Data Utilization and Data Space Guide 2nd Edition

IPA Information-technology Promotion Agency, Japan

Data Spaces Group

Data Engineering Department
Digital Infrastructure Center (DISC)

Contents

1. Introduction

In today's digital society and economy, data has become the key to improving the quality of products and services as well as to creating new business models. To develop the digital economy, it is not enough to simply use data. There is a growing need to protect data sovereignty, promote data sharing across industries, and build a reliable ecosystem.

In the current situation where large IT companies in the US collect and control data, Europe is working to secure its economic and political independence by managing its own data and taking the lead in the distribution and use of data within Europe. To this end, Europe is building a framework that will allow it to share data with others in a reliable way while maintaining data sovereignty.

In Japan, the importance of digital transformation (DX) is increasingly being recognized, but progress in data utilization lags behind other developed countries, and data sharing across companies and industries is not advancing fast enough. Data-driven innovation remains sluggish. The IMD World Competitiveness Ranking puts Japan as having low capabilities in data utilization, and there are other major issues such as the following.

- Interoperability issues: Data is shared within organizations and among a limited number of companies, but data sharing with other countries and industries is lacking.
- Data sovereignty issues: Unlike in Europe, there is not yet a well-established, standard framework for protecting data sovereignty.
- Human resource issues in data utilization: Companies and organizations lack the human resources to use data strategically.

A data space[1] is a new approach to addressing these issues and is expected to drive innovation and competitiveness.

2. Purpose and Intended Audience

The purpose, structure, and intended audience of this document are as follows.

Purpose of this document

This document is intended to help organizations maximize their competitiveness and data value. It does this by clarifying not only the general process of data utilization but also the procedures and details of data space utilization.

---

[1] A data space is a novel concept that has attracted attention in recent years, especially in Europe, as an economic and social space that transcends national borders and sectoral differences. By establishing rules and frameworks that allow data to be collaborated across national and organizational boundaries, and by making available a larger volume of "diverse" and "reliable" data than ever before, it can be used to create new services and improve existing ones.

Structure of this document

Based on the intended audience, this document is divided into two parts: "Part 1: Data Users" and "Part 2: Data Providers." These two parts of the document define the general process for utilizing the data space and explain each phase. Those interested in learning about the processes and tasks for data providers should refer to Part 2, Chapter 9 and beyond. * Note that sections that overlap with Part 1 are indicated with references to Part 1. Please refer to Part 1 as needed.

Intended audience

This document is intended for those with a basic knowledge of data spaces. For basic knowledge, please refer to "Introduction to Data Spaces"[2] published by the IPA.

Those who seek to acquire data and use it for service delivery and other business purposes are defined as "data users," while those who seek to provide data are defined as "data providers," and these are the intended audiences for this document. This document also provides separate explanations for those in the following three areas of the organization.

① Executive management (CEO, CDO (Chief Data Officer), etc.)
The executive management makes management and business decisions. The CDO takes those decisions and formulates data strategies.

② Business departments
Business departments formulate and execute specific business plans based on data strategies.

③ IT departments
IT departments develop the functions and services required for data collaboration in data spaces.

Each section of this document is indicated as follows according to the degree of involvement of the three above.

- Primary responsible department: Shown in blue with a star (★)
- Not a primary responsible department, but related: Shown in blue
- Department with little relationship: Grayed out

[2]Information-technology Promotion Agency, Introduction to Data Spaces, https://www.ipa.go.jp/en/digital/data/data-spaces.html

[Legend]

Executive management

Business departments

IT departments

3. Overview of Data Utilization and Data Space Utilization

There are many phases in the data utilization process when using data spaces. The following process diagram shows the outline and the tasks required at each phase for data users and data providers.

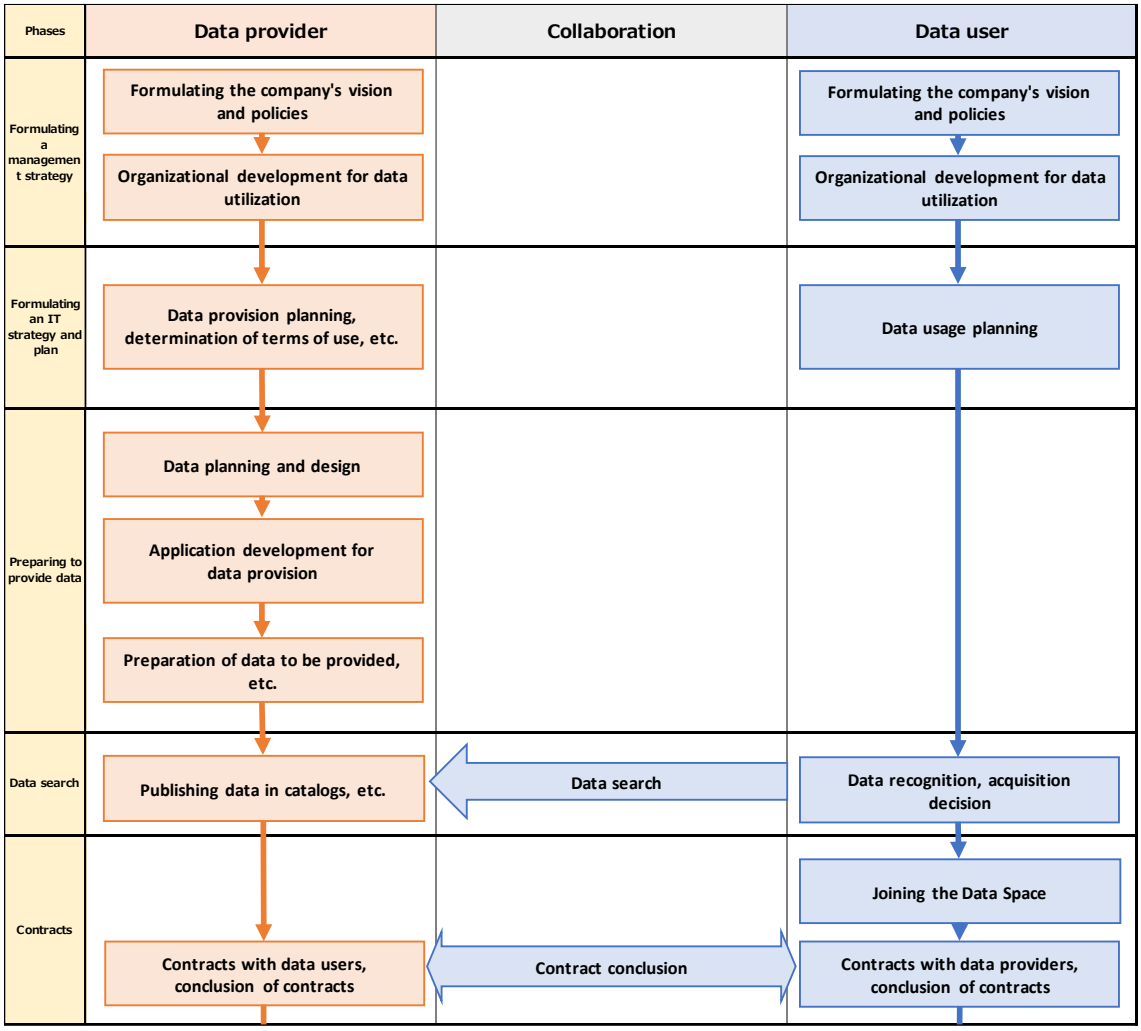

Figure 0.1 Overall process

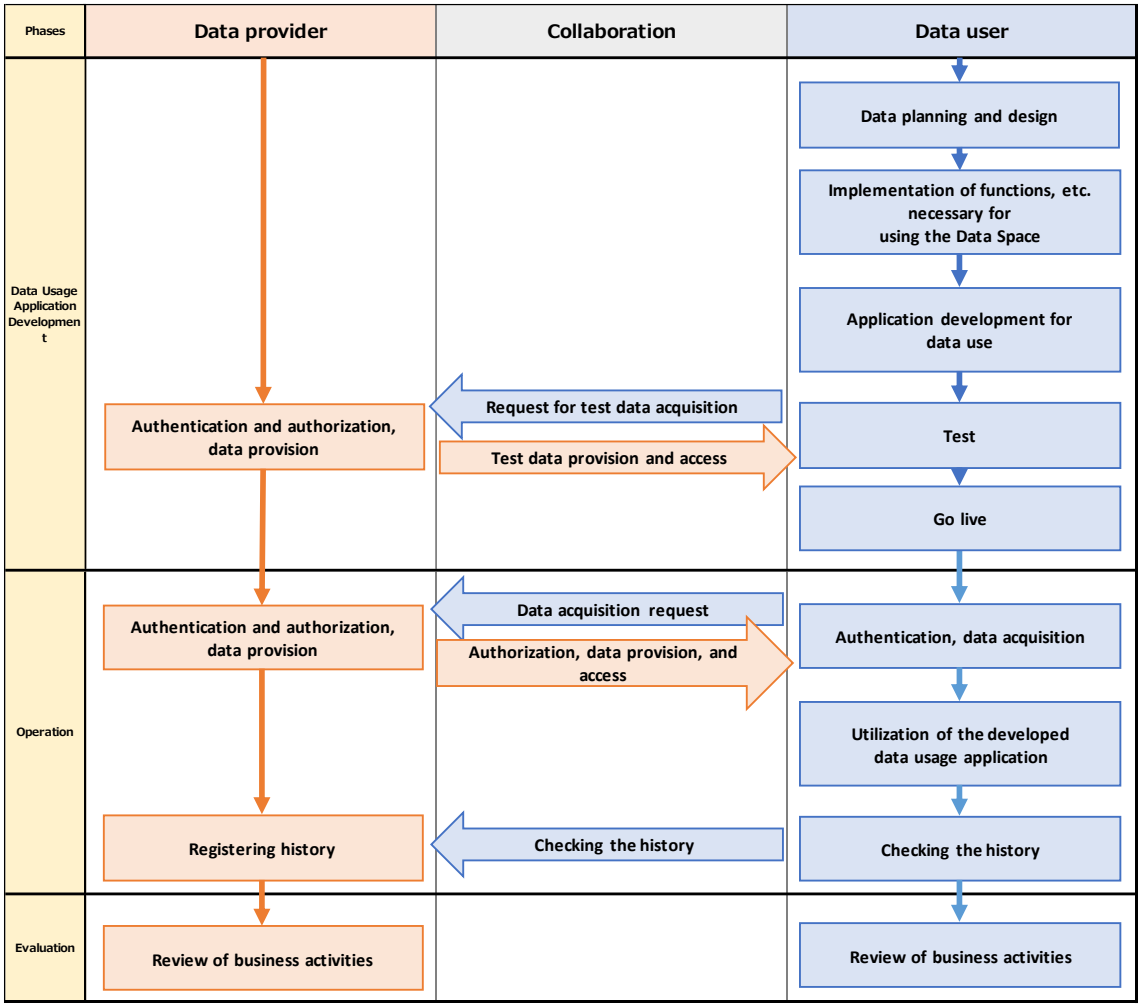

Figure 0.2 Overall process

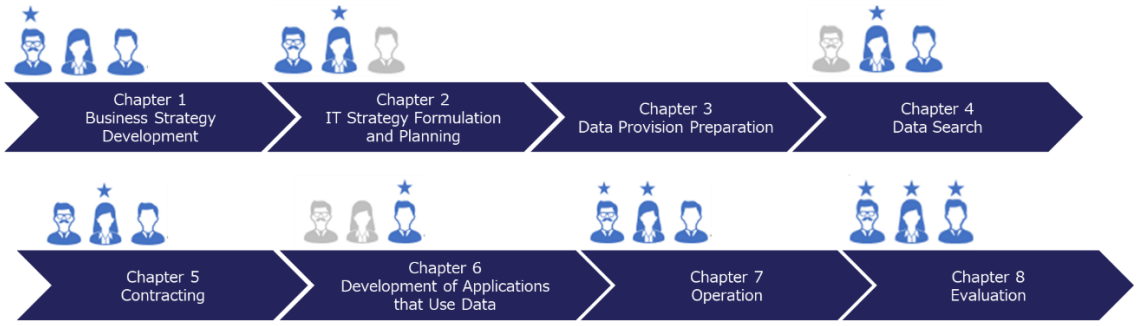

Figure 0.3 Overall process - data users (simplified version)

In this document, each phase on the left is set as a chapter, and the tasks within each phase are explained. In some cases, a data marketplace acts as an intermediary between users and providers in the processes shown in Figures 0.1 and 0.2. For more information about data marketplaces, see the Key Points section in "Chapter 5 Contracting Phase."

[Part 1] Data Users

Chapter 1: Business Strategy Development Phase

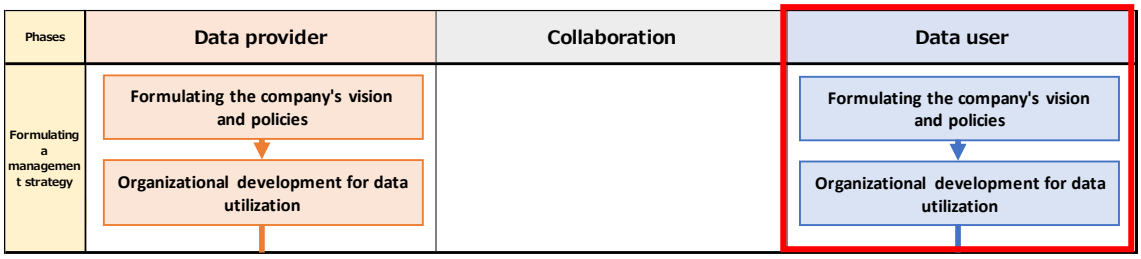

Figure 1.1 Process diagram for the Business Strategy Development Phase

This chapter, "Business Strategy Development Phase," explains the tasks for the data users shown in the red box in Figure 1.1 of the overall process diagram.

In this phase, the company's vision and policies are formulated and how data will be used to drive the business is explored. First, the management needs to have a basic understanding of data (the importance of data utilization, the role and value of the CDO, the overall data utilization process, etc.). Then, the CDO is appointed, and various data-related initiatives are implemented.

1.1.Formulating the company's vision and policies

To promote businesses focused on the utilization of data, it is first necessary to formulate a company-wide vision and policies and then to communicate this mindset throughout the company. This form of fostering data-driven management and organizations is closely related to achieving DX. The following are excerpts from Chapter 1: "Approach to Achieving DX" of the Practical Guidelines for Companies Promoting DX (DX Practical Guide IT System Construction Edition) (hereinafter referred to as the "DX Practical Guidelines")[3] published by the IPA on the basic elements necessary for formulating a vision and policies.

Communicate the company's vision

The first step in implementing DX as a company-wide initiative is to rigorously examine the future of the organization and its vision of how the use of digital technology will transform the business, and to communicate this vision with a clear understanding by all members of the management team. Communicate this vision repeatedly to the entire

[3]Information-technology Promotion Agency, Practical Guidelines for Companies Promoting DX (DX Practical Guide IT System Construction Edition), https://www.ipa.go.jp/digital/dx/hjuojm000000gx4n-att/000094497.pdf (Japanese)

organization so that the organization as a whole can demonstrate its strength. If there is resistance to changing the current way of doing business, it may be necessary to keep trying to convince them.

Formulate a roadmap

Formulate a roadmap of what needs to be accomplished by when, and what needs to be done now to realize the company's vision. The roadmap should not be a fixed set of goals but should be reviewed, refined, and revised as needed to meet the changing needs and demands of the business environment in order to move toward achieving the vision.

Rather than simply examining within the organization how data can be utilized (i.e., starting with the means), the priority should be to thoroughly consider and discuss what the organization should look like and what kind of world it wants to realize.

1.2. Organizational development for data utilization

The utilization of data should not be limited to individual departments. The entire organization should use data effectively to create business value. This is called data maturity. An organization with a high level of data maturity can gain a competitive advantage by using data-driven results for business strategies in data collection, analysis, and decision-making. For more information on data maturity, see the *Data Maturity Handbook*[4].

1.2.1: Appointing a Chief Data Officer (CDO)

In recent years, organizations have come to view data as a strategic asset, so the need to effectively manage and utilize this data has become increasingly important. Managing and utilizing data requires the involvement and collaboration of many departments within an organization. Therefore, it is recommended that a Chief Data Officer (CDO) be appointed as a key professional position to lead the entire organization to success. The goal is to transform data from a mere resource to a strategic asset that can improve competitiveness.

For more information on the CDO, see the *CDO Handbook: Leadership for Data Driven Management*[5].

[4]Information-technology Promotion Agency, Data Maturity Handbook, https://www.ipa.go.jp/digital/data/data-spaces-academy.html (Japanese)
[5]Information-technology Promotion Agency, CDO Handbook: Leadership for Data Driven Management, https://www.ipa.go.jp/digital/data/data-spaces-academy.html (Japanese)

1.2.2: The role of the CDO and organizational development

Often a member of the management team, the CDO is a key figure in the digital age, responsible, along with the CEO, COO, and CFO, for managing and maximizing the value of an organization's critical data assets. The CDO may be an individual or a team (CDO office), depending on the nature of the organization.

The role of the CDO is diverse, but the most important part is to formulate a data strategy that governs the management and use of data in the organization, based on the vision to be communicated by the management and in agreement with the management. The CDO's role also includes the following, starting with No. 2. The most important of these is working with the business departments (DataOps) to facilitate the use of data within the organization.

Table 1.1 Examples of roles the CDO should play

No.	Role	Description
1	Formulate a data strategy	Oversee the management and use of data in the organization, based on the vision to be communicated by the management and in agreement with the management.
2	Establish data governance	Establish consistent rules across the organization for how data should be used, rather than allowing data to be used in disparate ways.
3	Improve data quality	Ensure data accuracy, completeness, and consistency. (See Key Points: [What is data quality?] in Chapter 4: Data Search Phase)
4	Promote data analysis	Enable management and business departments to perform data analysis on their own.
5	Manage data security	Decide who has access to what data. Consult with the CISO[6] and implement data leakage prevention, data breach prevention, and BCP[7] measures.
6	Foster a data culture throughout the organization	Allow the entire organization to use data, not just specific departments.
7	Develop data infrastructures	Develop the infrastructure to store, collaborate, and analyze data.
8	Restrict the purposes of data use	Establish usage rules within the organization to ensure that data with specific usage conditions, such as external data, is used appropriately.
9	Manage privacy	Ensure that data is appropriately anonymized, that personal information is protected, and that laws and regulations, such as the General Data Protection

[6]A Chief Information Security Officer A person responsible for overseeing and managing an organization's information security.

[7]A business continuity plan (BCP) A plan that outlines the actions to be taken in the event of a natural disaster, major fire, terrorist attack, or other emergency situation to continue core business operations or facilitate rapid recovery while minimizing damage to the organization's assets. (Source: The Small and Medium Enterprise Agency, Guidelines on Formulating and Implementing BCPs for Small and Medium Enterprises, https://www.chusho.meti.go.jp/keiei/antei/download/110728JapanBCP_SME_Eng.pdf)

		Regulation (GDPR), are followed.
10	Contracts	When entering into a contract with a data provider for the use of external data, be sure to comply with the scope and limitations of data use in the contract. In some cases, it may be necessary to work with a legal department to minimize legal risks.

By fulfilling these roles, the CDO can facilitate the collection and analysis of data and the use of data-driven results in the development of business strategies. This requires data management. DMBOK2[8] is a widely used source of comprehensive information on data management. The following is the overall structure of the data management functions in DMBOK2.

[8]Data Management Body of Knowledge. Created by the Data Management Association (DMA), a nonprofit organization that promotes data management, DMBOK is a systematic collection of data management knowledge.

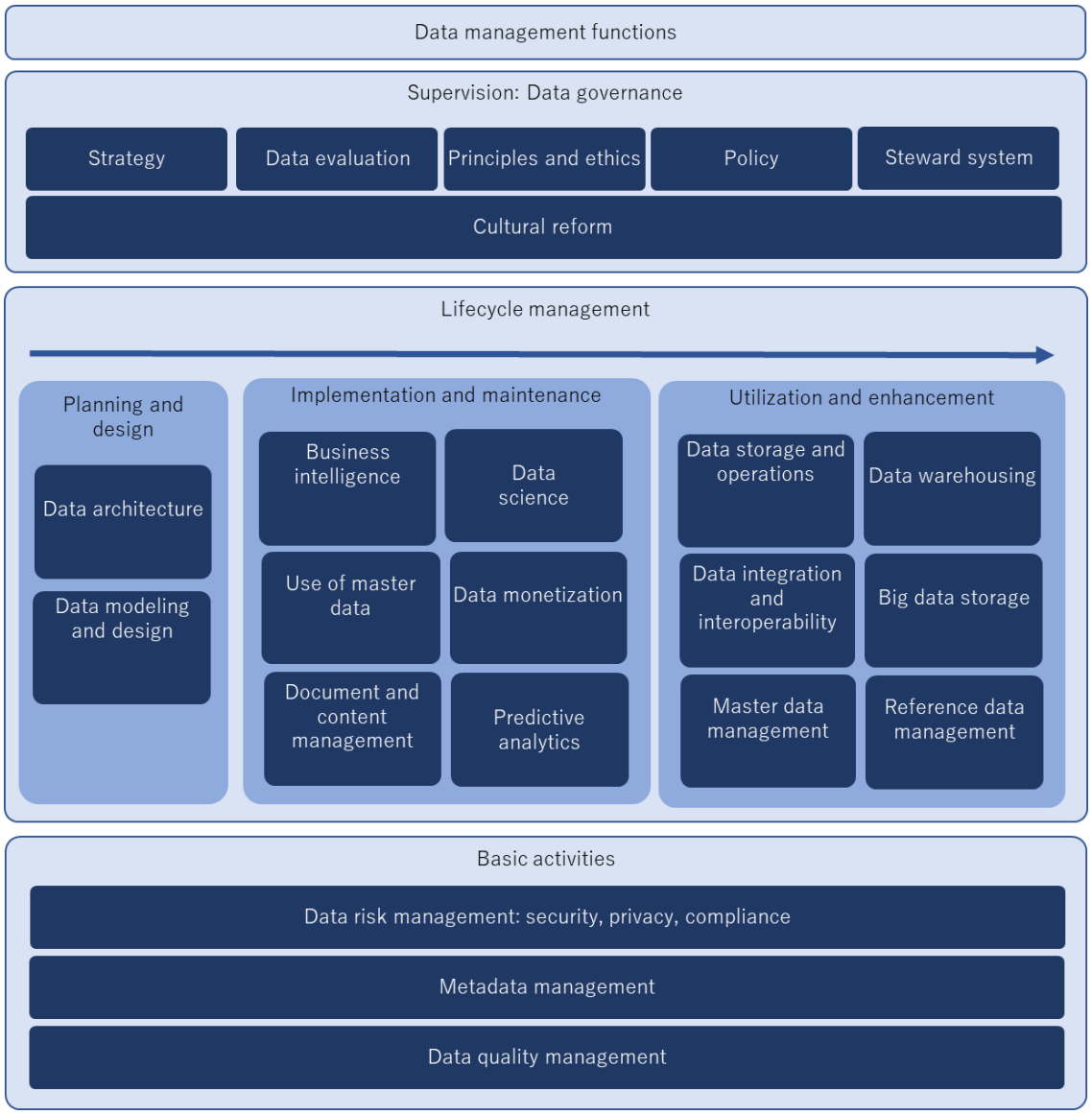

Figure 1.2 Excerpt from "4.4.2. Elements necessary to promote data utilization" in the DX Practical Guidelines

For more information, see "4.4.2. Elements necessary to promote data utilization" in the DX Practical Guidelines.

Chapter 2: IT Strategy Formulation and Planning Phase

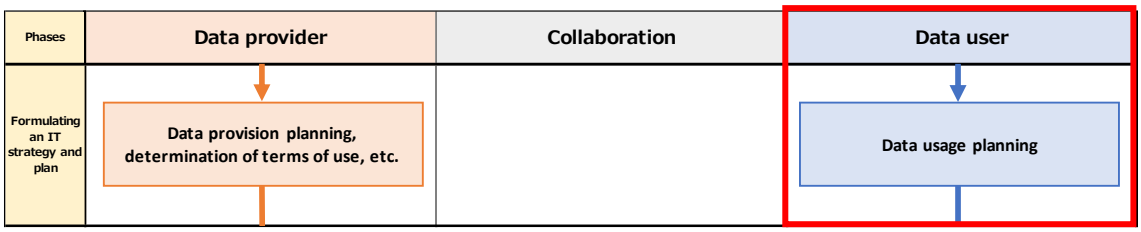

Figure 2.1: Process diagram for the IT Strategy Formulation and Planning Phase

This chapter, "IT Strategy Formulation and Planning Phase," explains the tasks for the data users shown in the red box in Figure 2.1 of the overall process diagram.

In this phase, data users examine specific plans for business projects that involve the use of data.

2.1 Data usage planning

Based on the vision communicated by management and the data strategy formulated by the CDO, the business department formulates a concrete business plan. The CDO provides the necessary data as the business department develops the business plan and, in some cases, may consider acquiring new data.

If data is not available within the company, the use of external data is also considered. Data acquisition is explained in Chapter 4, Data Search Phase.

2.1.1 Approaches to using data for business

Using data for business purposes involves collecting the data needed to create new business or improve business operations. Collecting data requires the cooperation of many stakeholders, including the management (CDO), business departments, and IT departments.

2.1.2 Review of data to be collected

The following steps are considered effective in determining the data to be collected when developing a data usage plan.

① **Clarify goals**

First, clarify the business goals. For example, if the goal is to launch a new product, you need to understand the target market and the competitive landscape.

② **Identify information needed**

List the information needed to accomplish the goal. For example, you may need the following information.

- Market data: Market size, growth rate, and trends
- Customer data: Age, gender, and purchasing behavior of target customers
- Competitor data: Market share, strengths and weaknesses, and pricing strategies of key competitors

③ **Search for data**

Data can be searched for in a number of ways, including data catalog sites. Data providers make data available to the public, and in addition to catalogs, data is sometimes made available on company websites and in the data marketplace.

④ **Examples of data types**

▼ Use of open data

For the benefit of society as a whole, government agencies and public organizations make open data available for widespread use. In principle, anyone can use open data, and there are presumably few restrictions on its use. Here are some of the benefits of using open data.

▽ Benefits

- Increases transparency and trust: Using publicly available data from public institutions and governments increases transparency and trust among citizens and businesses.
- Facilitates innovation: By using population data, traffic data, and GDP data that can serve as a benchmark for any organization, there is the potential to develop new solutions, businesses, etc. by combining data from inside and outside the organization.

▽ **Open data-related information**

Examples of types of open data:

- Traffic data (operation status, traffic volume, accident information, etc.)
- Environmental data (weather information, energy consumption, etc.)
- Economic data (GDP, price index, etc.)
- Societal data (population statistics, health statistics, education statistics, crime statistics, etc.)

Examples of where to get open data:

Examples of where to get open data are provided as links on the IPA website.

- Links to data-related initiatives in Japan and abroad[9]

▼ Use of unstructured data

In addition to structured data, it is also possible to use unstructured data (e.g., text, images, videos, and social media posts).

However, there are many types and quantities of unstructured data, and it is generally considered difficult to use as is. Unstructured data has no structural definition and is difficult to process in its raw form. In some cases, it may need to be converted or processed to make it useful for analysis.

The actual data acquisition is done in "Chapter 4: Data Search Phase," but the feasibility of the data usage plan cannot be guaranteed without examining the actual situation, such as what kind of data can be acquired. Therefore, in most cases, the examination of what data can be acquired is done in this phase. See the information in Chapter 4 for details on how to use catalog sites that are useful for this purpose.

[9]Information-technology Promotion Agency, Links to data-related initiatives in Japan and abroad, https://www.ipa.go.jp/digital/data/link.html (Japanese)

Chapter 3: Data Provision Preparation Phase

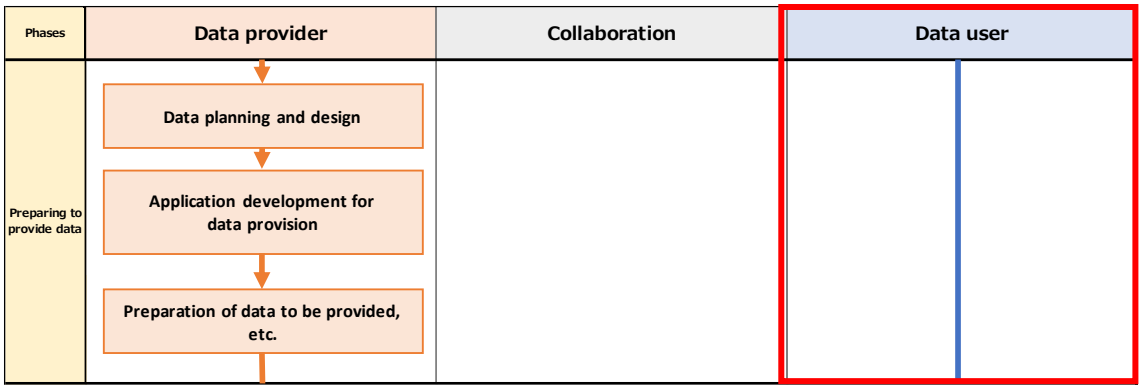

Figure 3.1 Process diagram for the Data Provision Preparation Phase

This chapter, "Data Provision Preparation Phase," is the phase in which the data provider prepares the data to be provided.

We will skip to the next phase as there are no specific tasks required of data users in this phase. The tasks required of data providers are explained in "Part 2: Data Providers."

Chapter 4: Data Search Phase

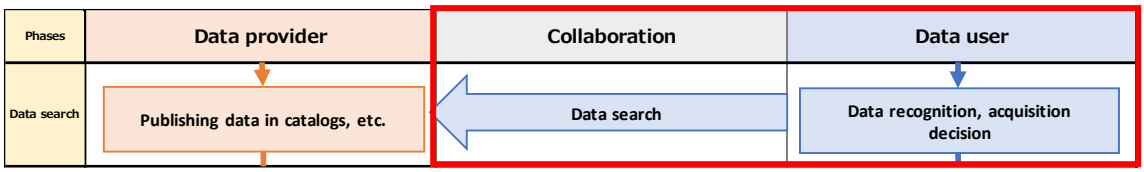

Figure 4.1: Process diagram for the Data Search Phase

This chapter, "Data Search Phase," explains the tasks for the data users shown in the red box in Figure 4.1 of the overall process diagram.

In this phase, data users learn about the data that data providers have disclosed in catalogs, etc., and make acquisition decisions about whether to actually use the data.

4.1 Data search and retrieval

Data users acquire the necessary data in accordance with the data usage plan formulated in Chapter 2. Data not available within their own organizations must be obtained through web searches or other means.

4.1.1 Use of data catalog sites

One of the many ways to search for data is to use data catalog sites.

▼Benefits

① More convenience: Enables you to easily search, browse, and compare data using keywords from a vast amount of data.

② Assured quality of the data being acquired: These sites typically provide detailed information (metadata) about the dataset, allowing confirmation of information about the data necessary to ensure quality, such as data source, creator, and update history. Ensuring data quality is essential for using and analyzing accurate data, and is a key point to keep in mind for making appropriate decisions. See the Key Points section in this chapter for an explanation of data quality.

4.2 Data recognition and acquisition decision

4.2.1 Data acquisition decision

After searching for data, determine whether to actually acquire and use the data. This decision should be based not only on the content of the actual data (data values), but also on information about the data, including metadata. Also consider what processing (merging data, masking unnecessary items, etc.) will be required if the data is actually acquired, and use this as a basis for defining the requirements for the application to be developed in Chapter 6: Data Utilization Application Development Phase.

In some cases, however, the data needed for the project may not be available, and you may need to take a different approach. For example, you could try a more detailed search or use different keywords, but you may also need to look at other sites or, in some cases, revise your data usage plan.

4.2.2 Examination of data spaces

Based on the data to be used, examine the data space to be used and decide whether to use it.

[What is data quality?]

When deciding whether to acquire data, it is also important to consider whether the data is of good quality. What is good-quality data? The following are some characteristics of good-quality data.

Table 4.1 Data quality assessment metrics

No.	Quality evaluation assessment metrics	Description
1	Accuracy	The degree to which data represent the true value of the intended concept or event. For example, whether the last name and first name are in the correct order.
2	Completeness	The degree to which data covers the entire set of data, if the subject data is a collection of multiple data items. For example, whether all required items are included.
3	Consistency	The degree to which data is consistent and coherent with other data. For example, whether there is a discrepancy between the date of birth and age.
4	Credibility	The degree to which data is regarded as true and believable by users. For example, whether information is provided about the reliability of the data, such as how it was collected and verified.
5	Currentness	The degree to which the data in or displayed by the system is kept current with changes in the real world. For example, whether data that is constantly changing, such as stock prices, is up to date.
6	Accessibility	The degree to which data can be accessed by people with disabilities. For example, whether the same information is communicated through multiple channels (text, images, audio, etc.).
7	Compliance	The degree to which data adheres to standards, conventions, or regulations and similar rules related to data quality. For example, whether data related to personal information and

		financial products subject to regulatory requirements complies with applicable regulations.
8	Confidentiality	The degree to which data is accessible only to authorized users. For example, whether data is encrypted.
9	Efficiency	The degree to which data can be processed using the appropriate amount and type of resources (e.g., memory) of a system. For example, whether unnecessary redundancy is causing excessive processing time and resource consumption.
10	Precision	The degree to which data is precise. For example, whether the error is within the expected range if the data is expected to have an error.
11	Traceability	The degree to which data provides an easy audit trail of access to and changes made to the data. For example, whether a history of access to the data (i.e., who accessed it and when) and a history of changes to the data (i.e., who changed it and when) are available.
12	Understandability	The degree to which data can be read and interpreted by users. For example, whether appropriate languages, symbols, or units are used.
13	Availability	The degree to which data can be retrieved as needed. For example, the extent to which data can be retrieved on certain days or at certain times, such as at night or during the holiday season.
14	Portability	The degree to which data can be installed, replaced, or moved from one platform/organization to another. For example, whether there is a way to easily exchange data between different systems.
15	Recoverability	The degree to which original data can be recovered in the event of a system failure. For example, whether the system can be restored to its pre-failure state using backup data or RAID devices.

Source: ISO, "data quality characteristics" in *Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model*[10]

[10] ISO/IEC 25012:2008, Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model, https://iso25000.com/index.php/en/iso-25000-standards/iso-25012

Table 4.1 is a list of metrics for assessing data quality based on ISO definitions. Data quality can be assessed from these perspectives, and data that meet the criteria would be considered high quality data. This is important because if data quality is poor, the data cannot be used or analyzed accurately, and this can lead to incorrect decision-making.

The extent to which data quality should be improved depends on the intended use. In many cases, the higher the quality of the data, the higher the cost of obtaining it. For example, if it is acceptable to use data with some degree of low accuracy, such as using customer surveys to gain a general understanding of trends, the cost of obtaining the data can be kept low. Therefore, when considering data quality, it is important to consider the balance between quality and cost.

Chapter 5: Contracting Phase

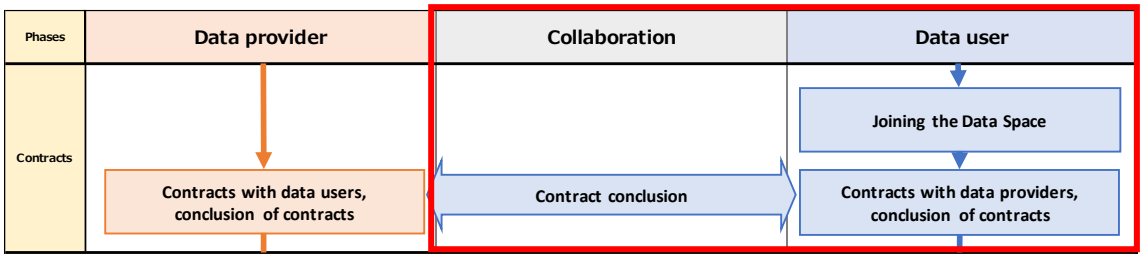

Figure 5.1: Process diagram for the Contracting Phase

This chapter, "Contracting Phase," explains the tasks for the data users shown in the red box in Figure 5.1 of the overall process diagram.

In this phase, data users decide which data spaces they will participate in, and enter contracts with data providers for the use of the data.

5.1 Participating in data spaces

Based on the data usage plan explored in Chapter 2, data users should evaluate the data spaces that provide the services, tools, and features their organization needs. Some data spaces have their own terms of use and service agreements. Review these before deciding whether to use the data space.

Some data spaces are already used by different industries, sectors, and regions. Check with the organizations that promote the use of the data space[11] before deciding whether to participate in or use the data space.

5.2 Contracts with data providers and conclusion of contracts

Procedures or contracts may be required in order to use data provided by a data provider. Their purpose is based on the concept of data sovereignty, such as preventing the use of data by users not intended by the data provider.

[11]Examples of organizations promoting the use of data spaces: Japan Business Federation (Keidanren), Japanese Bankers Association (JBA), Japan Machinery Federation (JMF), Japan Retailers Association (JRA), and Automotive and Battery Traceability Center Association (ABtC)

If the data is provided based on the terms of the contract, the provision of the data may be determined mechanically based on the terms of the contract, rather than having a human decide whether to provide the data each time there is a request for the data. Some digital infrastructures may offer contract management capabilities that can be used to review the terms and conditions presented by the provider before entering into a contract.

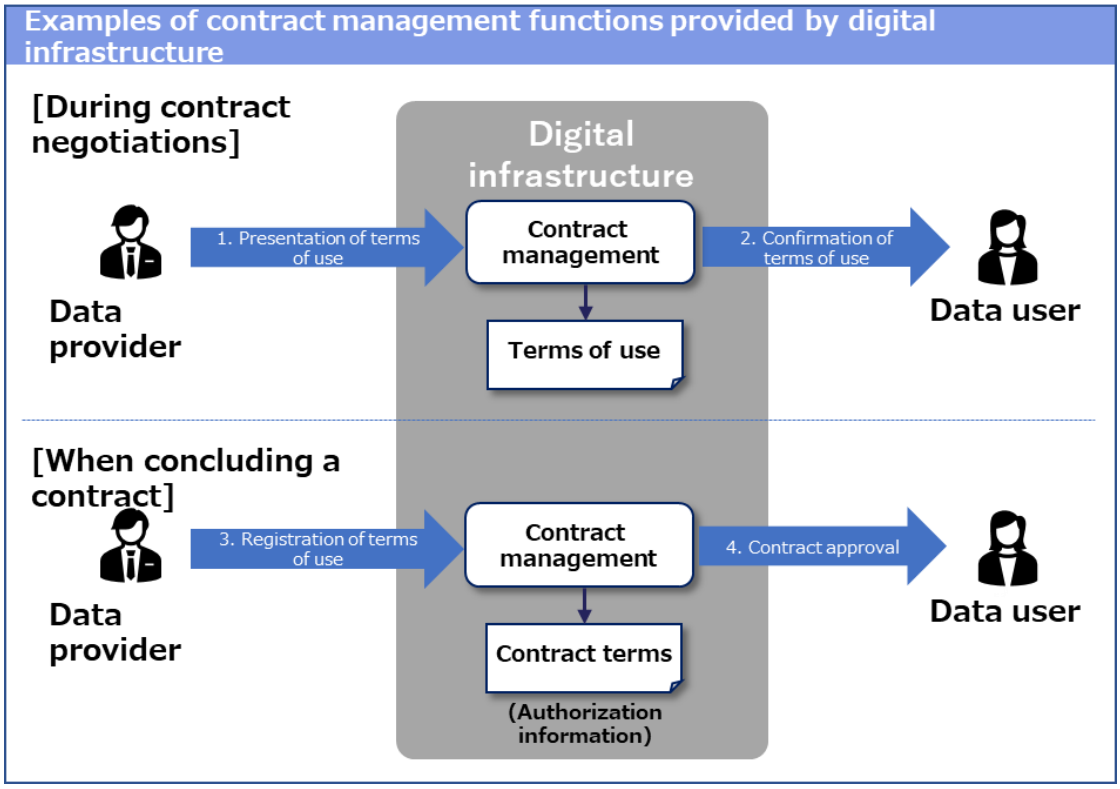

Figure 5.2: Contract management process offered by a digital infrastructure[12]

1. Presentation of terms of use

   The data provider presents the terms of use for the data to be provided to the data user.

2. Confirmation of terms of use

   The data user confirms the terms of use presented in step 1. Then, if necessary, the data user negotiates the contract details with the data provider.

3. Registration of terms of use

   The data provider registers the contract details in the contract management function based on the terms of use agreed in step 2.

4. Contract approval

   The data user checks and approves the contract details.

[12]Information-technology Promotion Agency, created by the IPA based on Introduction to Data Spaces, https://www.ipa.go.jp/en/digital/data/data-spaces.html

5.2.1 Points to consider when concluding a contract

The following should be considered when entering into a contract with a data space or organization that provides data.

- Review the license terms, purpose and scope of use, and disclaimer.
- Open data often does not require a contractual agreement, but check first because there are cases where terms of use are specified.
- From a security and cost perspective, try to keep contract terms to the minimum necessary, such as duration and scope of use.
- When conducting a data transaction through a data marketplace, ask if the marketplace operator has a prepared contract document.

[What is a data marketplace?]

In the section on what to consider when concluding a contract for the use of data, the term "data marketplace" was used. Below is a brief explanation of what this is.

▼ What is a data marketplace?

A data marketplace is a place where many buyers and sellers come together to buy and sell data. Like a stock exchange, it allows data to be traded efficiently and securely.

▼ Benefits of using a data marketplace

- Streamline transactions: The time and cost of research to obtain data can be reduced.
- Conduct transactions at fair prices: The results of multi-party transactions will help shape market prices and stabilize transactions.

There are many benefits to using data marketplaces. When exploring business opportunities that use data, consider using data from a data marketplace.

Chapter 6: Development Phase of Applications that Use Data

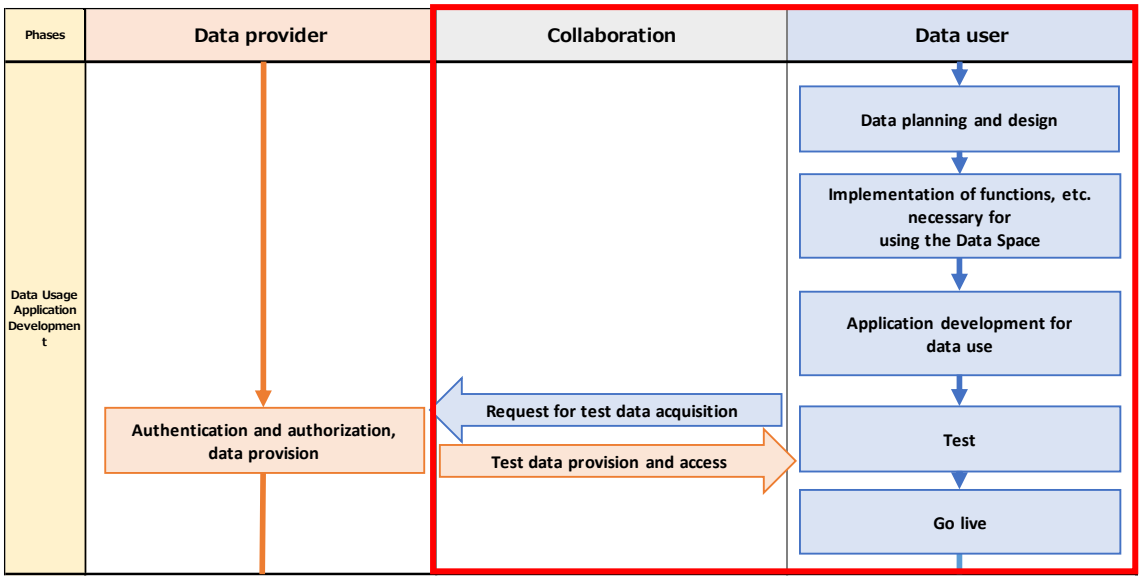

Figure 6.1 Process diagram for the development phase of applications that use data

This chapter, "Development Phase of Applications that Use Data," explains the tasks for the data users shown in the red box in Figure 6.1 of the overall process diagram.

In this phase, an application is developed to use the data acquired by the data user in the business. It is recommended that the same data be used in the final tests as in the actual production run.

6.1 Data planning and design

The data provided by data spaces is usually standardized because it is used by different organizations. To ensure that the data provided can be used smoothly within the organization, the organization must decide in advance on its data structure and data management methods.

① Data modeling: Clarify the structure and relationships of data.

② Data design: Design how data is used, managed, and maintained.

③ Data standardization: Standardize master data using GIF, IMI, etc. to enable efficient and accurate data exchange between different systems and organizations. For more information, see the Government Interoperability Framework (GIF) [13] and the

[13] Digital Agency, Government Interoperability Framework (GIF), https://www.digital.go.jp/en/policies/data_strategy_government_interoperability_framework

Infrastructure for Multilayer Interoperability (IMI)[14].

6.2 Implementation of functions, etc., necessary for using data spaces

Some data spaces provide useful functions and services to data users.

A prime example is the "connector." A connector is a basic module of data spaces that enables secure data transfer and access between systems.

Some also provide a testbed[15] for verification and evaluation. A testbed can provide benefits such as increased development efficiency and reduced implementation risk.

For the services and functions provided, check to see if there are specifications or other documents released by the data space, and refer to them if available. If they are not publicly available, try to obtain them from the organization providing the service. Using these available services and functions and avoiding developing them as much as possible will speed up the application development process and improve the quality of the application.

However, if a service or function is not provided, the minimum necessary software development will be required.

Even in such cases, it is desirable to try to minimize the amount of in-house development by using the services and functions that are publicly available.

6.3 Application development for using data

If it is determined that a minimum level of development is required, we recommend developing the application for using the data in phases, such as "collecting data," "organizing and processing collected data," "accumulating organized and processed data," "extracting accumulated data," and "using extracted data for analysis, services, and applications." For more information, see "4.4.1 Process of creating a data usage environment" in the DX Practical Guidelines.

---

[14] Information-technology Promotion Agency, What is IMI, https://imi.go.jp/goi/imi-about-en/

[15] A platform or proof-of-concept infrastructure used in system development to test and verify in a production-like environment.

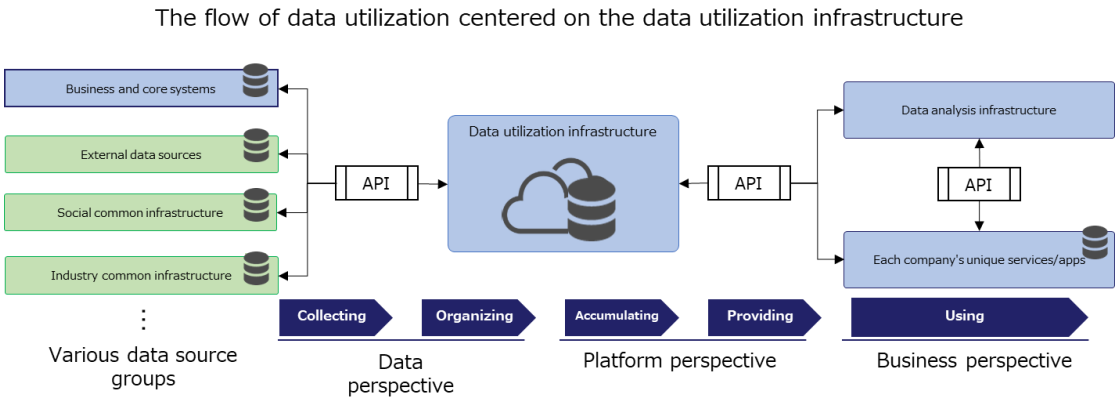

Figure 6.2 Flow of data utilization centered on the data utilization infrastructure (Source: DX Practical Guidelines)

Below are examples of the application development required for each phase.

① **Data collection**

- Develop APIs for retrieving data from external data sources
- Retrieve only the minimum amount of data necessary, using Graph[16] and other technologies, rather than retrieving all the data.

② **Data organization and processing**

- Data cleansing: Correct errors and missing parts of data to improve data quality
- Data conversion: Convert data into a format that can be used for analysis and between systems
- Data integration: Integrate multiple data sources into a single data set.
- Data anonymization: Mask personally identifiable information in data to protect personal information

③ Data accumulation

- Store organized and processed data in a designated location

④ **Data provisioning (collaboration with data utilization)**

- Retrieve accumulated data and collaborate it with data analytics platforms and data utilization applications

⑤ **Data usage (analysis and data usage applications)**

- Data tabulation: Group data by specific criteria (e.g., by category) and calculate statistics for each group
- Data filtering: Extract only the necessary data
- Data sampling: Extract a portion of the overall data

[16] A data operation method that allows users to get the information they need directly and go deeper. It is highly scalable, fast, and has the advantage of low transmission load.

- Data normalization: Reduce data variation to provide consistent data
- Data augmentation: Create new data by processing existing data to increase the learning data for a model
- Data anonymization: If data containing personal information is accumulated without being processed in ②, consider anonymizing the data in this phase

6.4 Tests

We recommend that data users test the developed application with production data before going live. Some data providers may not provide production data until a contract is signed, even if the data is to be used for testing purposes. Therefore, we recommend that the testing period be included in the contract period.

If you can test with production data, send a data acquisition request to the data provider that holds the relevant data and acquire the data.

The data can be obtained from the data provider in a variety of ways, such as receiving a file containing the production data or obtaining access rights to the location where the data is stored.

6.5 Go live

When testing is complete, run the application for using the data in the production environment. Next is the operation phase.

Chapter 7: Operation Phase

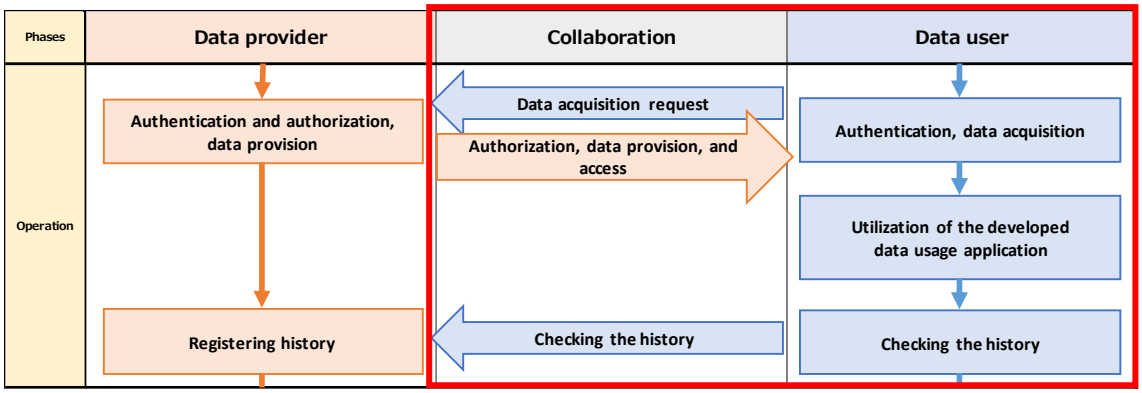

Figure 7.1 Process diagram for the Operation Phase

This chapter, "Operation Phase," explains the tasks for the data users shown in the red box in Figure 7.1 of the overall process diagram.

In this phase, identification and contract terms are verified (authentication and authorization) before the data provided is obtained and the developed application is used to perform business activities and analysis. If necessary, a history check is performed to increase the credibility of the data.

During this phase, proper data management is also conducted in accordance with data governance and other rules established by the CDO. If there are problems with the data or its use, feedback is provided to the department managing the data for improvement.

7.1 Authentication, data acquisition request, and authorization

When data is acquired, authentication and authorization are required to verify the identity of the data recipient and to confirm that the acquisition meets the terms of the data use agreement.

7.1.1 Authentication and data acquisition request

The identity of a data user as a contracted user is proven by registering the contracted user's information in advance and authenticating the user prior to data collaboration. The authentication process shown here is only an example, and user authentication, which is the most basic process, is explained here. (The user, in this case, is not necessarily an individual, but can be an organization.) Other examples include endpoint authentication and application authentication.

There are several ways to perform authentication. The following is an example of using an authentication function provided by a digital infrastructure[17].

Example of using the authentication function provided by a digital infrastructure

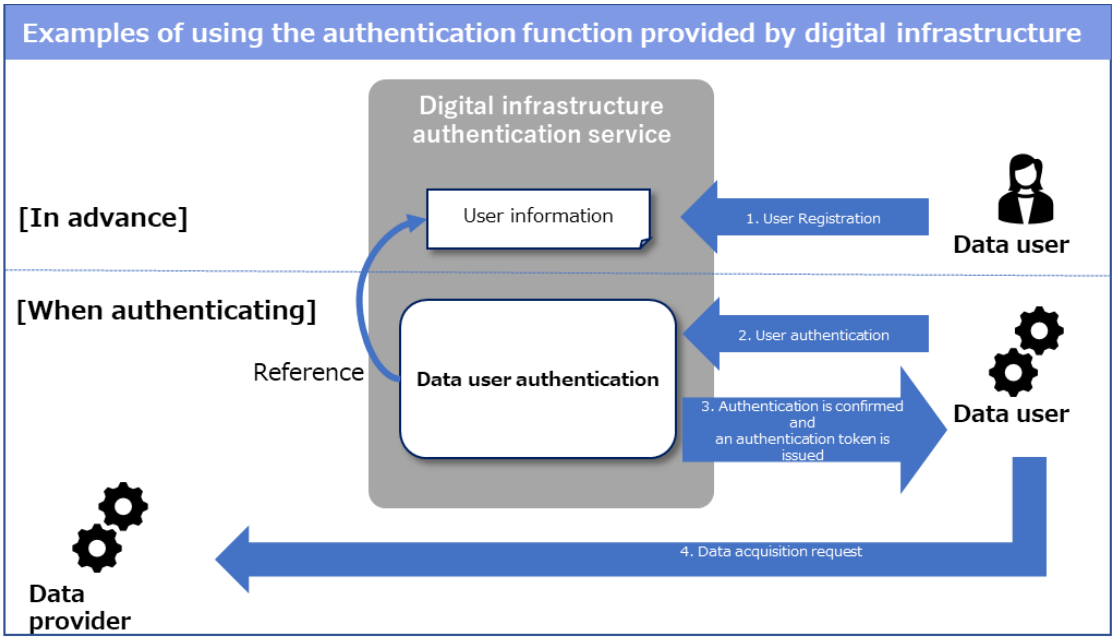

Figure 7.2 Authentication process using a digital infrastructure[18]

1. User registration

   The data user registers in advance as a user of the digital infrastructure and obtains the credentials, including a user ID, to use the digital infrastructure.

2. User authentication

   When the data user submits an authentication request to the authentication service, the authentication service prompts the user for the user ID and other credentials. When the user provides this information, the authentication process is performed.

3. Authentication confirmed and authentication token issued

   If authentication is successful, an authentication token associated with the user ID is issued.

4. Data acquisition request

   The data user submits a data acquisition request with an embedded authentication token to the data provider. The data provider then uses the authentication token to determine whether the user is legitimate.

---

[17]Information-technology Promotion Agency, sourced from the section on digital infrastructures in Introduction to Data Spaces, https://www.ipa.go.jp/en/digital/data/data-spaces.html

[18]Information-technology Promotion Agency, created by the IPA based on Introduction to Data Spaces, https://www.ipa.go.jp/en/digital/data/data-spaces.html

7.1.2 Authorization

The data user can access the data if the data provider authorizes the user's request for data acquisition. Some digital infrastructures also provide an authorization function, in which case the user is allowed to access data in accordance with the terms of the contract registered with the digital infrastructure at the time the contract is concluded.

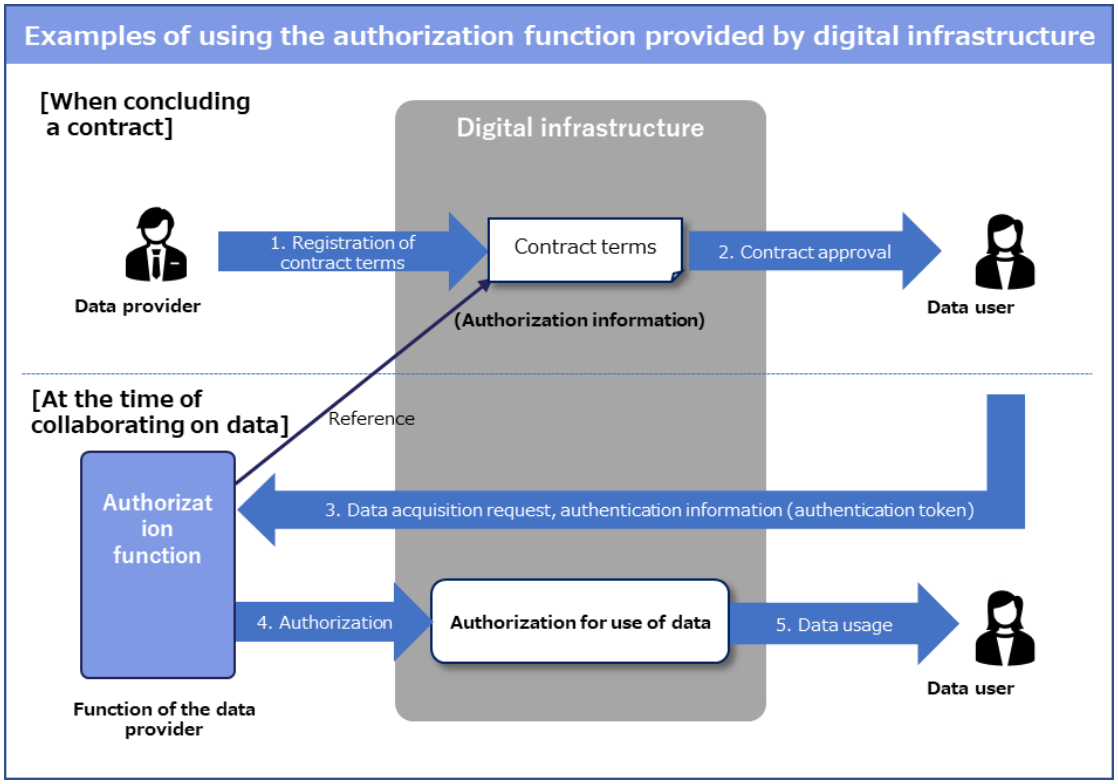

Figure 7.3 Authorization process using a digital infrastructure[19]

1. Contract terms registration
   The data provider registers the terms of use for the data it provides to the data user with whom it has a contract.
2. Contract approval
   The data user reads and accepts the contract terms presented in 1. These contract terms will be used as the basis for authorizing the data user when the user submits a data acquisition request.
3. Data acquisition request and credentials (authentication token)
   The data provider receives a data acquisition request and credentials (authentication token) from the data user.
4. Authorization

[19]Information-technology Promotion Agency, created by the IPA based on Introduction to Data Spaces, https://www.ipa.go.jp/en/digital/data/data-spaces.html

The data user will be granted access to the data if the user is found to be in compliance with the contract terms.

5. Data usage

   The data user can now use the acquired data.

7.2 Use of the application developed for the use of data

Based on the data acquired in 7.1., the data user will use the application developed in Chapter 6: Development Phase of Applications that Use Data > 6.3 Application development for using data > (1) Data organization and processing and thereafter to perform business activities and analyses.

7.3 History check

During operation, the credibility of the data is verified by checking its history. A history check function may be provided by a digital infrastructure. Using this function, both data users and data providers can check the records of who registered, processed, or acquired the data and when, and verify that the data has not been tampered with. This section explains the basics, but note that there are different types of history, such as data conversion history, data modification history, and data usage history. It is also sometimes used in conjunction with a timestamping function provided by a digital infrastructure and version management.

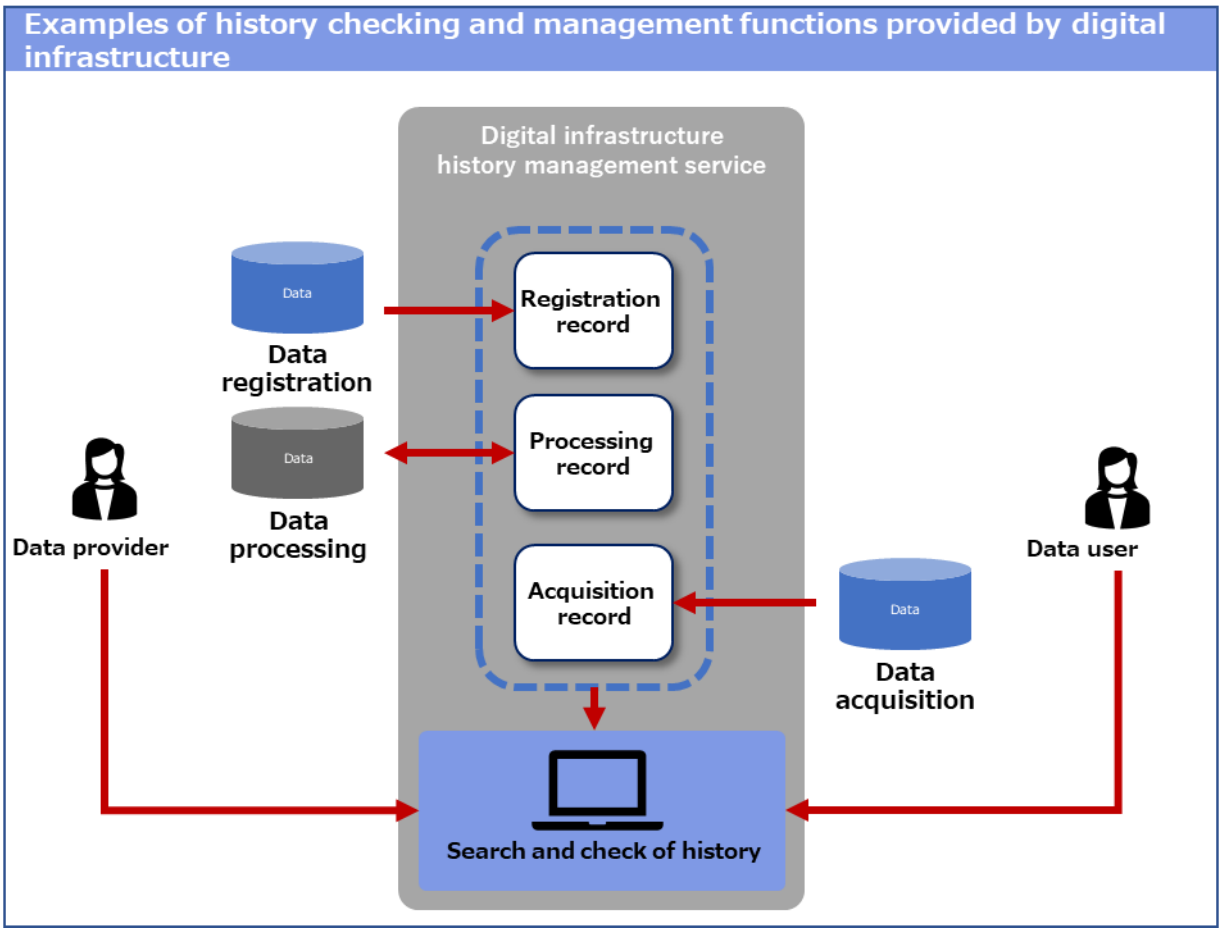

Figure 7.4 Overview of history management and history check functions provided by the digital infrastructure

History management may have the following functions. It manages data records for both data users and data providers, and both parties can view the records.

1. History records
   - Data registration: Original information is recorded
   - Data processing: Data processing history is recorded
   - Data acquisition: Data transmission and reception history is recorded
2. History search and history check
   - Search for a specific set of data and view the entire history

note

[Differences between authentication and authorization]

Authentication and authorization are two different functions and concepts, but they are sometimes confused because they are similar words. A simple explanation of their functions is as follows.

- Authentication: The process of verifying and identifying a user.
- Authorization: The process of granting permission to use something.

To understand authentication, imagine entering your ID and password to log in to a website or application.

For authorization, imagine a situation where you cannot access a folder or file on an internal server. This means you do not have permission to access that folder or file, or you are "not authorized" to do so.

See the example in Figure 7.5. Only data user A was able to pass the authentication and authorization processes and access the provider's data. Data user B's credentials were correct (the user was identified as a legitimate user), but the user was unable to pass the authorization process (the user was identified as someone who did not have access rights to the data) and was therefore denied access to the provider's data. Data user C was unable to pass the authentication process (the user was not identified as a legitimate user) and was, therefore, also denied access to the provider's data.

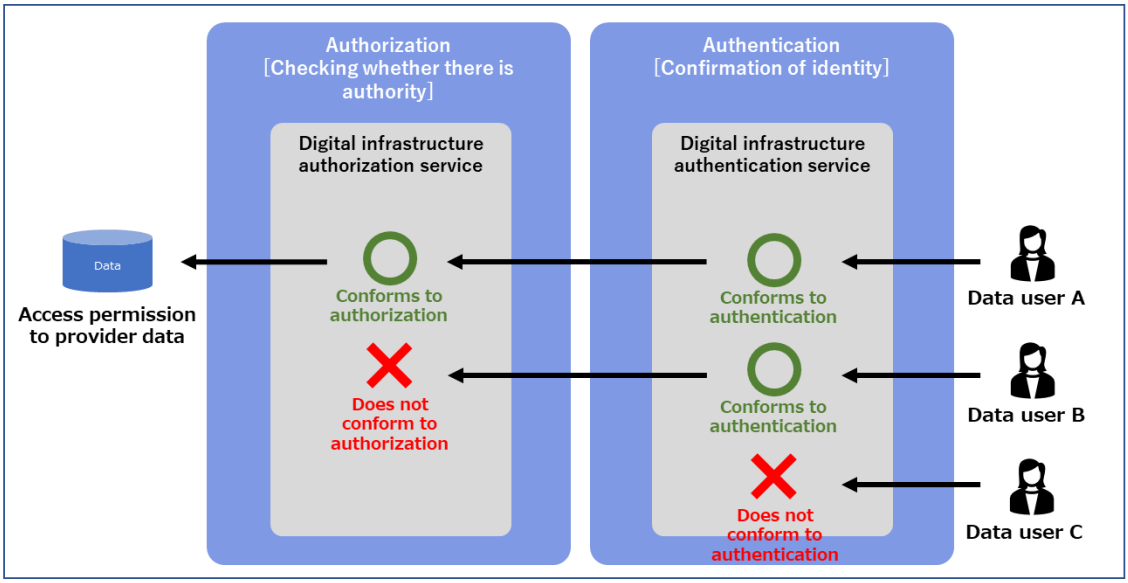

Figure 7.5 Example of using the authentication and authorization functions provided by the digital infrastructure

The use of both authentication and authorization, as described above, makes it possible

to verify the identity of users and properly control their access privileges.

[Expected authentication and authorization cases]
This chapter used the terms data user and data provider, both generic terms, to explain the concept of authentication. This section will provide some additional explanations using specific expected cases to help with understanding.
Let's look at the following case.

Table 7.1 Expected cases

Business	Purpose	Data to be acquired	Data acquisition source
Real estate brokerage	Accurately appraise real estate sales prices	- Commercial facility usage data (gender, age, product price, etc.) - Human flow data (location, gender, age range, etc.) - Property data (price, location, etc.) etc.	- Service providers in the commercial facility (retail, food and beverage, railway, etc.) - Other real estate agents (sales brokerage, rental brokerage, management)

When applied to the data acquisition process, it looks like Figure 7.6 below.

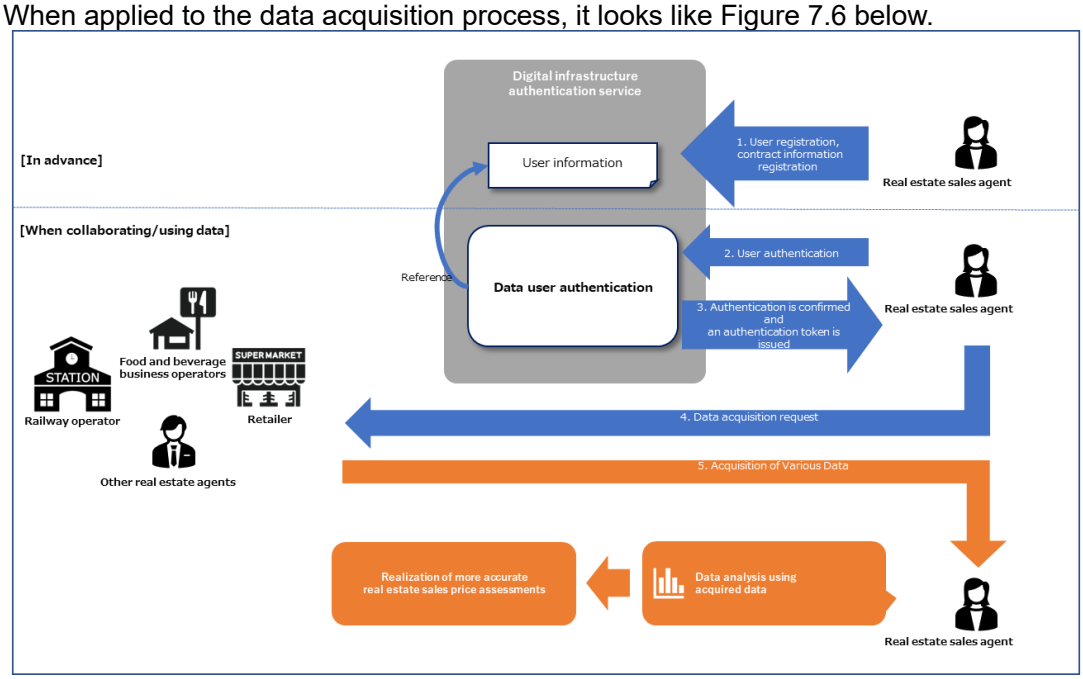

Figure 7.6 Example of authentication process using a digital infrastructure

In this hypothetical case, a real estate broker can acquire data from each business and analyze the data to more accurately estimate property sales prices.

Chapter 8: Evaluation Phase

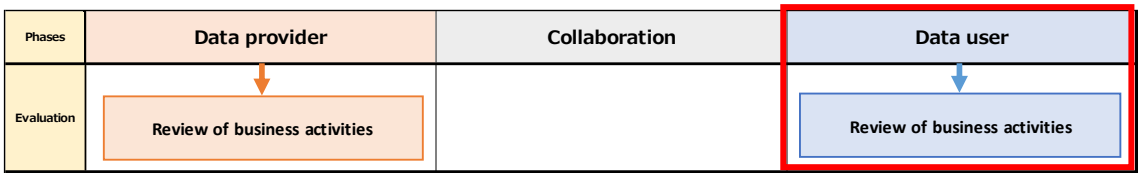

Figure 8.1 Process diagram for the Evaluation Phase

This chapter, "Evaluation Phase," is a phase in which the results of the business activities and data analysis conducted based on the data strategy in Chapter 7, "Operation Phase," are reviewed and improvements are implemented.

8.1. Review of business activities

Evaluate business activities from a Chief Data Officer (CDO) perspective to improve data-driven business activities and data analysis. If a data space is being used, we advise including this perspective in the evaluation. Depending on the results of the evaluation, various actions may be required, such as acquiring additional data or changing the perspective of the analysis.

Below are some examples of evaluations based on the characteristics of the data space and the roles of the CDO (see "Chapter 1: Business Strategy Development Phase").

▼Evaluation focused on the use of data space

① **Interoperability**

Evaluate how effectively data is now shared across organizations and how cooperation has progressed through the use of the data space.

② Data sovereignty

Assess whether data complies with a country or region's data sovereignty laws and regulations. Also, verify access rights to data held by an organization or country and assess whether a third party has accessed the data.

▼Evaluation based on the roles of the CDO

① Data strategy performance level

Assess the extent to which the formulated data strategy has contributed to the company's business goals (vision). Measure the impact of cost reduction, revenue, customer satisfaction, and other KPIs, and assess areas of success and areas for improvement.

Also, assess the extent to which the use of data has helped drive innovation and the creation of new products and services.

② **Data governance effectiveness**
Evaluate whether the data governance framework and data management processes are working as intended. In terms of legal compliance, assess whether the organization is complying with laws and regulations related to data privacy and security.

③ **Data-quality status**
Evaluate whether the accuracy, completeness, and consistency of the data are assured.

④ **Data-analysis promotion status**
Determine if any data necessary for data analysis are missing. If so, determine if necessary data can be added. If an analysis from a new perspective is needed, determine the data needed for that analysis.

⑤ **Security and risk management status**
Evaluate whether security policies and risk management are properly implemented, and whether the risks of data leakage and cyberattack are minimized.

⑥ Extent to which data culture has spread throughout the organization
Assess the extent to which a data-driven culture has taken root throughout the organization. Assess the extent to which employees' data literacy has improved and the extent to which they use BI[20] and other tools to make data-driven decisions in their daily work.

⑦ **Effectiveness of data infrastructure**
Evaluate whether the development of data utilization infrastructure and data analysis infrastructure is progressing as planned and whether the performance of these infrastructures is adequate. Evaluate the extent to which cloud migration and data lake construction are contributing to data processing efficiency and accessibility.

Thus, continuously understanding the differences between plans created based on data strategies and actual results from the perspectives of the data space and the CDO, responding appropriately, and being flexible to changes in the world and new needs will enable the organization to progress toward the vision set by management, grow, and succeed.

[20]Business intelligence (BI) tools are used by organizations to collect, analyze, and visualize data. For example, graphing sales data or analyzing customer behavior can help executives and employees make data-driven decisions.

This document provided a comprehensive explanation of data spaces and data utilization in general, from planning in the upstream phase to operations in the downstream phase. The effective use of data by management is essential to maintaining a competitive edge in modern business. It is hoped that this document will help organizations to extract even more value from their data.

[Part 2] Data Providers

In Part 2: Data Providers, you will learn about the benefits of providing data, and for companies that want to use data for business such as service development, the processes and tasks on the data provider side are explained. As with Part 1: Data Users, the chapters are based on the overall process diagram, with each phase of the diagram forming a chapter and the tasks within each phase explained.

*However, where there is overlap with the content of Part 1, the references in Part 1 are listed, so please refer to Part 1 as appropriate as you read through the guide.

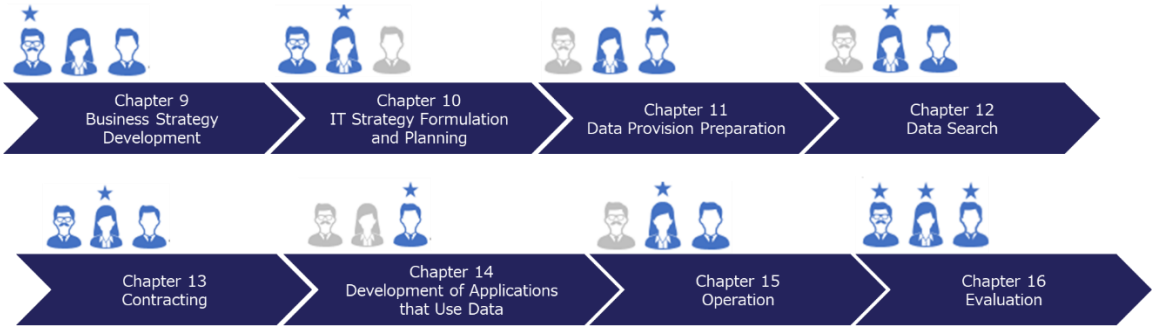

Figure 9.0 Overall Process - Data Providers (simplified version)

Chapter 9: Business Strategy Development Phase

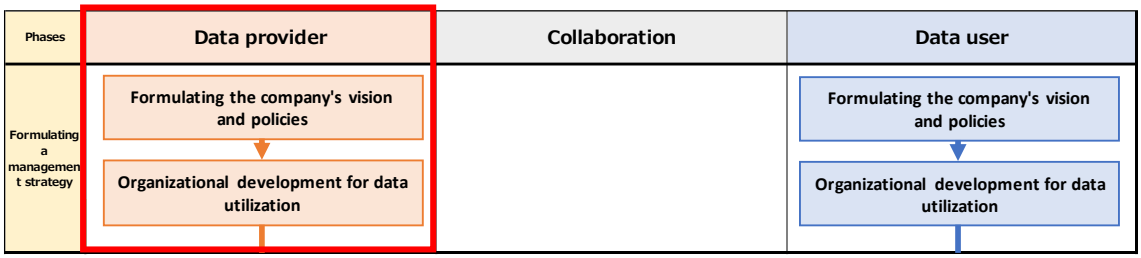

Figure 9.1 Process diagram for “Business Strategy Development Phase”

This chapter, “Business Strategy Development Phase,” explains the tasks on the provider side in the red box in Figure 9.1 of the overall process diagram.

In this phase, the company's vision and policies are formulated, and how to provide data and proceed with the business is considered. To do this, first the management needs to understand the basic knowledge about data (the importance of data utilization, the role and value of the CDO, the overall process of data utilization, etc.). On that basis, the position of CDO is established within the organization, and various initiatives related to data utilization will be carried out. We hope you will deepen your understanding by reading through this document, including Part 1: Data Users, and referring to various reference materials.

9.1. Formulating the company's vision and policies

The explanation for this section is the same as the content described in “Part 1: Data Users: Chapter 1,” so please refer to that.

9.2. Organizational development for data utilization

9.2.1. Appointing a Chief Data Officer (CDO)

As described in Part 1: Data Users: Chapter 1, we recommend that data providers also appoint a Chief Data Officer (CDO). The CDO is responsible for everything from formulating data provision strategies to managing the data provided.

9.2.2. The role of the CDO and organizational development

For more information on the role of the CDO, please refer to “Part 1, Chapter 1, 1.2. Organizational development for data utilization” and “Data-Centric Commander: CDO Handbook.”

In some cases, from the perspective of providing data, the roles that the CDO should play as the data provider are added compared to the case of data use.

Table 9.1 Examples of roles the CDO should play in providing data

No.	Role	Role content
1	Formulation of data strategy	Based on the vision to be achieved as communicated by the management, while reaching agreement with the management, clarify the purpose and strategy of data provision, formulate a policy to maximize the value of data, and share it throughout the organization.
2	Establishing data governance	Same as in “Data Users”
3	Improving data quality	Same as in “Data Users”
4	Promotion of data provision	Examine how to provide data to more users in an easy-to-use format
5	Data security management	Same as in “Data Users”
6	Fostering a data culture throughout the organization	Same as in “Data Users”
7	Development of data infrastructure	Develop infrastructure to provide data efficiently and safely

8	Restrictions on the purpose of use of data	Set restrictions to prevent inappropriate use, based on assumptions about how the data provided will be used
9	Privacy management	Ensuring that data is appropriately anonymized and personal information is protected, and complying with laws and regulations such as the GDPR (General Data Protection Regulation)
10	Contracts	Clarify contracts that relate to the provision of data, and define the scope and restrictions of data use. In some cases, it may be necessary to work with the legal department to minimize legal risks

By keeping these points in mind, you can ensure that the data is highly reliable, and to build a beneficial relationship for both the recipient and the provider of the data.

As mentioned above, the role of the CDO is wide-ranging, but before moving on to the next chapter, “IT Strategy and Planning Phase,” it is necessary to make the following preparations.

① **Clarification of the purpose and goals of data provision**

First, the purpose and goals should be clarified. For example, this could be to create new business opportunities for both the data provider and the data recipient, to improve existing services, or to improve customer satisfaction. The CDO should take the lead in this process, and should work with the management to reach agreement.

② **Legal and ethical measures**

Consider the legal and ethical issues involved in providing data. For example, consider issues such as data privacy protection, terms of use, and contract conditions, and take appropriate measures.

③ **Data value assessment and analysis**

Assess the value of the data to be provided. Consider key points regarding data quality, such as data quality, quantity, and update frequency, and analyze what benefits there are for the recipient.

Chapter 10: Formulating an IT Strategy and the Planning Phase

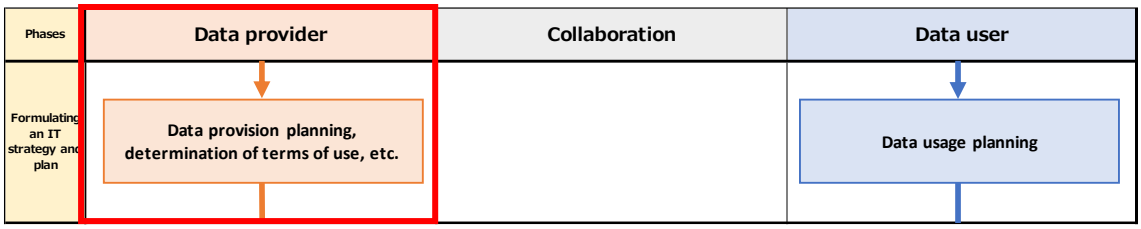

Figure 10.1 Process diagram for “Formulating an IT Strategy and the Planning Phase”

This chapter, “Formulating an IT Strategy and the Planning Phase,” explains the tasks on the provider side in the red box in Figure 10.1 of the overall process diagram.

In this phase, as a data provider, the business plan by providing data and the terms of use are examined.

10.1. Data provision planning, determination of terms of use, etc.

10.1.1. Data provision planning

Before planning data provision, we will introduce some of the benefits of data provision.

10.1.1.1. Benefits of data provision

Consider what benefits can be enjoyed by providing data held by the company. First, we will introduce an example of a data provider and the benefits that can be brought about.

① **Utilizing data that was not being used for internal improvements**

In addition to the data that companies generally manage (such as sales data and inventory data in the case of a retail business), there are cases where data that was not being used internally can be used for the improvement of the company's business through data collaboration with other companies. For example, in the case of a transportation business that operates trucks throughout Japan, you can collect telematics [21]data such as that from ETC cards, car navigation systems, and drive recorders on the routes taken. Then, in order for the transport company to create new value using this data, they can provide the data to a systems development company to help them develop a system that optimizes the most efficient delivery routes. By using

[21] Telematics is a technology that enables the collection and transmission of information in real time by installing a communication system in a vehicle, etc. For example, by using a drive recorder with telematics functions, you can automatically link location information and accident footage when an accident occurs, and accurately grasp the situation surrounding the accident.

the developed system, the transport company can reduce fuel costs and improve service by shortening delivery times, and as a result, they can gain benefits such as increased customer satisfaction. The systems development company also gains the opportunity to create a new business.

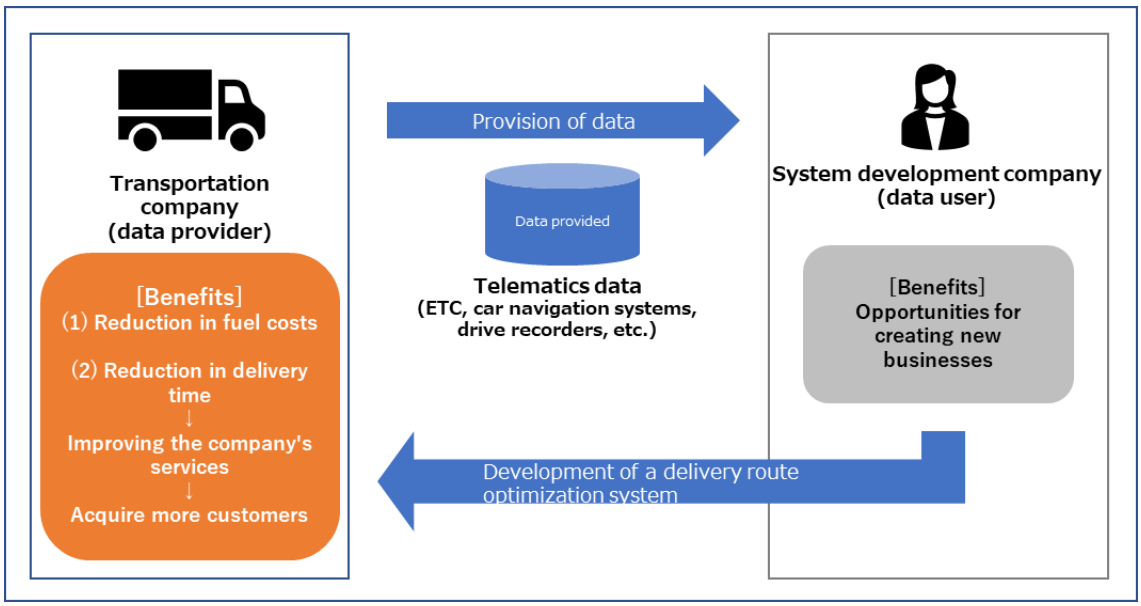

Figure 10.2 Benefits of data provision (1)

② **Mutual provision (sharing) of data within the same industry**

Not only providing data unilaterally, but also having partners provide data (sharing) will help to mature and develop the industry as a whole. However, care is needed from the perspective of the Anti-Monopoly Act. [22]To give an example, in the automobile industry, multiple companies could provide each other with test data on safety devices, which would be useful for improving the functionality of safety devices and the like. This will enable each company to provide the market with safer products, which will in turn lead to an improvement in the value of each company's products and boost the Japanese automobile industry as a whole. At the same time, it will also contribute to solving the social issue of traffic accidents.

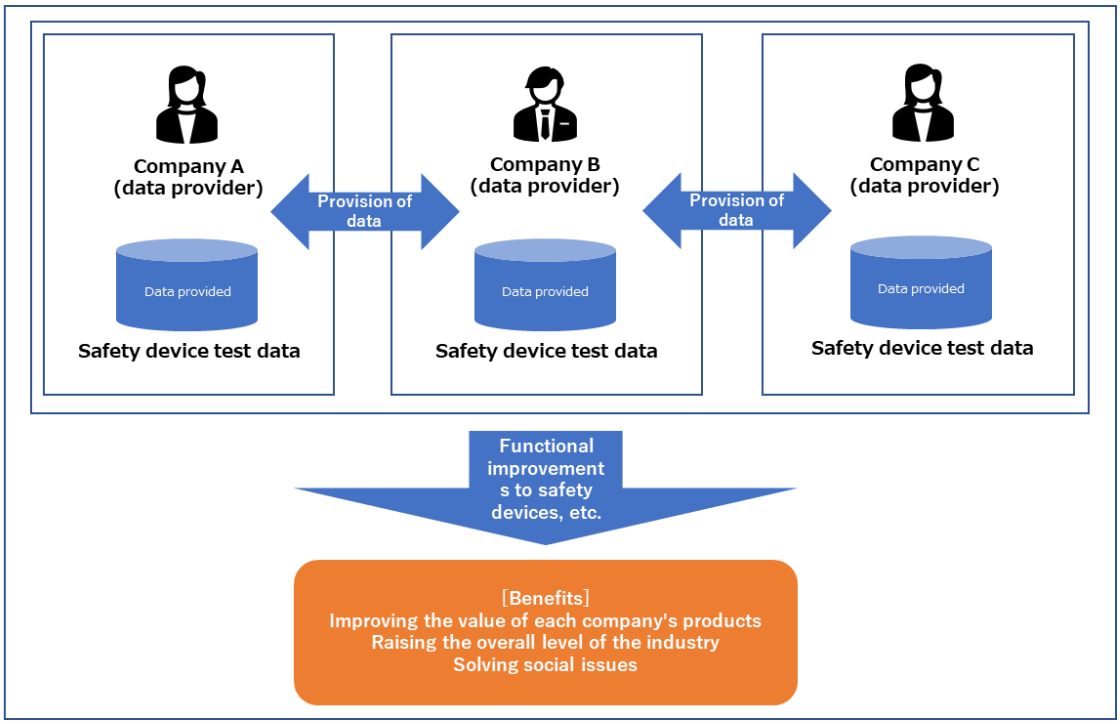

Figure 10.3 Benefits of data provision (2)

[22] From the perspective of the Anti-Monopoly Act, particular care is needed in cases where there is a high degree of market dominance. We recommend liaising with the legal department as necessary to check and consult as needed.
Reference: Japan Fair Trade Commission, “Guidelines Concerning the Activities of Enterprises, etc. Toward the Realization of a Green Society Under the Antimonopoly Act”: https://www.jftc.go.jp/dk/guideline/unyoukijun/green_gl.pdf

③ **Mutual provision (sharing) of data between different industries**

For example, even between different industries such as the automotive and medical industries, data sharing can bring benefits. Using technology such as telematics, as introduced in Example (1), you can link data on the situation and location of an automobile accident with data from medical institutions, and to construct a system that automatically notifies the nearest medical institution when an accident occurs, enabling rapid emergency response. This will speed up response times in the event of an accident, and is expected to improve the rate of survival.

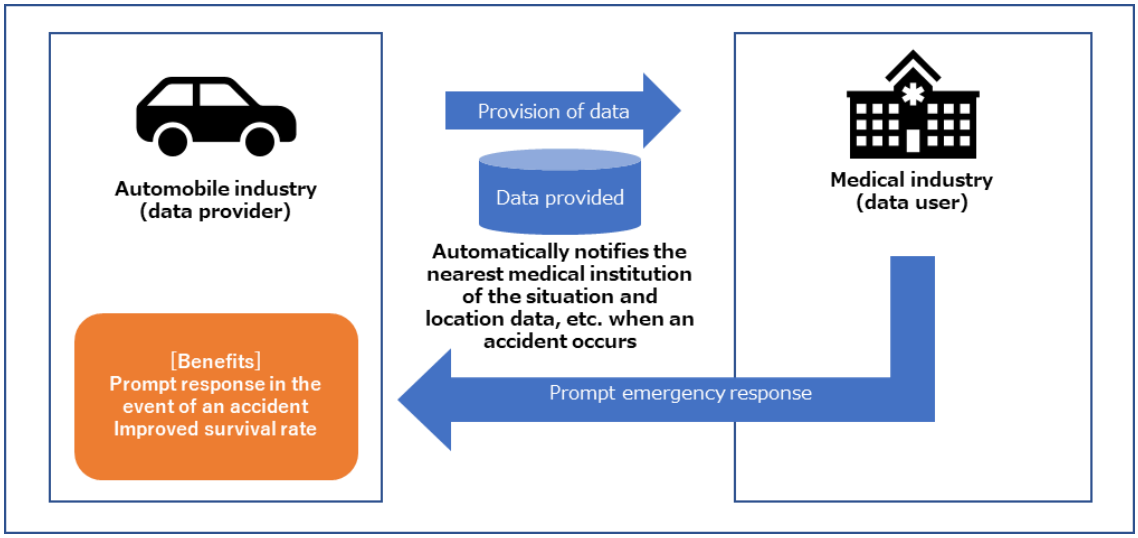

Figure 10.4 Benefits of data provision (3)

From examples like this, data provision can bring a variety of benefits, such as “realizing cost reductions for the organization's business,” “creating new business opportunities,” “contributing to the maturation and development of the industry as a whole,” and “solving social issues.”

10.1.1.2. Planning data provision projects

When planning to provide data to other organizations, the business division is primarily responsible for planning, based on the goals and vision clarified by the management of the organization. For example, the following steps are taken to proceed with planning.

① **Clarify the purpose and confirm compliance with laws and contracts**

First, the purpose of providing the data should be clarified. For example, if the data is being provided to a business partner, it is being provided to improve the services of the partner and the organization itself. If it is being provided to a government agency, it is

being provided for public policy purposes. If it is being provided to a research institution, it is considered to be provided for the creation of new knowledge.
In addition, the data to be provided should be checked from a legal perspective (such as the Act on the Protection of Personal Information and the Anti-Monopoly Act) and contracts should be checked to ensure that there are no problems with providing the data.

② **Understanding the value of the data provided**

Evaluate what data the organization has and what value or benefit that data has for the recipient.

[Examples of data characteristics evaluation]

• Uniqueness: Is the data difficult to obtain from other organizations or markets? (e.g. data on the usage of the organization's products, detailed customer feedback, etc.)
• Standardization: Is the data standardized so it is easy to use in collaborations with other organizations?
• Accuracy and quality: Is the data accurate and free from errors? Is it highly reliable?
• Timeliness: Can it be used quickly and in a timely manner?
• Scale: Is the amount of data large enough to allow for multifaceted analysis?
• Regulatory compliance: Does the use of the data conflict with legal or privacy regulations?

[Examples of business-oriented evaluation]

• Increased revenue: Can the data be used to create new revenue streams? (e.g. development of new services)
• Cost reduction: Can efficiency be improved and waste reduced? (e.g. optimization of supply chains)
• Decision-making: Can it facilitate more accurate and faster decision-making by the management?
• Third-party value: Is it data that other companies would want to purchase or data that would encourage them to form partnerships?

③ **Selecting targets**

Narrow down the target by analyzing the needs of the target group, similar past cases, and the trends of other companies in the same industry. If the target is more or less clear, you can consider providing data priority to organizations with which there is a close relationship at present or with which there is a plan to form a long-term partnership in the future.

[Examples of perspectives for selecting targets]

• Industry: manufacturing, medical, distribution, non-profit organizations, etc.

• Region: domestic, overseas, etc.

• Organization size: small and medium-sized enterprises, large corporations, etc.

• Data maturity: organizations that are beginners in data utilization, organizations that are using data in an advanced way, etc.

④ **Formulation of an implementation plan**

Until the data catalog and terms of use are published, consult with the CDO to clarify the processes for each phase, such as the process from the time of contract to the time of data provision after the contract has been signed, as well as the allocation of resources within the organization, and formulate an implementation plan for data provision.

By following these steps, your organization will be able to effectively plan the provision of data and build cooperative relationships with other organizations.

10.1.2. Determining the terms of use

10.1.2.1 Examples of points to consider regarding the terms of use

In this phase, the terms of use for providing the data should also be considered. The terms of use should be considered with particular care from the perspective of security and legal compliance, for example to prevent the data from being used by unintended users in future contracts.

There are many points to take into account when examining the terms of use, but some examples of points to consider are introduced below.

Table 10.1 Examples of the terms of use that should be noted when collaborating on data

No.	Terms of contract	Examples to note
1	Ownership and usage rights of data	• Who holds ownership of the data? (In the case of data spaces, the data provider holds the ownership, etc.) • What kind of usage rights are granted to data users?

		• Restrictions on reuse and redistribution of data, prohibition of commercial use, Creative Commons Attribution 4.0 International Public CC BY 4.0[23], etc.
2	Data quality and assurance	• How much is guaranteed in terms of data quality (see Part 1, Chapter 4, Point Memo)? • What is the responsibility and compensation in the case of poor quality?
3	Security and privacy	• How to handle personal information (masking of information that can identify individuals, etc.) • Notification and response process in the event of a security breach
4	Access control and authentication	• How is data access authentication implemented? • What are the conditions for managing and granting access privileges?
5	Restrictions and conditions for data use	• Are there any prohibited uses or usage guidelines? • Are there any restrictions on use for specific purposes or in specific industries?
6	Fee structure and payment terms	• What is the method for calculating data usage fees? • What is the payment schedule and method?
7	Responsibilities and compensation	• What is the scope of liability for damages arising from the use of data? • What is the maximum amount of compensation? • What is the maximum amount of compensation?
8	Contract period and termination of contract	• How long is the contract valid for? • What are the conditions for renewal and contract termination? • What will happen to the data after the contract ends (deletion, return, etc.)?

The above points are just examples, but it is important to clarify the terms of use for provided data.

It is important to guarantee safety and security by ensuring that the data provided is handled appropriately.

[23] Creative Commons Attribution 4.0 International Public (CC BY 4.0) is a license that allows the free sharing and modification of copyrighted works. https://creativecommons.org/licenses/by/4.0/deed.ja

As for how to define fair usage rights for data and background knowledge, the “Contract Guidelines for Data Usage Rights”[24] published by the Ministry of Economy, Trade and Industry can be used as a reference when considering these issues.

10.1.2.2. Publication of the terms of use (contract information)

One way to publish the terms of use that have been decided is to use a data catalog. A data catalog is a tool for publishing data, but in some cases the terms of use are also published. In some cases, the terms of use are published in a simplified form in the data catalog itself, with the detailed terms of use being presented separately in a contract such as a non-disclosure agreement (NDA)[25] or license agreement.

Data catalogs are explained in detail in Chapter 13, but there are data catalog sites on the web where data intended to be provided is posted, and in some cases the digital infrastructure provides contract management functions. By including the terms of use explained in this chapter, you can contact users.

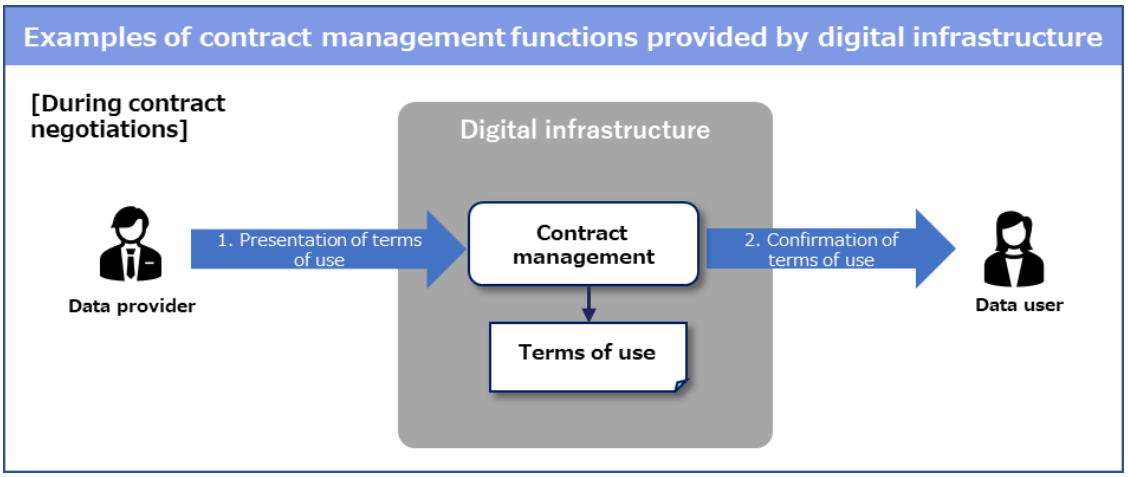

Figure 10.5 Example of the presentation of terms of use by a digital infrastructure-based contract management function[26]

As shown in Figure 10.5, by using the contract management functions provided by the digital infrastructure, you can present the terms of use for the data provided, contact users who wish to use the data, and negotiate the terms of the actual data usage contract.

---

[24] Ministry of Economy, Trade and Industry, Contract Guidelines for Data Usage Rights” : https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization.html

[25] A non-disclosure agreement (NDA) is a legal contract that promises not to divulge confidential information to a third party when sharing confidential information between two or more parties.

[26] Created by the IPA with reference to “Introduction to Data Spaces” by the Information-technology Promotion Agency, Japan: https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/dataspaces-gb.pdf

Chapter 11: Data Provision Preparation Phase

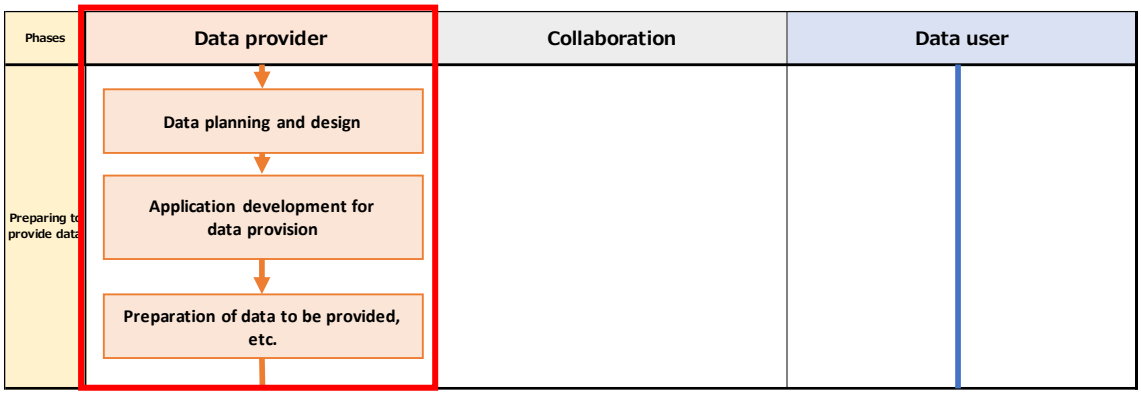

Figure 11.1 Process diagram for “Data Provision Preparation Phase”

This chapter, “Data Provision Preparation Phase,” explains the tasks on the provider side in the red box in Figure 11.1 of the overall process diagram.

In this phase, the data provider prepares the data to be provided.

11.1. Data planning and design

Based on the data provision plan from the previous chapter, make a plan for how to actually collect and manage data within the organization, etc. As an example, we recommend making a data provision plan and design for the following items.

① **Review of data to be provided**

Based on the content reviewed in the data provision plan from the previous chapter, clarify the type, format, quality standards, etc. of the data to be provided, and review the necessary data.

② **Review of data collection methods, etc.**

Consider the method of collecting the necessary data (e.g. manual collection, automatic collection from the organization's own systems or external data, etc.), frequency and timing of data collection, and timeliness, etc.

③ **Consideration of data management methods**

Consider data management methods, including storage location and storage format.

The following two patterns of management method are examples.

• Separate management: Separately manage the data for provision based on the original data, apart from the original data

• Original management: Manage only the original data

In the case of separate management, since the data for provision is managed separately, the risk of leakage of confidential information can be reduced by including only the minimum necessary information, but there is a possibility that the management costs will increase, such as for maintenance to preserve consistency. In the case of original management, it is efficient, with simple management and excellent timeliness, but there is a risk of confidential information leaking if appropriate filtering and masking are not done in advance.

Each management method has its advantages and disadvantages, so it is important to choose a management method that is suitable for each data provision strategy.

As a storage format, consider databases, JSON, CSV, XML, etc. In addition, for database formats, you need to consider whether to use an RDB[27], which is good at advanced searches and aggregation, or a NoSQL[28] or vector database[29], which can hold unstructured data and is easy to handle large volumes of data.

In addition, when managing data, consider data backup and recovery plans.

④ **Consideration of data collaboration methods**

Consider how to collaborate on data. Referring to the DMBOK, the following are examples.

• Point-to-point (P2P): A method of directly collaborating on data between systems. Although it is difficult to create a single point of failure (SPOF), as the number of systems increases, the efficiency of interface management decreases.

Examples: Email, API, Data Space, etc.

• Hub and spoke: Shared data is aggregated in a central data hub, and data is then collaborated via a central common system. This is efficient because it reduces the need for inter-system collaboration, but it is prone to SPOFs and requires high performance.

Examples: database systems, EAI, etc.

• Publish-subscribe: A system for exchanging messages asynchronously between multiple devices and systems. Because the coupling is low, the sender is not easily affected, and multiple recipients can receive data in the same format. On the other hand, the configuration is likely to become complex, and development costs will be high.

Examples: News distribution systems, IoT device monitoring, etc.

---

[27] RDB (Relational Database) is a type of database that manages data in tabular form.

[28] NoSQL (Not Only SQL or Non-relational) is a type of database that uses a data model that differs from that of relational databases.

[29] A vector database is a database that stores data as numerical vectors and performs searches and inquiries based on the similarity between vectors. It is particularly good at managing unstructured data (text, images, audio, etc.) that is handled in the fields of generative AI and machine learning, and is used for efficient searching and management of large data sets.

⑤ **Establishing data governance**

Clarify the ownership of data and the scope of responsibility of the data provider, and consider policies and rules for internal use of data. Also plan how to manage the data lifecycle (generation, use, disposal).

⑥ **Security assurance**

Implement appropriate security measures (access permission settings, encryption, etc.) for the data being managed.

⑦ **Data quality management**

Consider quality control processes to ensure data quality, such as data accuracy, consistency, and integrity. For example, we recommend considering the introduction of data verification tools to automatically check data.

11.1.1. Consider and participate in Data Spaces

Based on the terms of use and the content of consideration of data provision plans to date, participate in Data Spaces if necessary, depending on the status of provision of services, tools, functions, etc. that are necessary for the organization. In addition, since there are cases where templates for contractual content and terms of use are prepared for participating Data Spaces, etc., consider use after checking in advance.

11.2. Application development for data provision

It is preferable to avoid development to the extent possible by making use of the tools and functions provided. However, if it is judged that a minimum level of development is necessary, consider development for each phase of the data provision process, such as “collecting data,” “organizing and processing collected data,” “accumulating organized and processed data,” “extracting accumulated data,” and “providing extracted data.”

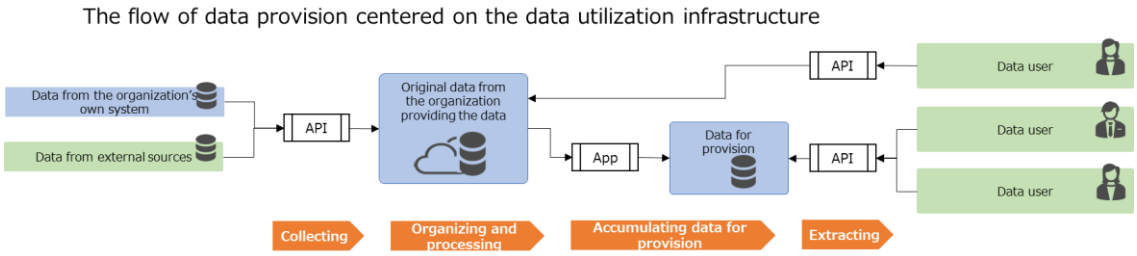

Figure 11.2 Flow of data utilization centered on the data utilization infrastructure

Next, we will look at examples of development required in each phase.

▼Examples of phases and development content associated with the provision of data itself

① Data collection

- Development of APIs for acquiring data from the organization's own system
- Development of APIs for collecting data from external data

② Data organization and processing

- Data cleansing: Correcting errors and omissions in data to improve data quality
- Data conversion: Converting data into a format that can be used as data to be provided
- Data integration: Integrating multiple data sources into one to create a unified data set
- Data anonymization: Masking information that can identify individuals in data to protect personal information

③ Accumulation of data to be provided

- Storing data that has been organized and processed in a specified location

For information on how to manage the data to be accumulated, see “11.1. Data planning and design, (3): Consideration of data management methods.”

④ Data extraction

- Extract the data necessary for providing data from the original data or data for provision
- Data filtering: Extract only the necessary data
- Data sampling: Extract a portion of the overall data
- Data normalization: Extract consistent data by reducing data variation
- Data augmentation: Create new data by processing existing data to increase the learning data for the model
- Data anonymization: If the original data is not anonymized in (2), and data including personal information is accumulated, anonymization is carried out in this phase

▼Development of other functions related to data provision

- Development of functions for creating data catalogs

  Development based on the Data Catalog Creation Guidelines[30], etc.

- Development of a contract management system for use between data providers and users

  Determine requirements based on the functional descriptions in Chapter 13, and develop functions

- Development of authentication and authorization functions for data provision

[30] DSA, “Data Catalog Creation Guidelines: V3.1”: https://data-society-alliance.org/wp-content/uploads/2023/03/230331-D97-DataCatalogGuidelineV31-gl-tecst.pdf

Determine requirements based on the functional descriptions in Chapter 14, and develop functions

• Development of a history management function for recording information such as data registration and processing

Determine requirements based on the functional descriptions in Chapter 15, and develop functions

11.3. Preparing the data to be provided

In this section, preparation is made for the data to be provided and for the publication of information about that data.

11.3.1. Generating the data to be provided

In “Application development for data provision,” the infrastructure for preparing the data to be provided is prepared, and the data to be provided is prepared using that platform.

11.3.2. Generating metadata

In the next chapter, we will publish information about the data we provide in a data catalog, but in many cases, some of the information published in the data catalog is based on metadata. This section explains what metadata is.

Note that it is not mandatory to include it in the data catalog, and companies can choose to publish it on their own website, etc. instead.

Metadata is “information about data.” [31]For example, in the case of a customer list, the following information would be metadata.

• Data name: “Customer List 2025”

• Unique ID: “customer_list_2025_v1”

• Creation date: “December 24, 2025”

• Creator: “Jyohou Tarou”

• Data format: “CSV”

The reason why metadata is being used is because it is difficult to utilize data if it cannot be easily determined what kind of data it is, even if the data exists.

There are various advantages to metadata, other than its use in catalogs for searching

[31]Reference: Information-technology Promotion Agency, Japan, “Practical Guidelines for Companies Promoting DX (DX Practical Guide IT System Construction Edition)” 4.4.2. Elements necessary for promoting data utilization: https://www.ipa.go.jp/digital/dx/hjuojm000000gx4n-att/000094497.pdf

data, and it plays an important role in efficiently promoting collaboration on data. Examples are given below.

① **Improved data searchability and visibility**
Benefit: It is easy to understand where the data is and what it means.
Example: You can quickly search for data sets such as customer lists through a data catalog.

② **Improving data quality**
Benefit: It is easier to maintain the accuracy, consistency and completeness of data.
Example: You can prevent incorrect input by defining in the metadata that the “purchase date” field should be recorded in date format.

③ **More efficient data management**
Benefit: You can standardize the process of saving, updating, and deleting data.
Example: By recording the retention period and archive scheduled date in the metadata, you can manage the data lifecycle efficiently.

④ **Improved data security and compliance**
Benefit: You can clearly manage who can access the data and how it should be used.
Example: Define access privileges using metadata to prevent unauthorized access and data leaks.

⑤ **Support for data integration and interoperability**
Benefit: It becomes easier to share and integrate data between different systems
Example: Including API endpoint information and data formats in metadata allows for smoother collaboration between systems

When writing metadata, it is desirable to conform to DCAT (Data Catalog Vocabulary)[32]. This is because DCAT is a set of rules for standardizing and describing data catalog information, developed by the W3C[33]. By writing metadata in accordance with DCAT, you can create unified descriptions even between different data catalogs. This makes it easier to share, use and search data between different organizations and systems, and supports the creation of new businesses and services. In recent years, services that automatically extract and generate metadata, such as descriptions of table and column item names, using technologies such as generative AI have also appeared, so it is likely that metadata can be created more efficiently than before by using these services.

---

[32] Data Catalog Vocabulary (DCAT) – Version 3, https://www.w3.org/TR/vocab-dcat-3/

[33] An abbreviation for World Wide Web Consortium. It is an international organization that creates web technology standards, and its aim is to make the web easy for everyone to use, safe, and compatible. For example, in addition to DCAT, it also creates web standards such as HTML and CSS.

Chapter 12: Data Search Phase

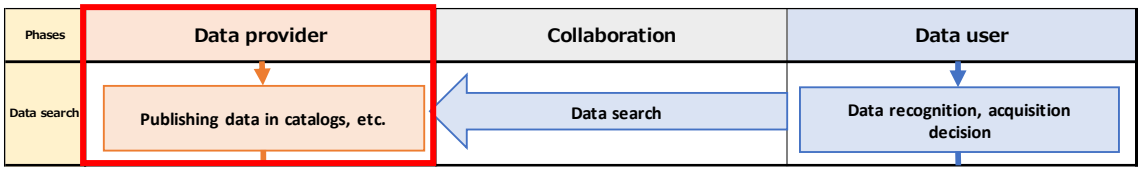

Figure 12.1 Process diagram for “Data Search Phase”

This chapter, “Data Search Phase,” explains the tasks on the provider side in the red box in Figure 12.1 of the overall process diagram.

In this phase, data providers publish the data to be provided in catalogues, etc. The information on the published data is searched for and recognized by data users.

12.1. Publishing data in data catalogs, etc.

Publish the data for provision prepared in Chapter 11 in data catalogs, etc. It is not mandatory to publish in data catalogs, and there are also other methods such as publishing on companies' own websites.

12.1.1. How to publish in catalogs

There are data catalog sites on the web, and many data are published there. In order to publish data catalogs on these sites, it is necessary to publish them in accordance with the rules of the site. In addition, the following are some of the key points to consider when creating data catalogs.

Table 12.1. Examples of key points to consider when publishing a data catalog

No.	Point	Content
1	Clearly state the content of the data and the assumed purpose of use	Clearly explain what data will be made public and how it is expected to be used. For example, include a statement such as “This data is a record of traffic volume in Bunkyo Ward, Tokyo in 2024, and can be used for traffic congestion analysis and improvements to public transportation,” etc.
2	Clearly indicate the data format	Specify in what format the data will be provided. For example, CSV, JSON, XML, etc.
3	Clearly state the metadata	Provides additional information (metadata) about the data set. For example, the creator of the data, the date it was created, the date it was last updated, etc.
4	Provision of sample data	Provide a portion of the actual data set or sample data that is not actual data to make it easier for users to

		understand the contents and format of the data. For example, using fictitious traffic volume data (date, time, location, number of vehicles, etc.) and providing it in CSV format, etc.
5	Data quality assurance	Assure that the data is accurate and reliable. For example, describe the method of data collection and frequency of updates, etc.
6	Clearly indicate the terms of use and license	Clearly indicate the terms of use and license for the data. This will make it clear to users how they can use the data. For example, clearly indicate that "this data can be used under a Creative Commons license," etc.
7	Assurances of privacy and security	Make it clear that the data to be released does not contain personal information and that security measures have been taken. For example, make it clear that "this data does not contain personal information and that security measures have been taken," etc.
8	Clearly indicate how to access and obtain the data	Clearly indicate how data can be accessed and obtained. For example, download links, API usage methods, and methods for obtaining data using Graph technology, etc.
9	Includes support and contact information	Provide contact information for questions or if there are any problems with the data. For example, include the address of the support center, etc.

For data users, the benefits of a data catalog include the ability to use it as a tool to determine which data is useful and easy to use for their organization, thereby promoting data utilization. For providers, it enables them to build a system that can quickly provide the data they need. For providers, it enables them to build a system that can quickly provide the data they need. It also allows them to centrally manage their data assets and gain a better understanding of the big picture. In addition, by making it easier for users to consider using the data, it can lead to not only the promotion of new services and product development using the published data, but also to an improvement in the brand image of the organization providing the data. For information on how to create a data catalog, please also refer to the [34]"Data Catalog Creation Guidelines"[35] published by the DSA.

[34] DSA is an abbreviation for Data Society Alliance, Japan. This is an organization that aims to promote the utilization of data and to realize a data-driven society.

[35] DSA, "Data Catalog Creation Guidelines": https://data-society-alliance.org/survey-research/datacatalogguideline/

Chapter 13: Contract Phase

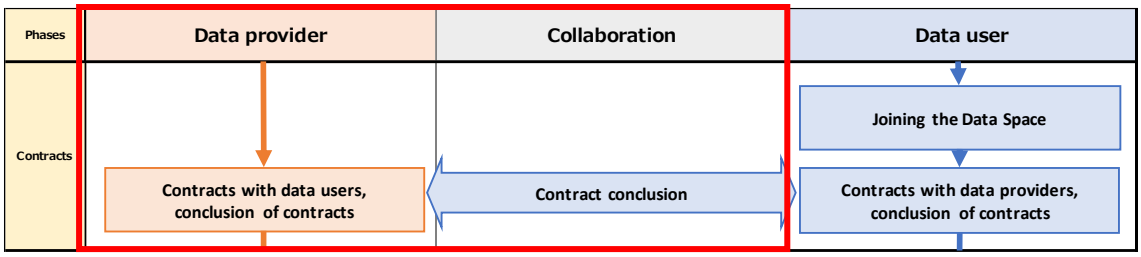

Figure 13.1 Process diagram for “Contract Phase”

This chapter, “Contract Phase,” explains the tasks on the provider side in the red box in Figure 13.1 of the overall process diagram.

In this phase, data users participate in the Data Space and conclude a contract with data providers regarding data use.

13.1. Contracts with data users, conclusion of contracts

In some cases, data providers enter into contracts with data users in order to prevent the data from being used by unintended users. In other cases, if they want the data to be used openly, they do not enter into contracts.

Depending on the digital infrastructure, etc., there are cases where contract management functions are provided that register the contract conditions during contract negotiations and when the contract is concluded, and then use that information as authorization information when providing the data.

Examples of contract management functions provided by digital infrastructure, etc., are shown below.

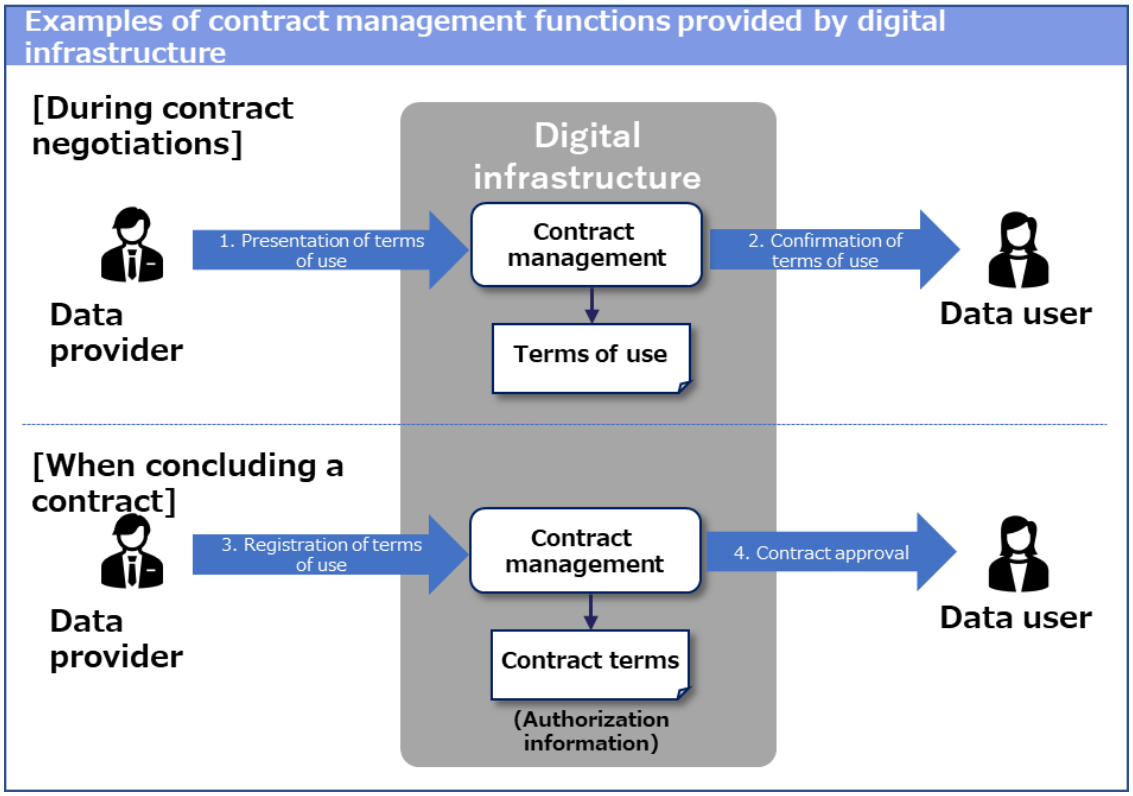

Figure 13.2 Contract management process using digital infrastructure[36]

1. Presentation of terms of use

   The data provider presents the terms of use for the data to be provided to the data user.

2. Confirmation of terms of use

   The data user confirms the terms of use in 1. Then, if necessary, the data user negotiates the contract details with the data provider.

3. Registration of terms of use

   The data provider registers the contract details in the contract management function based on the terms of use agreed in 2.

4. Contract approval

   The data user checks and approves the contract details.

For other points to note regarding contracts, please refer to “Chapter 10, 10.1.2. Determining the terms of use.”

[36]Created by the IPA with reference to “Introduction to Data Spaces” by the Information-technology Promotion Agency, Japan: https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/dataspaces-gb.pdf

Chapter 14: Development Phase of Applications that Use Data

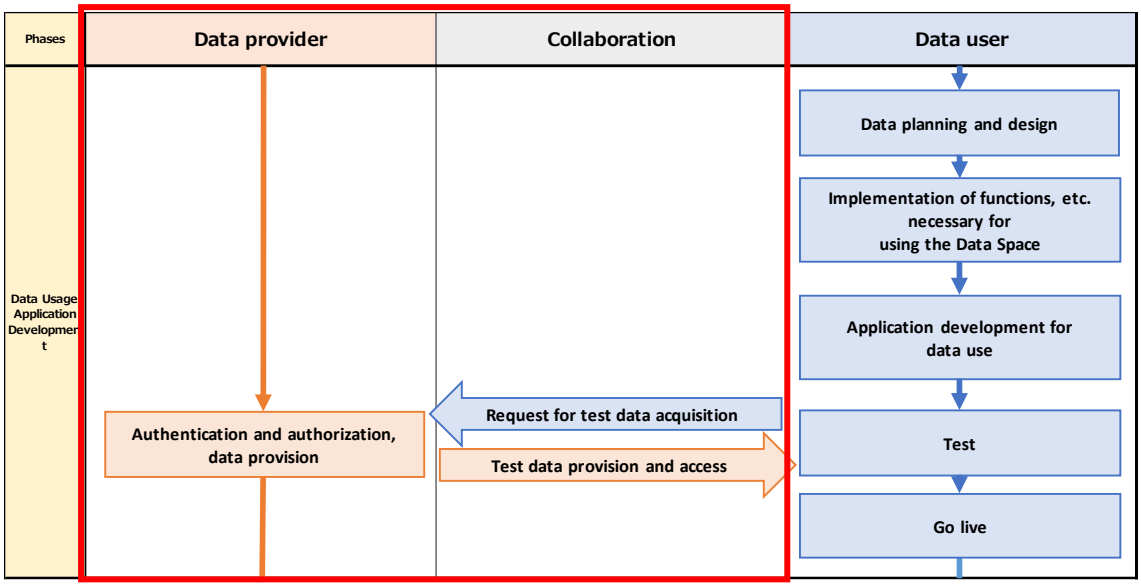

Figure 14.1 Process diagram for “Development Phase of Applications that Use Data”

This chapter, “Development Phase of Applications that Use Data,” explains the tasks on the provider side in the red box in Figure 14.1 of the overall process diagram.

In this phase, actual data is provided for testing the applications developed by data users. This is to deal with cases where provision is required by contract, etc.

14.1. Implementation of functions, etc. necessary for using the Data Space

In some cases, functions and services are provided for data providers from the participating Data Spaces, etc. The explanation for this section is the same as the content described in “Part 1, Chapter 6, 6.2. Implementation of functions, etc. necessary for using the Data Space,” so please refer to that for details.

14.2. Authentication, authorization, data provision

Data providers grant users access to data by authenticating and authorizing them, and then providing the data.

14.2.1. Authentication

When providing data, it is necessary to confirm that the other party is who they say they

are and that they meet the terms of the data usage agreement. This is done by cross-checking the information of previously registered contract users before providing the data, to prove that the other party is a contracted user. The authentication method shown here is just one example, and it explains the most basic case of user authentication. (In this case, the user does not necessarily represent an individual, but may also represent an organization.)Other examples include endpoint authentication and application authentication.

There are several authentication patterns, but we will introduce an example that uses the authentication function provided by the digital infrastructure[37] (similar to the figure in "Part 1, Chapter 7, 7.1.1. Authentication, data acquisition request").

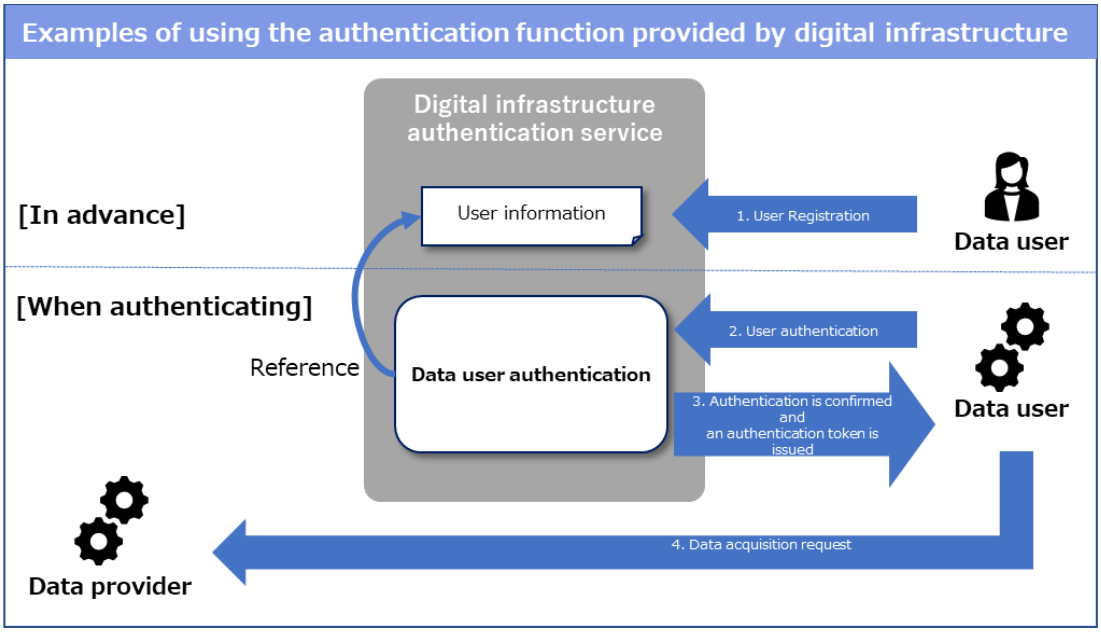

Figure 14.2 Authentication process using digital infrastructure[38]

At the time of "4. Data acquisition request," the data provider will detect that authentication has been completed and will receive a request for data acquisition from the user.

In steps 1, 2, and 3, the content of each step is basically carried out between the authentication service and the user based on the user information pre-registered with the authentication service, so in this case, the provider itself is not particularly involved.

14.2.2. Authorization, test data provision and access

When receiving a data acquisition request from a data user and the provider gives authorization, the user will be granted access to the data. As introduced in "Chapter 7, Section 7.1.2. Authorization," some digital infrastructures provide an authorization function to

[37] Reference: Information-technology Promotion Agency, Japan, Description of the digital infrastructure in "Introduction to Data Spaces": https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/dataspaces-gb.pdf
[38] Created by the IPA with reference to "Introduction to Data Spaces" by the Information-technology Promotion Agency, Japan: https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/dataspaces-gb.pdf

enable automatic authorization. In this case, the user will be granted access to the data in accordance with the terms and conditions registered with the digital infrastructure when the contract was concluded. If the function is provided by the digital infrastructure at the stage of “Chapter 11, 11.2. Application development for data provision,” the authorization function is incorporated into the company's own system, and if it is not provided, the authorization function is developed on its own and incorporated.

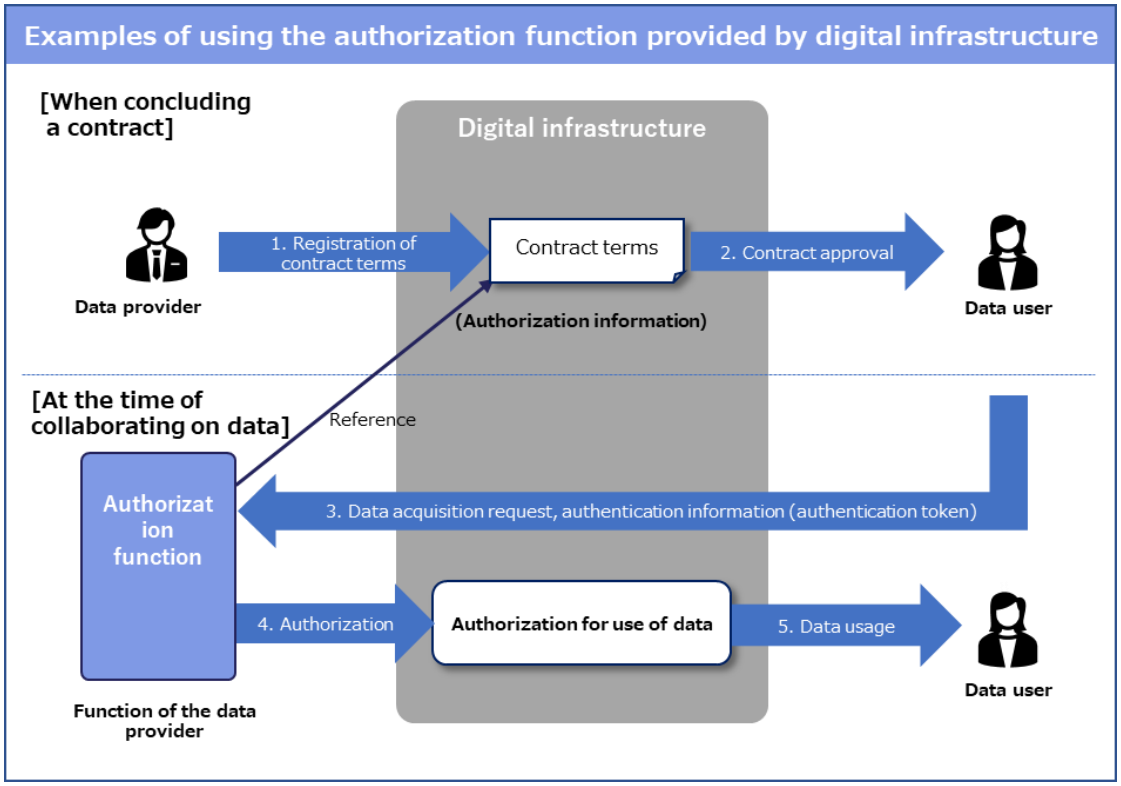

Figure 14.3 Authorization process using digital infrastructure[39]

The authorization function receives a request for data acquisition from the user and, in accordance with the terms of the contract (authorization information) agreed in advance, if there are no problems, the authorization function automatically grants access to the data.

[39]Created by the IPA with reference to “Introduction to Data Spaces” by the Information-technology Promotion Agency, Japan: https://www.ipa.go.jp/digital/data/jod03a000000a82y-att/dataspaces-gb.pdf

Chapter 15: Operation Phase

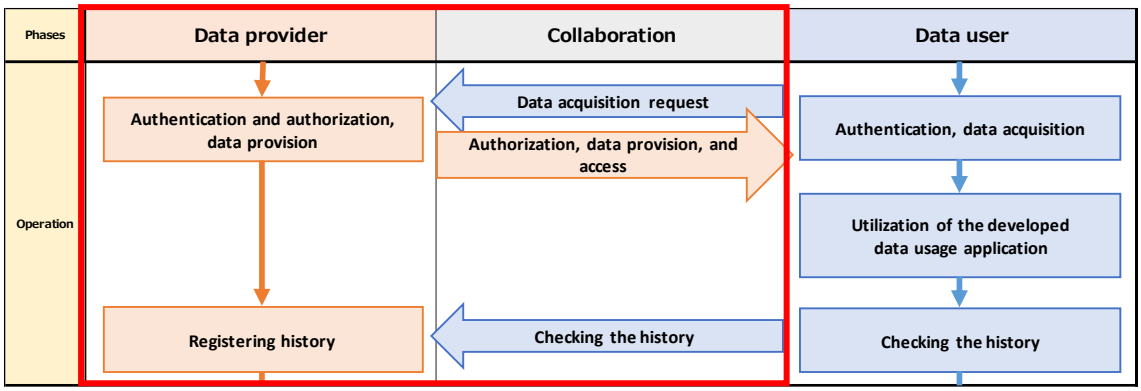

Figure 15.1 Process Diagram for “Operation Phase”

This chapter, “Operation Phase,” explains the tasks on the provider side in the red box in Figure 15.1 of the overall process diagram.

In this phase, data is provided to data users after verifying their identity and checking it against the terms of the contract (authentication and authorization), and the history is recorded and managed.

In addition, during operation, data will be handled appropriately in accordance with data governance and other rules set by the CDO. If there are any inadequacies in the data or issues with its use, feedback will be provided to the department managing the data for improvement.

15.1. Authentication, authorization, data provision

The process here is the same as that described in “Chapter 14: Development Phase of Applications that Use Data,” so please refer to that.

15.2. Registration of history

If history management is provided by a digital infrastructure, etc., by making use of this, the data provider can record who registered, processed, or acquired the data, when, and can guarantee that the data has not been tampered with inappropriately.

This ensures the reliability of the data, and from the user's perspective, they can use the

data with confidence. Although the basic content is explained here, there are different types of histories, such as data conversion history, data modification history, and data usage history. They are also sometimes used in combination with the time stamp function and version management provided by digital infrastructure.

The following is an example of using the history management and confirmation functions provided by digital infrastructure, etc. (same as the figure in “Chapter 7: Operation Phase, 7.3. History confirmation”).

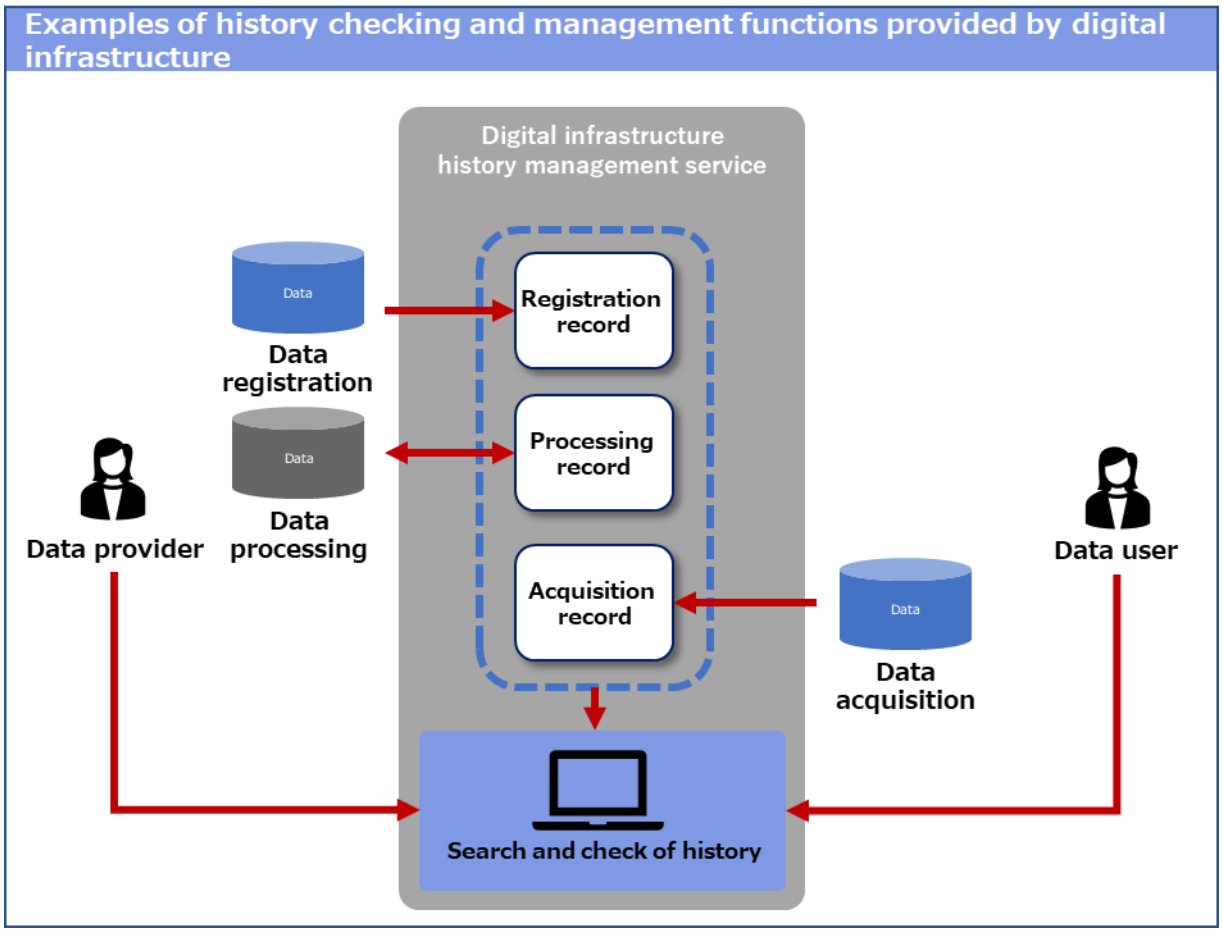

Figure 15.2 Overview of history management and confirmation provided by the digital infrastructure

Data providers can register history when registering and processing data.

Both data providers and data users can search and confirm the recorded history.

Chapter 16: Evaluation Phase

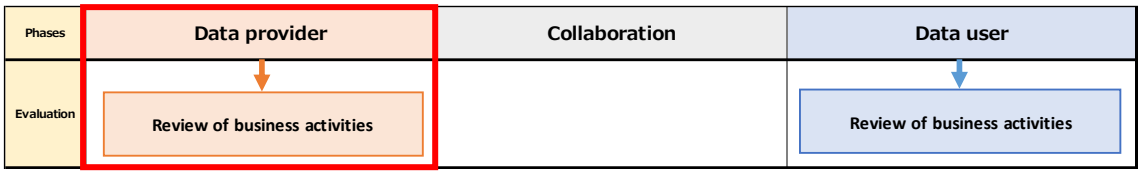

Figure 16.1 Process Diagram for “Evaluation Phase”

This chapter, “Evaluation Phase,” is a phase in which the results of the business activities and data analysis actually carried out according to the data strategy in Chapter 15: Operation Phase, are reviewed, and put improvements from this are put into practice.

16.1. Review of business activities

Basically, the same as the content described in “Part 1, Chapter 8, 8.1. Review of business activities,” in order to improve business activities and results through data provision, evaluate business activities from the perspective of the CDO, and if the Data Space is being used, include that perspective in the evaluation.

For perspectives based on the characteristics of the data space, please refer to “Chapter 9: Business Strategy Development Phase.” The following are some examples from the perspective of the CDO's role as a data provider.

▼Examples of evaluation criteria as the role of the CDO on the data provider

① **Achievement of data strategy**

Check the extent to which the data strategy formulated contributes to the company’s business goals (vision). The provider measures the revenue, customer satisfaction, etc. that have been set as KPIs (key performance indicators) and evaluates the points of success and shortcomings. In addition, by evaluating the extent to which the data provided has been utilized by the user (promotion of innovation, creation of new products and services, etc.) and the extent to which it has contributed to this, it can be used as material for reviewing the data provided.

② **Data governance effectiveness**

See “Data users.”

③ **Data quality status**

See “Data users.”

④ **Data provision status**

Evaluate how many users used the data prepared for provision, etc. Try to improve the data by obtaining feedback through means such as questionnaires when providing the data.

⑤ **Security and risk management status**

See “Data users.”

⑥ **Extent to which data culture permeates the entire organization**

Evaluate the extent to which a culture of promoting data provision is permeating the entire organization, rather than just specific departments.

⑦ **Effectiveness of data infrastructure**

Evaluate whether the development of the data provision infrastructure is progressing as planned, and whether the performance of the infrastructure is appropriate.

⑧ **Status of restrictions on data usage purposes**

Check how the data provided is being used, and evaluate whether it is being used for purposes other than those intended.

⑨ **Status of privacy management**

Evaluate whether the data provided is being anonymized appropriately and used appropriately while protecting personal information.

⑩ **Status of contract details**

Evaluate whether the scope of data usage, etc., in the content of the contract regarding data provision is appropriate.

In this way, by continuously understanding the differences between plans and actual results based on data strategies from the perspective of the Data Space and the CDO, and responding appropriately, as well as flexibly responding to changes in the world and new needs, etc., we can approach the vision set out by the management and realize the growth and success of the organization.

In Part 2, we comprehensively explained everything from upstream management strategy formulation to downstream operation and evaluation, from the perspective of the data provider. Further development of data utilization is often achieved not only through the data owned by the organization itself, but also through the various data provided by other organizations. In such cases, the role of the data provider goes beyond simply supplying the data itself, and can become a source of social change and new value creation.

We sincerely hope that many organizations will understand the potential of data through

this guidebook and actively promote the provision and use of data, and we will conclude with these words.

Data Utilization and Data Space Guidebook

[Copyright and production] Digital Infrastructure Center, Information-technology Promotion Agency, Japan (IPA)

[Editorial responsibility] Kenji HIRAMOTO, Digital Infrastructure Center

[Supervision] Mirai ODAGIRI and Masaru DOBASHI, Digital Engineering Department
Kimitaka TAMURA, Digital Architecture Design Center

[Author] Yasushi NISHIMOTO and Fumiyasu TSUCHIYA, Digital Engineering Department

[Revision history] November 13, 2024: 1st Edition published
January 29, 2025: 2nd Edition published
Part 2: “Data Providers” added.

[Publisher] Data Spaces Group,
Digital Engineering Department, Digital Infrastructure Center,
Information-technology Promotion Agency, Japan (IPA)
Bunkyo Green Court Center Office,
2-28-8 Honkomagome, Bunkyo-ku, Tokyo, Japan
113-6591
URL : https://www.ipa.go.jp/

Information-technology
Promotion
Agency, Japan