses in other areas of its business, which typically earn a return on equity of around 15%. This is the opportunity cost to Company A, which also needs to be taken into account. The cost for the return component is estimated at 15% * S$5,826,000 = S$873,900
Step 4 Calculate the value of the subject data.	Applying the determined profit margin to the cost will yield the estimated value of the subject data.	**S$** **Total cost** 5,826,000 **Add: Profit margin** 873,900 **Total cost incl. profit margin** **6,699,900**

VALUE RANGE FROM THREE APPROACHES

From the worked examples, using the Market Approach, Cost Approach and Income Approach provided an estimated range of potential values between S$6.7 million and S$8.7 million. This range of values would be useful for the data consumer and data provider when negotiating the final value of the data, which will also be influenced by other organisational and market-related factors as such the organisations' bargaining power, financial ability and ability of the data provider to market the data.



Case Study #2: Multiple Data Providers and Consumers



Hospitals generate patient treatment data as a by-product of day-to-day operations.

They have been approached by Company C which wanted to purchase anonymised patient treatment history generated at their hospitals.

Company C is a Data Service Provider that intends to aggregate the hospitals' patient treatment data.

Company C aggregates, cleans and processes data from hospitals.

The data is then marketed to pharmaceutical companies and hospitals.

Hospitals and pharmaceutical companies looking to optimise patient treatment and gain business insights consider purchasing the aggregated data from Company C.

The hospitals estimate that they will be able to deploy resources more efficiently with the aggregated data.

The hospitals would like to evaluate the value of their data to Company C.
Company C thinks its data can help the hospitals and considers how to value its data.

MARKET APPROACH

Some other considerations when applying the Market Approach for aggregated data lie in the estimation of appropriate adjustments of the aggregated data compared to the value of the standalone datasets, as certain datasets may gain value when combined with other data.

Case Study #2: Multiple data providers and consumers
Having identified previous transactions of similarly anonymised patient treatment data, Company C discovers that there have been a series of transactions that have taken place over the last year. Based on these past transactions, Company C determines that the average price-per-patient record based on previous transactions is S$50 per record.

When comparing the subject data asset to the transacted data asset, Company C notes that:

- its data is much more complete, as it captures data from all hospitals within the country. This makes recognising trends and efficiency easier; and
- its data is more consistent than the data in the comparable transaction as it has already cleaned the data, ensuring syntax consistency across all the aggregated datasets and removing any errors.

For the reasons above, Company C determines that because the data is much more complete and consistent, the value to hospitals looking to use the data to improve treatment efficiencies is significantly higher.

In this case, Company C estimates that the aggregated dataset should be adjusted by 10 times (i.e. the aggregated dataset is worth S$500 per record).

INCOME APPROACH

The key consideration that a data service provider who aggregates data would need to take into account is the estimation of the incremental value that aggregated data could bring compared to value of the input data (i.e. the data that the hospitals held in the first place).

Case Study #2: Multiple data providers and consumers
To apply the Income Approach to estimate the value of the hospital's data to Company C, the hospitals will have to estimate the incremental impact of its data for Company C.

- In this case, the hospitals estimate that by aggregating their data, Company C would be able to generate a certain amount of incremental revenue, based on Company C's product history. They also estimate that Company C would incur incremental costs, mostly from the costs of marketing and processing the data. Given that their data asset forms only a part of the aggregated dataset, Company C estimates the incremental cash flow attributable to its data asset by applying a fraction corresponding to the estimated share of its data as a proportion of total records in the dataset.
- The subsequent steps for the Income Approach can be applied as per the previous example in Case Study #1.

Company C intends to apply the Income Approach to estimate the value of the aggregated data to the hospitals, given that the hospitals would be the main target consumers for the data.

- In estimating the potential revenues and costs savings, Company C has to estimate the potential incremental benefits that using a more complete and consistent dataset would yield for the hospital.
- In this case, Company C estimates that hospitals would be able to better market aftercare services to patients using the data, yielding a revenue upside for the hospitals. In addition, Company C also estimates that the hospitals would be able to optimise the medication and treatment prescribed by the hospital to be more cost efficient. The revenue upside and the cost savings would have to be summed up to arrive at the total incremental gross benefits for the hospitals.
- The subsequent steps for the Income Approach can be applied as per the previous example in Case Study #1.

COST APPROACH

A key consideration for organisations using the Reproduction Cost approach is how they can identify which costs are relevant for the purposes of creating and managing the dataset. This is a because as many businesses generate data as an incidental by-product of their day-to-day activities, taking the cost of running daily operations as part of the costs of creating and managing the dataset would result in an overinflated valuation of the dataset.

The hospitals generate patient data as a by-product of their main activity of providing healthcare services. Therefore, in valuing the patient treatment data, the costs of creating the data (i.e. providing treatment to the patients) should not be taken into account.

Case Study #2: Multiple data providers and consumers
In valuing the hospital's data assets, the key cost items would be as follows:

Key cost item	Worked example with illustrated values
Input costs	Input costs would be marginal as the data is generated incidentally. Some marginal inputs costs for the hospitals would include the costs incurred in ensuring that the data is properly anonymised and in compliance with all the relevant regulations, as well as any additional costs incurred in tracking and recording the data.
Software costs	This should only take into account the costs of the software directly related to the storing and management of the data. This would not include, for example, the cost of the other software that would be required for the hospital's data day-to-day operations but are not relevant for the purpose of this specific dataset.
Hardware costs	Hardware costs should only include the costs of hardware required for the storage and management of data and would not include any of the medical equipment.
Overheads	Overhead data would be marginal given the limited resources dedicated to the creation of the dataset. Marketing costs would also be limited given that the hospital spends relatively little effort in marketing the data due to the sensitive nature.
Returns	For returns, the hospital would apply a rate of return that is typical to its main business activities.

Alternatively, in valuing the aggregated patient data asset from the data service provider's perspective, the key cost items would be as follows:

Key cost item	Worked example with illustrated values
Input costs	Input costs for the data service provider would include the sum of all the costs incurred in acquiring the raw data. For the data service provider, this would also include human capital costs, as the raw data acquired would require significant time and resources to clean, process and combine into a complete dataset.
Software costs	Software costs for the data service provider would include all software costs incurred in processing and transforming the raw datasets as well as the software required to manage and store the combined data asset.
Hardware costs	Hardware costs would include the costs incurred in acquiring and maintaining the hardware required to store and distribute the data.
Overheads	Assuming the data service provider business consists primarily of the sale of the subject data, a significant share of the overheads would need to be accounted for in the cost of building and maintaining the data. In addition, any sales and marketing costs would also have to be accounted for given the nature of the aggregator's business model.
Returns	For returns, the data service provider would apply a rate of return that is typical to its other business activities.

The subsequent steps for the Cost Approach can be applied as per the previous example in Case Study #1.

#SGDIGITAL

Singapore Digital (SG:D) gives Singapore's digitalisation efforts a face, identifying our digital programmes and initiatives with one set of visuals, and speaking to our local and international audiences in the same language.

The SG:D logo is made up of rounded fonts that evolve from the expressive dot that is red. SG stands for Singapore and :D refers to our digital economy. The :D smiley face icon also signifies the optimism of Singaporeans moving into a digital economy. As we progress into the digital economy, it's all about the people - empathy and assurance will be at the heart of all that we do.

BROUGHT TO YOU BY

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

pdpc PERSONAL DATA PROTECTION COMMISSION SINGAPORE

IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE