Open Data Guide

－Rules and Technical Guide for Open Data－

Version 1

4, July, 2014

Open Data Promotion Consortium

Table of Contents

Part I Getting Started: Open Data

Chapter 1. Introduction

1.1 Purpose of This Guide

Since the formulation of the Open Government Data Strategy on April 7, 2012 by the Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society (IT Strategic Headquarters), there has been growing momentum for open data among the national and local governments, incorporated administrative agencies, public corporations (including utilities and railways), and other entities. Expectations are high for these entities to disclose their public data to citizens in order to enable information users to exploit such data for development of applications and other activities, as well as contributing to revitalization of the economy and enhancement of the transparency of government.

Previously, governments and other public agencies permitted access to information owned by them in accordance with the Act on Access to Information Owned by Administrative Organs (1999 Law no.42), the Act on Access to Information Owned by Incorporated Administrative Agencies, etc. (2001 Law No.140), and an information public disclosure system under relevant regulations of each local government. An information public disclosure system, however, is intended to explain the activities of central and local governments or other public entities to citizens and achieve a fair democratic government practice with the understanding and criticism of the citizens. Those laws and ordinances do not set for a provision concerning the use of disclosed information

By comparison, an open data system aims for the national and local governments, incorporated administrative agencies, public corporations, and other entities to disclose their data to the public in order to "enhance the transparency and reliability of management," "promote public participation in government and in government-private coordination," and "revitalize the economy and streamline management," or for other purposes especially in business. Open data are edited, processed, modified, and other otherwise handled on a machine (computer). For this reason, it is important to permit secondary use in accordance with an established rule for use and provide open data in a format easy for use (machine-readable format).

This guide is designed for reference by national and local governments, incorporated administrative agencies, public corporations, and other entities to disclose their public data as open data. Compiled by Open Data Promotion Consortium (Data Governance Committee and Technical Committee), this Guide discusses matters to be considered in creating, reshaping, or disclosing open data to the public, in two perspectives—"rules on use" and "technical matters."

1.2 Target Readers of This Guide

This guide is targeted at persons who intend to disclose to the public data currently held or to be created in the future by them. Such persons will mainly include those at central and local governments and incorporated administrative agencies. The guide may also be useful for reference by public corporations and other private entities

Open data are crated and disclosed to the public by information providers and accessed, edited, modified, or otherwise processed by information users (Figure 1-1). Part I and Part II are targeted at persons engaged in the data creation phase through the public disclosure phase. On the other hand, Part III is aimed at those who create or modify their data into machine-readable data.

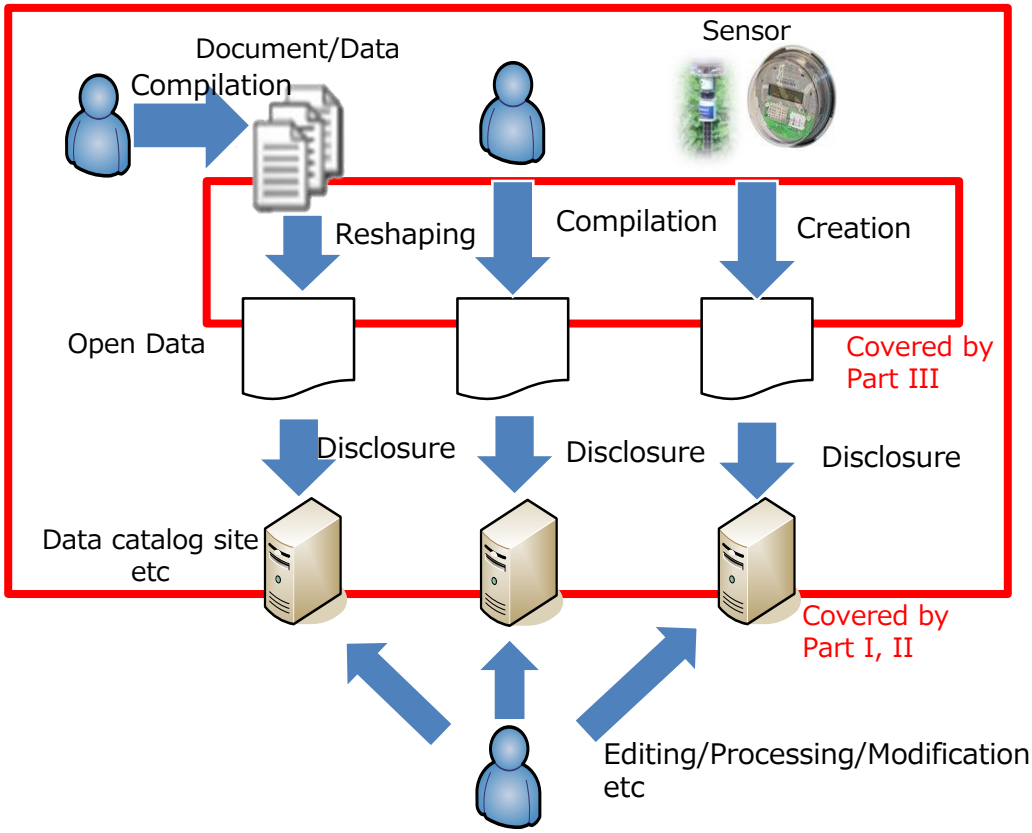


Figure 1-1 Flow of open data through public disclosure

1.3 Composition of This Guide

This guide is composed as shown in the flow table.

Table 1-1 Content of This Guide

Composition	Description
Part I Getting Started: Starting with Open Data	Part I explains main trends of open data at governments in Japan and foreign countries and considers the significance of open data for the purpose of understanding the background of the open data concept. ● Chapter 1 describes the purpose, target readers, and composition of this Guide, and defines terms used in this Guide. ● Chapter 2 explains trends of open data in Japan and foreign countries, and discusses the significance of open data. ● Chapter 3 explains a procedure in six steps for creation and public disclosure of open data.
Part II Rules on Use : Setting Rules on Use of Data	Part II explains rules on use of open data. ● Chapter 4 explains the importance of rules for using open data, and discusses international and domestic trends of rules on use, including a rule set up by the Japanese government. ● Chapter 5 describes three rules on use－CC-BY and CC0, which are being adopted by an increasing number of foreign countries and the Government Standard Terms of Use (Version 1.0) to be adopted by the Japanese government. ● Chapter 6 compares CC-BY, CC0, and Japan's Government Standard Terms of Use (Version 1.0) from a perspective of information users and information providers. In addition, we discuss which rule of use would be the most suitable for public disclosure of open data. ● Chapter 7 discusses the future direction of the Government Standard Terms of Use (Version 1.0)
Part III Technical Matters: Making Machine-readable Data	Part III explains technical matters to be considered in creating machine-readable data. ● Chapter 8 discusses technical matters related to the creation and public disclosure procedures for open data as discussed in Chapter 3. ● Chapter 9 provides a technical guideline for creating machine-readable data. The discussion focuses on identifiers, file formats, and data with respect to tabular data, document data, geospatial data, and real-time data.

Composition	Description
Appendix	Appendix explains standards and tool to be referred to in creation, editing, and public disclosure of data. ● Chapter 10 provides standards and tools that can be used for reference in creating and editing machine-readable open data. ● Chapter 11 outlines CKAN, a data catalog system, and explains how to use it.

In Table 1-2, each main topic in the left column corresponds to each chapter in the right column.

Table 1-2 Topics and Related Chapters

Topics	Chapter
1. Definition, background, and significance of open data	Chapter 2
2. Trends of open data in Japan and in foreign countries	Chapter 2
3. Process through the creation of open data, including organizational system, arrangement, and planning	Chapter 3
4. Background and principles of rules on secondary use of open data	Chapter 4
5. Specific types of rules on use and their characteristics	Chapter 5
6. Perspectives in applying an appropriate rule on use and evaluation of rules on use based on the perspective.	Chapter 6
7. Rule on use to be applied to specific open data	Chapter 6
8. Possibility of the government's future review of its rule on use	Chapter 7
9. Data format, identifier system, and data transmission protocol useful in enhancing machine-readability	Chapter 8 Chapter 10
10. Technical levels required to create and edit open data	Chapter 8
11. Technical matters concerning creation and editing as open data of tabular data, document data, geospatial data, and other data	Chapter 9
12. Tools useful for creation, editing, and public disclosure of open data, including Web services and GIS tools	Chapter 10
13. CKAN and its use as a representative data catalog system	Chapter 11

1.4 Definition of Terms

Key terms used in this Guide are defined as shown in the following table.

Table 1-3 Definition of Key Terms

Term	Definition
Data	means information that can be used as open data, irrespective of whether such information is copyrighted or not.
Public data	means data owned by the national government, local governments, incorporated administrative agency, public corporations, or other public organizations.
Content	means the same as data. This guide does not use this term except when making a citation.
Open data	means data in a machine-readable format made open under a rule on use permitting secondary use, including commercial purpose. For details, see Section 2.3.
Information provider	means an individual or organization that provides data as open data.
Information user	means an individual or organization that makes secondary use of open data.
Secondary use	means editing, processing, modification, or other use by information users of data disclosed by information providers to create new data. It also means simple reproduction or redistribution of data provided by information provider without changing them.
Mashup	means creation of data by information users by combining their own data and/or data provided by information providers.
License	means a rule on use specified by information providers when distributing their data. This is one form of agreement between information providers and information users. This term is not used in this Guide except where it is used for citation or as a proper noun.
Rule on use	means a rule on use specified by information providers when distributing data, including contracts not based on copyrights and a declaration from an information provider.
Falsification	means claiming falsely that certain data that have been modified are original data.
Machine reading	means automatic editing, processing, modification, etc. of values on a structure (e.g. figures and text in a table) by a computer program that can interpret the logical structure of data.

Term	Definition
Machine-readability	means that target data can be read with a machine.
Metadata	means data that give information on the properties of data disclosed to the public.
Data catalog	means a list, catalog, or index of metadata describing the location, type, name, and other properties of data disclosed to the public.
Tabular data	means data in columns arranged vertically and horizontally in a two-dimensional form.
Document data	means data composed mainly of one-dimensionally arranged characters, partially including figures and tables, and intended to act on humans who read such data.
Geospatial data	means Data with a spatial component. For example, data connected with geographic location in a 2D map or sphere
Real-time data	means data in which values change with time.
Vocabulary	means a set of meanings and definitions of data in attribute or type used to describe matters and data in a certain area.

Chapter 2. Trends and Significance of Open Data

This chapter explains main trends of open data among the Japanese government, local governments, and foreign countries for better understanding of the background of open data, and discusses the significance of open data.

2.1 Major Trends of Open Data

2.1.1 Trends of open data at the Japanese government

With an open government initiative already in place, the launching on July 4, 2012, of the Open Government Data Strategy by the IT Strategic Headquarters has been prompting the Japanese government to proceed with its open-data projects (Table 2-1).

On June 14, 2013 the cabinet decided to set up the Japan Reconstruction Strategy and issue the Declaration to be the World's Most Advanced IT Nation, to demonstrate that open data is one of the most essential government policies.

Table 2-1 Major Trends of Open Data at the Japanese Government

Date	Names/URL	Led by
2009.10.14	Electronic METI Idea Box made open	Ministry of Economy, Trade and Industry
2010.07.29	Open Government Labo made open http: //www.openlabs.go.jp/	Ministry of Economy, Trade and Industry
2011.03.15 to date	TEPCO's planned power shortage and electricity data made open	Tokyo Electric Power Company (TEPCO)
2011.07.01	"Data Box" made open http: //databox.openlabs.go.jp/	Ministry of Economy, Trade and Industry
2012.01.17	Recovery and Reconstruction Support Program Database (Open Government) launched https: //www.r-assistance.go.jp/	Cabinet Secretariat; Reconstruction Agency; Ministry of Economy, Trade and Industry
2012.07.04	Government Open Data Strategy http: //www.kantei.go.jp/jp/singi/it2/denshigyousei.html	Decision at the IT Strategic Headquarters
2012.07.27	Open Data Promotion Consortium established http: //www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000047.html	Open Data Promotion Consortium
2012.09 to date	Open data demonstration experiment (development of information distribution cooperation infrastructure, etc.) http: //www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/opendata03.html	Ministry of Internal Affairs and Communications
2013.01.18	"Open DATA METI"(βVersion) public disclosure http: //datameti.go.jp/	Ministry of Economy, Trade and Industry

Date	Names/URL	Led by
2013.03.28	Government Open Data Experts' Meeting set up http: //www.kantei.go.jp/jp/singi/it2/densi/	Decision of Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society (IT Strategic Headquarters) (operating under Planning Committee from 2012.11.30 to 2013.03.27)
2013.04.19	**White Paper** on **Information** and **Communications** in Japan" **and "Information** & **Communications Statistics Database**" made open http: //www.soumu.go.jp/johotsusintokei/open.html	Ministry of Internal Affairs and Communications
From 2013.06.10 in gradual stages	Development of statistical open data (including provision of API functions and reinforcement of statistical GIS functions) http: //www.soumu.go.jp/menu_news/s-news/01toukei01_02000024.html	Ministry of Internal Affairs and Communications Statistics Bureau
2013.06.14	Japan Reconstruction Strategy (Disclosure of public data to the public and construction of and Innovative e-Government Services) http: //www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf	Cabinet decision
2013.06.14	Declaration to be the World's Most Advanced IT Nation (Promotion of open data and big data) http: //www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf	Cabinet decision
2013.06.14	Roadmap for promotion of e-government open data http: //www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf	Decision at IT Strategic Headquarters
2013.06.18	Open Data Charter (Original) https: //www.gov.uk/government/publications/open-data-charter (Translation) http: //www.mofa.go.jp/mofaj/gaiko/page23_000044.html	Agreement at G8 summit meeting (at Lough Erne, Britain)
2013.06.25	Fundamental principles on disclosure of public data for promoting secondary use (guideline) http: //www.kantei.go.jp/jp/singi/it2/CIO/dai52/kihon.pdf	Decision at a liaison meeting of information experts (CIO) from ministries and agencies
2013.10.29	Japan's action plan in response to Open Data Charter http: //www.kantei.go.jp/jp/singi/it2/CIO/dai53/plan_jp.pdf	Decision at a liaison meeting of information experts (CIO) from ministries and agencies
2013.12.20	Trial version of government data catalog site "DATA.GO.JP" opened http: //data.go.jp/	Cabinet Secretariat
2013.04.25	Action Plan for Construction of an Open Environment in the e-Government Fields http: //www.kantei.go.jp/jp/singi/it2/CIO/dai56/seibi2.pdf	Decision at a liaison meeting of information experts (CIO) from ministries and agencies

2.1.2 Open Data Initiatives at Local Governments

Prior to the July 2012 introduction of the Government Open Data Strategy, some local governments already initiated open data activities. Their activities have been gathering momentum since the introduction. Many of them disclose their data to the public on their own data portal or by other means. These include Fukui City, which makes its entire homepage open, and Fukui Prefecture, which applies

a unified data format to all municipalities within the prefecture Table 2-2).

Table 2-2 Main Open Data Initiatives at Local Governments

Local government	Name of initiative (URL)	Outline
Fukui Pref.; Sabae City	Data City Sabae http: //www.city.sabae.fukui.jp/pageview.html?id=12765	A pioneering initiative by a local government for open data. Cooperation with local enterprises to develop various applications. In 2013, open data activities were extended in cooperation the Ministry of Internal Affairs and Communications in an open data demonstration experiment.
Chiba Pref., Nagareyama City	Nagareyama City open data Trial http: //www.city.nagareyama.chiba.jp/10763/ Nagareyama City Assembly Open Data Trial http: //www.nagareyamagikai.jp/opendata/	The municipal government and assembly joined effort in line with a website renewal. Votes at each assembly resolution are made public by assembly member.
Kanagawa Pref., Yokohama City	Yokohama Open Data Portal http: //data.yokohamaopendata.jp/	Information of libraries, etc. has been provided to private groups since FY2012. In FY2013, an open date project was established within the municipal government. Municipal data have been made open in cooperation with Ministry of Internal Affairs and Communications in an open data demonstration experiment.
Shizuoka Pref.	Fujinokuni Open Data Catalog http: //www.pref.shizuoka.jp/kikaku/ki-330/opendata/	The first municipality to start a data portal. The portal can be used by other municipalities within the prefecture (used by Susono City).
Shizuoka Pref.; Yamanashi Pref.	Fugaku 3776 Kei http: //fugaku3776.okfn.jp/	Anyone may contribute photographs of Mt. Fuji together with location information. Contributed photographs are disclosed to the public as open data. In a disaster (heavy snow), the portal was used for sharing disaster information.

Local government	Name of initiative (URL)	Outline
Fukui Pref.	Open Data Library http: //www.pref.fukui.lg.jp/doc/toukei-jouhou/opendata/ Format unification of public data at municipalities http: //www.fukuishimbun.co.jp/localnews/politics/46384.html	Open data and applications using open data are disclosed to the public. An initiative is in progress toward unification of data formats used by municipalities within the prefecture.
Fukui City	City Homepage Standard Terms of Use revised http: //www.city.fukui.lg.jp/sisei/kohou/hp/site-p.html	CC-BY-SA license is applied to the city's entire homepage.
Aomori Pref.	Aomori Image Material library http: //amcp-aomori.jp/	Image materials produced by prefectural government personnel are made public as open data.
Fukushima Pref. Aizu-wakamatsu City	Public disclosure of data under an open data license http: //www.city.aizuwakamatsu.fukushima.jp/docs/2009122400048/	In addition to an open license (CC-BY), an open document format (ODF) is used to make public data open.
Kanagawa Pref. Yokosuka City	Public disclosure of open data for disaster prevention information http: //yokosuka-opendata.ubin.jp/	Disaster-related data disclosed to the public by Yokosuka City are converted into a machine-readable format for public disclosure. Ideathon and Hackathon events were held using the data.

2.1.3 International Trends

Open data initiatives were started for the first time in foreign countries, mostly in the U.S. and Europe, during the latter half of the 2000s.

In the U.S., President Barak Obama disclosed Memorandum in Transparency and Open Government on January 21, 2009, directly following his inauguration day, to clarify three principles of open government－transparency, private participation, and public-private cooperation. On May 21, 2009, a portal site "Data.gov" was launched as a core project in connection with the memorandum. The presidential memorandum "Building a 21st Century Digital Government," released on May 23, 2012, provides that the federal government shall make open, in principle, all data owned by it by

permitting access with any sorts of device at any time and by providing an API for all data. The memorandum also required main government agencies to provide a special page for developers on their homepage.

In EU, in response to the 2003 Europe Directive on the Re-use of Public Sector Information (PSI), member countries launched open data initiatives. Especially active in its effort, the British government enforced a PSI reuse regulation in July 2005. Britain's open data promotion system was established in October 2006, and the British Office of PSI was merged with The National Archives. In 2007, the Power of Information Task Force was set up. Its initiatives gave rise to open data for initiatives in the U.S. and other countries. In line with these moves, discussions were carried out in Japan in 2008 at a workshop "Government CIO Forum" by CIOs from the national, prefectural, and municipal governments. The workshop proposed a recommendation for "making government information open." In 2009, the UK government opened a data portal site "data.gov.uk," and in May 2010, Prime Minister Cameron announced “Transparency Agenda" to set forth three principles on open data: disclosure to the public of reusable and machine-readable data, adoption of the same open license for commercial use, and data availability through a single online access point. In September 2010, Britain introduced an open government license. Meanwhile, the Dutch government has also been strenuously dealing with open data issues. In March 2010, the Dutch government site was the first to adopt Creative Commons **CC0 (Public Domain Declaration) to place part of the** government data in the public domain (for details, see 4.2)[1]. A number of other European countries, including France, Germany, and Italy, are advancing open data policies.

Responding to these moves in European countries, European Committee announced the European Open Data Strategy in 2011 together with plans for establishment of an EU data portal and other actions, while carefully watching economic effects of open data.

At a G8 summit meeting of June 8, 2013, the G8 leaders agreed upon Open Data Charter. Setting forth five open data principles (Table 2-4), the summit meeting required G8 countries to formulate an action plan by the end of October 2013 in compliance with the Open Data Charter and report their performance status in October 2014 and in 2015. In response to these principles, Japan announced the Japan Open Data Charter Action Plan on October 29, 2013[2].

Table 2-3 Main Trends of Open Data in Foreign Countries

Time	Initiative implemented	Country
2003	Directive for reuse of PSI (public owned data)	European Committee
July 2005	Rule for reuse of PSI	Britain

[1] http://wiki.creativecommons.org/Case_Studies/Netherlands_Government
[2] http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf

Time	Initiative implemented	Country
2007	Power of Information Task Force installed	Britain
January 2009	Memorandum on Transparency and Open Government	U.S.
May 2009	Data portal site Data.gov set up	U.S.
September 2009	Data portal site data.gov.uk set up	Britain
March 2010	CC0 adopted for the first time by the government site	Netherlands
May 2010	"Transparency Agenda" announced	Britain
November 2010	Cabinet decision for establishing Etalab	France
December 2011	Data portal site data.gouv.fr set up	France
December 2011	European Open Data Strategy	European Committee
June 2013	Open Data Charter	G8 countries

Table 2-4 Open Data Principles

1. Open Data by Default

   We will

- establish an expectation that all government data be published openly by default, as outlined in this Charter, while recognizing that there are legitimate reasons why some data cannot be released.

2. Quality and Quantity

   We will

- release high-quality open data that are timely, comprehensive, and accurate.
- ensure that information in the data is written in plain, clear language, so that it can be understood by all, though this Charter does not require translation into other languages;
- make sure that data are fully described, so that consumers have sufficient information to understand their strengths, weaknesses, analytical limitations, and
- release data as early as possible.

3. Usable by All

   We will

- release data in open formats wherever possible, ensuring that the data are available to the widest range of users for the widest range of purposes; and
- release as much data as possible.

4. Releasing Data for Improved Governance

   We will

- share technical expertise and experience with each other and with other countries across the world so that everyone can reap the benefits of open data; and
- be transparent about our own data collection, standards, and publishing processes.

5. Releasing Data for Innovation

   We will

- work to increase open data literacy and encourage people, such as developers of applications and civil society organizations that work in the field of open data promotion
- empower a future generation of data innovators.

Source: G8 summit meeting Open Data Charter[3]

[3] http://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf

2.2 Significance of Open Data

As discussed above, open data activities are gathering momentum in Japan and foreign countries. The Open Government Data Strategy (IT Strategic Headquarters decision of July 4, 2012) provided three significant purposes for open data－"Enhancement in Transparency and Confidence", "Promotion of Public Participation and Collaboration between the Public and Private Sectors," and "Economic Stimulus and Higher Efficiency in Government."[4]

Following this, the Liaison Conference of Chief Information Officers (CIOs) of Public Offices and Ministries released the Basic Stance on Public Release of Ministry Information to Encourage **Secondary Use** (Guidelines) (decision of June 25, 2013) to propose the following:[5]

a. Revitalization of economy and creation of new business

An open data system will enable automatic collection of machine-readable data and cross-sectional use of collected data by using a code. This will help reduce operating costs and provide new business services (for example, collection and analysis for commercial purpose of data concerning meteorological, geological, traffic, and other observations and survey findings).

b. Realization of public services through government-private coordination (including disaster prevention and mitigation)

By combining multiple data from government agencies and private businesses, it is possible to create and offer services useful in improving the living environment and in responding to disasters (for example, by offering information on childbearing, education, medical care, welfare, and other public services available in the neighborhood and the quality of such services in a way easily understandable by users or by distributing disaster information compiled from multiple data immediately in the event of a disaster).

c. Improvement in the transparency and reliability of the government

Changes in and characteristics and appropriateness of government policies can be better understood and evaluated by studying, totalizing, and comparing, in a cross-sectional manner, government policies and project plans, decision processes, and policy results (for example, subsidies and government expenditures can be analyzed by ministry or agency, field, region, recipient, etc.)

[4] http://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf
[5] http://www.kantei.go.jp/jp/singi/it2/densi/

Source: Basic Stance on Public Release of Ministry Information to Encourage **Secondary Use** (Guidelines) (decision of June 25, 2013 by the Liaison Conference of Chief Information Officers (CIOs) of Public Offices and Ministries)

2.3 Definition of Open Data in This Guide

In open data, "The Five Stars of Open Data"[6] is a document frequently referred to in foreign countries. This defines a five-step open data development scheme, as follows.

1. make your stuff available on the Web (whatever format) under an open license
2. make it available as structured data (e.g., Excel instead of image scan of a table)
3. use non-proprietary formats (e.g., CSV instead of Excel)
4. use URIs to denote things, so that people can point at your stuff
5. link your data to other data to provide context

The above scheme provides as the first step (★1) of open data that data should be provides under an open license, and provides as the second step (★2) that data should be made open in a structured way. At this step, compatible software can extract data from files in a compatible format. This means that data are machine-readable.

Open Data Handbook, edited by Britain's Open Knowledge Foundation[7], defines "open data" address data that can be used and reused freely by anyone. All that the handbook requires is that the author's credit should be maintained and that data should be distributed under the same terms and conditions.

Japan's Open Government Data Strategy sets forth four basic principles for promoting use public data: (i) the government should be aggressively engaged in public disclosure, (ii) public data should be made open in a machine-readable format, (iii) public data utilization should be encouraged whether for commercial or non-commercial purposes, and (iv) public data should be made open as soon as possible to accumulate disclosure results.

In response to this initiative, the IT Strategic Headquarters released **Roadmap for** the e-Government **Open Data Promotion** (June 14, 2013) to provide that "machine-readable data" should be disclosed to the public as "open data" under a rule that permits secondary use for "commercial and non-commercial purposes."

Under these circumstances, this Guide defines "open data" as follows in compliance with the Roadmap **for** the e-Government **Open Data Promotion:**

[6] Defines five steps to open data. (http://5stardata.info/ja/)
[7] What is open data? (Open Data Handbook) (http://opendatahandbook.org/ja/what-is-open-data/index.html)

"Open data" means "machine-readable data" disclosed to the public under "a rule that permits secondary use for commercial or non-commercial purposes."

Chapter 3. Procedures for Open Data Creation and Public Disclosure

This chapter provides a creation and public disclosure procedure for open data.

To create open data and disclose them to the public, it is desirable to follow the procedure as shown below. Following the small-start principle, one important way is to start with data that can be made into open data relatively easily, instead of waiting for all necessary arrangements to be completed, including the establishment of an open data promotion organization and the study of the current status. For example, departments and sections may start separately with information already disclosed to the public.

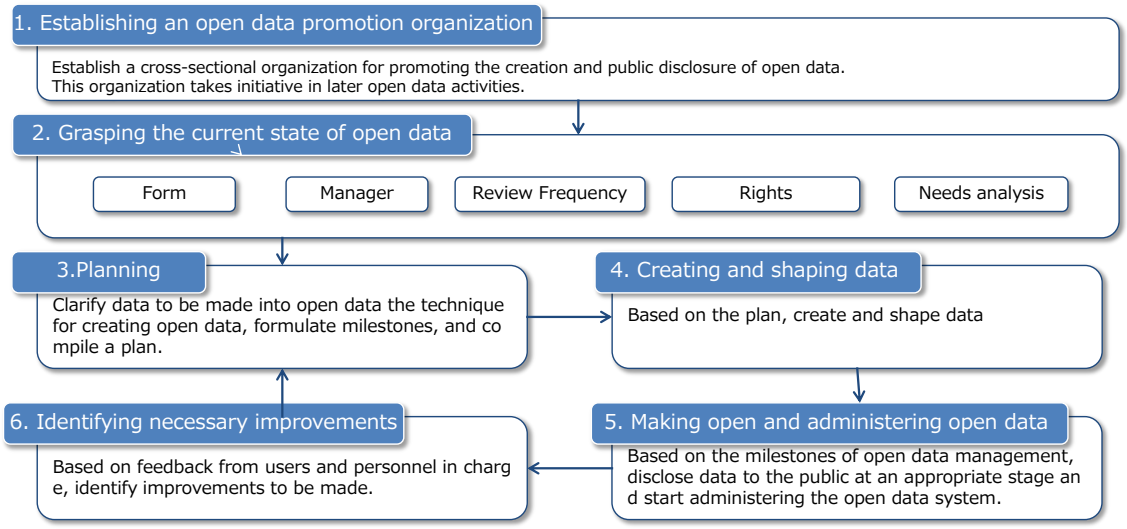


Figure 3-1 Recommended Procedures for Open Data Creation and Public Disclosure

The following paragraphs describe the implementation and review of the above steps:

3.1 Establishing an open data promotion organization

Work involving creation and public disclosure of open data needs to be shared among departments and sections owning relevant data. Cooperation and coordination must be formed among those organizations.

To proceed with creation and public disclosure of open data, it is desirable to set up a new unit to be tasked with promotion of open data, independently from those organizations.

3.2 Grasping the current state of open data

To understand the current status of open data, it is important to gather and join together data managed by each department from the following perspectives (it is desirable to conduct stocktaking of data):

- Departments in charge of data,
- Types of data (e.g. data on budgets, reports, statistics, or public relations)
- Amount of data
- Needs analysis

In grasping the current status of data, a focus should be placed on the following points:

① Data format

Confirm the format of each data.

- Paper (Confirm the existence of electronic data corresponding to the same information.)
  - In the absence of electronic data, relevant materials need scanning for public disclosure.

② Electronic data (Confirm the file format)

Identify the information management system at each department that is managing data.

- Is a manager in place?
- Are all data controlled by a single manager?

③ Data update frequency

Confirm how frequently data are updated.

- Once a year, once a month, as necessary, etc.

④ Relationships between data and rights

Confirm data, for example, in respect of the following. For details, see Part II.

- Whether the data is protected by a third-party copyright or other rights
- Presence of legal, regulatory, or other limitations

⑤ Needs analysis

There is a strong need for information and data if they are frequently referred to by users or if they are disclosed to the public by similar organizations. It is useful to start disclosing such highly needed open data.

3.3 Planning

In the planning stage, target open data should be specified, and the technique for creating and disclosing such data should be clarified, based on the current status or feedback. In such work, it is desirable to create milestones and formulate a schedule in accordance with the milestones. It should be noted that planning is more important for larger organizations.

In the planning stage, special attention should be directed to the following:

① Preliminary planning of a data format and a system

- Based on Section 8.4, establish a policy for determining the level of "data" and "data catalog."
- Based on Section 9.1, specify necessary metadata and an identifier system.

② Establishment of an operation rule

- Specify a procedure and frequency of collecting data from organizations controlling the data.
- If data are to be updated as necessary, establish a rule for data updating.
- Based on Section 8.5, establish a data registration policy.

③ Establishment of a rule on use

- Determine a rule on use of open data.
- If there are third party rights or legal or regulatory limitations involved, adjust the content of the rule on use or the scope of open data to such rights or limitations.

④ Small-start principle

- Proceed with work in stages, and set milestones so that data are disclosed sequentially from the data first completed.
- It is desirable to establish objectives and a plan on an annual basis.
- The Open Government Data Strategy adopts the small-start principle as it provides that public data that can be made into open data should be disclosed to the public sequentially without delay in order to consistently accumulate results.

3.4 Public Disclosure Work

In preparations for public disclosure, necessary tools should be procured and open data created or modified, in accordance with the established plan.

Target data should be registered on the data catalog system in compliance with the plan.

In public disclosure of data, special attention should be directed to the following:

- Information to be specified for public disclosure
- Effects of public data disclosure
- Matters to be considered concerning the data disclosure server

- Reliability of data
- Privacy and anonymization

The details of the above are given in the next paragraph.

3.4.1 Information to be Specified at Disclosure

With respect to individual data to be disclosed to the public, the following three points should be clarified:

① Metadata: What does "metadata" mean?

② Access to metadata: How can such data be obtained?

③ Rule on use: What rule is applied to acquisition and use of the data?

The above is explained in the following paragraph.

3.4.1.1 Metadata

For example, the government data catalog site "DATA.GO.JP" (trial version) provides the following items as metadata:

- Title
- Name of organization
- Disclosing party (department/section)
- Creator
- Update frequency
- Tag
- Release date
- URL
- File size
- Latest date of update
- Languages used
- Supplementary information

3.4.1.2 Access Method

An access method is a technique to collect data. For example, the method specifies an API[8] used to collect Web addresses (URL) and data.

When different data formats are used for different uses, the access method chosen is desired to be capable of collecting data in multiple formats. This case is discussed in Section 8.1.3.

[8] Short for Application Programming Interface, an API is a code setting forth procedures and data formats for using and managing data by retrieving them from other external programs.

3.4.1.3 Rule on Use

For example, the following matters should be considered in a rule on use. For details, see Part II.

- Is secondary use permitted?
- Is commercial use permitted?
- Are there any legal and regulatory limitations on use?
- Are any terms and conditions provided concerning sources and other matters for secondary use?

3.4.2 Effect of Disclosure

Open data disclosed to the public are disclosed to the world. For this reason, it is assumed that public disclosure of data generates inquires from overseas.

3.4.3 Matters to be Considered Concerning Data-disclosing Servers

Prior to public disclosure of any data, it is necessary to estimate the amount of access to the data. Unforeseeable amounts of access concentrating on a server beyond its capacity may lead to access failure for the data. In addition, real time data on a data catalog system may cause access failure because it occupys more space of server storage capacity.

More details of these problems are omitted in this document. Prior to public disclosure of data, it is desirable to discuss with the disclosure service provider.

3.4.4 Reliability of Data

In open data, it is assumed that secondary use will be permitted, and it is desirable to permit secondary use. However, data can be falsified by information users in the distribution process. Also, data can be edited or changed in a manner not intended by the information provider, without the specification of the credit of the information provider. These cases include an intentional change of names of place on a map and a change to the name of a region in statistical data.

The following explains two methods for information providers to claim the legitimacy of their data. The one is that data once disclosed should be kept in the public domain in compliance with “open data principle”. By specifying the source of their open data, information providers can prove that falsified data are not theirs (that the falsified data are different from the data disclosed by the information provider). The other is technical measures against data falsification, which is explained in detail at the end of this chapter.

3.4.5 Privacy and Anonymization

In public disclosure of data, it is necessary to ensure that the data does not contain any information

that identifies individuals. Where necessary, privacy should be protected by using an anonymization technique.

More details of this problem are omitted in this document. Prior to public disclosure of data, it is desirable to consult professionals or the department in change.

3.5 Making Open and Administering Open Data

Following the established milestones, data should be disclosed as open data as they are registered.

During operation of the open data system, it is desirable to set up a questionnaire page and an inquiry contact in order to collect feedback from information users.

3.6 Identifying Necessary Improvements

Feedback from information users and operational problems should be reviewed at regular intervals to identify improvements to be made. It is also important seek to find problems to be improved each time new data are disclosed.

In the next stage, a plan should be formulated for implementing necessary improvements. It is recommended to enhance the technical levels as specified in Section 8.4 and review the rule on use each time problems are found.

NOTE: Technical Measures against Data Falsification

The following are citations from Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use, Appendix"[9] concerning technical measures to deal with falsifications.

> *There is no mechanism contained in software to completely prevent data. A practical technical measure is to build a scheme for detecting data falsification and identifying the person who has falsified the data. This will enable users to acquire unfalsified data and reduce data falsification.*
>
> *Technical anti-falsification measures tend to reduce the ease of using data and increase calculation load in encryption processing. For this reason, it is appropriate to apply such measures only to data that really need one. Basically, it is desirable to handle falsification*

[9] http://www.kantei.go.jp/jp/singi/it2/densi/

by formulating a rule or by improving literacy.

(i) Falsification detection techniques

Checksums, electronic signatures, and time stamps are among effective techniques to determine the presence/absence of falsification of original data.

Table 6 Falsification detection technique

Falsification detection technique	*Falsification detection method*	*Detectable description*
Checksum (CRC/SHA-256)	*Data owners calculate figures in the data to be disclosed, by using a detection function (a rule), and disclose the error detection function used and the numeric value obtained, together with the data to be disclosed. Users (including data owners) confirm that the disclosed data have not been falsified, by using the error detection function and the numeric value.*[10]	*Presence/absence of a falsification of original data*

[10] Reference URL：http://www.atmarkit.co.jp/fsecurity/rensai/inci03/inci01.html

Falsification detection technique	*Falsification detection method*	*Detectable description*
Electronic signature	*Data owners place an electronic signature on the data to be disclosed, and disclose the data together with a public key. Users (including data owners) confirm that the disclosed data have not been falsified, by verifying the electronic signature with the public key.*[11]	*An individual or an organization that has created original data; presence/absence of a falsification of the original data (however, falsification cannot be detected if the original data has been falsified by the issuer of the electronic signature)*

[11] Reference URL：http://www.jipdec.or.jp/esac/intro/shikumi.html

Falsification detection technique	*Falsification detection method*	*Detectable description*
Time stamp	*Data owners disclose data after obtaining a time stamp from a professional organization, in addition to a time stamp usually recorded when data are saved. Users (including data owners) confirm that the disclosed data have not been falsified, by checking with the professional organization that the time stamp is genuine.*[12]	*Latest time of original data update; presence/absence of a falsification of the original data (a falsification by the issuer of the electronic signature can be detected when different persons issue an electronic signature and place a time stamp)*

Of the above three falsification detection techniques, our attention is focused on electronic signature (a function using an encryption technique), which is intermediate in terms of security and cost, when discussing specific techniques and applied systems, as follows:

a) Falsification detection technique using an encryption technique

To detect data falsification, it is effective to introduce an encrypted electronic signature or hush value of data. Specifically, an electronic signature supported by public-key cryptographic techniques is known to be highly effective in ensuring security.

Specifically, original data should be disclosed together with a hush value or an electronic signature (In using a hush or an electronic signature, the CRYPTREC Ciphers List may be used. Hush values should be disclosed in an environment that can disable falsification. In using electronic signatures, the Government Public Key Infrastructure (GPKI) should be employed.). These measures facilitate falsification detection as falsified data are different

[12] Reference URL：http://www.dekyo.or.jp/tb/system/system_7.html

from original data in hush value or in electronic signature

It is known that falsification by calculating hush values or electronic signatures are nearly impossible.

b) Use of a mechanism furnished on application software

Currently, various data formats permit setting an electronic signature. For example, the following data formats are provided with such a mechanism:

docx, xlsx, pptx: Microsoft Office format
ods: Tabular format of open document

Application software that mainly handles these data formats is compatible with such mechanism and capable of detecting falsified data. Thus, the adoption of these applications enables easy use of electronic signature and other mechanisms.

Part II Rules on Use: Setting a Rule on Use on Open Data

Chapter 4. Rules on Use Necessary for Data

For any data are to be "open data," it is important to permit the secondary use of public data by information users, by adopting an official rule to permit such secondary use. This chapter explains the importance of a rule on use in open data and discusses trends at the Japanese government and in foreign countries with respect to rules on use.

4.1 Importance of Rules on Use in Open Data

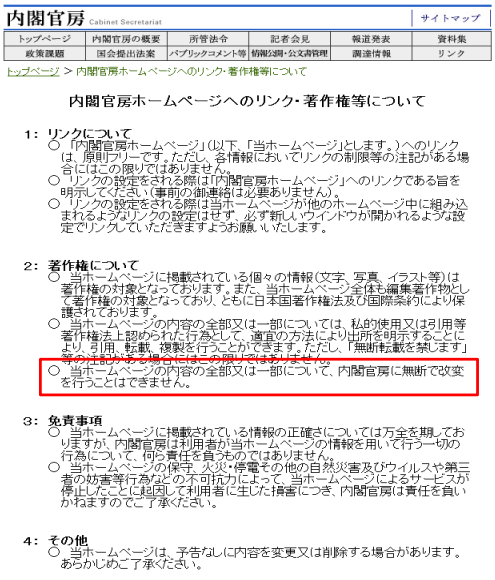
A copyright is generated when public data owned by the national government, local governments, incorporated administrative agency, public corporations, and other public organizations have creativity.[13] If a copyright has been generated, users who desire to make secondary use of public data must obtain a license from the national government, local government, etc. that has created the public data.

As of March 2014, the homepages of the national government, local governments, etc. do not permit public data users to modify such data without their approval. Users are currently in an environment that restricts free use of public data (Figure 4-1, Figure 4-2)[14].

Information users would be free to make secondary use of public data if homepages on the national government, local governments, etc. clearly indicate that such users may make secondary use of their public data.

[13] Data do not generate copyrights if they lack creativity. These data include numeric values and simple tables and graphs. For details of copyrights, see http://www.bunka.go.jp/chosakuken/gaiyou.html

[14] Many homepages of other ministries and agencies impose limitations on use.



内閣官房 Cabinet Secretariat　サイトマップ

トップページ | 内閣官房の概要 | 所管法令 | 記者会見 | 報道発表 | 資料集
政策課題 | 国会提出法案 | パブリックコメント等 | 情報公開・公文書管理 | 調達情報 | リンク

トップページ ＞ 内閣官房ホームページへのリンク・著作権等について

内閣官房ホームページへのリンク・著作権等について

1：リンクについて
○「内閣官房ホームページ」（以下、「当ホームページ」とします。）へのリンクは、原則フリーです。ただし、各情報においてリンクの制限等の注記がある場合にはこの限りではありません。
○「リンクの設定をされる際は「内閣官房ホームページ」へのリンクである旨を明示してください（事前の御連絡は必要ありません）。
○「リンクの設定をされる際は当ホームページが他のホームページ中に組み込まれるようなリンクの設定はせず、必ず新しいウインドウが開かれるような設定でリンクしていただきますようお願いいたします。

2：著作権について
○ 当ホームページに掲載されている個々の情報（文字、写真、イラスト等）は著作権の対象となっております。また、当ホームページ全体も編集著作物として著作権の対象となっており、ともに日本国著作権法及び国際条約により保護されております。
○ 当ホームページの内容の全部又は一部については、私的使用又は引用等著作権法上認められた行為として、適宜の方法により出所を明示することにより、引用、転載、複製を行うことができます。ただし、「無断転載を禁じます」等の注記がある場合にはこの限りではありません。
○ 当ホームページの内容の全部又は一部について、内閣官房に無断で改変を行うことはできません。

3：免責事項
○ 当ホームページに掲載されている情報の正確さについては万全を期しておりますが、内閣官房は利用者が当ホームページの情報を用いて行う一切の行為について、何ら責任を負うものではありません。
○ 当ホームページの保守、火災・停電その他の自然災害及びウイルスや第三者の妨害等行為などの不可抗力によって、当ホームページによるサービスが停止したことに起因して利用者に生じた損害につき、内閣官房は責任を負いかねますのでご了承ください。

4：その他
○ 当ホームページは、予告なしに内容を変更又は削除する場合があります。あらかじめご了承ください。

リンクについて

総務省ホームページはリンクフリーとなっており、リンクのご連絡も不要です。
また、総務省シンボルマークを使用して、総務省ホームページへリンクする場合は、下のバナー画像をご利用ください。
ただし、この画像を総務省ホームページへのリンク以外の用途に使用することはできません。
（総務省シンボルマークについて）

画像	リンク先
総務省 MIC Ministry of Internal Affairs and Communications	http://www.soumu.go.jp/

著作権について

「総務省ホームページ」に掲載されている個々の情報（文字、写真、イラスト等）は著作権の対象となっています。また、「総務省ホームページ」全体も編集著作物として著作権の対象となっており、ともに日本国著作権法及び国際条約により保護されています。
当ホームページの内容の全部又は一部については、私的使用又は引用等著作権法上認められた行為として、適宜の方法により出所を明示することにより、引用・転載複製を行うことが出来ます。
ただし、「無断転載を禁じます」等の注記があるものについては、それに従ってください。
当ホームページの内容の全部又は一部について、総務省に無断で改変を行うことはできません。

他社所有商標に関する表示

Microsoft、WindowsおよびWord、Excel、PowerPointは、米国Microsoft Corporationの、米国およびその他の国における登録商標または商標です。
Adobe、Adobeロゴ、Flash、Flash Liteは、アドビシステムズ社の米国およびその他の国における登録商標または商標です。
その他の会社名および製品・サービス名は、それぞれを表示するためだけに引用しており、各社の登録商標あるいは出願中の商標である場合があります。
当ホームページに記載されているシステム名、製品などには、必ずしも商標表示（(R)、TM）を付記していません。

ウェブアクセシビリティ

総務省では、多くの利用者に確実かつ正確に情報伝達が行われることを目指してホームページを制作し運用を行っています。

総務省ウェブアクセシビリティ方針

Figure 4-1 Example of a Rule on Use of the Homepage (Cabinet Secretariat)[15]

Figure 4-2 Example of a Rule on Use of the Homepage (Ministry of Internal Affairs and Communications)[16]

As explained in Section 2.3, the 5 Star Open Data provides as the first condition that open data should be provided under an open license, while Open Data Handbook refers to a license. As seen from these moves, it is essential to adopt a rule on secondary use prior to establishing a technical requirement for machine-readability.

For public data owned by the national government, local governments, incorporated administrative agency, public corporations, and other public organizations, three methods are available to disclose such public data in such a way as to permit secondary use, as shown below:

(i) In principle, public data do not generate copyrights;

(ii) Public data generate copyrights, but they are waived; or

(iii) Public data are disclosed in such a way as to permit secondary use.

Most desirably, public data should not generate copyrights as in (i). The Copyright Act is designed to grant an exclusive right in the form of copyright as an incentive to promote creation. However, copyrights do not work as an incentive when it comes to creation of public data with taxpayers' money. In the U.S., public data owned by the federal and local governments are being used widely as no copyright is generated in such public data. To adopt the U.S. way, however, the Copyright Act need to

15 http://www.cas.go.jp/jp/tyosakuken/index.html
16 http://www.soumu.go.jp/menu_kyotsuu/policy/tyosaku.html

be amended to exclude government-created data from the scope of the application of the same law.

If copyrights are generated in public data but are waived in option (ii), the same effect as in option (i) is generated as the national and local governments waive such copyrights. In this case, it is required to carefully determine the feasibility of such waiver in relation to the National Property Act, the Public Finance Act, the Local Autonomy Act, the Subsidy Budget Rationalization Act, and other relevant laws, because copyrights are part of the property of the national government, local governments, etc.

By contrast, option (iii) asserts that authors should keep their copyrights in public data but permit free secondary use by adopting a rule on use, in order to promote open data.

The state in which no copyright is generated as in option (i) or in which copyrights extinguish as in option (ii) is called "public domain." Once data are in the public domain, copyright holders are not entitled to exercise their right to claim damages, injunction, etc., and users may make free use of such data. For this reason, it is desirable to place open data in the public domain. Nevertheless, it takes a substantial amount of time to amend related laws and interpret the National Property Act and other laws. It is desirable then to start with option (iii) on a short-term basis, while reviewing legal matters from a long-term perspective

This guide focuses its attention to option (iii) in the paragraph to follow.

4.2 International Trends of Rules on Use of Open Data

As discussed in Part I, open data initiatives have been carried out in foreign countries ahead of Japan.

In the U.S., no copyrights are generated in public data created by the federal government (public domain). Once such data are disclosed to the public, citizens may make secondary use of them freely. In the Netherlands, copyrights and database rights are generated in government-created public data,[17] but are waived by the government, which places their data in the public domain for secondary use. To waiver its copyrights, the Dutch government relies on "Public Domain Declaration" (hereinafter referred to as "CC0") issued by an international non-commercial organization Creative Commons (hereinafter referred to as "CC"). CC0 aims to waive copyrights in certain data and permit secondary use of the data without imposing any limitation or condition (for more details, see Section 5.4). The Netherlands adopted the CC rule on use, as many other western nations, including the U.S, were introducing a CC license following its first issuance in 2002. Initially, a CC license was used for data created by individuals, such as books, music, and videos, in addition to data owned by universities. With the spread of open data, the license came to be adopted for government-owned data.

While the U.S. and Netherlands retain public data in the public domain, other countries permit

[17] EU Database Directive (96/9/EC) provides that special rights (sui generic) may be granted to un-copyrighted databases if they have been created with substantial investment.

secondary use under a rule on use, as copyrights and database rights are generated in public data created by the government. Such rule on use adopted may be either (i) an existing rule on use of open data or (ii) a new rule on use formulated on their own initiative.

Australia, New Zealand, and Germany are among the countries that chose (i). These countries adopt a Creative Commons license (hereinafter referred to as "CC license") issued by CC. A CC license provides six different rules on use.[18] Among those rules, these countries have introduced a Creative Commons Attribution license (CC-BY), which is best suited to open data (for details of rules on use, see Section 5.1 and Section 5.2). Basically, a CC-BY license permits free secondary use to the extent that the user specify the source of data. Since the formulation of this rule on use in 2002, this license has been used for a large number of data, spreading worldwide.

Britain, France, Italy, and some other countries chose (ii). However, these countries are proceeding with a rule on use compatible with a CC license. Britain compiled Open Government License[19], and France License Ouverte[20], both of which are supposed to be compatible with a CC-By license. In Italy, Italian Open Data License (IODL) was created. Its 2011 Version 1.0[21] is compatible with Creative Commons Attribution-ShareAlike License (CC-BY-SA), and its 2012 Version 2.0[22] with CC-BY.

In these countries, their rules on use were previously incompatible with the CC rules on use. Since 2013, when data rights were dealt with in CC-BY version 4, there has been discussion on whether CC-BY should be adopted directly.

Table 4-1 Rules on Use Adopted by Western Countries

採用利用ルール		国名
Existing rules on use adopted	Public Domain Declaration(CC0)	Netherlands
	Attribution License (CC-BY)	Germany, Australia, New Zealand, etc.
Voluntary rules on use adopted	Compatible with Attribution License (CC-BY)	Britain, France, Italy (Version 2.0), etc.
	Compatible with Attribution-ShareAlike License (CC-BY-SA)	Italy (Version 1.0)

As discussed above, many foreign countries have adopted a CC-BY license or a rule on use

[18] A CC license applies to data having copyright or other rights. CC0 is a declaration, not a license, to waive copyrights and other rights. Thus CC0 is not included among CC licenses.
[19] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/
[20] http://www.etalab.gouv.fr/pages/licence-ouverte-open-licence-5899923.html
[21] http://www.formez.it/iodl/
[22] http://www.dati.gov.it/iodl/2.0/

compatible with a CC-BY license. CC-BY is now a *de facto* international standard on open data.

4.3 Review of Rules on Use for Open Data by the Japanese Government

The Open Government Data Strategy, released on July 4, 2012 (IT Strategic Headquarters), provides as a basic principle for promoting the use of public data that the use of open data should be promoted whether for commercial or non-commercial purposes.

In response to the release of the strategy, the Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use (Liaison Conference of Chief Information Officers (CIOs) of Public Offices and Ministries decision, June 25, 2013) was announced at the **e-Government Open Data Conference** of Working-level Personnel in order to disclose its position toward a rule on secondary use, as shown below.

Table 4-2 Basic Guideline Principles for Secondary Use Promotion (Excerpt)

- For data that is not copyrighted, it shall be made clear that it is outside the scope of protection of copyright, and that secondary use shall not be restricted on the grounds of the presence of a copyright.
- For works of which the national government is the author, its stance toward use of the work shall be displayed in advance, in the form of widely permitting secondary use.
- If secondary use of any part of open data is to be restricted based on a copyright, the government shall clearly specify the part restricted in secondary use together with the reason for such restriction. Such reasons may include the fact that such part is copyrighted by a third party and the fact that the copyright holder cannot be identified.
- Upon establishment of these guidelines, ministries and agencies shall endeavor to reach agreement with other stakeholders on data created or acquired by the public organizations to ensure that they permit secondary use of their data disclosed to the public on the Internet. In preparing or executing any outsourcing or other agreements, ministries and agencies are required to take these guidelines into consideration (for example, ministries and agencies are to disclose on the Internet a report resulting from a contracted study, the content of such report should not interfere with the secondary use of such report).
- To restrict the secondary use of open data based on specific reasonable grounds other than copyrights, as under an individual law or regulation, the scope of limitation should be minimized to the extent possible, and the details and the grounds of such limitation should be described clearly. Such description should be unified and clearly understandable.
- Ministries and agencies should state clearly that they shall not be held responsible for any damage incurred by a user in using data that have been created by a third party for secondary use from data disclosed to the public by ministries and agencies.

Source: Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use (Liaison Conference of Chief Information Officers (CIOs) of Public Offices and Ministries decision, June 25, 2013)

The Data Governance Committee of Open Data Promotion Consortium conducted a case study in 2012 on the basis of **White Paper** on **Information** and **Communications** in Japan (Ministry of Internal Affairs and Communications) and other government publications. Upon request of the Cabinet Secretariat IT Strategy Planning Office, the committee discussed the Review of Rule on Use of Homepages of Ministries and Agencies (Draft), in accordance with the guidelines. The draft focused on issues to be discussed by the end of the second half of FY2013 under the **Roadmap for** the e-Government **Open Data Promotion** (IT Strategic Headquarters decision, June 14, 2013) concerning the rule on use on homepages of ministries and agencies, which provides that, basically, public content shall be disclosed with a secondary use license and that restricted content shall be provided separately.

The recommendations based on the discussions were presented to the Rule Promotion Working Group under the **e-Government Open Data Conference** of Working-level Personnel (January 17, 2014). Based on the recommendations, the conference worked out and approved the Government Standard Terms of Use (Version 1.0) on April 1. CIO Conference also approved it on June 19 2014 (for details of the Government Standard Terms of Use (Version 1.0) , see Section 5.5)

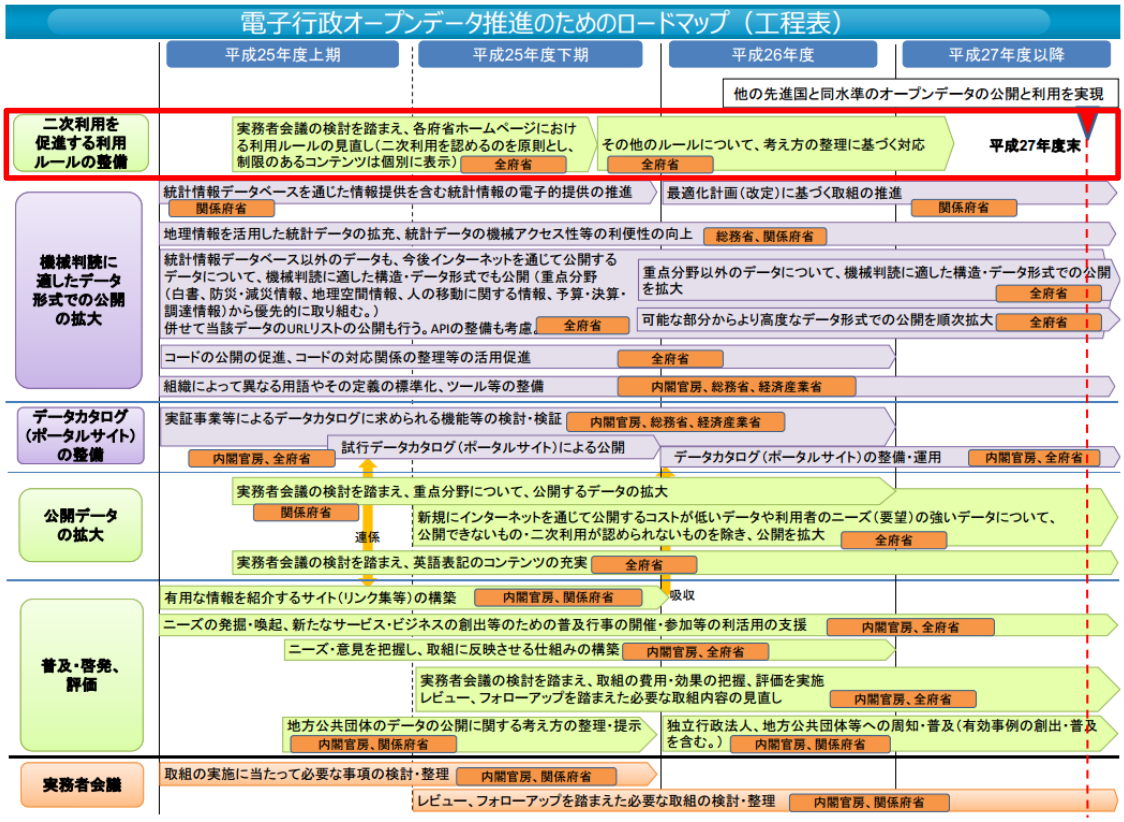


Figure 4-3 Roadmap for **the e-Government** Open Data Promotion

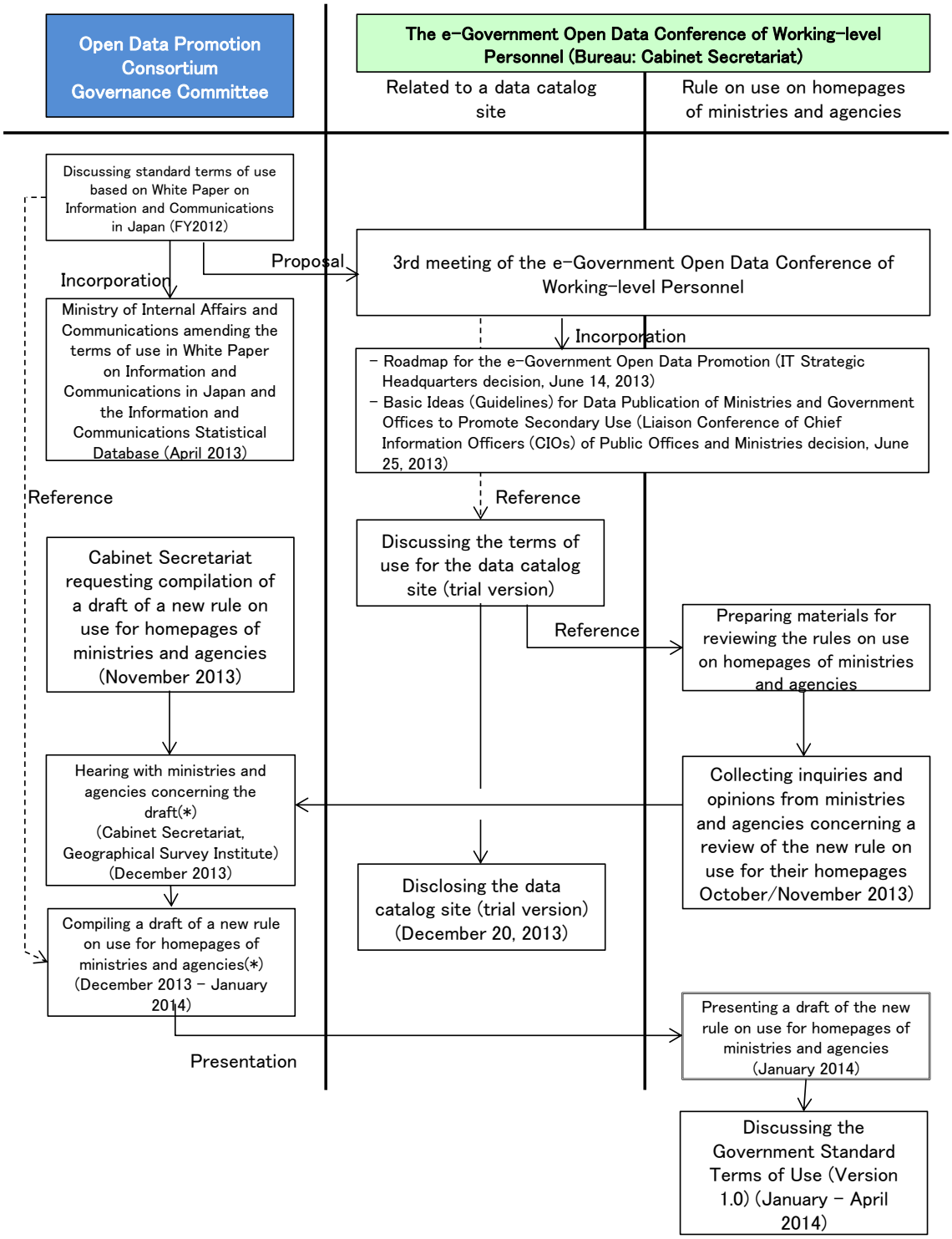


Figure 4-4 Revision Flow of the Government Standard Terms of Use

If the Japanese government adopts a rule on use different from those adopted in other countries, overseas users would have to take trouble to understand differences between them. In addition, they could face complicated rules on use when the users attempt to mash up domestic and foreign data. For

these and other reasons, the Government Standard Terms of Use (Version 1.0) (Draft) were compiled based on opinions collected from ministries and agencies, in such a way as to ensure that the terms apply to the widest range of public data, although they require a statement of sources as in CC-BY. As a result, the terms differ from CC-BY as they contain a provision prohibiting "the use of public data in violation of laws and regulations or ordinances or against public order" and "the use that may threaten the security of the nation and its people."

Table 4-3 Incorporation of Opinions of Ministries and Agencies into the Government Standard Terms of Use

Opinion	Incorporation in the Government Standard Terms of Use (Version 1.0)
Edited or processed data should accompany a statement of the information source, a statement of the fact that the data have been modified, and information identifying the person responsible for editing or processing the data.	A provision will be included to require the user to state that the data have been edited or processed by them.
The terms should prohibit the user from falsifying public data or causing others misunderstanding.	A provision will be included to prohibit users from using or disclosing to the public any information edited or processed by them in such a way as if such information had been created by the national government (including ministries and agencies).
Content copyrighted by a third party should be clearly specified.	It is practically impossible to specify all content copyrighted by a third party. Instead, some examples will be provided to suggest or state the presence of any third-party right.
The terms should clearly state that any misuse of maps, marine charts, aerial charts, warnings or predictions, disaster prevention information, and some other information could threaten the security of the nation and its people, and that the government may implement appropriate measures against any use of public data in violation of laws and regulations or ordinances or against public	These data constitute a major part of public data. A unified rule should be applied as widely as possible to these types of public data. Thus, in order to cover this rule, these terms of use will prohibit these acts.

order.	
Especially important applicable laws and regulations should be clearly indicated and made readily accessible.	A relevant indication will be added to major data that are subject to limitations under specific laws or regulations.
A statement should be included to inform the user that any content may be changed or deleted without notice.	A statement will be included to inform the user that any content is subject to change or deletion without notice.

The Government Standard Terms of Use (Version 1.0) will be reviewed in FY2015. At the review, the necessity of changing prohibited uses will be discussed in light of the use of content after the rules on use are adjusted to conform to the Government Standard Terms of Use (Version 1.0) (see Table 4-5).

Table 4-4 Main Reasons for Not Adopting CC-BY

(Requirements to be Included in the Government Standard Terms of Use (Version 1.0))

- The rules on use may impose and indicate their own policy, including a requirement for indication of the fact that relevant content has been edited, processed, or otherwise changed, if any, and the prohibition of use of content deemed by disclosing ministries and agencies to be against public order.
- CC-BY is a license to cover copyrighted content. It should be noted that there are conditions and matters to be set forth uniformly regardless of whether certain content has the nature of a literary work.
- CC-BY licenses set forth professional conditions for a copyright statement to be used to redistribute CC-BY licensed content or to combine such content with other content. To make a rule on use simpler and easier to understand, it is not always necessary to adopt these professional conditions.

Source: Explanatory Notes on the Government Standard Terms of Use (Version 1.0)

Table 4-5 Review of the Government Standard Terms of Use (Version 1.0)

・ (These the terms of use) further draws the attention of users to the provision that the rule on use is subject to change. The paragraph also provides that the rule on use will be reviewed by the end of FY 2015 as specified in the Declaration to be the World's Most Advanced IT Nation (Cabinet decision, June 14 2013) and **Roadmap for** the e-Government **Open Data Promotion** (IT Strategic Headquarters decision, June 14, 2013) in order to achieve the same level of public disclosure as in other advanced countries. ・ Cosidering how wide contents in compliance with updated Government Standard Terms of Use (Version 1.0) are utilized, this review will include a discussion on the necessity of Section 1.3) (Prohibited use) in order to make the rule on use more compatible with globally adopted CC-BY licenses.

Source: Explanatory Notes on the Government Standard Terms of Use (Version 1.0)

In parallel with the review of the Government Standard Terms of Use (Version 1.0), the Cabinet Secretariat IT Strategy Planning Office opened a government data catalog site (trial version) "DATA.GO.JP" on December 20, 2013. This site adopted a CC-BY license based on discussions at the **e-Government Open Data Conference** of Working-level Personnel. One of the charateristics of this term of use is that it refers to the protection of third party rights (For details, see Section 5.3).

Chapter 5. Outline of Open Data Rule on Use

As discussed in Chapter 4, with respect to rules on use for open data, foreign governments have been introducing CC-By or CC0, while the Japanese government is expected to adopt CC-BY (as in the government data catalog site (trial version) DAT.GO.JP and *White Paper on Information and Communications in Japan*) or the Government Standard Terms of Use (Version 1.0).

This chapter explains CC0, CC-BY, and the Government Standard Terms of Use (Version 1.0) , "DATA.GO.JP"(Trial Version) Term of Use as main rules on use for open data.

5.1 CC License

5.1.1 Outline of CC Licenses

CC refers collectively to the international non-commercial organization providing CC licenses and its project[23]. Established in 2001, the organization released its first version of rule on use in 2002 in the U.S. In Japan, the first version was disclosed in 2004.

CC licenses are a tool for data creators to express their intention to authorize free use of their works under the condition that users comply with certain requirements for the purpose of helping advance copyright right rules in the new age of the Internet. By adopting a CC license, data creators are able to distribute their works freely while maintaining their copyrights, and users are permitted to distribute, modify, or otherwise use their works at their own will within the extent authorized by a specified rule on use.

5.1.2 Features of CC Licenses

A CC license is composed of three elements:

① Commons Certificate－an easy-to-understand one-page explanation of rights,

② License－a document describing the subject to matter of the Commons Certificate in legal terms for the legal profession, and

③ Metadata－explanatory information accompanying works (content) to be used by search engines.

These three elements enable a rule on use described for general users in an easy, clear manner. In addition, the metadata enables retrieval of CC-licensed data via a machine. The metadata is a piece of information embedded in the source code on a homepage to mechanically obtain names for a credit statement. Metadata are retrieved together with the data by a search engine.

[23] Creative Commons Japan homepage (http://creativecommons.jp/)

5.1.3 Types of CC Licenses

There are six different types of CC licenses, as shown in Table 5-1.

Each rule on use under these licenses specifies (1) whether commercial use is permitted or not and/or (2) whether modification is permitted or not. All rules on use obligate the indication of sources.

Table 5-1 Outline and Types of CC Licenses

Symbol	Name of rule on use	Rule on use		
		Indication of source	Commercial use	Modification
	Attribution 2.1 Japan (CC-BY 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Permitted	Permitted (*)
	Attribution-NonCommercial 2.1 Japan (CC-BY-NC 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Not permitted (Also, commercial use of modified content is not permitted)	Permitted (*)
	Attribution-NoDerivs (CC-BY-ND 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Permitted	Not permitted
	Attribution-NonCommercial-NoDerivs 2.1 Japan 2.1 Japan (CC-BY-NC-ND 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Not permitted	Not permitted

Symbol	Name of rule on use	Rule on use		
		Indication of source	Commercial use	Modification
	Attribution-ShareAlike 2.1 Japan (CC-BY-SA 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Permitted	Modification is permitted, but secondary literary works produced by such modification shall be subject to the rule on use as this rule on use (*).
	Attribution-NonCommercial-ShareAlike 2.1 Japan (CC-NC-SA 2.1 Japan)	Mandatory (Specify title, all authors, and URL)	Not permitted (Also, commercial use of modified content is not permitted)	Modification is permitted, but secondary literary works produced by such modification shall be subject to the rule on use as this rule on use (*).

*Modification that infringes the human rights of the author is not permitted.

Source: Created by the Data Governance Committee based on the Creative Commons Japan homepage. (http: //creativecommons.jp/licenses/).

The pages to follow explain CC-BY and CC0, which are widely used in open data.

5.2 CC-BY License

5.2.1 Outline of CC-BY License

Among CC licenses, a CC-BY license imposes the rule on use that has the least limitation and permits free use of content to the extent that users state the source of the content. Initially, rules on use were created to suite the legal system of each country. Since the issuance of its version 4.0[24], CC has been applied as a unified rule on use. In Japan, Version 4.0 is yet to be translated, and currently Version 2.1[25] is in place.

In Germany, Australia, New Zealand, and some other countries, the governments have adopted the CC-BY license as a standard rule on use of open data. The Japanese government has also adopted the CC-BY license as its standard rule on its data catalog site (trial version) "DATA.GO.JP." Britain, France, and Italy are among countries that apply a rule on use compatible with the CC-BY license.

CC-BY is characterized by permitting free use of data to the extent that the source of the data is clearly stated, in such a way as reproduction, adaptation, distribution, dramatization, filming, public transmission, dictation, display, voice and image recording, broadcasting, cable broadcasting, making transmittable, and communication. Commercial use is also permitted.

To indicate sources, it is required to describe the credit of the original author or performer in a reasonable manner, specify the title of the original work, and indicate any specified URI. Any secondary literary work must be accompanied by a statement that the original literary work has been used (CC-BY license, Article 5 h).

If a notice was given by the licensor, the user must delete references to the licensor or the original author to the extent possible (Article 5 i). CC-BY also provides that users shall not conduct copy control so as to narrow the licensed scope (Article 5 f).

The following provides the whole text of the CC-BY rule on use (legal code). The Commons Certificate provides a summary of the legal code made understandable to the general public, and does not substitute the code.

As mentioned above, the Japanese government adopted CC-BY as its standard terms of use for its data catalog site (trial version) "DATA.GO.JP." In addition CC-BY, this catalog site draws the attention of users to content carrying a third party right, which may be used as reference to disclose data to the public as open data (for details, see Section 4.3).

[24] http://creativecommons.org/licenses/by/4.0/deed.ja
[25] http://creativecommons.org/licenses/by/2.1/jp/

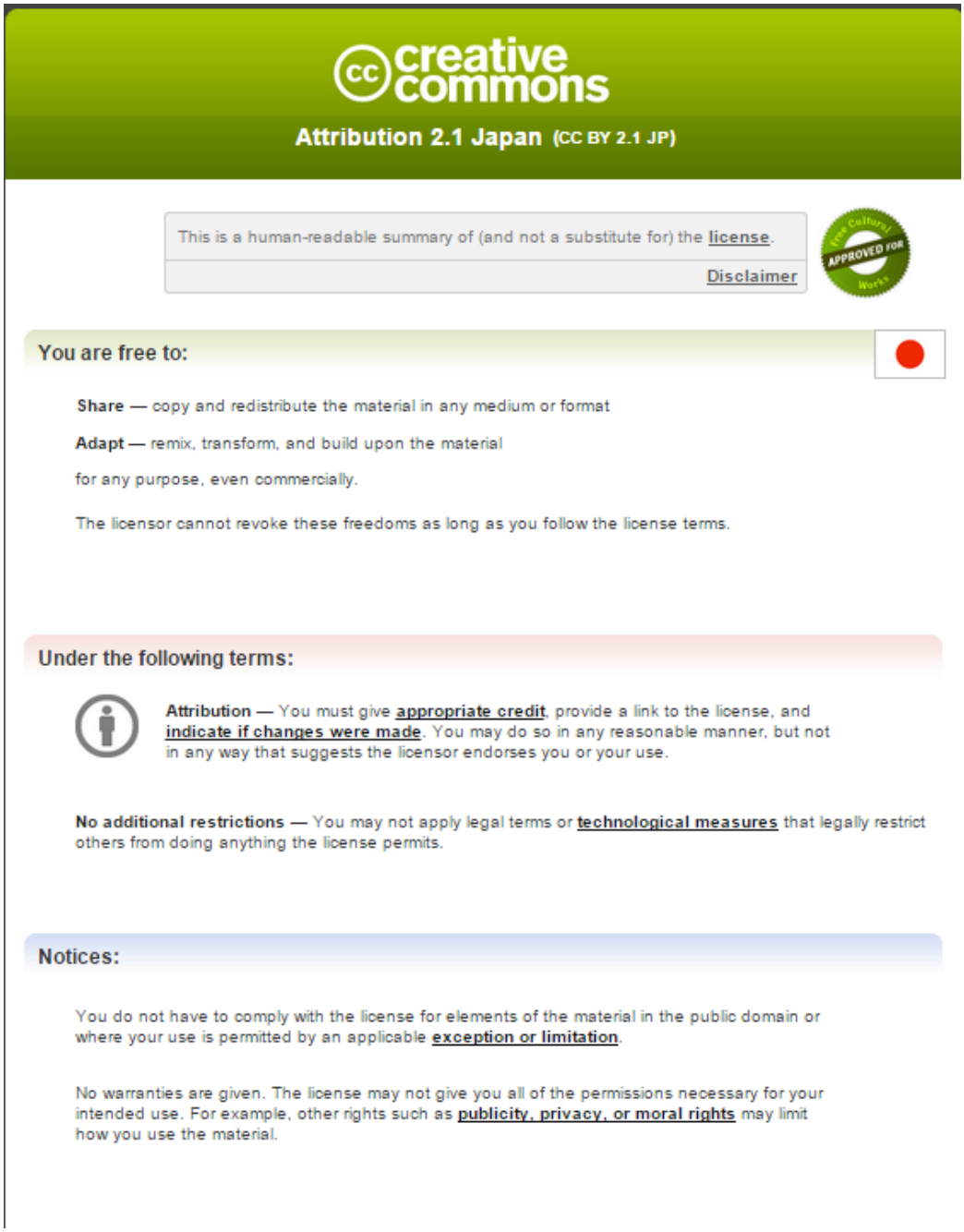


Figure 5-1 CC-BY Commons Certificate

Table 5-2 Text of CC-BY use of rule (Legal Code)

creative commons

LEGAL CODE

Attribution 2.0

CREATIVE COMMONS CORPORATION IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS LICENSE DOES NOT CREATE AN ATTORNEY-CLIENT RELATIONSHIP. CREATIVE COMMONS PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. CREATIVE COMMONS MAKES NO WARRANTIES REGARDING THE INFORMATION PROVIDED, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM ITS USE.

License

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions

a. **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this License.

b. **"Derivative Work"** means a work based upon the Work or upon the Work and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work will not be considered a Derivative Work for the purpose of this License. For the avoidance of doubt, where the Work is a musical composition or sound recording, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered a Derivative Work for the purpose of this License.

c. **"Licensor"** means the individual or entity that offers the Work under the terms of this License.

d. **"Original Author"** means the individual or entity who created the Work.

e. **"Work"** means the copyrightable work of authorship offered under the terms of this License.

f. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

2. Fair Use Rights. Nothing in this license is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

a. to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b. to create and reproduce Derivative Works;

c. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works;

d. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission Derivative Works.

e. For the avoidance of doubt, where the work is a musical composition:

i. **Performance Royalties Under Blanket Licenses**. Licensor waives the exclusive right to collect, whether individually or via a performance rights society (e.g. ASCAP, BMI, SESAC), royalties for the public performance or public digital performance (e.g. webcast) of the Work.

ii. **Mechanical Rights and Statutory Royalties**. Licensor waives the exclusive right to collect, whether individually or via a music rights agency or designated agent (e.g. Harry Fox Agency), royalties for any phonorecord You create from the Work ("cover version") and distribute, subject to the compulsory license created by 17 USC Section 115 of the US Copyright Act (or the equivalent in other jurisdictions).

f. **Webcasting Rights and Statutory Royalties**. For the avoidance of doubt, where the Work is a sound recording, Licensor waives the exclusive right to collect, whether individually or via a performance-rights society (e.g. SoundExchange), royalties for the public digital performance (e.g. webcast) of the Work, subject to the compulsory license created by 17 USC Section 114 of the US Copyright Act (or the equivalent in other jurisdictions).

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4. Restrictions. The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a. You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this License, and You must include a copy of, or the Uniform Resource Identifier for, this License with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this License or the recipients' exercise of the rights granted hereunder. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this License Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this License. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested. If You create a Derivative Work, upon notice from any Licensor You must, to the extent practicable, remove from the Derivative Work any reference to such Licensor or the Original Author, as requested.

b. If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Derivative Works or Collective Works, You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied; to the extent reasonably practicable, the Uniform Resource Identifier, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and in the case of a Derivative Work, a credit identifying the use of the Work in the Derivative Work (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Derivative Work or Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5. Representations, Warranties and Disclaimer

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

6. Limitation on Liability. EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Derivative Works or Collective Works from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.

b. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

8. Miscellaneous

a. Each time You distribute or publicly digitally perform the Work or a Collective Work, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

b. Each time You distribute or publicly digitally perform a Derivative Work, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.

c. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

d. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

e. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

Creative Commons is not a party to this License, and makes no warranty whatsoever in connection with the Work. Creative Commons will not be liable to You or any party on any legal theory for any damages whatsoever, including without limitation any general, special, incidental or consequential damages arising in connection to this license. Notwithstanding the foregoing two (2) sentences, if Creative Commons has expressly identified itself as the Licensor hereunder, it shall have all rights and obligations of Licensor.

Except for the limited purpose of indicating to the public that the Work is licensed under the CCPL, neither party will use the trademark "Creative Commons" or any related trademark or logo of Creative Commons without the prior written consent of Creative Commons. Any permitted use will be in compliance with Creative Commons' then-current trademark usage guidelines, as may be published on its website or otherwise made available upon request from time to time.

Creative Commons may be contacted at https://creativecommons.org/.

« Back to Commons Deed

5.2.2 Using CC-BY

CC-BY is used as follows:[26]

1. On the CC homepage, open the Select Rule page (http: //creativecommons.org/choose/). The latest version 4.0 of CC-BY appears on the first screen displayed. This version is yet to be translated into Japanese.

2. To use the Japanese version, on top of the page, click "Looking for earlier license versions, including ports?" (the red frame in Figure 5-2). Then, on the pop-up window, select "Show earlier licenses" (the red frame in Figure 5-3).

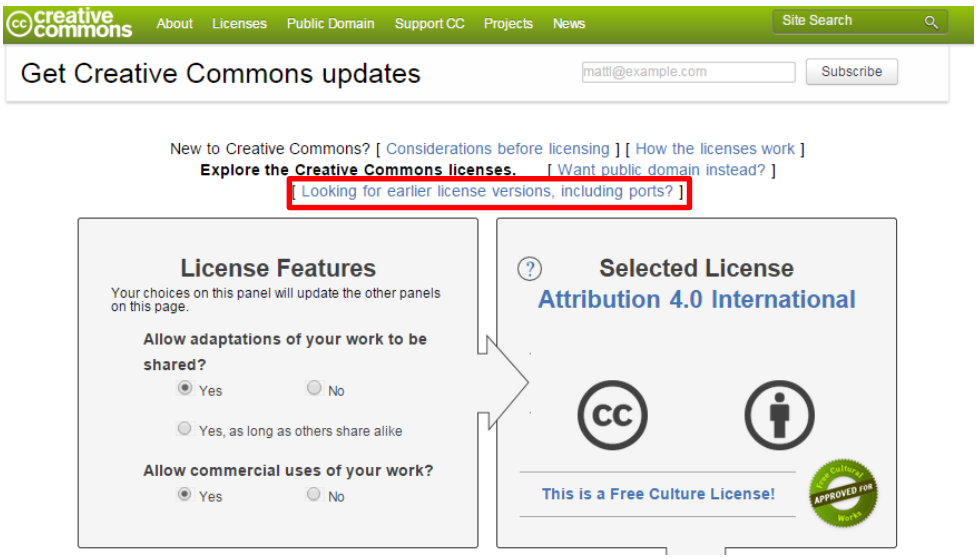


Figure 5-2 Select Rule screen

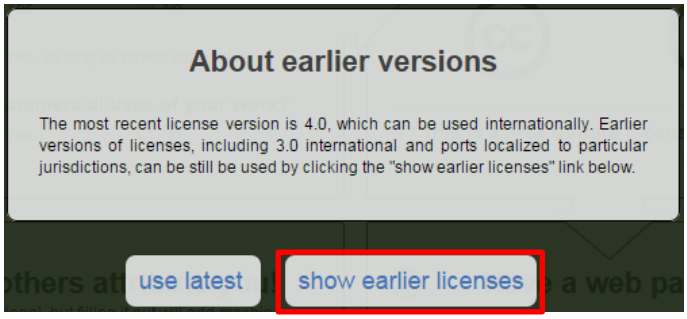


Figure 5-3 Pop-up window

[26] Compiled based on the License Selection page of Creative Commons homepage (http://creativecommons.org/choose/)

3. On the left of the screen appears the drop box "Jurisdiction of License" (the red frame in Figure 5-4). Then select "Japan" to select the latest version CC-BY 2.1.

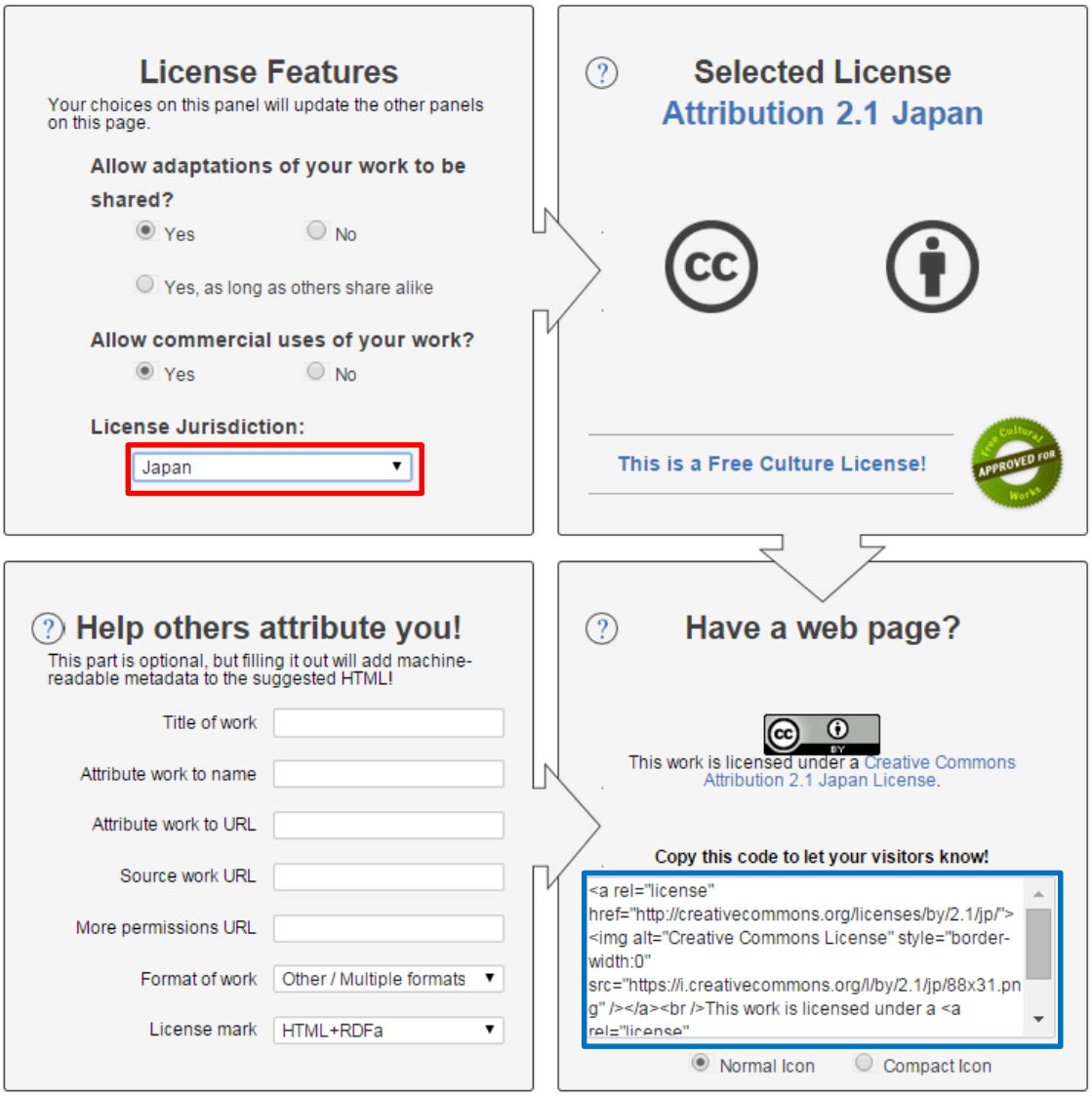


Figure 5-4 Select License Screen (Previous Version)

4. Paste the code provided on right bottom of Figure 5-4 to the source code on the homepage, and then you can start using the homepage. This code is composed of an image of CC-BY, a link to the CC-BY license, and the metadata of the CC-BY license. When they are pasted on the homepage, the screen displays the image of the CC-BY license and a statement to the effect that the work (or the names of the author and the document when a credit is inserted) is provided under the Creative Commons Attribution 2.1 Japan license.

5.3 Government Data Catalog Site “DATA.GO.JP” (Trial Version) Term of Use

5.3.1 The Outline of Government Data Catalog Site “DATA.GO.JP” (Trial Version) Term of Use

As stated above, IT strategic headquarters launced the trial version of “DATA GO JP”, a government data catalog site, on December 20, 2013. The trial verson of “DATA GO JP” adopted CC-BY in compliance with the discussion at the e-Government Open Data Conference of Working-level Personnel.

“DATA.GO.JP” (Trial Version) term of use, where CC-BY is used as a rule on use, tells users to pay attention to a third party right, no guarantee regarding the content disclosed on this website, and governing law and jurisdiction. “DATA GO JP” requests users to state the source of figures and other data that are not subject to the provisions of CC-BY because no copyright is generated. It also provides that data disclosed on this site is subject preferentially to the standard terms of use set forth on the site.

5.3.2 Notes on Adopting the Term of Use

“DATA GO JP” pays special attention to contents not containing a third party right, which means it has more consideration toward users.

Since “DATA.GO.JP” (Trial Version) term of use refers to license IDS used only for “DATA.GO.JP” (Trial Version), it is better to use universally applicable version (Table 5-4) to avoid any confusion.

Table 5-3 Standard Terms of Use for Government Data Catalog Site (Trial Version)

Article 1 (Useof the content of this website)

The relevant content disclosed on this website may be used in accordance with the conditions set forth in these Terms of Use. By using the content of this website, users agree to these Terms of Use.

Article 2 (Use of data pursuant to CC-BY)

1. Data falling under (1) and (2) below may be used in accordance with the copyright license Attribution 2.1 Japan of the Creative Commons License (hereafter referred to as the CC License). For details ,see http://creativecommons.org/licenses/by/2.1/jp/legalcode.
   (1) (a) All metadata in the data catalog; (b) the resources for which metadata is disclosed in the data catalog (hereafter referred to as "actual data disclosed in the catalog") and for which "CC-BY" is displayed in the license column and (c) article contents (content disclosed in parts of this website other than the data catalog)
   (2) Data (excluding logos, symbols, and so forth) for which the publisher owns the copyright (In the case of metadata in the data catalog and the actual data disclosed in the catalog, the publisher refers to the party shown in the publisher column for the metadata of the data-set of the relevant data. In the case of article contents, the publisher refers to the state of publication. The same applies hereinafter.)

Note, however, that numerical data and data in simple tables, graphs,and so forth are not subject to copyright. Accordingly, even if "CC-BY" is displayed in the license column for the applicable resource, the CC License does not apply to such data, and said data may be used freely.

2. In some cases, a third party other than the publisher (hereafter referred to as a "third party") owns the copyright or other rights for actual data disclosed in the catalog or for article contents. In such cases where a third party holds the copyright or other rights (such as portrait rights or the right of publicity for photographs), the user is responsible for obtaining permission to use such data or article contents from said third party. Note that, in general, there is no specification or indication whether any given portion of the actual data disclosed in the catalog or article contents is
subject to rights held by a third party. (In the case of actual data disclosed in the catalog that is provided by a national
ministry or agency, see the website of the applicable ministry or agency shown in the publisher column. These websites
provide examples of how to display or indicate third-party ownership of such rights.)

3. Users must observe the relevant laws and ordinances when using the content of this website. (For laws that require
particular attention when using actual data disclosed in the catalog that is provided by a national ministry or agency,
see the website of the applicable ministry or agency shown in the publisher column.)

4. Even if you are using data that is not subject to the CC License, such as numerical data or data in simple tables, graphs,
and so forth, we ask that you indicate the source of such data. (For details about data that is not subject to the CC
License, see Article 2, Item 1.) Indicate the source as follows: For metadata in the data catalog or article contents,
indicate the name of the data catalog site (DATA.GO.JP) as well as the URL of the relevant page (if it is difficult to
indicate the specific page, indicate the home page). For actual data disclosed in the catalog, indicate the following items
for the data-set or resource containing the actual data disclosed in the catalog: (a) the name of the organization shown
in the Publisher column for the metadata of the dataset, (b) the name of the resource, and (c) the URL of the resource.

Article 3 (Data use pursuant to the Government of Japan Standard Terms of Use)

If "Government of Japan Standard Terms of Use" is displayed in the license column for the Actual data disclosed in the catalog, such data can be used pursuant to Appendix : Government of Japan Standard Terms of Use (Version 1.1; the relevant content Publishers Expansion Version) (http://www.data.go.jp/terms-of-use/terms-of-use#label-link1). For information about the Government of Japan Standard Terms of Use (Version 1.1; the relevant content Publishers Expansion Version), such as how to cite sources, examples of how third-party ownership of rights can be displayed or indicated, and the relevant content for which use is limited pursuant to individual laws, see the website of the ministry or agency that provides such information.

Article 4 (Resources to which a third party owns the copyright in full)

In the case of a third party owning the copyright to everything for an item of Actual data disclosed in the catalog, with the exception of quotations permitted under the Copyright Act, the user of such data must, upon his or her own responsibility, obtain permission to use such data from any relevant third parties. In the case of a third party clearly owning the copyright to everything for an item of Actual data disclosed in the catalog, the license column will be blank and the column for the copyright of the metadata of the resource will indicate that the data is a work of a third party.

Article 5 (Logos, symbols, etc.)

Regarding independent use of marks that are used in this website or by a ministry or agency, and which represent organizations or businesses such as logos, symbols, etc., the user must consult with the responsible person of the relevant organization or business to check the conditions of use.

Article 6 (No guarantee)

The publisher provides no guarantee regarding the accuracy, completeness, fitness for a particular purpose, etc. of the content disclosed on this website. The publisher bears no liability for any acts done by users using the content disclosed on this website (including the use of information edited, processed, etc. based on said content). Even if the publisher displays or indicates third-party ownership of rights in said relevant content, such a display or indication is not comprehensive.

Article 7 (Relation to Terms of Use of other websites)

In the case that Actual data disclosed in the catalog is also disclosed on other governmental websites, and if the Terms of Use of such other websites (referring to the Terms of Use set forth on such respective websites rather than the Terms of Use stipulated by laws) differ from those of these Terms of Use, these Terms of Use will prevail.

Article 8 (Changes to these Terms of Use)

These Terms of Use might be subject to changes without prior notice. However, in the case that such changes will have a major impact on the use of the relevant content, we will provide prior notification of the substance of such changes.

Article 9 (Migration version of these Terms of Use)

In the case that a later version of the Terms of Use is designated as eligible for migration from these Terms of Use (such version is hereinafter referred to as a "migration version"), the user of the content disclosed on this website can select said migration version in lieu of these Terms of Use as the Terms of Use that govern the use of the content.

Article 10 (Governing law and jurisdiction)

These Terms of Use will be construed in accordance with the laws of Japan. Regarding the court of first instance, any disputes arising in connection with these Terms of Use or the use of the content disclosed on this website will be subject, in the case of the metadata in the data catalog and the Actual data disclosed in the catalog, to the exclusive jurisdiction of the court that has jurisdiction over the address of the party shown in the publisher column for the metadata of the data-set of the relevant data; and in the case of article contents, to the Tokyo District Court.

5.4 CCO

5.4.1 About CC0

CC0[27] is a project being carried out by CC and an attempt to encourage authors to voluntarily waive their copyrights in their literary works or data and place them in the public domain.

Other CC licenses grants a license while maintaining copyrights. By contrast, CC0 pledges (i) to waiver copyrights, (ii) to grant rights unconditionally and permanently if they cannot be waived, and (iii) not to exercise rights if a license cannot be granted[28].

It is to be noted that, once this declaration is issued for certain data, it cannot be retracted as a number of users will have begun using the data in a variety of ways.

As of July 2014, public comments are being gathered concerning a Japanese version. A formal version is to be disclosed to the public in the near future.

The adoption of CC0 means the waiver of data-related rights, including copyrights, neighboring rights, and portrait rights and the permission of unconditional free use.

At the same time, CC0 declares that CC will not be responsible for any works or data nor make any representation or warranty.

The following is the text of the CC0 rule on use (Legal Code)[29]. The Commons Certificate is a summary of the legal code made understandable to the general public, and does not substitute the code.

---

[27] http://creativecommons.org/choose/zero/

[28] If the waiver of a copyright in certain data is prohibited under law, CC0 enables such data to have the same effect as the waiver of the copyright. For example, the transfer of a moral right is prohibited under the Copyright Act of Japan. Data may be placed in the public domain if the copyright holders pledge that they will not exercise their copyright permanently. However, if ministries and agencies are to grant a CC0, there still remains the problem whether they may waive their right in data at their own discretion where the National Property Act prohibits the transfer of national property without reasonable compensation.

[29] As of April 2014, public comments are being collected. The wording may be changed. http://wiki.creativecommons.org/File:CC0v1_pubcom_JP.pdf

Table 5-5 CC0 Commons

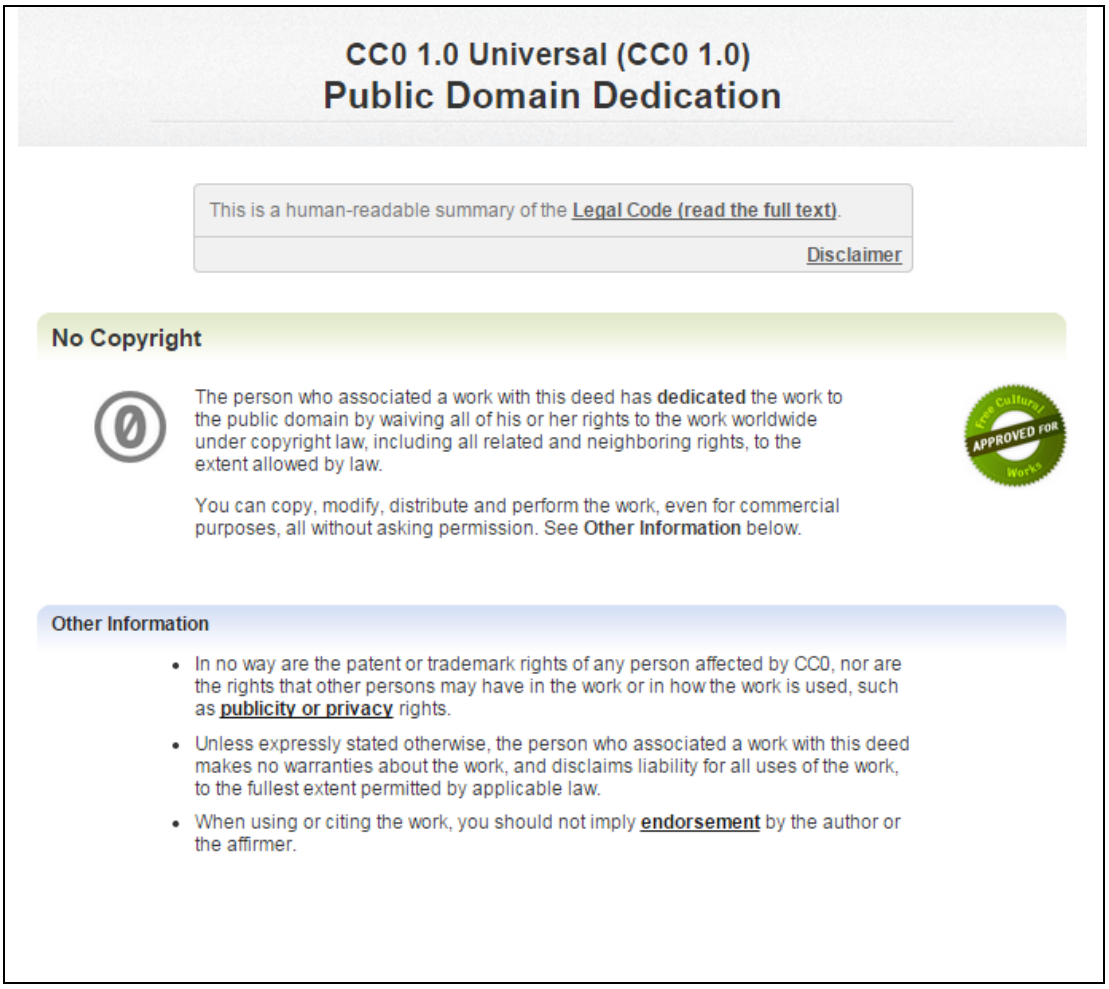


Table 5-6 Text of the CC0 Rule on Use

creative commons

CC0 1.0 Universal

Official translations of this legal tool are available in other languages.

CREATIVE COMMONS CORPORATION IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS DOCUMENT DOES NOT CREATE AN ATTORNEY-CLIENT RELATIONSHIP. CREATIVE COMMONS PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. CREATIVE COMMONS MAKES NO WARRANTIES REGARDING THE USE OF THIS DOCUMENT OR THE INFORMATION OR WORKS PROVIDED HEREUNDER, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM THE USE OF THIS DOCUMENT OR THE INFORMATION OR WORKS PROVIDED HEREUNDER.

Statement of Purpose

The laws of most jurisdictions throughout the world automatically confer exclusive Copyright and Related Rights (defined below) upon the creator and subsequent owner(s) (each and all, an "owner") of an original work of authorship and/or a database (each, a "Work").

Certain owners wish to permanently relinquish those rights to a Work for the purpose of contributing to a commons of creative, cultural and scientific works ("Commons") that the public can reliably and without fear of later claims of infringement build upon, modify, incorporate in other works, reuse and redistribute as freely as possible in any form whatsoever and for any purposes, including without limitation commercial purposes. These owners may contribute to the Commons to promote the ideal of a free culture and the further production of creative, cultural and scientific works, or to gain reputation or greater distribution for their Work in part through the use and efforts of others.

For these and/or other purposes and motivations, and without any expectation of additional consideration or compensation, the person associating CC0 with a Work (the "Affirmer"), to the extent that he or she is an owner of Copyright and Related Rights in the Work, voluntarily elects to apply CC0 to the Work and publicly distribute the Work under its terms, with knowledge of his or her Copyright and Related Rights in the Work and the meaning and intended legal effect of CC0 on those rights.

1. Copyright and Related Rights. A Work made available under CC0 may be protected by copyright and related or neighboring rights ("Copyright and Related Rights"). Copyright and Related Rights include, but are not limited to, the following:

i. the right to reproduce, adapt, distribute, perform, display, communicate, and translate a Work;

ii. moral rights retained by the original author(s) and/or performer(s);

iii. publicity and privacy rights pertaining to a person's image or likeness depicted in a Work;

iv. rights protecting against unfair competition in regards to a Work, subject to the limitations in paragraph 4(a), below;

v. rights protecting the extraction, dissemination, use and reuse of data in a Work;

vi. database rights (such as those arising under Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, and under any national implementation thereof, including any amended or successor version of such directive); and

vii. other similar, equivalent or corresponding rights throughout the world based on applicable law or treaty, and any national implementations thereof.

2. Waiver. To the greatest extent permitted by, but not in contravention of, applicable law, Affirmer hereby overtly, fully, permanently, irrevocably and unconditionally waives, abandons, and surrenders all of Affirmer's Copyright and Related Rights and associated claims and causes of action, whether now known or unknown (including existing as well as future claims and causes of action), in the Work (i) in all territories worldwide, (ii) for the maximum duration provided by applicable law or treaty (including future time extensions), (iii) in any current or future medium and for any number of copies, and (iv) for any purpose whatsoever, including without limitation commercial, advertising or promotional purposes (the "Waiver"). Affirmer makes the Waiver for the benefit of each member of the public at large and to the detriment of Affirmer's heirs and successors, fully intending that such Waiver shall not be subject to revocation, rescission, cancellation, termination, or any other legal or equitable action to disrupt the quiet enjoyment of the Work by the public as contemplated by Affirmer's express Statement of Purpose.

3. Public License Fallback. Should any part of the Waiver for any reason be judged legally invalid or ineffective under applicable law, then the Waiver shall be preserved to the maximum extent permitted taking into account Affirmer's express Statement of Purpose. In addition, to the extent the Waiver is so judged Affirmer hereby grants to each affected person a royalty-free, non transferable, non sublicensable, non exclusive, irrevocable and unconditional license to exercise Affirmer's Copyright and Related Rights in the Work (i) in all territories worldwide, (ii) for the maximum duration provided by applicable law or treaty (including future time extensions), (iii) in any current or future medium and for any number of copies, and (iv) for any purpose whatsoever, including without limitation commercial, advertising or promotional purposes (the "License"). The License shall be deemed effective as of the date CC0 was applied by Affirmer to the Work. Should any part of the License for any reason be judged legally invalid or ineffective under applicable law, such partial invalidity or ineffectiveness shall not invalidate the remainder of the License, and in such case Affirmer hereby affirms that he or she will not (i) exercise any of his or her remaining Copyright and Related Rights in the Work or (ii) assert any associated claims and causes of action with respect to the Work, in either case contrary to Affirmer's express Statement of Purpose.

4. Limitations and Disclaimers.

a. No trademark or patent rights held by Affirmer are waived, abandoned, surrendered, licensed or otherwise affected by this document.

b. Affirmer offers the Work as-is and makes no representations or warranties of any kind concerning the Work, express, implied, statutory or otherwise, including without limitation warranties of title, merchantability, fitness for a particular purpose, non infringement, or the absence of latent or other defects, accuracy, or the present or absence of errors, whether or not discoverable, all to the greatest extent permissible under applicable law.

c. Affirmer disclaims responsibility for clearing rights of other persons that may apply to the Work or any use thereof, including without limitation any person's Copyright and Related Rights in the Work. Further, Affirmer disclaims responsibility for obtaining any necessary consents, permissions or other rights required for any use of the Work.

d. Affirmer understands and acknowledges that Creative Commons is not a party to this document and has no duty or obligation with respect to this CC0 or use of the Work.

Additional languages available: **français**, **Nederlands**. Please read the **FAQ** for more information about official translations.

5.4.2 Using CC0

CC0 is used as follows:[30]

1. On the CC homepage, open the Select Rule page (http: //creativecommons.org/choose/). On top of the page, click "Want public domain instead?" (the red frame in Figure 5-5).

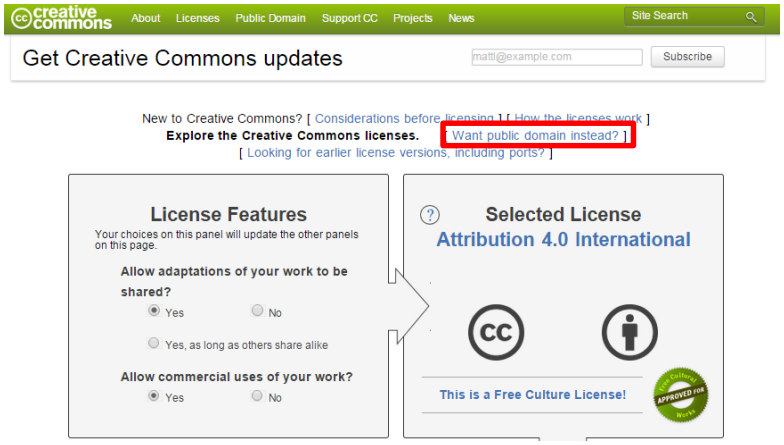


Figure 5-5 Select Rule Page

2. On CC0, click "Use this tool" (the red frame in Figure 5-6).

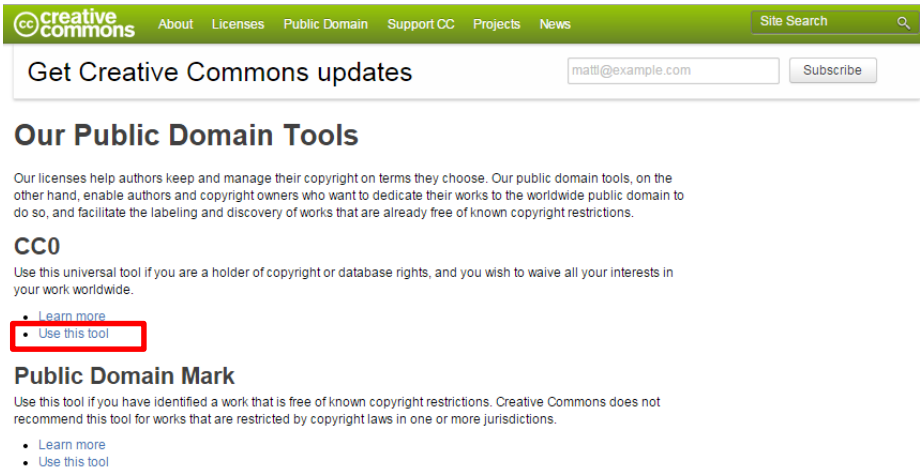


Figure 5-6 Select Public Domain Screen

3. On the CC0 Start screen, click "Start" (the red frame in Figure 5-7).
   - Check that all rights in the CC0 data are owned by the user.

[30]Compiled based on the Select License page of Creative Commons homepage (http://creativecommons.org/choose/).

- If the data contains any third-party right, the user needs to obtain the consent of the third party.

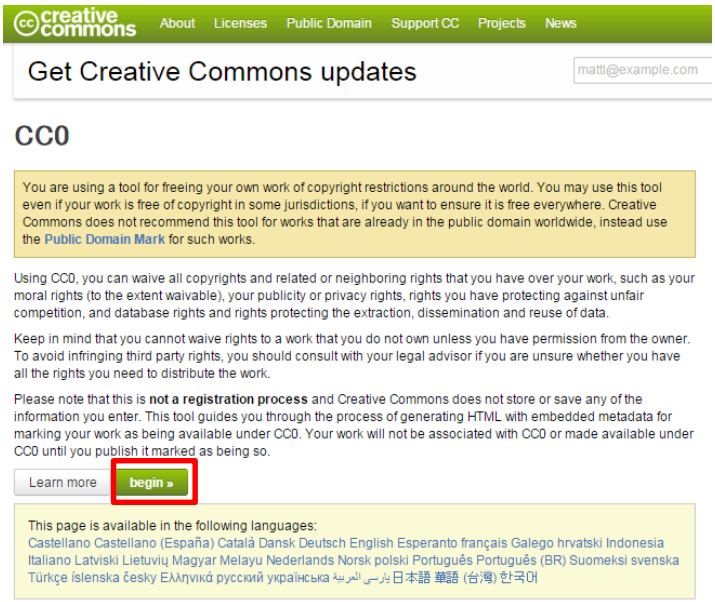


Figure 5-7 Start CC0 Screen

4. In Figure 5-8, check the two options in the red frame, and click "Next."
   - "I hereby waive all copyright and related or neighboring rights together with all associated claims and causes of action with respect to this work to the extent possible under the law."
   - "I have read and understand the terms and intended legal effect of CC0, and hereby voluntarily elect to apply it to this work."

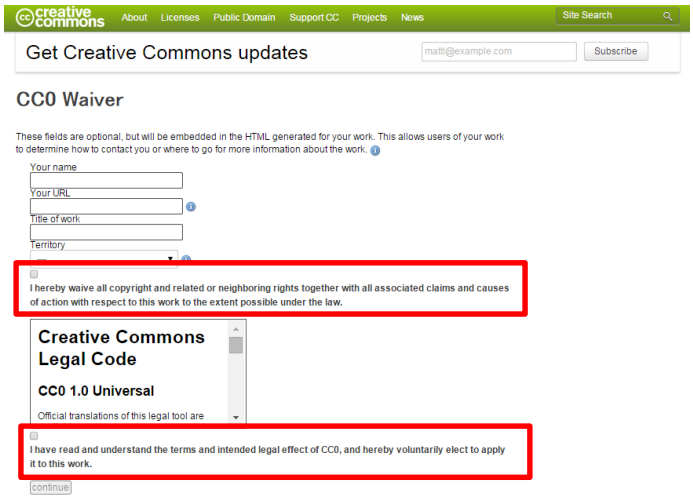


Figure 5-8 Select CC0 Screen

5. In Figure 5-9, select "Yes, I waive" (the red frame in Figure 5-9).

Figure 5-9 Check Screen

6. Then, paste the code (the red frame in Figure 5-10) on the next page to the source code on the homepage.

Figure 5-10 Code of Rules on Use

5.5 Government Standard Terms of Use (Version 1.0)

5.5.1 About Government Standard Terms of Use (Version 1.0)

As already discussed, the Government Standard Terms of Use (Version 1.0) were complied by the **e-Government Open Data Conference** of Working-level Personnel to be applied widely to official homepages, database sites, individual service sites, and other Internet sites of ministries and agencies of the national government (including government institutions, their local offices, and subsections of ministries and agencies). Opinions of experts and ministries and agencies were collected and reflected to compile these standard terms.

Basically, the standard terms of use permit users to make free use of data to the extent that they state the sources of those data, in any manner, including reproduction, public transmission, translation, modification, and transformation. Ministries and agencies may determine how users should state sources, but they are required to give examples of describing sources. They are advised to use easy-to-understand expressions for the text of their rules on use.

The standard terms of use are different from CC-BY in two main points: (i) the former sets forth a provision prohibiting use that is against laws and ordinances or public order and a provision that prohibits use that may threaten the security of the nation and the people; in addition, (ii) the standard terms of use obligates a statement to the effect that the user has edited, processed, or otherwise modified public data and set forth a provision that prohibit the user from disclosing or using their edited or processed data in such a way as if such data had been created by the government (or ministries and agencies). These provisions were included in the standard terms of use, based on opinions from ministries and agencies, which expressed their concerns that there are potential threats from abuse to the security of the nation and people and that falsified data could be misunderstood as original data created by the national government (or ministries and agencies etc.). Taking these opinions into consideration, the standard terms of use apply these requirements as widely as possible to public content owned by the national government.

The Government Standard Terms of Use (Version 1.0) require ministries and agencies to explain to user in an easy-to-understand manner if secondary use of public data is restricted by any specific law or regulation.

For data carrying other rights than copyrights of third parties (e.g. publicity rights), users must obtain a license from such third parties on their own responsibility except where their rights are appropriately processed. Ministries and agencies are required to state the presence or absence of any third party right in a clear and easy-to-understand manner. At the same time, however, ministries and agencies may determine how rights in the content should be expressed, as they are already using various ways of expression.

Ministries and agencies are required to at least adopt the Government Standard Terms of Use

(Version 1.0) as the rule on use of their homepage. If they impose more stringent rule, they are responsible for describing specific and reasonable grounds for taking such policy. If CC-BY or CC0 is adopted, ministries and agencies do not need to state specific reasons, as these licenses provide less limited rules on use than those under the Government Standard Terms of Use (Version 1.0).

The materials provided in Table 5- and Table 5- were presented at the 6th meeting of CIO Conof Working-level Personnel on June 19, 2014[31].

Table 5-7 Government Standard Terms of Use (Version 1.0)

Government Standard Terms of Use (Version 1.0)

Note: Blue sections are to be filled by the relevant government ministry or agency.

Note: Red sections are to be filled with a description of the item (not wording of the rule on use)

1. Use of content of this homepage

Unless a different rule on use is otherwise applied, information (hereinafter referred to as "content") disclosed to the public on this **homepage** may be reproduced, transmitted, translated, transformed, or otherwise handled in accordance with the terms of 1) to 7) below. Such content may be used also for commercial purposes. (For content subject to a different rule on use, see "2. Content subject to a different rule.")

When you start using the content, you are deemed to have agreed to this rule on use."

*** The wording of "homepage" may be changed by ministry or agency to "website," "site," etc.**

1) Indicating sources

a. You must state the source of the content used, in such a way as shown below:
(Example)
Source: AAA Ministry homepage (URL of the page)
Source: "XXX trend survey "(AAA Ministry) (URL of the page) (Used on mm/dd/yy), etc.

b. If you use the content after editing, processing, or otherwise modifying it, you must state that you have edited, processed, or otherwise transformed it, in addition to the source as described above. You are prohibited from disclosing or using your edited/processed information in such a way as if such information had been created by the national government (including ministries and agencies).
(Example of using content after editing, processing, or otherwise handling it)
"XXX trend survey" (AAA Ministry) (URL of the page) processed to create " XXX trend survey " (AAA Ministry) (URL of the page) used and by YYY Company to create this content

2) Avoiding infringement of a third party right

a. Some content may carry copyrights or other rights of a third party (except the national government; the same shall apply hereinafter). For such content having such rights (including portrait rights and publicity right, etc.), the user is required to obtain a relevant license from the third party, except where such right is explicitly waived.

b. For any content having a third party right that is stated directly or indirectly but not stating the specific part of the content subject to such right, users are required to confirm on their own responsibility which part of the content is subject to such right.
(Example of content indicating the presence of a third party right [provided in a separate sheet]

c. Content provided through an API (Application Programming Interface), etc. with an external database is subject to the rule on use of the original provider.
(Example of content provided through an API, etc. with an external database) [provided in a

[31] http://www.kantei.go.jp/jp/singi/it2/densi/

separate sheet]

***Remove this paragraph if the content is not applicable.**

d. Some content having a copyright or any other right may be used without the approval of the copyright holder if the use of such content, including citations, is permitted by the Copyright Act.

3) Prohibited use

a. Users are prohibited from using the content by

(A) using it in violation of laws, regulations, ordinances, or public order, or

(B) threatening the security of the nation and its people.

4) Use of content limited by a specific law or regulation

a. The use of some content is limited udder a specific law or regulations. Specifically note the following laws and regulations. For more details, see the relevant link destination page.

For AAA (restrictive provision) related to use of BBB (name of content) in accordance with the CCC Law (name of law) (link to the relevant page)

For DDD (restrictive provision) related to use of EEE (name of content) in accordance with the FFF Law (name of law) (link to the relevant page)

***Remove this paragraph if the content is not applicable.**

5) Governing law and jurisdiction

a. This rule on use is interpreted in accordance with the law of Japan.

b. Any dispute related to this rule on use or the use of content subject thereto will be brought to the exclusive jurisdiction of the district court governing the location of the organization that discloses to the public its content or its rule on use related to such dispute.

6) Disclaimers

a. The national government will not be held responsible for any act of the user conduced by using the content (including information edited, processed, or otherwise handled by the user).

b. The content of this rule on use is subject to change, relocation, deletion, etc. without prior notice.

7) Other

a. This rule on use does not limit any use, including citation, permitted by the Copyright Act.

b. This rule on use became effective on June 19, 2014. This rule on use conforms to the Government Standard Terms of Use (Version 1.0). This rule on use is subject to future change and will be reviewed by fiscal year 2015.

2. Content subject to a different rule

The following content is subject to a rule on use different from this rule on use. For details, the user should see the relevant link destination page.

Use of XXX (name of content) (link to the relevant page)

***To set forth a rule on use not based on any specific law or regulation, the relevant ministry or agency shall be responsible for clearly giving, on the link destination page, its justification for applying such separately rule on use.**

***Remove this paragraph if the content is not applicable.**

*Each ministry or agency may, at its own discretion, set forth provisions on links, privacy policy, accessibility, and disclaimers for the entire homepage (except use of content) unless they conflict with the above rule on use.

Table 5-8 Explanatory Notes on the Government Standard Terms of Use (Version 1.0)

Explanatory Notes on the Government Standard Terms of Use (Version 1.0)

June 19, 2014

IT Strategic Headquarters

<Composition and basic principles>

The Government Standard Terms of Use (Version 1.0) was been compiled as model for the rule on use of the homepage of ministries and agencies in reviewing their rules for their current homepage, in response to the release of the Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use (Liaison Conference of Chief Information Officers (CIOs) of Public Offices and Ministries decision, June 25, 2013) (hereinafter referred to as "Guidelines"). The Guidelines state that, since the national government is entitled to determine the rule on use of works copyrighted by the authors, it should express its intention concerning the use of its works in a unified form of statement in such a way as to permit wider secondary use (except where the government limits secondary use based on reasonable grounds other than the copyright).

The Creative Commons licenses (hereinafter referred to as "CC-BY") or many other licenses compatible with CC-BY are widely used globally as a rule on secondary use of open data. Different data can easily be combined for use if they are made open in accordance with the same rule on use. For this reason, it is desirable for users to adopt CC-BY (or its compatible licenses) if they intend to combine data collected from international sources.

During the course of discussion, some ministries and agencies suggested that a different rule on use should be used by separate government organizations because a wide diversity of data are made open on their homepages, which may possibly interfere exclusive application of CC-BY for secondary use. Others expressed concerns that any information that any information edited, processed, or otherwise created by using data disclosed by the government might be interpreted to have been created by the government itself.

To provide an easy-to-understand unified rule on use in compliance with the Guidelines, the Government Standard Terms of Use (Version 1.0) use simple wording easy for general users to understand. Based on opinions from ministries and agencies, these standard terms are designed to apply as widely as possible to content owned by the government.

If ministries and agencies are unable to apply the unified rule on use under Section 1 of the Government Standard Terms of Use (Version 1.0), the they may set up a rule appropriate to the nature of such content (provided that, if they set up a different rule on use imposing any limitation that are not based on an individual law or regulation, then they shall clearly define the scope of the applicable content and provide reasonable grounds of applying such different rule).

To comply with the Guidelines, ministries and agencies may adjust their rule on use of the content on their homepage (including copyright and disclaimer provisions) to the Government Standard Terms of Use (Version 1.0).

The Government Standard Terms of Use (Version 1.0) are so named in order to make clear the government intent to have users recognize that a unified rule on use is applied to the homepages of ministries and agencies. In addition, the naming will make it easier for users to recognize any change to the version by a citation of the original rule on use.

It is important for ministries and agencies to indicate their rule on use on their homepage in a concise, easy-to-understand manner. On their homepage, they are required to clearly indicate the link to their rule on use and otherwise design their homepage in such a way that users can access their rules on use at any time.

<Explanation of paragraphs>

1. Use of content of this homepage

Unless a different rule on use is otherwise applied, information (hereinafter referred to as "content") disclosed to the public on this **homepage** may be reproduced, transmitted, translated, transformed, or otherwise handled in accordance with the terms of 1) to 7) below. Such content may be used also for commercial purposes. (For content subject to a different rule on use, see "2. Content subject to a different rule.")

When you start using the content, you are deemed to have agreed to this rule on use."

*** The wording of "homepage" may be changed by ministry or agency to "website," "site," etc.**

[Explanation]

This part provides as a unified rule of the Government Standard Terms of Use (Version 1.0) that content may be reproduced, distributed, translated, transformed, modified, or otherwise changed freely if such content conforms to terms 1) to 5), except content that is subject to the provision of "2. Content subject to a different rule on use."

The Government Standard Terms of Use (Version 1.0) are expected to be applied widely to homepages, database sites, individual service sites, and other Internet sites operated by ministries and agencies of the national government (including facilities, institutions, local branches, and other organizations operated by ministries and agencies). For "homepage," ministries and agencies may instead use such terms as "website," and "site."

In order to prevent content users from claiming that they were not aware of a rule on use, a provision shall be included that users will be deemed to have agreed to the rule on use before they use such content.

The Copyright Act permits free use of content not having the nature of a literary work (including tabular data, figures and tables, and simple graphs). Thus, such content would not need covering by the Government Standard Terms of Use (Version 1.0). These standard terms of use basically intend to permit free use if the source is indicated, and do not lead to any limitation on use of such content if they are applied. It is considered desirable to have a source indicated for content not having the nature of a literary work. There are matters to be set forth, regardless of whether any content has the nature of a literary work. It is extremely difficult to confirm the presence of the nature of a literary work, and thus unified treatment of varying types of content would be more advantageous to users. For those reasons, the Government Standard Terms of Use (Version 1.0) will apply to content without the nature of a literary work. Rules on use are designed to grant a license to use copyrighted content owned by the government, while they have only an obligatory effect on content that is not copyrighted by the government.

1) Indicating sources

a. You must state the source of the content used, in such a way as shown below:
(Example)
Source: AAA Ministry homepage (URL of the page)
Source: "XXX trend survey "(AAA Ministry) (URL of the page) (Used on mm/dd/yy), etc.

b. If you use the content after editing, processing, or otherwise modifying it, you must state that you have edited, processed, or otherwise transformed it, in addition to the source as described above. You are prohibited from disclosing or using your edited/processed information in such a way as if such information had been created by the national government (including ministries and agencies).
(Example of using content after editing, processing, or otherwise handling it)
"XXX trend survey" (AAA Ministry) (URL of the page) processed to create " XXX trend survey " (AAA Ministry) (URL of the page) used and by YYY Company to create this content

[Explanation]

Paragraph a explains how to indicate sources and provides that users should indicate a source of the content used.

The method of stating a source may be different among ministries and agencies. This paragraph gives examples of ministries and agencies indicating a source that may be easily transcribed by the user. Ministries and agencies are required to provide these examples of indicating a source in the blue section.

Paragraph b requires the users to state that they have edited, processed, or otherwise changed relevant content if they have actually done so as secondary use. This paragraph also prevents users from disclosing or using their edited/processed information in such a way as if such information had been created by the national government, or its ministries or agencies. For example, users are not permitted from stating as if the data had been disclosed to the public by a ministry or agency after the users falsify figures in statistical data created by such ministry or agency

(Reference) Relationship with CC-BY

Although the government rules on use set forth the same source citation procedure as required in CC-BY, the former differs from CC-BY in the following terms:

- The rules on use may impose and indicate their own policy, including a requirement for indication of the fact that relevant content has been edited, processed, or otherwise changed, if any, and the prohibition of use of content deemed by disclosing ministries and agencies to be against public order.
- CC-BY is a license to cover copyrighted content. It should be noted that there are conditions and matters to be set forth uniformly regardless of whether certain content has the nature of a literary work.
- CC-BY licenses set forth professional conditions for a copyright statement to be used to redistribute CC-

BY licensed content or to combine such content with other content. To make a rule on use simpler and easier to understand, it is not always necessary to adopt these professional conditions.

2) Avoiding infringement of a third party right

a. Some content may carry copyrights or other rights of a third party (except the national government; the same shall apply hereinafter). For such content having such rights (including portrait rights and publicity right, etc.), the user is required to obtain a relevant license from the third party, except where such right is explicitly waived.

b. For any content having a third party right that is stated directly or indirectly but not stating the specific part of the content subject to such right, users are required to confirm on their own responsibility which part of the content is subject to such right.

(Example of content indicating the presence of a third party right [provided in a separate sheet]

c. Content provided through an API (Application Programming Interface), etc. with an external database is subject to the rule on use of the original provider.

(Example of content provided through an API, etc. with an external database) [provided in a separate sheet]

***Remove this paragraph if the content is not applicable.**

d. Some content having a copyright or any other right may be used without the approval of the copyright holder if the use of such content, including citations, is permitted by the Copyright Act.

[Explanation]

Some content disclosed to the public on the homepage of ministries and agencies carries rights of other persons than the national government (hereinafter referred to as “third party"). Content carrying any third party right requires a license from the third party, unless otherwise permitted by the Copyright Act.

Most of the content currently posted on the homepage of ministries or agencies is not intended as open data. The national government (ministries and agencies) has insufficient information on any third-party right in data disclosed by the government. In addition, the majority of data disclosed by the government have not been processed for secondary use.

For this reason, Item a provides that users themselves should be responsible for obtaining a license from the third party for content carrying a third party right, unless the data state that users do not require a license because the license has been properly processed.

It is impossible for ministries and agencies to identify all content carrying a third party right. To help users to determine whether certain content carries a third party right, Item b. provides some examples in Appendix of descriptions stating that the content carries a third party right (including examples of citations in white papers to indicate the presence of a third party right).

It is desirable for the government, upon any inquiry from users, to conduct a survey to the extent possible to determine whether the data carry a third party right and provide the users with relevant information.

Some homepages of ministries and agencies provide content obtained from external databases through API coordination, including SNS content shown in the window in real-time. Item c provides that such content should be subject to a rule on use adopted by the information provider.

3) Prohibited use

a. You are prohibited from using the content by

(A) using it in violation of laws, regulations, ordinances, or public order, or

(B) threatening the security of the nation and its people.

[Explanation]

This paragraph explains the ways of using content made open by the national government (ministries and agencies) that are deemed inappropriate by the government, and states that the rules on use prohibit such use. Any use of content in a prohibited manner will lead to retraction of the license.

Maps, marine charts, aerial charts, warnings/predictions, disaster prevention information, etc. could threaten the security of the nation and its people, depending on how such information is used. Experts insist that any use of such information should be subject to appropriate actions if such information is used in violation of laws, regulations, or ordinances or against public order. Such content accounts for a substantial part of public data, and thus it is

considered necessary to apply a unified rule to such content.

Data are considered to be against public order if they involve a crime or a work that is against marital or sexual ethics, involve gambling, extremely restrict the freedom of individuals, involve unfair transactions or transactions seeking excessive profits, contain sexually explicit materials, involve threatening use, or otherwise involve other illegal acts.

4) Use of content limited by a specific law or regulation
a. The use of some content is limited udder a specific law or regulations. Specifically note the following laws and regulations. For more details, see the relevant link destination page.
For AAA (restrictive provision) related to use of BBB (name of content) in accordance with the CCC Law (name of law) (link to the relevant page)
For DDD (restrictive provision) related to use of EEE (name of content) in accordance with the FFF Law (name of law) (link to the relevant page)
***Remove this paragraph if the content is not applicable.**

[Explanation]
Some data disclosed on homepages of ministries and agencies are subject to limitations under specific laws or regulations. For example, the secondary use of some maps (showing basic measurements) is subject to approval of Geographical Survey Institute pursuant to the Survey Act when they are to be reproduced, distributed, or otherwise processed.

This paragraph directs users' attention to content that is subject to limitations under specific laws or regulations.

It is desirable for ministries and agencies to specify important content that is subject to restrictions under specific laws or regulations.

5) Governing law and jurisdiction
a. This rule on use is interpreted in accordance with the law of Japan.
b. Any dispute related to this rule on use or the use of content subject thereto will be brought to the exclusive jurisdiction of the district court governing the location of the organization that discloses to the public its content or its rule on use related to such dispute.

[Explanation]
Item a provides that these terms of use are governed by the law of Japan.

Item b provides that ministries and agencies or content users should file any suit in connection with the use of content and these terms of use exclusively with the district court in the jurisdiction where the ministry or the agency is located.

In these rules on use, "dispute" means one between the ministry or the agency that provide the content and the content user.

6) Disclaimers
a. The national government will not be held responsible for any act of the user conduced by using the content (including information edited, processed, or otherwise handled by the user).
b. The content of this rule on use is subject to change, relocation, deletion, etc. without prior notice.

[Explanation]
This paragraph sets forth disclaimers by the content provider with respect to the use of content.

Item a provides that the national government (ministries and agencies) shall not be held responsible for any act of the user by using the content, as content disclosed on homepages of ministries and agencies can be used in a variety of ways and thus because the government is unable to predict how the user will use such content.

For example, if any content is not accurate, the national government (ministries and agencies) will not be responsible for any damage incurred by the user.

Item b informs the user that content disclosed to the public on homepages of ministries and agencies is subject to change, relocation, deletion, etc. without notice.

7) Other
   a. This rule on use does not limit any use, including citation, permitted by the Copyright Act.
   b. This rule on use became effective on mm, dd, 2014. This rule on use conforms to the Government Standard Terms of Use (Version 1.0). This rule on use is subject to future change and will be reviewed by fiscal year 2015.

[Explanation]
This paragraph describes matters to be recognized by users when they are to use content disclosed to the public on homepages of ministries and agencies.

Item a provides that this rule on use does not impose the limitations on acts that are restricted under the Copyright Act (Article 30 to Article 47.9).

Examples of use of content subject to limitation under the Copyright Act include reproduction for personal use, citations made within a reasonable extent in compliance with fair practices for the purpose of reporting, criticism, or research, or for other purposes, and reproduction to the extent necessary to use in classes at schools or other non-profit educational institutions. (In some cases, the Copyright Act requires indication of the source of a literary work to a reasonable extent, depending on how the content is to be reproduced or used.)

(Reference) Agency for Cultural Affairs homepage " Outline of the Copyright System,"
http: //www.bunka.go.jp/chosakuken/gaiyou.html

This term of use also applies to content not having the nature of a literary work and provides that content not having the nature of a literary work may be used in such ways as described above.

Item b requires ministries and agencies to indicate on their homepage the initial date of applying the Government Standard Terms of Use (Version 1.0).

Ministries and agencies are also required to state on their homepages that their respective rule on use is based on the Government Standard Terms of Use (Version 1.0).

Item b further draws the attention of users to the provision that the rule on use is subject to change. The paragraph also provides that the rule on use will be reviewed by the end of FY 2015 as specified in the Declaration to be the World's Most Advanced IT Nation (Cabinet decision, June 14 2013) and **Roadmap for** the e-Government **Open Data Promotion** (IT Strategic Headquarters decision, June 14, 2013) in order to achieve the same level of public disclosure as in other advanced countries.

This review will pick up discussions on Section 1.3) (Prohibited use) as a n important subject in order to make the rule on use more compatible with globally adopted CC-BY licenses, cosidering how wide contents in compliance with updated Government Standard Terms of Use (Version 1.0) are utilized

2. Content subject to a different rule
The following content is subject to a rule on use different from this rule on use. For details, the user should see the relevant link destination page.
Use of XXX (name of content) (link to the relevant page)

***To set forth a rule on use not based on any specific law or regulation, the relevant ministry or agency shall be responsible for clearly giving, on the link destination page, its justification for applying such separately rule on use.**
***Remove this paragraph if the content is not applicable.**

[Explanation]
For some content disclosed on homepages of ministries and agencies that is not subject to limitations under specific laws and regulations, it would be necessary to establish a rule on use different from the unified rule under Section 1 of the Government Standard Terms of Use (Version 1.0).

This paragraph provides that the scope of any such content should be specified clearly for users and that the details of the new rule on use and the reasonable grounds for imposing such rule on use, on a link destination page separately created.

Some content may be disclosed pursuant to CC-BY or CC0 (Note), within the extent permitted under Section 1. In such cases, applicable content should be specified and a statement should be included to indicate that CC-BY or CC0 is applied. Such content may be specified with such wording as "Content indicating a CC-BY mark."

(Note) A Creative Commons license indicating that no right is owned. As of March 24, 2014, Creative Commons Japan is working on a Japanese version of its licenses based on public comments collected in a survey.

The rule on use under Section 1 does not apply to content that is subject to a different rule on use under Section 2. Where necessary, the different rule on use should set forth matters as provided in 4) and 5) of Section 1.

The content subject to the different rule on use and the descriptions in the different rule should be reviewed as necessary, depending on changes in the content or in the environment where relevant data are used.

*Each ministry or agency may, at its own discretion, set forth provisions on links, privacy policy, accessibility, and disclaimers for the entire homepage (except use of content) unless they conflict with the above rule on use.

[Explanation]
The Government Standard Terms of Use (Version 1.0) were compiled as a rule concerning the use of content (matters described as "copyrights" and "disclaimers" on current homepages of ministries and agencies homepage).

Links, privacy policy, accessibility, and disclaimers (excluding disclaimers with respect to the use of content) are configured and described in different ways on homepages of ministries and agencies. As these do not need unifying, and ministries and agencies may be free to set forth provisions concerning these matters at their own discretion, unless such provisions conflicts with relevant provisions under the Government Standard Terms of Use (Version 1.0).

Chapter 6. Comparison of Rules on Use, and the Most Desirable Rule on Use

In Chapter 5, we presented three types of rules on use of open data; namely CC0, CC-BY, the Government Standard Terms of Use (Version 1.0), and the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site).

In this Chapter, we make a comparison of CC0, CC-BY and the Government Standard Terms of Use (Version 1.0) from the perspectives of both information users and the information provider so that the information provider can grasp the difference among different use rules at the time of reviewing which rule to adopt (the Standard Terms of Use of "DATA.GO.JP" is included in CC-BY on Table 6-1). In addition, based on the results of the comparison, we elucidate the most desirable rule on use at the time of publicly disclosing data as open data.

6.1 Comparison from the perspective of information users

By comparing the three types of rules on use of open data from the perspective of information users, as to (i) whether the data can be applied to secondary use freely, and (ii) whether the mash-up with data from foreign countries is easy, we get the following results.

Table 6-1 Comparison from the Perspective of Information Users

	CC0	CC-BY (※)	The Government Standard Terms of Use (Version 1.0)
(i) Whether the data can be applied to secondary use freely.	Possible	Possible, by indicating the source of data/information.	In addition to the source indication, there are certain prohibited matters.
(ii) Whether the mash-up with data from foreign countries (many of them are based on CC-BY) is easy.	Easy	When the number gets larger, the number of source indications will increase.	It is necessary to understand the difference with CC-BY.

※This is adopted in the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site)

In the case of CC0, since the copyright is abandoned, secondary use of the data is possible for information users without any restriction. In addition, the mash-up with data from foreign countries is easy.

In the case of CC-BY, since a condition is imposed that users are to indicate the source at the time of making secondary use, it is required of information users to abide by the condition. There is no limit about the range of the use-mode by which secondary use is permitted. In respect of the mash-up with data from foreign countries, since there are many cases that CC-BY is adopted in foreign countries, as stated before, there are many cases where data can be used by combining them under the same conditions.

In the case of the Government Standard Terms of Use (Version 1.0), secondary use is permitted if information users indicate the source of data/information, like the case of CC-BY. However, provisions are incorporated in the Terms to prohibit "any use that goes against laws and regulations, ordinances, and public order and morality" as well as "any use that gives threat to the safety and security of the nation and its people", and because the concrete use modes so prohibited are not necessarily clear to information users, there is a possibility that a chilling effect is generated at the time of using published data. Furthermore, with respect to the mash-up with data from foreign countries, it is necessary to comprehend the rules on use of both CC-BY and the Government Standard Terms of Use (Version 1.0), as the use conditions in the Terms and those in CC-BY adopted in foreign countries are not the same.

6.2 Comparison from the perspective of the information provider

From the perspective of the information provider, the following three points are important; namely, (i) there is no necessity of providing any guarantee concerning the published data (no guarantee), (ii) such a situation where any data tampered/altered under the false name of the information provider are publicly disclosed can be prevented, and (iii) any data use that the information provider generally considers not desirable can be prohibited.

Table 6-2 Comparison from the Perspective of the Information Provider

	CC0	CC-BY (※)	The Government Standard Terms of Use (Version 1.0)
(i) There is no necessity of providing guarantee concerning the published data. (no guarantee)	There is a provision about no guarantee.	There is a provision about no guarantee.	There is a provision about no guarantee.
(ii) Situation where data tampered/altered under the false name of the information provider are publicly disclosed can be prevented.	There is no provision.	There is a provision (legal code Article 5i), but its effectiveness is dubious.	There is a provision (1.1.a), but its effectiveness is dubious.
(iii) Data use that the information provider generally considers not desirable can be prohibited.	There is no provision.	There is no provision.	There are provisions (prohibition of use that goes against laws and regulations, ordinances, and public order and morality, as well as prohibition of use that gives threat to the safety and security of the nation and its people), but its effectiveness is dubious.

※This is adopted in the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site).

As regards (i) above, a provision stipulating that no guarantee is provided to the

published data is incorporated in all of CC0, CC-BY and the Government Standard Terms of Use (Version 1.0).

As regards (ii) above, there exists a provision in CC-BY and the Government Standard Terms of Use (Version 1.0) that prevents the publication of falsified data using the name of the data provider. In the case of CC-BY, deletion of credit information within the possible limit of implementation may be requested, by notifying information users (legal code Article 5i). In the case of the Government Standard Terms of Use (Version 1.0), it obligates information users to describe the fact that some editing and/or processing was performed, if the contents are to be used after their being edited and/or processed, in addition to the description of the source of data/information, and also, it prohibits information users that the edited and/or processed information is to be publicly disclosed or used in a manner as if it were created by the Government (or governmental ministries/agencies, etc.) (1 (a)). In the case of CC0, it does not contain a provision to specifically prohibit such matters.

In the case where CC-BY or the Government Standard Terms of Use (Version 1.0) is used, and if falsification is made by assuming the false name of the information provider, the information provider can request correction based on the rules on use, but there is a possibility that information users do not respond to such request in good faith. Bringing a lawsuit against such a case is among possible countermeasures, but if you consider legal expenses and the time and labor involved, there are many cases where we encounter difficulties in resorting to such measures in practice. Furthermore, in respect of the data with no copyright, rules on use merely have claimable effects, and it is difficult to request correction based on the rules of use when falsification has been conducted by various points in the chain of users. Therefore, the most realistic and effective way will be that we create a situation where anyone can confirm the presence of falsification, by making the original data publicly open.

As regards (iii) above, the Government Standard Terms of Use (Version 1.0) contains provisions prohibiting the “use of data that goes against laws and regulations, ordinances, and public order and morality”, as well as the “use of data that gives threat to the safety and security of the nation and its people”. On the other hand, both CC0 and CC-BY do not contain such prohibitive provisions. In the case of the Government Standard Terms of Use (Version 1.0), it assumes mainly such data with high probability of causing certain disadvantage against the nation and/or its people unless certain specific uses are prevented by concrete measures (e.g. it assumes such things

as maps that could be used or published as if they were created by the Japanese Government despite the fact that the original data were intentionally altered with respect to the description of the territory, and various kinds of warnings, predictions, disaster prevention-related information, and the like, which have possibilities of giving negative effects on the safety of the people as the result that the timing of evacuation is delayed, for instance, due to the dissemination of the data whose contents have been altered as if the degree of disaster were smaller than actually expected).[32]

6.3 Most Desirable Rule on Use at the Time of Publicly Disclosing Data

From the perspective of information users (i.e. usability of open data), it is desirable that CC0 is applied to the types of public data (such as numeric values, simple tables and graphs) to which no copyright occurs and CC-BY is applied to the types of public data to which copyrights occur, respectively. And if and when CC0 or CC-BY is applied, it is necessary to use them after thoroughly examining the full contents of the legal code, and confirming that the data in question satisfy the conditions of CC0 or CC-BY. (Among various cases, there are cases where certain conditions beyond those stipulated in the legal code are later-on added (for example, prohibition of use that goes against public order and morality), but this kind of measure is not thought to be appropriate.) And, although it is possible to apply CC0 to the public data to which copyrights occur, it is necessary to carefully examine the right or wrong of waiving copyrights.

Meanwhile, if we want to continue to incorporate the provisions prohibiting "any use of public data that goes against laws and regulations, ordinances, and public order morality", and "any use of public data that gives threat to the safety and security of the nation and its people", which have been frequently used in conventional rules on use of public data, as well as in the rules on use of open data, we can think of applying the Government Standard Terms of Use (Version 1.0). However, as we pointed out previously, its effectiveness at the time when a violative act has taken place is dubious even if these provisions are stipulated, and as the negative impacts from the chilling effect on the use of data could be rather significant, it is desirable not to adopt such provisions, if there isn't any concrete act which you specifically want to prohibit, and if the provisions are to be incorporated just for the sake of prevention as the reason for incorporating them.

---

[32] Reference material "Comments from various Ministries/Agencies on the contents to which separate rules on use are considered to be required, and ideas about the coordination of such comments" (draft)", 4th meeting of the Rules and Dissemination Working Group of the e-Government Open Data Conference of working-level personnel (February 28, 2013)
http://www.kantei.go.jp/jp/singi/it2/densi/rwg/dai4/siryou4.pdf

It should be noted here that if the incorporation of the above-said provisions is adopted, the rules with such provisions lose the compatibility with the internationally prevailing CC-BY. Moreover, it is necessary to bear in mind that the Government Standard Terms of Use (Version 1.0) is to be reviewed and revised, depending on its operational status in future.

In the event that the data includes any portion to which any third party reserves any right(s), it is necessary to distinguish the portion and to make it easy to see that the portion is not subject to the application of the rules on use such as CC-BY, at the time of disclosing public data as open data , or if it is difficult to distinguish the portion, it is necessary to give warning to information users, indicating that they need to obtain consent from the relevant third party on their own responsibility. And, if there is any legal restriction on the use of data, it is important to make the description of the restriction on use and the underlying laws and regulations easily understandable to information users. On these points, the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site) (see Section 5-3) adopted CC-BY as the rule on use, and calls for attention concerning the third party right separately. Thus, it is desirable to look at this material as reference[33]

The Government Standard Terms of Use (Version 1.0), too, is created, by taking the third party right and legal restrictions into account, but it is not appropriate to adopt the Government Standard Terms of Use (Version 1.0) instead of CC-BY, merely on the account that the warnings concerning the third party right and legal restrictions are required, because that will narrow the range of use by information users.

And among public data, there exist such data as symbol marks and logos of organizations or projects, for which we may not be able to apply any of CC0, CC-BY or the Government Standard Terms of Use (Version 1.0). As regards these data, it is necessary to stipulate a provision, reading like, "2. Contents to which separate rules on use are applicable", in the case of adopting our independent rule (i.e. the Government Standard Terms of Use (Version 1.0)), when any of the rules is to be adopted.

[33] As the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site) does not describe such details as the concrete contents of restriction and the underlying laws and regulations, it is necessary to additionally describe these points.

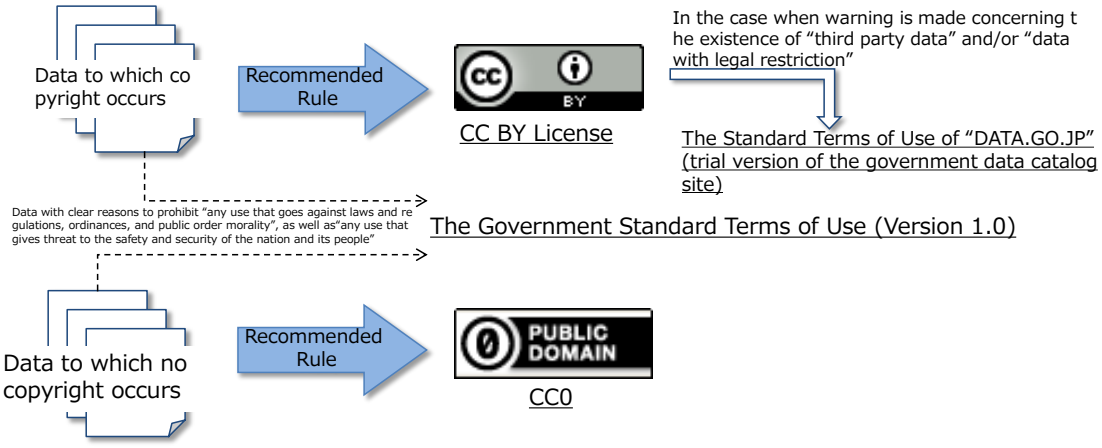


図 6-1 Desirable Rules on Use of Open Data

Meanwhile, in the event where certain inconvenience takes place to any third party due to inappropriate use of the data disclosed as open data, the responsibility lies with the information user who utilized the data in an inappropriate manner, and the responsibility should not be owed to the Government or local governments. However, as of Fiscal Year 2013, because there were concerns that the Government or local governments would be asked to assume the responsibility when problems occurred due to publicly disclosed data, the Government Standard Terms of Use (Version 1.0) incorporated provisions to prohibit such inappropriate use in order to clearly indicate that the Government and local governments do not approve inappropriate use of publicly disclosed data. Therefore, by enlightening the fact that the Government and local governments by no means assume the responsibility even if certain problems have occurred due to publicly disclosed data, this kind of prohibitive provision will become unnecessary (see Section 2.2).

(Supplement) Abuse of publicly Disclosed Data and Responsibility Therefor

As the number of publicly disclosed data increases, the cases of their abuses may also increase. Therefore, promotion of open data will necessarily lead to an increase in the cases of their abuses. However, the reason why open data policies have been adopted in foreign countries is, needless to say, because the merit from open data far exceeds the demerit from the abuse of publicly disclosed data.

It is basically impossible for the information provider to control the use of the data when they are once publicly disclosed, and to prevent abuses thereof. Although it is possible to prohibit the data use which the information provider considers to be undesirable, by imposing rules on use, it is realistically unthinkable to have a situation where anyone who intends to abuse data comes to refrain from conducting the abuse after reading the relevant rules on use, even if such prohibitive provision is established.

As regards the data subject to public disclosure, reasonable screening should be performed by the information provider as to whether or not they would infringe personal rights, or whether or not they might create dangerous results (such as fabrication of dangerous materials), and if certain risks were realized when such screening had not been properly performed, there would be good reason that the information provider is criticized. However, without such a situation, it is not appropriate to judge that the information provider who publicly disclosed the data that had the possibility of being abused has the responsibility, merely from the de facto relations of cause and effect. In the case of open data, because the information provider discloses data, not for his own benefit, but for the universal benefit of information users, he would come to take defensive behavior, should he come to be criticized due to his disclosure of data. As the most effective way (and probably the only way) to reduce the risk of data abuse is not to disclose any data, the information provider would necessarily come to refrain from disclosing any data, when he encounters such judgment about the responsibility.

As described above, it is strongly hoped for that the judgment of the responsibility for cases of data abuses should be a reasonable and unemotional one based on the purport of open data.

Reference: Important Points to note, Concerning the data Containing certain Right(s) of Third Parties

As regards the data containing certain right(s) of third parties, it is basically not possible to apply any rules on use without obtaining consent from the relevant third party

As the means of dealing with such cases, the ideal ways from the standpoint of information users are, (A) the information provider obtains approval from the third party for the secondary use of the data (In this case, it is necessary to obtain approval for the secondary use by unspecified large number of users.), and (B) the information provider distinguishes the portion to which a third party reserves the right as the portion not subject to any secondary use and so clearly indicates. However, if it is difficult for the information provider to obtain approval from the third party or to distinguish the portion not subject to the secondary use due to financial reasons, for instance, (C) the information provider directly or indirectly indicates or suggests that a third party possesses the right to certain portion by describing the source, and arouses the attention of information users that in respect of the portion to which a third party possesses the right, information users need to obtain approval from the relevant third party on their own responsibility.

Concerning the data for which approval for the secondary use was obtained from the third party through the method (A) above, they can be provided under CC0 or CC-BY. And concerning the data which can be distinguished and made subject to the secondary use through the method (B) above, too, they can be provided under CC0 or CC-BY as well. In the case where disclosure of data is made under CC-BY through the method (C) above, it is necessary for the information provider to give warning separately because such warning about the third party right is not included in the legal code. As an example of the warning to be given by the information provider separately when CC-BY is applied, we can name the Standard Terms of Use of "DATA.GO.JP" (trial version of the government data catalog site) (see Section 4.3 and 5-3). On the other hand, in the case of the Government Standard Terms of Use (Version 1.0), warning about the third party right is incorporated in the rule on use, and thus, this matter is already taken into account.

As regards the dealing methods through (A) or (B), a case study conducted by the Data Governance Committee in Fiscal Year 2012, on the "Whitepaper on Information and

Communications in Japan", will be useful as reference[34].

In the meantime, in the case of entrusting certain investigation companies or design companies with data creation in future, if the third party right required for secondary use has been obtained by the consigned companies in advance, or if it is stipulated in the entrustment contracts that the created data are to be delivered in a way that the portion to which the third party has right is clearly distinguished, work to disclose the data in question as open data becomes fairly easy. On this point, it is also stated in the Guideline "Basic principles on the public disclosure of data owned by government ministries and agencies for the promotion of their secondary use" (Decision made at the liaison conference of chief information officers (CIO) of all government ministries and agencies) as follows (see Table 6-3)[35]. The Data Governance Committee has also drawn up a draft specimen of the entrustment contract to be used for that occasion (see Table 6-4)[36].。

表 6-3 Important Points to Note, When Creating/Obtaining Data

With respect to the data which each Ministry or Agency creates or obtain anew after the establishment of this Guideline, efforts must be devoted to reach agreement with all stakeholders in advance so that the secondary use of the data in question which were publicly disclosed through the Internet by each Ministry or Agency is approved. To this end, when examining or executing entrustment or consignment contracts after the establishment of this Guideline, measures giving thought to this principle must to be pursued (for instance, in the case that the contents of an entrustment contract for investigation or the report describing results of a certain study are publicly disclosed by a government ministry/agency through the Internet, they should be created in a way that they do not become an obstacle for the acquisition of approval for the secondary use of the open data in question).

Source: Guideline "Basic principles on the public disclosure of data owned by government ministries and agencies for the promotion of their secondary use" (Decision made at the liaison conference of chief information officers (CIO) of all government ministries and agencies)

[34]Reference is made to http://www.kantei.go.jp/jp/singi/it2/densi/dai3/siryou3.pdf "Approaches and recommendations by the Open Data Promotion Consortium" (Reference material for the meeting of the e-Government Open Data Conference of working-level personnel, held on March 21, 2013, or to "Reference: Case study conducted on the Whitepaper on Information and Communications".

[35] http://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl_honbun.pdf

[36] http://www.kantei.go.jp/jp/singi/it2/densi/dai3/siryou3.pdf

表 6-4 Draft Wording for Provisions to be Incorporated in Contracts[37]

Article __ Copyrights and Moral Rights of Author

1. BBB assigns AAA all copyrights gratuitously, including the rights stipulated in Articles 27 and 28 of the Copyright Act on the literary works newly created by BBB (hereinafter called the "newly created literary works") in BBB's pursuing this business.

[1. BBB reserves all copyrights including the rights stipulated in Articles 27 and 28 of the Copyright Act on the literary works newly created by BBB (hereinafter called the "newly created literary works") in BBB's pursuing this business, but licenses AAA to use them gratuitously, including AAA's granting any third party the right to make secondary use thereof.]

2. BBB never exercises any moral rights of author against AAA as well as against any third party who uses the newly created literary works, as well as the literary works possessed by BBB from the past (hereinafter called the "existing literary works").

3. Should any existing literary works be included in the newly created literary works, BBB reserves the copyright thereof, too, but licenses AAA the right to use them gratuitously, including BBB's granting AAA the right to sublicense any third party the right to make secondary use thereof, to the extent possible. In the event that any literary works of any third party are included in the newly created works, the third party reserves the copyright thereof, but BBB endeavors to obtain license to use the literary works from the relevant third party, including BBB's granting AAA the right to sublicense other third parties the right to make secondary use thereof, to the extent possible. At the time of delivering products to AAA, BBB pays utmost attention so that the portion which third parties can make secondary use and the portion which third parties cannot make secondary use are clearly identified, and in respect of the portion which third parties are not allowed to make secondary use, BBB should also provide AAA with the reason therefor.

Source: "Approaches and recommendations by the Open Data Promotion Consortium" (Reference material for the meeting of the e-Government Open Data Conference of working-level personnel, held on March 21, 2013)

[37] AAA indicates Orderer, and BBB indicates Consignee.
The portion in parentheses describes draft wording for "the case where BBB does not assign AAA its copyright, and BBB solely grants AAA the license to use".

Reference: Case Study Conducted on the "Whitepaper on Information and Communications"

As regards the data containing the right of any third party, it is basically not possible to apply any rules on use without obtaining approval from the relevant third party. As the means of dealing with such cases, the ideal way from the standpoint of information users is that, (A) the information provider obtains approval from the third party for the secondary use of the data, or otherwise, (B) the information provider distinguishes the portion to which the third party reserves the right and indicates the relevant portion as not subject to any secondary use

As a reference for the case of our adopting the above-said method, we present below a case study conducted by the Data Governance Committee in 2012, on the "Whitepaper on Information and Communications in Japan". This Whitepaper, as the nature of this kind of document, contains not only the portions to which the publisher of this Whitepaper possesses the right, but also the portions to which third parties possess the right, as well as various kinds of data including statistical data to which no copyright occurs. In this case study, different types of data were classified, and the applicability of CC-BY when disclosing publicly owned data as open data, was examined.

1. Extraction and Classification of the Portions to Which Any Third Party May Have the Right:

   With respect to each data described in the "Whitepaper on Information and Communications" (such as sentences, figures, tables, graphs, pictures, and statistical data), the following different types of data were extracted and classified. Namely, "A: Data created by the Ministry of Internal Affairs and Communications (information provider) independently", "B: Data created under the entrustment by the Ministry of Internal Affairs and Communications (information provider)", "C: Data published based on the approval from third parties", "D: Data published and used in accordance with the citing rules permitted under the Copyright Act", and "E: Data not subject to any copyright, such as numerical data and laws and regulations".

2. Conformation of Approval by Third Parties

   For each extracted data, examination was performed to see whether or not they contain any third party right. "A" and "B" are essentially the data that have the possibility of containing the third party right, and "C" and "D" are the data that contain the third party right with certainty. And "E", including commercial database,

is the data that have possibility of being subject to the rules on use.

With respect to the portions that contain any third party right, it was checked anew, whether approval for secondary use is obtained, and also, to which parts approval for secondary use is not given, by contacting right-holders. With respect to old data, two kinds of data, namely, “C” (Data published based on the approval from third parties) and “D” (Data published and used in accordance with the citing rules permitted under the Copyright Act), were essentially categorized as the data not subject to the open data in this case study, because it was generally difficult to identify the contact addresses of right-holders, except for certain cases where confirmation was made easily.

3. Distinction and Publication of the Portions to Which CC-BY Cannot Be Applied

The portions which contain the third party right and to which approval for secondary use is not given were listed in a table so that the locations of such portions can easily be identified. The data which are not listed in this Table can publicly be disclosed as open data, after applying CC-BY.。

Meanwhile, in the case where this kind of classification is difficult, we can think of a way to directly or indirectly indicate or suggest that a third party or third parties possess(es) the right, by describing the source, and then, give warning to information users that approval for secondary use must be obtained from the relevant third party (source) on their own responsibility.

図 6-2 Example of Examination of the Whitepaper on Information and Communications[38]

[38] Open Data Promotion Consortium “2012 Data Governance Committee Report” http://www.opendata.gr.jp/committee/docs/20130331_1_datagov.pptx.pptx

表 6-5　Example of Examination of the "Whitepaper on Information and Communications" [39]

分類	第三者の権利（著作権、肖像権、商標権等）の有無	区分設定
総務省が独自に作成しているデータ	第三者の権利を含んでいる可能性がある。	A
総務省の委託調査で作成したデータ		B
第三者から掲載の許諾（著作権、肖像権、商標権等）を受けて利用	確実に第三者の権利が存在するため、確認の必要がある。 （簡易に確認する場合には、第三者に確認をせずに「CC-BY適用不可能」と整理する）	C
著作権法上認められた引用ルールに従って掲載・利用		D
数値データや法令など、著作権の対象外のデータ	商用DB等の利用規約の権利が働いている可能性がある	E

区分	表記
CC-BY適用可能	○
要確認	☆
CC-BY適用不可能	★
CCを付与できないが自由に利用できる（著作権無し）	―

表 6-6　Rules on Use in the "Whitepaper on Information and Communications"

Rules on use applied to "2013 Whitepaper on Information and Communications"

○**"2013 Whitepaper on Information and Communications" may be used freely in principle.**

・"2013 Whitepaper on Information and Communications" (HTML version including Excel data, PDF version and CSV data) may be used freely by anyone, for any purposes including commercial uses, and including reproduction, modification, distribution and public transmission, except for the figures and tables listed on the List of Figures and Tables below, or the sentences, etc. for which the sources of third parties are indicated.

・When you use the information and data, we request you to describe the source.。

Example of description of the source:

【In the case that the figures/tables not listed on the "List of Figures and Tables" and the sentences for which no source of third party is indicated.】

Source: "2013 Whitepaper on Information and Communications" (Ministry of Internal Affairs and Communications)

http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/html/XXXXXX.html (Description of URL of the relevant page, or link to URL of the relevant page) licensed under CC-BY 2.1 JP

[39] Same as above.

http://creativecommons.org/licenses/by/2.1/jp/

【Example of description for cases in which approval for use was obtained separately: In the case that the figures/tables are listed on the "List of Figures and Tables" and the sentences for which the source of the third party is indicated.】

Source: "2013 Whitepaper on Information and Communications"; Original source: "___Report" (___ Company)

http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/html/XXXXXX.html

(Description of URL of the relevant page, or link to URL of the relevant page)

※Among the figures contained in "2013 Whitepaper on Information and Communications", there are those using clip arts of Microsoft Corporation. Although it is not allowed to extract raw materials alone and sell them, because that goes against the rules on use of Microsoft Corporation, secondary use thereof, including their reproduction, modification, distribution and public transmission, is permitted.

(→Rules on use of Microsoft Corporation: http://office.microsoft.com/ja-jp/help/HA001089706.aspx)

○**As for the detailed use method, please refer to the description below**

【About the figures/tables listed on the "List of Figures and Tables" and the sentences for which the source of the third party is indicated】

・With respect to the figures/tables listed on the "List of Figures and Tables" and the sentences for which the source of third party is indicated, there is possibility that third parties possess copyrights and/or other rights (e.g. right of portrait about pictures, right of publicity). When using them, therefore, careful attention must be paid so that no rights of third parties are infringed.

・It is noted that even if the copyright of the relevant information is possessed by a third party, there are cases where secondary use, such as citation, is permitted without consent of the copyright-holder, under the Copyright Act.。

Use method of the data for which no approval of the copyright-holder is required.

・Reproduction for personal use

・Citation

・Reproduction by educational institutions,　and so forth

As for the details, please refer to the website of the Agency for Cultural Affairs.

http://www.bunka.go.jp/chosakuken/gaiyou/chosakubutsu_jiyu.html

Regarding the concrete use method, a document entitled "Copyright text: For those who study for the first time" is useful as good reference.

http://www.bunka.go.jp/chosakuken/text/pdf/chosaku_text_100628.pdf

【About the figures/tables not listed on the "List of Figures and Tables" and the sentences for which no source of the third party is indicated】

・Numerical data, simple tables and graphs may be used freely, as these data have no copyright, but we urge you to describe the sources when using them.

・The copyrights of sentences and figures having copyrightability are possessed by the Government and administered by the Ministry of Internal Affairs and Communications, but their use is permitted under the "Creative Commons License: Indication 2.1, Japan", which allows free use thereof. In using them, we request you to reprint the license indication as shown below.

"2013 Whitepaper on Information and Communications" published by the Ministry of Internal Affairs and Communications is licensed under the Creative Commons Indication 2.1. Japan License.

http://creativecommons.org/licenses/by/2.1/jp/

○Disclaimer

・Although we are making doubly sure about the accuracy of the information contained, should you find any errors or notice any dubious points, please feel free to contact the Office shown below.

・Meanwhile, it should be noted that the Ministry of Internal Affairs and Communications does not assume any responsibility, even if any loss or problems occurred due to your use of any information contained in "2013 Whitepaper on Information and Communications"

○Contact point concerning the "Whitepaper on Information and Communications"

Economic Research Office, Division of Information & Communications Policy, Global ICT Strategy Bureau, the Ministry of Internal Affairs and Communications

TEL： 03-5253-5720 FAX:03-5253-6041

E-MAIL： hakusho@soumu.go.jp

Source: "Guidance on use of 2013 Whitepaper on Information and Communications," published by the Ministry of Internal Affairs and Communications[40]

[40] http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/word/h25riyou.docx

表 6-7　List of Figures and Tables of the "Whitepaper on Information and Communications"

List of Figures and Tables2013 Whitepaper on Information and Communications

The below listed figures and tables have the possibility that third parties have copyrights and/or other rights (e.g. right of portrait about pictures, right of publicity, etc.). Therefore, when using them, utmost attention must be paid so that no right of the third party may be infringed.

Page	No. of Figures	Title
7	Fig. 1-1-1-8	Comparison of the number of shipments between laptop-computers and tablets
8	Fig. 1-1-1-9	Change in the domestic value of shipments of digital appliances and large household electric appliances
27	Fig. 1-1-2-15	Outline of GIS utilization at Kumano construction office in Mie Prefecture, under the disaster of Typhoon No. 12 in FY 2011
40	Fig. 1	Logos of fabrication laboratories
43	Fig. 1-1-3-15	Size of the electronic transaction market in the world (top 5 countries)
45	Fig. 1-1-3-19	Forecast of the advertisement market using smartphones in Japan
46	Fig. 1-1-3-25	Investigation on the risk of Amazon's showrooming in American companies
47	Fig. 1-1-3-26	Breakdown of shops for customers in the world
52	Fig. 1-1-3-41	Ratio of companies allowing use of personal terminals by BYOD in each industry
56	Fig. 1-2-1-5	Share of the number of shipments of tablet terminals in the world
59	Fig. 1-2-1-10	Number of investment, change in investment amount, and stage of investment in domestic venture capital
59	Fig. 1-2-1-11	Change in the number of IPOs in Japan
62	Fig. 1-2-1-14	Market size of crowd funding in the world
65	Fig. 1-2-1-19	Support structure of and participating teams in KDDI∞Labo
66	Fig. 1-2-1-21	Amount of investment in venture capital in the world by country
67	Fig. 1-2-1-22	Number of exists by venture companies in Japan and the United States (unit: number of exits)
67	Fig. 1-2-1-23	Comparison of IPO amounts by venture companies in Japan and the United States

67	Fig. 1-2-1-24	Comparison of M&A amounts by venture companies in Japan and the United States
68	Fig. 1-2-1-27	Ratio of CVC in venture investment in the United States
68	Fig. 1-2-1-28	Breakdown of CVC investment in the United States
（The rest is omitted）		

Source: “Guidance on the use of 2013 Whitepaper on Information and Communications”, published by the Ministry of Internal Affairs and Communications[41]

[41] http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/word/h25riyou.docx

Chapter 7. Future Prospect Concerning the Rules on Use

7.1 Direction of the Review in the Future

When publicly disclosing public data as open data, it is basically desirable to apply CC-BY or CC0 from the standpoint of information users, because these have been widely used as the rules on use of open data internationally. However, when incorporating prohibitive provisions concerning the use that goes against public order or morality meets the principle of small start (i.e. embarking upon things that can be done easily one by one as promptly as possible), or when disclosure of as many data as possible as open data is desired, we can also think of applying the Government Standard Terms of Use (Version 1.0) as the second best approach, taking the desire on the side of the information provider into account.

The Government Standard Terms of Use (Version 1.0) is slated for review, based on the results of data use after the shifting of rules on use of individual Ministries and Agencies to the Government Standard Terms of Use (Version 1.0). The provisions prohibiting the "use that goes against laws and regulations, ordinances, and public order and morality" and the "use that gives threat to the safety and security of the nation and its people", in particular, are not clear in respect of their concrete use modes, and there is a possibility that they would create a chilling effect on the use of publicly disclosed data. Therefore, the Data Governance Committee should closely watch the operational status of the Government Standard Terms of Use (Version 1.0), and if and when it is judged that there will be no problem even if such provisions on prohibition of use are eliminated, it should reconsider the future direction, taking into account the options of deleting these provisions on prohibition of use, or shifting to CC-BY or CC0.

When applying the Government Standard Terms of Use (Version 1.0) outside Japan, it is desirable to apply it, after due consideration that there is a possibility of change in the direction in future.

表 7-1 Use Modes Prohibited by the Government Standard Terms of Use(Version 1.0)

1） Indication of the source (A) (First part omitted) and it is prohibited to publicly disclose and use the compiled or processed information as if it had been created by the Government (or ministries or agencies).

３）Prohibited use:

ア Concerning the contents, usage in the following ways is prohibited.

（ア）Use that goes against laws and regulations, ordinances, and public order and moral.

（イ）Use that gives threat to the safety and security of the nation and its people

Source: The Government Standard Terms of Use (Version 1.0)

表 7-2 Review of the Government Standard Terms of Use (Version 1.0) (Cited Again)

・ Furthermore, besides it's calling users' attention in advance about the possibility that the rules on use could be changed in future, it is stipulated that the review of the rules on use will be made by the end of Fiscal Year 2015, in light of the situation that both the "Declaration to be the World's Most Advanced IT Nation" (Cabinet decision made on June 14, 2013) and the "Road Map for the Promotion of e-Government Open Data" (Decision made by Comprehensive IT Strategy Headquarters on June 14, 2013) are stating that the same level of public disclosure as that of other advanced nations will be realized by the end of Fiscal Year 2015.
・ Cosidering how wide contents in compliance with updated Government Standard Terms of Use (Version 1.0) are utilized at the time of conducting the review, "1.3) Prohibited use" is thought to become one of the important review items, from the viewpoint of making compatibility with internationally widely used CC-BY

Source: Explanation about the Government Standard Terms of Use(Version 1.0)

In the meantime, it is also necessary to enlighten information users that if and when any problem occurred to any third party due to improper use of the data publicly disclosed as open data, the responsibility lies with the information user who used the data in an improper manner, and that the Government or local governments as the information provider will never assume the responsibility therefor. At the same time, while it is assumed that there are cases where information publicly disclosed as open data has certain errors, it is also necessary to enlighten information users that even if certain problems occurred to information users or third parties owing to the errors, the Government or local governments will never assume any responsibility therefor because the information is disclosed publicly without any guarantee under either of CC license or the Government Standard Terms of Use (Version 1.0). What is most important is, among others, that data possessed by the Government and local governments are publicly disclosed, and as regards the reliability of the data, judgment thereof by individual persons who use the data is called for.

Disclosure of public data as open data will be promoted, by eliminating obstacles so that the Government and local governments do not hesitate to publicly disclose their possessed data.

Part III. Technical Information: Let's Make Your Open Data Machine-readable.

Chapter 8. Technical Levels of Data

This chapter explains technical details of procedures for creating and disclosing open data described in Chapter 3.

First of all, machine readability, data catalog, and identifiers are explained. Next, the "technical levels of open data" are defined on their basis. Some supplementary remarks are then added concerning open-data management policies and meta data.

The whole details of technical guidelines concerning open data are given in Chapter 9

8.1 Explanation of Machine Readability

8.1.1 What Machine-readable Data Are

The "Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use"[42] describes "machine readability" as follows.

> *"Machine readable" here means that a computer program (hereafter referred to simply as "computer" in this footnote) is able to reuse (process, edit, etc.) data automatically. There are several levels of the "degree of machine-readability" depending on to what extent a computer can reuse data without requiring human intervention. In order for a computer to reuse data automatically, it is required that a computer can identify (read) the logical structure of data in question and process values in the structure (numerical values in a table, text strings, etc.).*

As is indicated in the beginning of this guide (Section1.1), it is a computer that edits, processes, modifies, etc. open data. Given open data are analyzed by using a computer for the purpose of making it more efficient to obtain new knowledge from those data.

For instance, suppose a statistical table is published on a home page in the form of image or PDF file. In order to feed those data to a computer to have them analyzed, it is required that a human operator enter data contained in the image into a spreadsheet program and save the relevant file or feed a computer with numerical values and text strings obtained from published data by technical means such as image recognition. This is a method that imposes a burden on an information user and is inefficient.

Therefore, in order to have a computer analyze data more efficiently, it is desirable for an information provider to convert data provided into those in a format that makes numerical values and text strings therein more readily available to a computer to lessen the cost required for an analysis

[42] http://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl_honbun.pdf See the footnote on p.1.

using a computer by an information user. Data in such a format from which numerical values and text strings are readily retrievable by a computer are called "machine-readable data."

8.1.2 Index Regarding Machine-readability

There are various levels of machine readability of machine readable data. This section describes the levels of machine readability on the basis of 5 ★ Open Data, which is one of evaluative indices for open data.

5 ★ Open Data is an evaluative index suggested by Tim Berners-Lee and consists of the following five levels[43](Figure 8-1).

1. *make your stuff available on the Web (whatever format) under an open license*
2. *make it available as structured data (e.g., Excel instead of image scan of a table)*
3. *use non-proprietary formats (e.g., CSV instead of Excel)*
4. *use URIs to denote things, so that people can point at your stuff*
5. *link your data to other data to provide context*

A list of file formats for tabular data, document data, and geospatial data falling under each level is given in the appendix (Section 10.1.1). The characteristic features of each level are here explained by citing data in a typical format.

File formats falling under ★ include gif, jpeg, and other image file formats as well as pdf. Technical means such as image recognition is required for a computer to retrieve data from files in these formats, which is not easy.

File formats falling under ★★ include .doc (Word files) and .xls (Excel files). Since files in these formats are structurized, a computer can retrieve data from them if relevant software programs are provided. Data at ★★ or above are generally regarded as "machine-readable data."

★★★ file formats include CSV, HTML, .odt/.ods (OpenDocument files), .docx/.xlsx (Office Open XML files). The methods for analyzing data in these file formats are disclosed. For this reason, it is easier to create a software program for analyzing data in a ★★★ file format than to create one for analyzing data in a ★★ format.

RDF (Resource Description Framework) based files fall under ★★★★ or above. Data in such file formats can be linked with each other. For this reason, it is easy for a computer to mash up those data.

---

[43] 5 ★ Open Data. (Source) http://5stardata.info/ (Japanese translation) http://5stardata.info/ja

Section 0 of this guide expounds technical guidelines for enhancing machine-readability of data in a ★★★ format.

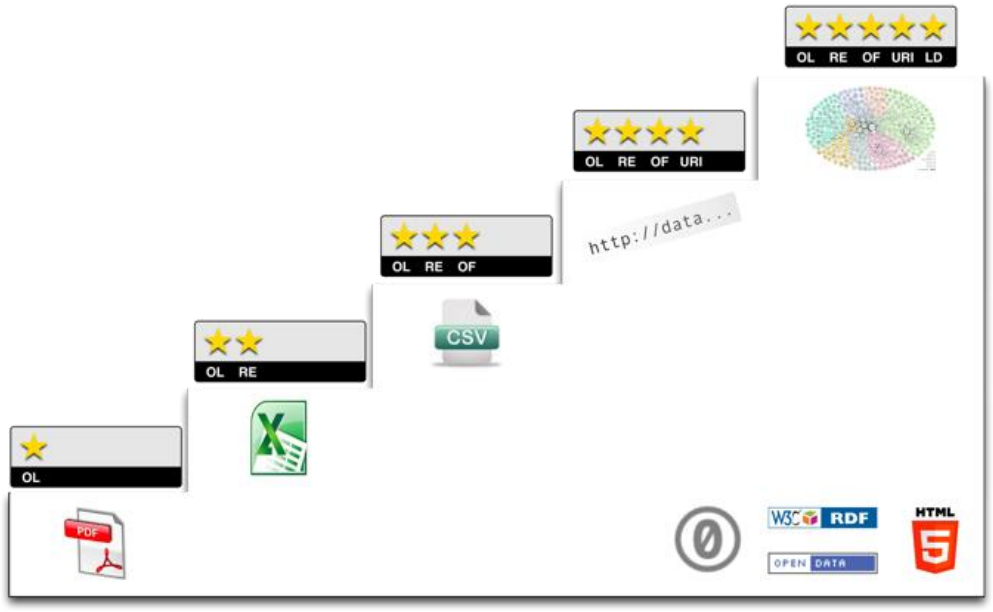


Figure 8-1 Levels of 5 ★ Open Data

8.1.3 Notes Regarding Handling of Machine-readable Data

It should be noted that machine-readable data shown in this guide are not necessaryily legible to a human being. It should thus be considered to disclose data in files in two types of formats: machine-readable and human-legible formats.

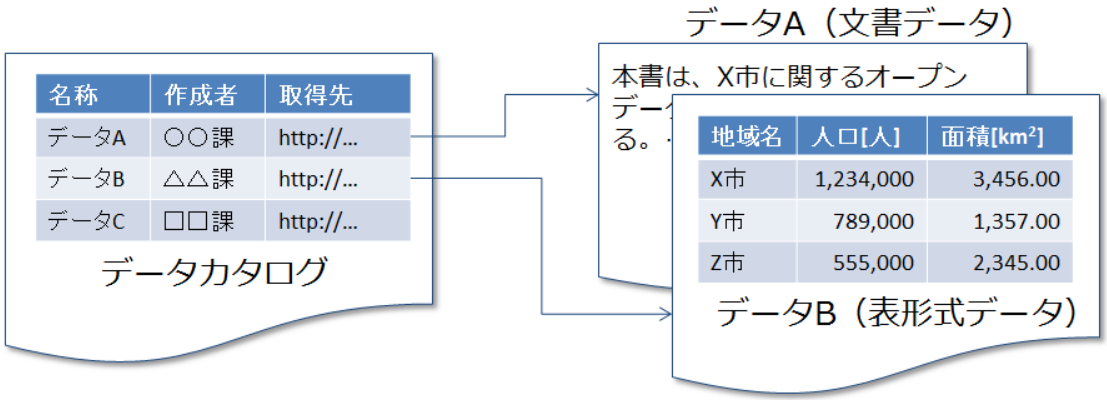
This is indicated as follows in the "Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use" (Decision by the Liaison Conference of Chief Information Officer (CIO) of Ministries and Agencies).[44]

> *After this guideline is established, efforts shall be made to disclose numerical values (tables), text, and geospatial data newly provided by ministries and agencies and disclosed through the Internet in the conventional data format (typically, pdf format) intended for human reading and printing as well as to disclose (structured) data containing them created in light of the attached caveats containing them in a machine-readable data format not dependent on a specific application program.*

[44] http://www.kantei.go.jp/jp/singi/it2/densi/

8.2 Explanation of Data Catalog

As data to be disclosed increase, a demand rises for a function that organizes, searches, and lists such data. A data catalog satisfies such a demand. Figure 8-2 shows the relation between "data" and "data catalog."



Data Author Source

Data A Division X http://

Data B Division Y http://

Data C Division Z http://

Data Catalog

Data A (document data)

Open data for City X …

Municipality Population Area [km^2]

City X

City Y

City Z

Data B (tabular data)

Figure 8-2 Relation between "Data" and "Data Catalog"

Tabular data which contain data names, data sources, etc. as illustrated by Figure 8-2 is also a kind of data catalog. It is desirable to introduce a data catalog system or provide data and metadata search functions using technology such as RDF, SPARQL, etc. in order to provide advanced data managing, searching, and listing functions. For details of data catalog system, see Section 10.3.2 and 11.

8.3 Open Data and Identifiers

Open data are data to be read and interpreted by a computer. Such data should be uniquely identified by a computer. It is also desirable that real objects, organizations, locations, etc. referred to in open data be uniquely identified. This is because a problem such as one given below can arise if the name of an organization or location is given in text string. First, an information user's computer can interpret expressions for an identical organization or location as referring to different ones due to notational fluctuations. For example, whereas "中央一丁目一番地一号" and "中央 1-1-1," which differ only whether they use Chinese or Arabic numerals, refers to the same location and whereas "システム管理課" and "ｼｽﾃﾑ管理課," which differ only whether they use full-width or half-width characters, refer to the same organization, a computer may interpret each pair of these expressions as referring to distinct locations or organizations. Second, an information user's computer cannot distinguish among organizations or locations which are given the same name only on the basis of text string denoting that name. For instance, locations with the address of "Chuo 1-chome," which exist in places over Japan, cannot be distinguished only on the basis of text string "Chuo 1-chome." Likewise, persons or organizations with the same name cannot be distinguished only on the basis of text string denoting its name.

An identifier (ID) is a number for enabling a computer to identify each such datum or a real object, organization, location, etc. referred to by that datum.

Let us note that there is a concept similar to that of identifier (ID): code. A code is a number assigned to a categorized concept or thing and is defined to encode the concept or thing subject to encoding by means of abbreviated expression. That is, whereas a code is assigned meaning, an identifier is not necessaryily assigned meaning. Although an identifier and a code differs from such a perspective, a code functions as an identifier in many cases.[45]

Properties that must be satisfied by open data identifiers are described in Section 9.1.

8.4 Technical Levels of Open Data

What has so far been stated is summarized as technical levels of open data, as shown in Table 8-1.

It is not required to align levels of data, data catalog, and identifiers in considering the application of a certain technical level. For instance, Level 2 data can be offered by using a Level 1 data catalog in the tabular format.

Providing data by using technology specified in the Level 1 enables an information user to acquire

[45] For instance, the Japanese Article Number (JAN) code is the code system that encodes the country in which a product is manufactured and its manufacturers by means of 13 decimal digits.

data directly without additional processing such as image analysis. It also enables him or her to obtain metadata such as the location of the relevant data electronically.

Going up to the Level 2 enhances machine-readability of data.

Moving then to the Level 3 enhances the interpretative efficiency and searchability of data to improve an information user's efficiency in utilizing data.

Going further up to the Level 4 makes easy cross search over multiple data sets and the like to widen the range of data use by information users.

Table 8-1 Technical Levels of Open Data

	Level 0	Level 1	Level 2	Level 3	Level 4
Data	Disclose PDF and image files on a website.	Create structured data and disclose them on a website. (XLS, DOC, etc.)	Disclose data in an non-proprietary (standardized) format. (CSV, etc.)	Create and disclose machine-readable data. (See Chapter 9.)	Create data by using technology such as RDF [46], XML [47], etc., implement API, and disclose those data.
Data catalogs	Non-existent	Create and disclose a data catalog as tabular data (CSV, etc.).	Same as the Level 1.	Introduce a data catalog system.	Offer a metadata search function using RDF, SPARQL[48], etc.
Identifiers	Identified by some means.	Same as the Level 0.	Same as the Level 0.	Identified by URL.	Identifiers based on a global system are used.
Necessary tools	Web server	Web server	Web server	Web server and CKAN, etc.	Web server, CKAN, information circulation and sharing platform, etc.

[46] RDF (Resource Description Framework) is a framework for describing information concerning "resources" (things referred to) on the Web. RDF describes information concerning resources by using three elements: subject, predicate, and object.
Fabien Gandon, Guus Schreiber, and Dave Beckett, RDF 1.1 XML Syntax. February 25, 2014. W3C Recommendation. http://www.w3.org/TR/rdf-syntax-grammar/

[47] XML (Extensible Markup Language) refers to specifications of a language that describes data and behavior of computer programs processing them as well as documents written in that Language.
Tim Bray, et al. Extensible Markup Language (XML) 1.1 (Second Edition). August 16, 2006. W3C Recommendation. http://www.w3.org/TR/xml11/

[48] SPARQL (SPARQL Protocol and RDF Query Language) is a query language for searching and manipulating data described in accordance with a RDF model.
Lee Feigenbaum, et al. SPARQL 1.1 Protocol. [Online] May 21, 2013. W3C Recommendation. http://www.w3.org/TR/sparql11-protocol/

8.5 Open Data Management Policy and Metadata Assignment Method

If metadata can be automatically assigned to a given set of open data in registering and managing it, it will reduce cost for management and registration. This section describes such a method.

A method for assigning metadata varies by a data registration policy. For instance, if a method is taken whereby a system administrator or an independent organization creating and editing open data gathers data from organizations, departments, and sections and disclose them (to be referred to as a "centralized registration method") it is desirable that the system administrator or the independent organization collect metadata along with data from organizations, departments, and sections when collecting those data. On the other hand, if a method is taken whereby each organization, department, or section directly registers and manage open data using some system (to be referred to as a "distributed registration method"), the competent organization, department, or section is required to register metadata in one way or the other. In this case, it is desirable that the competent organization be able to create metadata along with open data when creating the latter.

Some software programs for editing tabular data or document data provide a way to assign metadata to them. If it is used, metadata such as the creator and the creation date of a file can be stored when a file is created. For instance, software programs such as Microsoft Office, OpenOffice, Acrobat have a function to edit file's property. A computer can acquire metadata registered by using this function by using a program such as Apache Tika [49] (free) (Figure 8-3). The Technical Committee is scheduled to examine tools for effectively registering metadata with a data catalog system

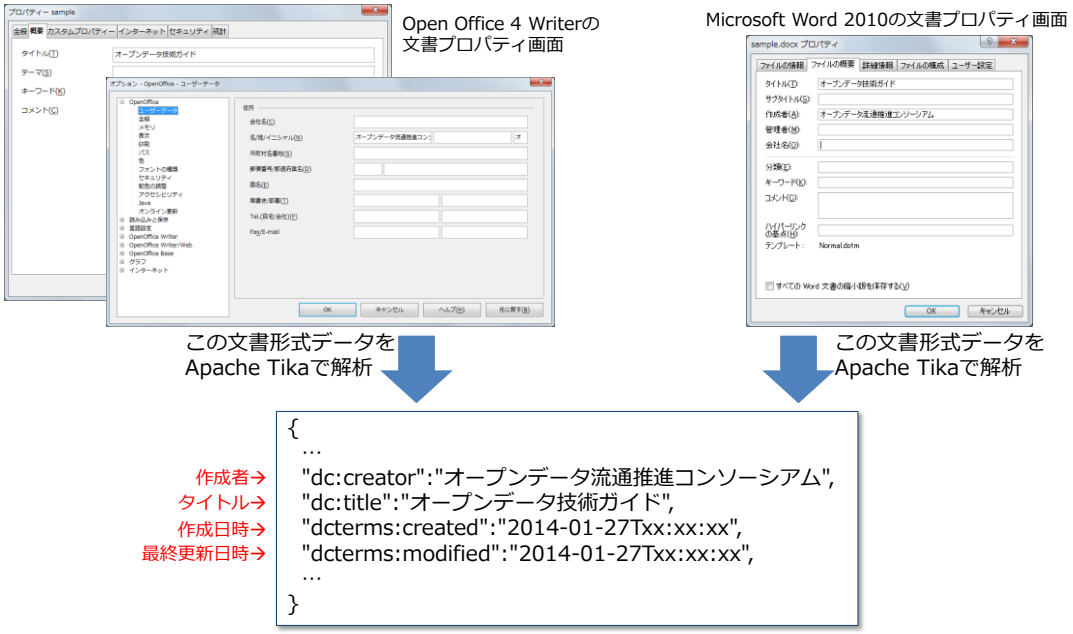


Figure 8-3 How to Register Metadata in a Document Editing Program

[49] http://tika.apache.org/

Chapter 9. Technical Guidelines for Open Data

This chapter shows technical guidelines for creating and editing machine-readable open data in terms of three elements: identifiers, file formats, and data.

This guide covers the following four data types.

- Tabular Data
- Document data
- Geospatial data
- Real-time data

9.1 Guidelines Concerning Identifiers

This section describes properties to be satisfied by identifiers for open data and explains how to satisfy them.

First, Section 9.1.1 shows properties to be satisfied by identifiers for open data. Available systems of identifiers which meet such properties are described in Section 9.1.2. 9.1.2 is available, it should be used.

Section 9.1.2 explains how to deal with a case where no identifier system mentioned in Section 9.1.3 is available.

9.1.1 Properties to Be Satisfied by Identifiers for Open Data

It is desirable that identifiers for open data satisfy the following properties.

① Being Unique

Suppose the name of a budget file to be disclosed by division X of City A and that of a budget file to be disclosed by City Y are both "123 csv." If an information user wants to download both of these open data files, a file name conflict occurs. For this reason, the information user must change their file name when downloading them, for instance, to "123_X.csv" and "123_Y.csv."

Likewise, if the identifier of open data file disclosed by City X overlaps with that of open data file disclosed by City Y, a computer cannot distinguish between them. 従って、少なくとも同一組織が公開するオープンデータの識別子は、一意でなければならない。なお、An information user may not belong to the same administrative district or organization to which an information provider belongs. For this reason, he wider the range covered by unique open data identifiers is, the more desirable it is.

② Constituting a System that Can Be Commonly Used

Open data are disclosed by multiple organizations and edited, processed, and modified by

people and computers over the world. They are expected to be used in diverse environments. It is possible that different sets of open data are interconnected by links. For this reason, the wider the range in which the system which open data identifiers are based on can be commonly used is, the more desirable it is.

9.1.2 Available Systems of Identifiers

Available systems of identifiers which meet requirements specified in Section 9.1.1 are given below.

① Globally unique systems of identifiers

② Systems of identifiers or codes prescribed by public institutions

③ Systems expressible as those of URIs (Uniform Resource Identifiers)

As indicated in Section 8.3, whereas an identifier is a number for enabling a computer to identify a datum or a real object, organization, location, etc. referred to by the datum, a code is a number assigned to a categorized concept or thing. Although both differ strictly, a code functions as an identifier in many cases. It is for this reason that systems of codes are included in the above list.

In addition, specific systems of identifiers and codes satisfying requirements specified in Section 9.1.1are gathered in Section10.2. For instance, there are ucode, DOI (Digital Object Identifiers), company code (ISO 6523), local government code of Japan, etc.

9.1.3 How to Deal with a Case Where There Is No Appropriate System of Identifiers

If there is no such useful system of identifiers as described in Section 9.1.2, it is desirable to proceed as follows.

1. If no numbers are assigned to real objects, organizations, and locations, assign a number to each of them.
2. Expand the range in which identifiers remain unique in the following two ways.
   - An organization can acquire and manage identifiers based on a global system such as ucode, DOI, etc. or a system of identifiers or codes prescribed by public institutions (Figure 9-1).
   - An organization can also globalize its system of identifiers by adding a url it chooses to a number it has assigned (Figure 9-2). It should, however, be noted that, if this method is adopted, such identifiers will change if the organization's domain name changes due to its integration with other organization(s), its abolition, etc.

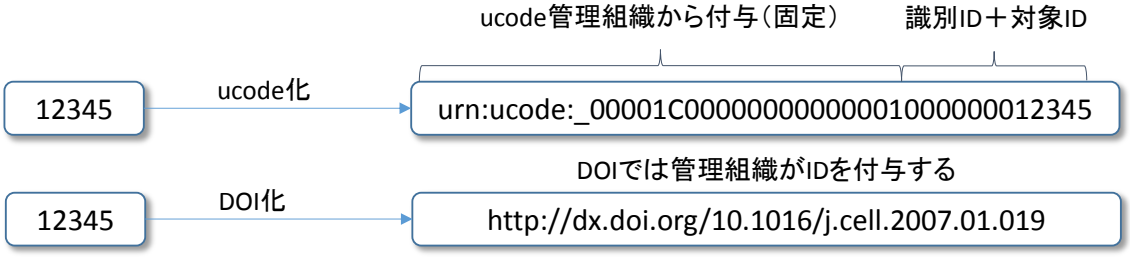


Figure 9-1 How to Globalize Identifiers by Using Global Identifiers

Figure 9-2 How to Globalize Identifiers by Using a URLl that an Organization Has

9.2 Guidelines Regarding File Formats

It is desirable that a highly machine-readable file format be used with data to be disclosed. In Table 9-1, typical file formats are classified in accordance with the technical levels of open data shown in Table 8-1.

Table 9-1 File Formats Classified in Terms of Machine-readability

	Level 1	Level 2 / Level 3	Level 4
表形式データ	xls (Microsoft Excel 形式)	CSV, xlsx (Office Open XML), ods (OpenDocument), JSON	RDF/XML, RDF/JSON, JSON-LD Notation3, Turtle 等の RDF 形式
文書形式データ	PDF (Acrobat 形式) doc (Microsoft Word 形式)	HTML, XML, docx (Office Open XML), odt (OpenDocument)	
地理空間データ	Shape	KML, GML	Recommended file format
リアルタイムデータ	Not exchange file format		

9.3 Guidelines Regarding Data

9.3.1 Grades of Guidelines

The following two grades are set for guidelines concerning data.

① Grade 1

Grade 1 Guidelines are those which open data are strongly recommended to meet whose purpose is to satisfy the following.

- Being consistent with a typical standards, if any, for data formats
- Enabling an information user acquiring data to write a program for correctly interpreting the essential content of the data without modifying or processing its content

② Grade 2

Grade 1 Guidelines are those which open data are recommended to meet whose purpose is to satisfy the following.

- Enabling a program acquiring data to interpret the items and structure of the data correctly

9.3.2 Guidelines Regarding Tabular Data

9.3.2.1 Definitions of Terms

First of all, following terms are defined(Figure 9-3).

- Field
  - A record is an element of a table consisting of a single row and a single column. It corresponds to a cell of a spreadsheet.
- Record
  - A record is an element of a table consisting of a single row. It is composed of one or more fields.
- Header
  - A header is one or more rows that contain names of columns and consists of one or more fields.
- File
  - A file refers to the entire table. It consists of records and headers.

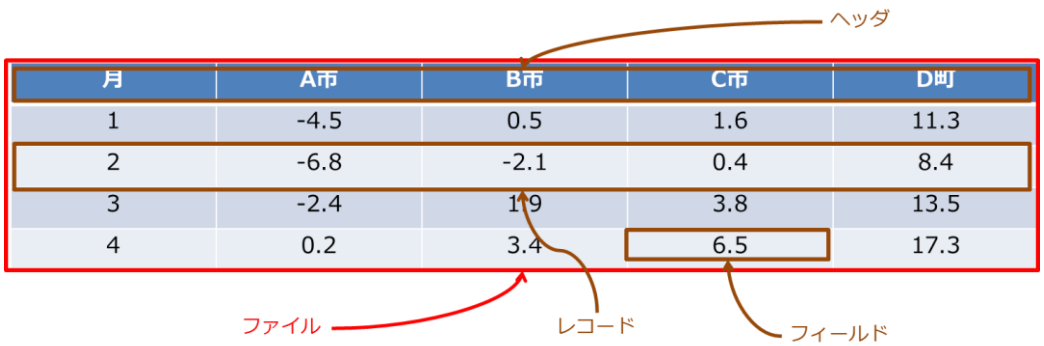


Figure 9-3 Definitions of Terms Used in the "Guidelines Regarding Tabular Data"

9.3.2.2 Grade 1 Guidelines

(Guideline 1) A tabular file should be composed of a single table.

[Explanation]

A file shown in Figure 9-4 contains multiple tables. A computer needs to identify the end of one table and the beginning of another table correctly in order to read such a file, which complicates reading procedures. For this reason, a single file should be composed of a single table.

This guideline will be satisfied if the file is split so that each of multiple tables contained in it is put into a different file (Figure 9-5).

File X

Month	City A	City B	City C	Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Month	City A	City B	City C	Town D
1	230	58	377	103
2	169	43	422	122
3	144	54	322	144
4	232	102	145	133

Figure 9-4 Example of a Single File Containing Multiple Tables (Failing to Meet Guideline 1)

File X

Month	City A	City B	City C	Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

File Y

Month	City A	City B	City C	Town D
1	230	58	377	103
2	169	43	422	122
3	144	54	322	144
4	232	102	145	133

Figure 9-5 The File in Figure 9-4is Split into Two (Meeting Guideline 1)

[Supplementary Remarks]

This guideline is intended to conform to RFC4180[50], which specifies the CSV (Comma-Separated Values) format. RFC4180 gives the following requirements 3 and 4 as a part of the definition of the CSV Format.

③ *There maybe an optional header line appearing as the first line of the file with the same format as normal record lines. This header will contain names corresponding to the fields in the file and should contain the same number of fields as the records in the rest of the file.*

④ *Within the header and each record, there may be one or more fields, separated by commas. Each line should contain the same number of fields throughout the file.*

In order to meet both requirements 3 and 4, a single file must be composed of no more than a single table.

[50] Shafranovich, Y. Common Format and MIME Type for Comma-Separated Values (CSV) Files. October 2005. RFC 4180. http://www.ietf.org/rfc/rfc4180.txt.

(Guideline 2) A header should be composed of a single row.

[Explanation]

A header for the file shown in Figure 9-6 consists of two rows. In order for a computer to read such a file, it needs to identify the end of the header and the beginning of data correctly, which complicates reading procedures. For this reason, a header should consist of a single row only.

This guideline will be satisfied if the two rows of this header are integrated into a single row (Figure 9-7).

Month	Temperature			
	City A	City B	City C	Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-6 Example of a Header Consisting of Multiple Rows (Failing to Meet the Guideline 2)

Month	Temperature of City A	Temperature of City B	Temperature of City C	Temperature of Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-7 The Integrated, Single-Row Header (Meeting the Guideline 2)

[Supplementary Remarks]

Just as the Guideline 1, this guideline is intended to conform to the requirement 3 of the CSV format definition in RFC4180.

9.3.2.3 Grade 2 Guidelines

(Guideline 3) It is desirable that fields do not contain information other than relevant data.

[Explanation]

The field for the value for City C in January in the file shown in Figure 9-8 contains the numerical value "1.6" and a link to the note "(*1)." A computer cannot read (*1), a link to the note, and the linked note written in a natural language. In order to enhance machine-readability of this file, it is desirable to remove (*1), which is a link to the note, to leave the numerical value "1.6" onlyFigure 9-9).

Let us note that a file containing notes such as the one shown in Figure 9-8 is required for a human reader to interpret data. For this reason, based on Section 8.1.3, it is desirable that such a file be provided in addition to a highly machine-readable file such as the one shown in Figure 9-9.

Month	City A	City B	City C	Town D
1	-4.5	-0.5	1.6 (*1)	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-8 Example of Record Containing Information Other Than Relevant Data (Failing to Meet Guideline 3)

Month	City A	City B	City C	Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-9 Information other than relevant data is removed (meeting Guideline 3).

(Guideline 4) It is desirable that any field be not combined with another field.

[Explanation]

All fields for "year" are combined in the file shown in Figure 9-10. A human reader can understand that data for these four months are for those months in 2013, which a computer cannot read off. In order to enhance machine-readability of the file, it is desirable to uncombine those fields and insert a value into each field (Figure 9-11).

Year	Month	City A	City B	City C	Town D
2013	1	-4.5	-0.5	1.6	11.3
	2	-6.8	-2.1	0.4	8.4
	3	-2.4	1.9	3.8	13.5
	4	0.2	3.4	6.5	17.3

Figure 9-10 Example of Fields Combined (Failing to Meet Guideline 4)

Year	Month	City A	City B	City C	Town D
2013	1	-4.5	-0.5	1.6	11.3
2013	2	-6.8	-2.1	0.4	8.4
2013	3	-2.4	1.9	3.8	13.5
2013	4	0.2	3.4	6.5	17.3

Figure 9-11 Those Fields Are Uncombined (Meeting Guideline 4).

(Guideline 5) It is desirable not to leave a field blank (omit a value for a field) unless any value for any field does not exist.

[Explanation]

A value for the "year" field is omitted in the second and subsequent rows in the file shown in Figure 9-12. A human reader can understand that data for these four months are for those months in 2013, which a computer cannot read off. In order to enhance the machine-readability of the file, it is desirable to fill each blank field with an value omitted Figure 9-13).

Year	Month	City A	City B	City C	Town D
2013	1	-4.5	-0.5	1.6	11.3
	2	-6.8	-2.1	0.4	8.4
	3	-2.4	1.9	3.8	13.5
	4	0.2	3.4	6.5	17.3

Figure 9-12 Example of the Omission of Field Values (Failing to Meet Guideline 5)

Year	Month	City A	City B	City C	Town D
2013	1	-4.5	-0.5	1.6	11.3
2013	2	-6.8	-2.1	0.4	8.4
2013	3	-2.4	1.9	3.8	13.5
2013	4	0.2	3.4	6.5	17.3

Figure 9-13 Blank Fields are Filled with Omitted Values (Meeting Guideline 5).

(Guideline 6) It is desirable to provide a value for year in accordance with the Gregorian calendar.

[Explanation]

Each value for a "year" field is given in accordance with the Japanese calendar in the file shown in Figure 9-14. Since it is easier for a computer to process values for years that are numerically comparable, years in accordance with the Gregorian Calendar, which increase monotonically, are easier to be handled by a computer. For this reason, it is recommended to add year values in accordance with the Gregorian Calendar (Figure 9-15).

Year	Month	City A	City B	City C	Town D
Heisei 25	1	-4.5	-0.5	1.6	11.3
Heisei 25	2	-6.8	-2.1	0.4	8.4
Heisei 25	3	-2.4	1.9	3.8	13.5
Heisei 25	4	0.2	3.4	6.5	17.3

Figure 9-14 Example of Years Given in the Japanese Calendar (Failing to Meet Guideline 6)

Year	Year	Month	City A	City B	City C	Town D
Heisei 25	2013	1	-4.5	-0.5	1.6	11.3
Heisei 25	2013	2	-6.8	-2.1	0.4	8.4
Heisei 25	2013	3	-2.4	1.9	3.8	13.5
Heisei 25	2013	4	0.2	3.4	6.5	17.3

Figure 9-15 Years in the Gregorian Calendar Are Added (Meeting Guideline 6).

(Guideline 7) It is desirable that specify a unit for number fields.

[Explanation]

No units for numbers are specified in the file shown in Figure 9-16. Since units for data (physical unit, monetary unit, etc.) are indispensable for data processing, it is desirable that units be specified.

A unit for number fields can be specified by adding unit to the header of this table (Figure 9-17) or attaching the description of this file based on the Simple Data Format[51] separately (Figure 9-18).

Month	City A	City B	City C	Town D
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-16 Example of Number Fields for Which a Unit Is Not Indicated (Failing to Meet Guideline 7)

Month	City A [℃]	City B [℃]	City C [℃]	Town D [℃]
1	-4.5	-0.5	1.6	11.3
2	-6.8	-2.1	0.4	8.4
3	-2.4	1.9	3.8	13.5
4	0.2	3.4	6.5	17.3

Figure 9-17A Unit Is Added to the Header (Meeting Guideline 7).

[51] Simple Data Format is explained in Section 10.1.3 of the Appendix.

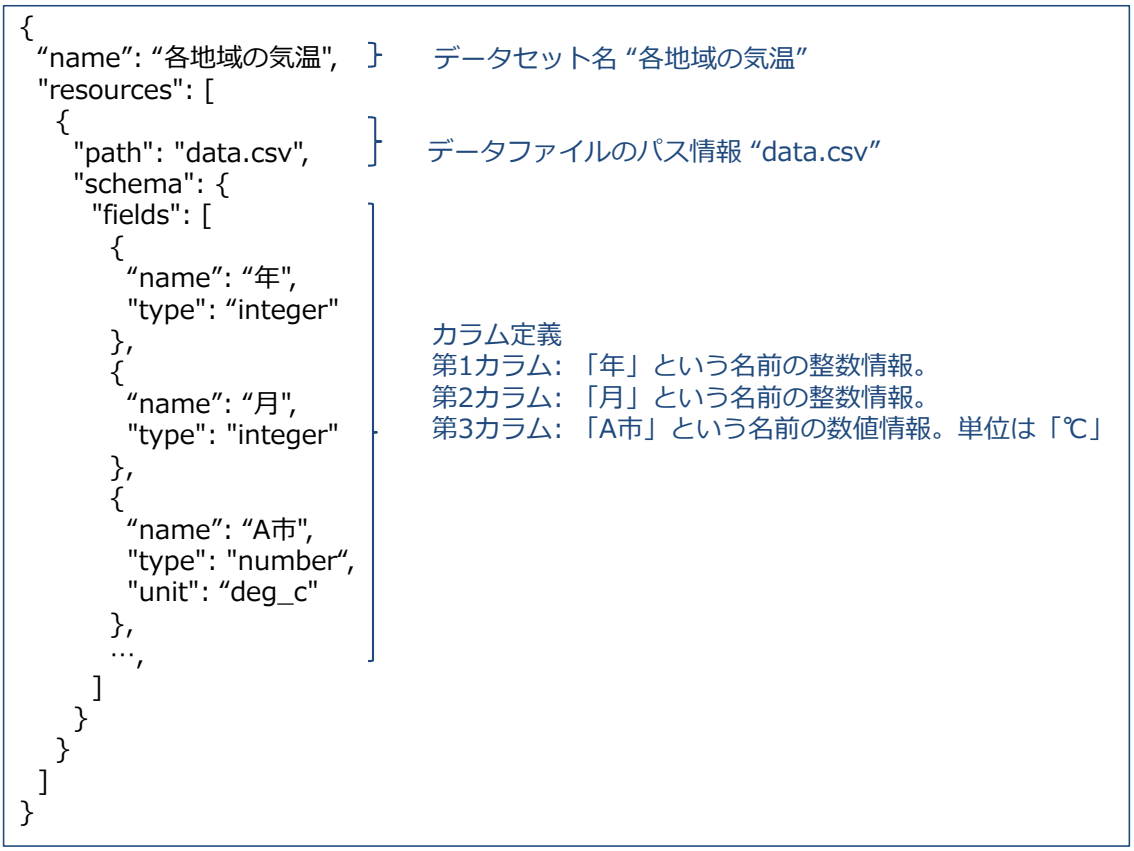


Figure 9-18 The Definition File for Figure 9-16 in Accordance with the Simple Data Format Is Added

(Meeting Guideline 7).

(Guideline 8) It is desirable to specify the character code used. It is also desirable to use a character code used widely internationally.

[Explanation]

There are multiple character codes for Japanese, including JIS (ISO-2022-JP), Shift-JIS, EUC, and UTF-8. It is thus difficult for a computer to read Japanese unless the character code used is specified. In addition, in light of the international distribution of data and consistency with other standards, it is desirable to use UTF-8.

The Japanese version of Microsoft Excel, which is currently widely used, outputs data in the CSV format using Shift-JIS encoding. A typical method for converting the CSV file containing those data to a UTF-8 file is shown below.

① Open the CSV file in Notepad and save it in the UTF-8 format.

② Open the CSV file in OpenOffice.org[52] and save it in the UTF-8 format.

③ Use a command-line tool (such as nkf[53]).

(Guideline 9) It is desirable that metadata representing file properties and descriptions be formally described. It is also desirable that metadata are linked to the data set which they describe.

[Explanation]

If metadata are described formally a computer can search and read data efficiently.

This guideline can be satisfied by using the Simple Data Format referred to with regard to Guideline 7 or registering metadata with a data catalog system.

(Guideline 10) The formal description of data using a formal language such as XML and RDF is desirable.

[Explanation]

The formal description of data using a formal language that can describe semantics such as XML, RDF, etc. enables the description of data including the meaning of each field and enhances machine-readability further.

[52] http://www.openoffice.org/

[53] http://sourceforge.jp/projects/nkf/

9.3.3 Guidelines Regarding Document Data

9.3.3.1 Grade 1 Guidelines

There is no grade 1 guideline regarding document data.

9.3.3.2 Grade 2 Guidelines

(Guideline 1) It is desirable that the structure existing in a document including parts, chapters, sections, figures, tables, etc. is described in a highly machine-readable format.

[Explanation]

A document has a structure consisting of parts, chapters, sections, paragraphs, figures, tables, etc. In order for a computer to extract such a structure, when using a document editing software program, use style functions provided by the program (heading, etc.), instead of font or character formatting, to represent the structure existing in a document such as parts, chapters, sections, figures, tables, etc. If a document is written in HTML, represent a structure using tags such as , <h3>, etc. as well as style notation.

(Guideline 2) It is desirable that text includes no formatting symbols or characters such as blanks and line breaks.

[Explanation]

A computer cannot determine whether blanks and line breaks contained in text are significant or not. Those blanks and line breaks become obstacles to its analyzing or reading out text. For this reason, remove blanks or line breaks which a computer does not have to read.

(Guideline 3) If document data include tabular data, it is desirable that tabular data of grade 1 or higher be attached to them.

[Explanation]

If a figure or table is contained in text, this guideline can be met by linking it to a file of tabular data of grade 1 or higher.

(Guideline 4) It is desirable to specify the character code used if document data in a text format are used. It is also desirable to use a character code used widely internationally.

[Explanation]

There are multiple character codes for Japanese, including JIS (ISO-2022-JP), Shift-JIS, EUC, and UTF-8. For this reason, if document data in a text format such as a text document, an HTML document, etc. are used, it is difficult for a computer to read them unless the character code used is specified. In addition, in light of the international distribution of data and consistency with other standards, it is desirable to use UTF-8.

For instance, with document data in HTML format, specify the character code used by using a meta tag given at Table9-2.

The Japanese version of Microsoft WORD, which is currently widely used, outputs data in the text format using Shift-JIS encoding. A typical method for converting the CSV file containing those data to a UTF-8 file is shown below.

① Open the text file in Notepad and save it in the UTF-8 format.

② Open the text file in OpenOffice.org[54] and save it in the UTF-8 format.

③ Use a command-line tool (such as nkf[55]).

(Guideline 5) It is desirable that an explanation of the text of a document which an information user can understand is given as metadata and linked to the document.

[Explanation]

If metadata are described formally a computer can search and read data efficiently.

This guideline can be satisfied by registering metadata with a data catalog system.

[54] http://www.openoffice.org/

[55] http://sourceforge.jp/projects/nkf/

9.3.4 Guidelines Regarding Geospatial data

9.3.4.1 Grade 1 Guideline

(Guideline 1) If data regarding location information is given, the geodetic system should be specified in addition to location information including latitude and longitude. For an outdoor location, it is desirable to use the World Geodetic System. For an indoor location, the coordinate system and the drawing scale (input precision) should be indicated.

[Explanation]

There are multiple geodetic systems for representing geospatial data and each of them gives a different value.

For instance, the latitude and longitude in accordance with the International Terrestrial Reference Frame (ITRF) and those in accordance with Tokyo Datum diverge by approximately 450 m on the ground surface in the vicinity of Tokyo. Therefore, a location cannot be identified unless the geodetic system on which geospatial data is based is specified.

Using a geographic information

system (GIS) enables you to edit geospatial data easily. In addition, the geodetic system used is often specified in data output.

9.3.4.2 Grade 2 Guidelines

(Guideline 2) It is desirable that geospatial data be in the vector format. Use Japan Profile for Geographic Information Standards (JPGIS) based on the latest ISO and JIS standards in creating data in the vector graphic format.

[Explanation]

For disclosing geospatial data, the vector format or the GML format is desirable, which requires less storage capacity than the raster format to represent the identical information. In disclosing such data, indicating the coordinate reference system which they are based on makes coordinate transformation easier in using them.

Standards for Geographic information prescribes rules regarding data design, quality, description methods, how to write a (technical) specification etc to ensure geospatial data, which is spatial data framework of GIS, is mutually used by different systems. They are deemed as the governmental technical standards by the liaison meeting of GIS related ministries and agencies (now is called Geospatial Data Utilization and Promotion Meeting).

JPGIS are domestic functional standards based on the latest ISO/JIS standards to improve and provide geospatial data based on Standards for Geographic Information

(Guideline 3) It is desirable that an explanation of geospatial data of a document which an information user can understand is given as metadata and linked to the document.

[Explanation]

If metadata are described formally a computer can search and read data efficiently.

JMP prescribes common specifications of metadata contained in geospatial data.

A wide reduction in processing time is expected by preparing and providing geospatial data and metadata in accordance with JPGIS and JMP. In addition, an environment enabling mutual use of data will then be created, which is expected to enable sharing of data prepared by different agents, reduce system dependency, eliminate redundant investment, etc.

This guideline can be satisfied by registering metadata with a data catalog system.

9.3.5 Guidelines Regarding Real-time Data

9.3.5.1 Grade 1 Guidelines

(Guideline 1) Specifications for data acquisition should be specified.

[Explanation]

The properties of real-time data and real-timeness required thereof depend on equipment which acquires such data and the system which provide them. In order for a computer to acquire and interpret those data, however, their acquisition methods and notational specifications need to be specified.

(Guideline 2) If tabular or geospatial data are to be acquired in a file format, grade 1 guidelines for each data type should be satisfied.

[Explanation]

Real-time data are normally acquired and interpreted by a computer. Therefore, if they are tabular or geospatial data, they should satisfy at least grade 1 guidelines for each.

9.3.5.2 Grade 2 Guideline

(Guideline 3) It is desirable that methods are provided for acquiring the latest and/or differential values of real-time data.

[Explanation]

The shorter transmission/reception time is, the better it is for ensuring the real-timeness of data transmission/reception. For this purpose, the less the volume of data transmitted/received, the more desirable it is. Providing methods for acquiring the latest and/or differential values of real-time data can reduce the volume of data transmitted/received.

Ways to make such an approach possible include providing RDF data real time and providing data by using a system based on the "External Specifications for the Information Circulation and Sharing Platform Systems," which uses Streams API.

9.3.6 Check list

Given below is a check sheet for confirming whether guidelines shown in this section are satisfied or not.

Meeting all conditions with ◎ on the check sheet satisfies Level 2 data requirements of open data technical levels specified in Section 8.4. Moreover, meeting all conditions with ○ in addition to ◎ satisfies Grade 1 of Level 3 data requirements, and furthermore, meeting other conditions satisfies Grade 2 of Level 3 data requirements.

Table 9-3 Checklist for Guidelines Regarding Tabular Data

Required	Conditions	Confirmation
◎	Whether a non-proprietary (standard) file format such as CSV, Office Open XML (.xlsx), OpenDocument (.ods) is used. ● If open data in multiple file formats are to be disclosed, it suffices if those formats include the above.	
○	Whether each single file disclosed consists only of a single table.	
○	Whether the head of a file disclosed consists only of a single row.	
	Whether no information other than relevant data is contained in any field of a file.	
	Whether no field is combined with another.	
	Whether every field contains a value except where no value is available.	
	Whether a field containing a year value represents it in accordance with the Gregorian Calendar or whether there is a field containing a year in the Gregorian Calendar that corresponds to one containing a year in the Japanese Calendar.	
	Whether the character code used is specified if the CSV file format is used.	
	Whether a character code used widely internationally such as UTF-8 is used if the CSV file format is used.	
	Whether meta data indicating a file's properties and descriptions are formally described and linked to the file.	

Table 9-4 Checklist for Guidelines Regarding Document Data

Required or not	Conditions	Confirmation
◎	Whether a non-proprietary (standard) file format such as Office Open XML (.docx), OpenDocument (.ods) is used. ● If open data in multiple file formats are to be disclosed, it suffices if those formats include the above.	
	Whether the structure existing in a document including parts, chapters, sections, figures, tables, etc. is described in a highly machine-readable format. ● Whether style functions provided by a document editing software program such as heading is used to represent structure if such a program is used. ● Whether tags such as and <h3> are used to represent structure if a document is written in HTML.	
	Whether text includes no formatting symbols or characters (such as blanks and line breaks).	
	Whether tabular data, if any, contained in text meet required conditions in the "Checklist for Guidelines Regarding Tabular Data" (Table 9-).	
	Whether the character code used is specified if document data in a text format such as an HTML document are used. ● Whether meta tags such as the following are included in an HTML document, if any. <meta charset=”UTF-8”> <meta http-equiv=”Content-Type” content=”text/html; charset=UTF-8” />	
	Whether a character code used widely internationally such as UTF-8 is used if document data in a text format such as an HTML document are used.	
	Whether an explanation of the text of a document which an information user can understand is given as metadata and linked to the document.	

Table 9-5 Checklist for Guidelines Regarding Geospatial Data

Required or not	Conditions	Confirmation
◎	Whether a non-proprietary (standard) file format such as GML, KML, etc. is used. ● If open data in multiple file formats are to be disclosed, it suffices if those formats include the above.	
○	Whether the geodetic system is specified in addition to location information including latitude and longitude if data regarding location information is given. ● For an outdoor location, it is desirable to use the World Geodetic System. ● Whether, for an indoor location, the coordinate system and the drawing scale (input precision) is indicated.	
	Whether cartographic information data is in the vector format. ● Whether Japan Profile for Geographic Information Standards (JPGIS) based on the latest ISO and JIS standards are used in creating data in the vector graphic format.	
	Whether an explanation of geospatial data of a document which an information user can understand is given as metadata and linked to the document.	

Table 9-6 Checklist for Guidelines Regarding Real-time Data

	Conditions	Confirmation
○	Whether specifications for data acquisition are specified.	
○	Whether a file to be acquired meets required conditions in the "Checklist for Guidelines Regarding Tabular Data" (Table 9-) or the "Checklist for Guidelines Regarding Geospatial Data (Table 9-) if data in tabular format or geospatial data are to be acquired in a file format.	
	Whether methods are provided for acquiring the latest and/or differential values of real-time data.	

Appendix

Chapter 10. (Appendix) Standards and Tools Regarding Open Data

This chapter describes standards and tools informative for creating and editing machine-readable open data.

Section 10.1 cites typical standards for each of file formats explained respectively in Sections 8.1, 9.2, and 0 — tabular data, document data, and geospatial data — as well as typical software programs capable of handling files in each format. Protocols for transmitting these data and Simple Data Format mentioned in Section 9.3.2 "Guidelines Regarding Tabular Data."

Section 10.2 enumerates identifiers that can be used to identify open data explained in Section 9.1.

Section 10.3 describes tools useful for creating, editing, and publishing open data.

10.1 Standards for Data Formats

10.1.1 Standards for File Formats

Typical standards for tabular data file formats and typical software programs that can handle each such file format are shown in Table 10-1.

Table 10-1 Typical Standards for Tabular Data

Standard	Developer and/or Publisher	Standard Number	Description	Software Programs Capable of Handling the File Format
		★2		
Microsoft Office Binary (.xls)	Microsoft		This is the file format for Excel workbooks up to Microsoft Excel 2003. Its specifications were disclosed under Microsoft Open Specification Promise in July, 2008.	Microsoft Excel, OpenOffice, etc.
		★3		
Office Open XML (.xlsx)	Microsoft	ISO/IEC 29500	This is one of file formats based on XML for an office suite. It is a standard file format for Microsoft Excel 2007 and later versions.	Microsoft Excel, OpenOffice (*), etc.
OpenDocument (.ods)	Organization for the Advancement of Structured Information Standards (OASIS)	ISO/IEC 26300	This is one of file formats based on XML for an office suite.	Microsoft Excel, OpenOffice, etc.

CSV (Comma-Separated Values) (.csv)	Internet Engineering Task Force (IETF)	RFC 4180	A CSV file is a text file containing text data in which fields are separated by a comma ",". Although no official specifications existed for this format for a long time, it was standardized as RFC 4180 in October, 2005.	Microsoft Excel, OpenOffice, etc.

Typical standards for document data file formats and typical software programs that can handle each such file format are shown in Table 10-2.

Table 10-2 Typical Standards for Document Data

Standard	Developer and/or Publisher	Standard Number	Description	Software Programs Capable of Handling the File Format
★1				
Portable Document Format (.pdf)	Adobe Systems	ISO 32000-1	This is a format for electronic document developed and advocated by Adobe Systems. It has a feature that allows you to view text, images, etc. in a PDF file in almost the same way in every environment without depending on any specific environment.	Acrobat, etc.
★2				

Standard	Developer and/or Publisher	Standard Number	Description	Software Programs Capable of Handling the File Format
Microsoft Office Binary (.doc)	Microsoft		This is the file format for Word documents up to Microsoft Word 2003. Its specifications were disclosed under Microsoft Open Specification Promise in July, 2008.	Microsoft Word, OpenOffice, etc.
Rich Text Format (.rtf)	Microsoft		This is a format created by adding a set of control characters for formatting and layout to the plain text format. It has a feature that allows you to designate the font, font color, font size, and font style such as bold as well as to perform simple document formatting by inserting images, centering text, creating lists and tables, etc.	Microsoft Word, OpenOffice, etc.
★3				
Office Open XML (.docx)	Microsoft	ISO/IEC 29500	This is one of file formats based on XML for an office suite. This is the standard file format for Microsoft Word 2007 and later versions.	Microsoft Word, OpenOffice (*) etc.
OpenDocument (.odt)	Organization for the Advancement of Structured Information Standards (OASIS)	ISO/IEC 26300	This is one of file formats based on XML for an office suite.	Microsoft Word, OpenOffice, etc.

Standard	Developer and/or Publisher	Standard Number	Description	Software Programs Capable of Handling the File Format
HTML (HyperText Markup Language)	World Wide Web Consortium (W3C)	ISO/IEC 15445	This is one of markup languages for document description on the web. It enables mutual reference to documents, figures, and tables by using hypertext links.	Microsoft Word, OpenOffice, etc.
XML (Extensible Markup Language)	World Wide Web Consortium (W3C)		These are specifications that can be applied generically to create a group of markup languages for individual purposes.	
XHMTL (Extensible HyperText Markup Language)	World Wide Web Consortium (W3C)		This is a markup language created by redefining HTML in accordance with the XML grammar.	

Typical standards for geospatial data are shown in Table 10-3. Tools that can handle data in these standards are enumerated in Section 10.3.3.

Table 10-3 Typical Standards for Geospatial Data

Standard	Developer and/or Publisher	Standard Number	Description
★2			
Shapefile	ESRI		This is a file format used as an open standard for the mutual use of data by different geographic information systems (GIS).
★3			
GML (Geography Markup Language)	Open Geospatial Consortium (OGC)	ISO 19136	An XML-based markup language for describing geographical features. Fundamental Geospatial Data provided by the Geospatial data Authority of Japan since April, 2008 are provided in this format.
KML	Open Geospatial Consortium (OGC)		This is an XML-based markup language developed for controlling the display of 3D geospatial data in application programs. It does not support the definition of the geodetic reference system presupposed by coordinate systems.
GeoJSON	None (This is developed and managed by developers over the world.)		This is a file format for encoding spatial data along with their non-spatial attributes using JSON (JavaScript Object Notation). The spatial data include points (therefore addresses and locations), line strings (therefore streets, highways, and boundaries), polygons (countries, provinces, tracts of land), etc.

10.1.2 Standards for Data Transmission Protocols and Formats

Typical data transmission protocols and standards for data transmission formats are shown in Table 10-4.

Table 10-4 Typical Data Transmission Protocols and Standards for Data Transmission Formats

Standard	Developer and/or Publisher	Standard Number	Description
FTP (File Transfer Protocol)	Internet Engineering Task Force (IETF)	RFC 959	This is a typical protocol for transferring (document, image, video, or any other) files between a terminal and a server.
HTTP (HyperText Transfer Protocol)	Internet Engineering Task Force (IETF)	RFC 2616	This is a communication protocol used for transmitting and receiving HTML files and other content between a web browser and a web server.
REST			This is a way to handle resources (data) on the web by using URL to identify target resources and using the four HTTP methods of "GET, POST, PUT, and DELETE" to perform operations of "retrieve, create, update, and delete" respectively.
SOAP	World Wide Web Consortium (W3C)		This is a protocol for software programs to exchange messages (objects). Messages exchanged are based on XML.
SPARQL	World Wide Web Consortium (W3C)		This is a query language for searching and manipulating data based on the RDF model.
JSON (JavaScript Object Notation)	Internet Engineering Task Force (IETF)	RFC 4627	This is a lightweight data description language based on the object notation of JavaScript.

Standards for real-time data transmission include the following.

- Streams API[56]
  - This is a mechanism for maintaining the HTTP connection between a server and a client and returning the updated value each time.
  - Standardized by World Wide Web Consortium (W3C).
  - Used by Twitter and others.
- GTFS（General Transit Feed Spec） Realtime[57]
  - GTFS is a common format used for public transportation schedules and associated geographic information.
  - GTFS-realtime is a feed specification that allows public transportation agencies to provide realtime updates about their fleet to application developers.
  - It is standardized by Google.
- External Specifications for the Information Circulation and Sharing Platform Systems
  - This is a standard for using Streams API to transmit real-time data.

10.1.3 Format for Describing Definitions for Tabular Data: Simple Data Format

Simple Data Format[58] is a JSON file format composed of following fields is one of standards which defines data in the CSV file format in a separate file.

- name (of the data)
- licenses
- datapackages_version (version)
- resources (CSV file definition)
  - url (of the data)
  - path (to the data)

---

[56] Moussa, Feras and Yoshino, Takeshi. Streams API. [Online] November 5, 2013. W3C Working Draft. http://www.w3.org/TR/streams-api/.
[57] Google. GTFS-realtime. July 26, 2012. https://developers.google.com/transit/gtfs-realtime/
[58] Simple Data Format. March 16, 2014. http://dataprotocols.org/simple-data-format/

- schema (definition of the data in the CSV file indicated by url or path)
- fields (definition of columns in the CSV file)
- name (of a column)
- type (data type／string, number, integer, date, time, datetime, boolean, binary, object, geopoint, geojson, array, any）
- description (of a column)

Information based on Simple Data Format is specified to be encoded in UTF-8. CSV files referred to in Simple Data Format are also specified to be encoded in UTF-8.

The standard for a new version of RFC 4180 is considered in the CSV on the Web Working Group Charter by W3C[59]. Simple Data Format is cited as a candidate for such consideration.

[59] http://www.w3.org/2013/05/lcsv-charter/

10.2 Standards for Identifiers

Typical identifiers that can be used for identifying open data are shown in Table 10-5.

Table 10-5 Standards for Identifiers Usable for Identifying Open Data

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
Generic	ucode [ITU-T H.642.1]	urn:ucode:_0123456789ABCDEF0123456789ABCDEF	uID Center	It is an identifier that can be assigned to anything including objects, places, concepts, etc. Since reusing the same ucode is prohibited, its uniqueness is guaranteed permanently.	128 bits	○	○
Physical distribution	EPC SGTIN (Serialized Global Trade Item Number)	urn:epc:id:sgtin:4512345.167890.2 urn:epc:tag:sgtin-96:2.4512345.167890.2	GS1	Code for product identification SGTIN-96, which is a 96-bit code, is composed of header (8 bits), filter value, which indicates a logistical type (3 bits), partition value (3 bits), company prefix (20-40 bits), item reference (24-4 bits), and serial number (38 bits). The total number of bits for the company prefix and the item reference is 44.	96 bits	Δ	○

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
Data	DOI (Digital Object Identifiers) [ISO 26234]	http://dx.doi.org/10.1021/jo0349227	The International DOI Foundation	DOI is an identifier permanently assigned to a document on the Internet. A user is redirected through the DOI directory to the address of the referred object in order to avoid the problem of a dead link due to a change in the server on which the object is stored. The DOI system is widely used in the area of academic publications, and a DOI is assigned to an article in academic journals. It can be assigned not only to a title of a book but also to any page, figure, or table in a book or even a tune in a CD.	Variable	○	○
	UUID (Universally Unique IDentifier) [ISO/IEC 11578]	urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6	None (random number)	A code intended to enable distributed systems to uniquely identify information without significant central coordination. Version 4 UUIDs based on random numbers are frequently used now. They are often used as content IDs for blogs and similar objects.	128 bits	×	○

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
Companies and organizations	Company code [ISO 6523]	urn:oid:1.3.170.201233049	ICD (International Code Designator) specified by ISO.	This is a method of assigning a code for identifying an organization (company) specified by ISO (ISO/IEC JTC1 SC32). It is a multicode system that can contain multiple corporate and/or organizational code systems. The first four digits of a code identify an ICD. The rest of the code is determined by the ICD. At present, about 150 ICDs are registered.	Variable	×	×
	Teikoku Company Code	urn:oid:1.3.170.201233049	Teikoku Databank	One of ICDs. This is an identifier issued by Teikoku Databank for identifying companies subject to its corporate credit research. It is also used, for example, when a company tries to acquire an electronic certificate. About 175 companies are registered.	decimal numeration, 9 digits	×	×

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
	Standard Company Code	urn:oid:1.3.147.123456	Japan Institute for Promotion of Digital Economy and Community (JIPDEC)	This is a company code for uniquely identifying a company which sends or receives information in an information sharing system using EDI or AIDC media. About 25,000 companies are registered.	12 numeric (0-9) and/or upper-case alphabetic (A-Z) characters	×	×
Municipality	Municipality Code (the standard regional code used in statistics)	http://statdb.nstac.go.jp/lod/sac/13101	Ministry of Internal Affairs and Communications	It is a standard code indicating prefectural and municipal divisions for representing and sharing statistical information. It was instituted in April, 1970 in light of the report by the Statistical Council.	decimal numeration, 5 digits	○(*)	×

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
Books	ISBN [ISO 2108]	urn:isbn:4-13-060800-2	International ISBN Agency Japan ISBN Agency (national administrative body)	ISBNs constitute a system of numeric book identifiers. There are 10-digit ISBNs (ISBN-10) whose form is X-A-B-C and 13 digits ISBNs (ISBN-13) prefixed by 978. X is the registration group element identifying a language-sharing country group. A is the registrant element identifying a particular publisher. B is the publication element identifying the particular edition and format of a specific title. C is a check digit. The number of digits of X, A, and B vary by the number of items to be identified. ISBN-13 is integrated with JAN/EAN code (book JAN code).	decimal numeration, 10 or 13 digits	×	×

Type	Standard	URI Examples	Administrative Body	Description	Length	Durability	Individual Identifiability
	ISSN [ISO 3279]	urn:issn:1560-1560	ISSN Network/National Diet Library (national administrative body)	This is a system of numeric identifiers for serial publications such as academic journals. An ISSN is composed of eight digits and normally represented as two groups of four digits separated by a hyphen. The first four digits are assigned by each country and the next three digits are assigned as a serial number. The last digit is a check digit and calculated by using the modulus 11.	decimal numeration, 5 digits	×	×
Others	OpenID	http://<username>.openid.ne.jp/	OpenID Foundation	A user identifier for single sign-on (logging in to multiple sites with the same ID and password).	Variable length	×	×

10.3 Tools Useful for Creating, Editing, and Publishing Open Data

10.3.1 Web Service

A web service refers to one that provides information to a browser installed in a PC or smartphone in accordance with the protocol known as HTTP. Most of rental server services provide the function of a web service. Table 10-6shows typical tools and where they are available.

Table 10-6 Typical Web Server Tools and Where They Are Available

Tool	Developer and Provider	Available From
Apache HTTP Server (Free)	Apache Foundation	http://www.apache.org/
Microsoft Internet Information Services (IIS) (Free)	Microsoft Corporation	http://www.microsoft.com/ja-jp/server-cloud/windows-server/

10.3.2 Data Catalog System

A data catalog system refers to a software program which provides the service of registering and managing data and publishing them as a portal site. CKAN is explained in Chapter 11. Table 10 7shows typical data catalog systems and where they are available.

Table 10-7 Typical Data Catalog Systems and Where They Are Available

The name of System		Website
CKAN (free)	Open Knowledge foundation	http://www.ckan.org/
DKAN (free)	Drupal Project	https://www.drupal.org/project/dkan
Socrata Open data Server (charge/free)	Socrata	http://www.socrata.com/products/open-source-development-community/

10.3.3 GIS System

A GIS system is a software program that creates and edits geospatial data.

Table 10- shows typical tools and where they are available.

Table 10-8 Typical GIS Tools and Where They Are Available

Tool	Developer and Provider	Available From
QGIS (Free)	QGIS Development Team	http://qgis.org/
Google Earth (Free)	Google	http://www.google.co.jp/earth/
GRASS GIS (Free)	GRASAS Development Team	http://grass.osgeo.org/
ArcGIS (Free)	ESRI	http://www.esrij.com/products/arcgis/

10.3.4 Information Circulation and Sharing Platform System

An Information Circulation and Sharing Platform System refers to an environment which has prepared versatile technical and operational rules for distributing and sharing applications for registering and using open data.

The "External Specifications for the Information Circulation and Sharing Platform Systems[60]" has been published as a standard for showing how to build applications and servers for registering and using open data in creating such an environment to facilitate building such applications and servers.

[60] http://www.opendata.gr.jp/cfc/

10.3.5 RDF Repository

An RDF repository refers to a database system which stores RDF data and allows search by SPARQL.

Table 10-shows typical repositories and where they are available.

Table 10-9 Where Typical RDF Repositories Are Available

Tool	Developer and Provider	Available From
AllegroGraph (Paid or Free)	Franz	http://www.franz.com/agraph/allegrograph/
Apache Jena (Free)	Apache Foundation	http://jena.apache.org/
Neo4j (Free)	Neo Technology	http://www.neo4j.org/
Sesame (Paid or Free)	Aduna	http://www.openrdf.org/
Virtuoso RDF (Paid or Free)	OpenLink Software	http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSRDF

Chapter 11. (Appendix) CKAN, a Data Catalog System

This chapter explains CKAN, which is widely used as a data catalog system for satisfying Level 3 requirements for data catalogs explained in Section 8.4 "Technical Levels of Open Data."

11.1 What CKAN Is

11.1.1 Overview

CKAN[61](Figure 11-1) is a web-based data management and distribution system and offered free of charge by Open Knowledge Foundation.

CKAN is used by many governmental organizations distributing open data including data.gov (U.S.), data.gov.uk (U.K.), publicdata.eu (EU), data.gov.au (Australia), DATA.GO.JP (trial version of the Data Catalog Site of the Japanese government), and datameti.go.jp (Open DATA METI).

Official documentation about how to install and configure CKAN is found at:

http://docs.ckan.org/en/latest/

How to install CKAN in different environments is described at:

https://github.com/okfn/ckan/wiki/How-to-Install-CKAN

[61] http://ckan.org/

Figure 11-1 Welcome Screen of CKAN

11.1.2 Explanation of Terms

Terms used in CKAN are explained below.

- User
  - A user is an agent who registers data with CKAN.
- Dataset
  - A dataset is a parcel of data:
  - for instance, the "statistical data for year N," "temperature data for the area A," etc.
- Organization
  - An organization is an agent which publishes and manages (controls access to) data:
  - for instance, "Ministry of M," "Section S," "Bureau B," etc.
  - Each organization can manage a dataset.
  - Users added to an organization is authorized to add or edit a dataset and to view a dataset.
- Group
  - A group is a collection of datasets for a particular community or topic.
- Tag

- A tag characterizes data.
- Examples of tags include "finance," "measurement," and "transportation."
- They serve as keywords for searching data.

11.2 Things to Consider and Prepare before Using CKAN

It is desirable to consider and prepare the following before using CKAN to maintain a data catalog.

11.2.1 Sorting Out Open Data to Be Published

List open data subject to publication. Chapter 3.

11.2.2 Determining a Policy for Managing Open Data

It is desirable to determine a policy described below in advance.

① How to configure datasets and organizations

Each dataset is controlled as to whether it is public or private. A private dataset can be viewed only by users who are members of the organization owning the dataset.

List organizations and datasets in light of the above and group open data to be published into datasets.

② Setting Up Groups and Tags

Groups and tags are something intended to enhance convenience. Decide what groups and tags to set up as well as which group each dataset of open data belongs to and what tags are to be added to each dataset.

③ Selecting a Rule for Using Data Provided

Select a rule for using open data that should be applied to each dataset of open data.

④ Establishing the Rule for Adding and Managing Data

Access CKAN and specify one or more users who are in charge of adding data as well as the procedures for doing so. Establish a rule documenting these. Such procedures include those for registering for an account, for adding data, etc.

In addition, a sysadmin user shall be specified for each organization.

11.2.3 Developing Requirement Specifications

knowledge of a server system is required for installing a CKAN system. Depending on how it is used, some configurations require using a console or changing system codes.

If these operations are to be outsourced, prepare the requirement specifications which at least

include the following.

- List of open data to be handled
- Datasets to be added, organizations to be created, and open data users to belong to them
- Group and tag settings
- Instruction to disable registering user from the web page, if this is intended

It is desirable to include the creation of an operation manual in the requirement specifications.

11.2.4 Data Preparation Plan

Create a plan for converting listed data into those that are more machine-readable in accordance with Chapter 3 and Chapter 8 of this guide and implement the plan.

It is desirable to disclose data even if they are less machine-readable or if there is no open rule for using them. It suffices add or renew data or a rule for using them as soon as more machine-readable data or an open rule for using data becomes available.

It is desirable to review the plan every fiscal year or at an appropriate interval in comparison with the state of its implementation.

11.3 Example of Adding Open Data to CKAN

As a result of consideration in light of procedures specified in エラー! 参照元が見つかりません。, City X decided to use CKAN to manage open data shown in Table 11-1.

Table 11-1 Example of Open Data to be Managed

Title	File Formats	Rule for Using Data	Managing Division
AED Locations	GeoJSON / KML	CC-BY	Division A
Demographic Statistic for FY2013	Excel(.xslx) / CSV	CC-BY	Division B
FY 2014 Budget	CSV	CC-BY	Division C

It was also decided that these open data were to be managed by each division and that personnel shown in Table 11-2 were to be involved.

Table 11-2 Staff Managing Open Data

Name	Division	Account
Taro Yamada	(Sysadmin user)	ckan_admin
Hanako Sato	Division A	div_a
Jiro Suzuki	Division B	div_b
Saburo Yamamoto	Division C	div_c

The procedure for adding the above data to CKAN, starting from the initial screen, is shown below.

11.3.1 Registering for an Account

Each staff member is to register for a user account for adding data to CKAN. The registration procedure is as follows.

1. Click the “Register” link in the upper right corner of the initial screen (Table 11-2, left). Then a window appears which asks you to enter username, e-mail address, and other information (Table 11-2, right).
2. Enter required information including username, e-mail address, and password.
3. If you click the "Create Account" button, registration will complete. You will have then been logged in.

Once your account is created, log out and notify the sysadmin user of account information registered.

Table 11-2 Registering for an Account

11.3.2 (For Sysadmin) Creating an Organization

While controlling access to data, a sysadmin user is to create organizations for relevant divisions so that each division may add open data to CKAN. The procedure for doing so is indicated below.

1. Click the "Organization" link at the top of the initial screen (screen after login/upper left).
2. Click the "Add Organization" button (upper right).
3. Click the "Edit" button below the title field (lower left)
   - Enter a title, an URL, description, and an image URL.
   - Only an URL is required.
   - If the title contains only alphanumeric characters, this step 3 is not required.
4. In the URL field, enter a string of alphanumeric characters that has almost the same meaning as the title (lower right).
5. Click the "Create Organization" button. The creation of an organization will then be complete.

A user who creates an organization becomes its "Admin." In this instance, the sysadmin becomes the organization's Admin.

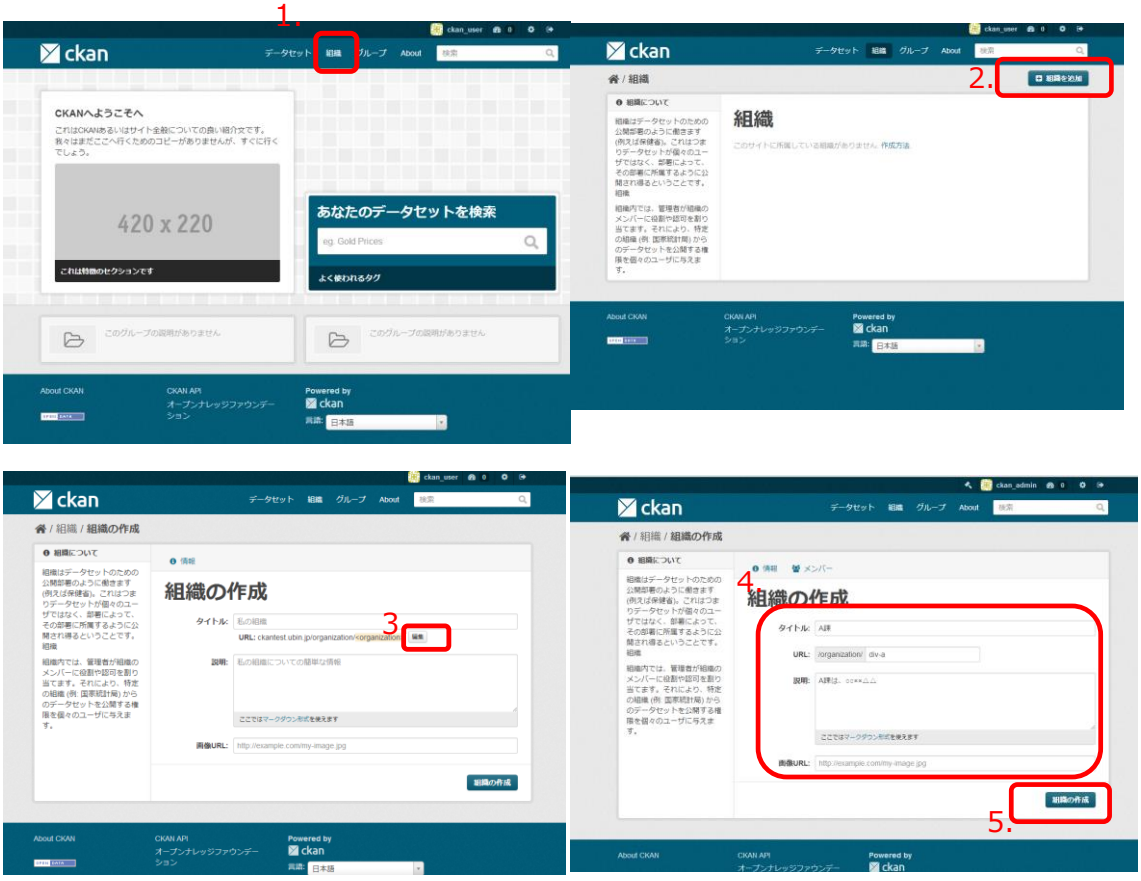


Table 11-3 Creating an Organization

11.3.3 (For Sysadmin) Adding Members to an Organization

Next, a sysadmin is to add members to the organization created. The procedure for adding a user "div_a" to Division A is shown below.

1. Click the "Organization" link at the top of the initial screen (screen after login/upper left).
2. Select "Division A."
3. Click the "Admin" button on the organization's page (upper right).
4. Select the "Members" tab (lower left).
5. Click the "Add member" button.
6. Click the username field and enter the username "div_a" (lower right).
   - If you enter the first few characters of the username, a list of all matching usernames pops up. You can select the one you wan to enter from the list.
7. Select an access role to be given to this user.

   There are three roles: "Admin," "Editor," and "Member." Their description is given on the left of the screen.

   In this instance, "Editor" is selected.
8. Click the "Add" button.

When this operation is complete, the syadmin is supposed to notify the staff members.

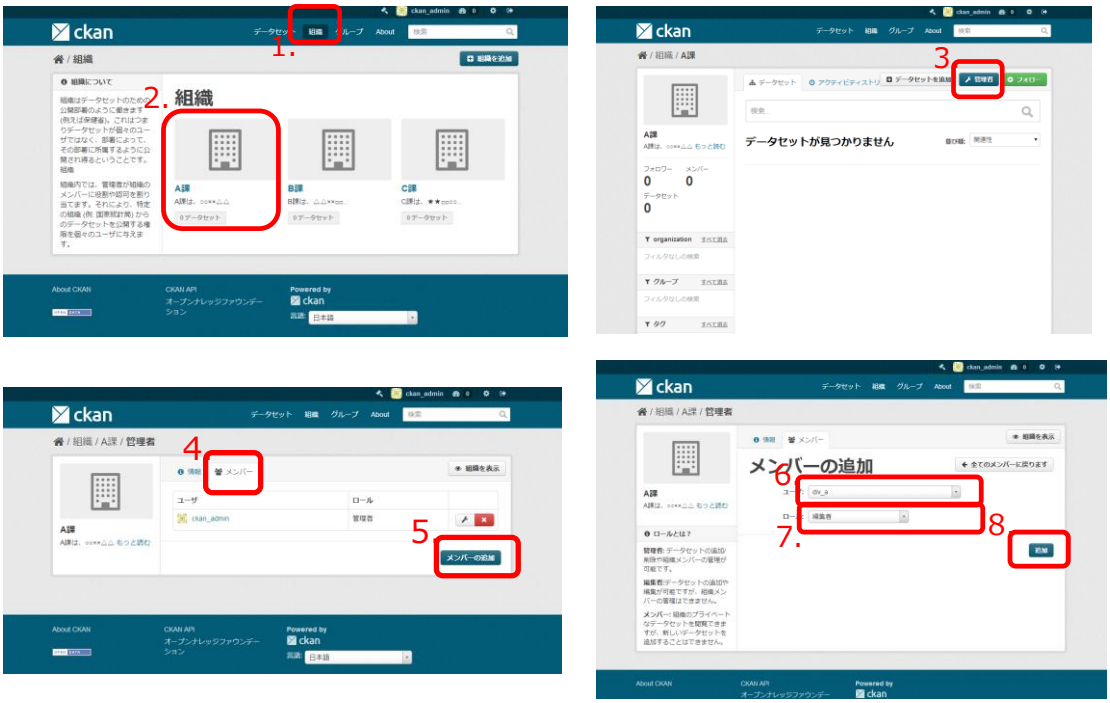


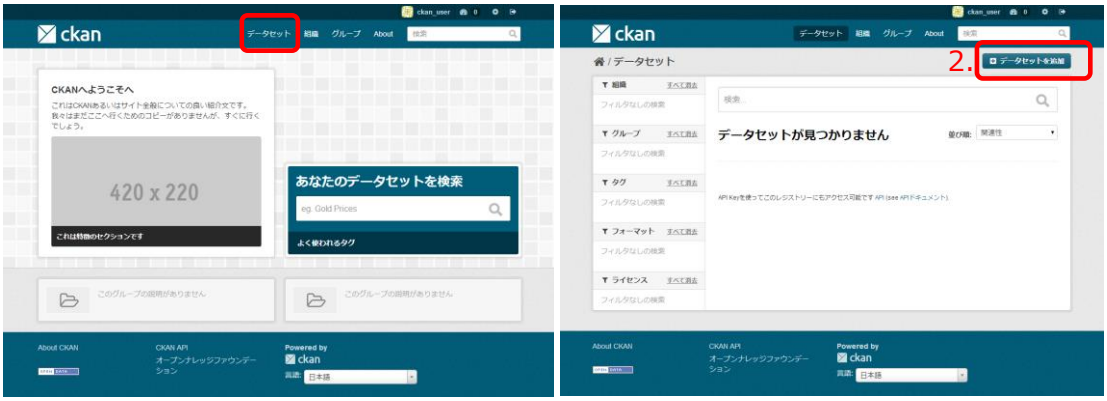
Table 11-4 Adding Members to an Organization

11.3.4 Creating a Dataset

A staff member notified by the sysadmin is to create and add a dataset.

Its procedure is illustrated by taking the example of "AED Locations" managed by "Division A."

1. Click the "Datasets" link at the top of the initial screen (screen after login/Figure 11-5, upper left).
2. Click the "Add Dataset" button (Figure 11-5, upper right).
3. Enter the title "AED Locations" as well as description and tags for the data (Figure 11-5, lower right).
   - Only an URL is required.
   - If the title contains any character other than alphanumeric characters, click the "Edit" button below the title field and enter the string of alphanumeric characters having almost the same meaning as the title into the URL field that appears.
4. For the "license" field, select the "Creative Commons Attribution" since the license for the dataset "AED Locations" is "CC-BY."[62]
5. Select an organization to which this dataset belongs. Since this user can only edit data owned by Division A, only "div_a," which denotes "Division A" is displayed in the drop-down list.
6. Click "Next: Add Data" button. You will then be taken to the "Add data" screen.



[62] If it is not "CC-BY," select an appropriate license from the drop-down list.

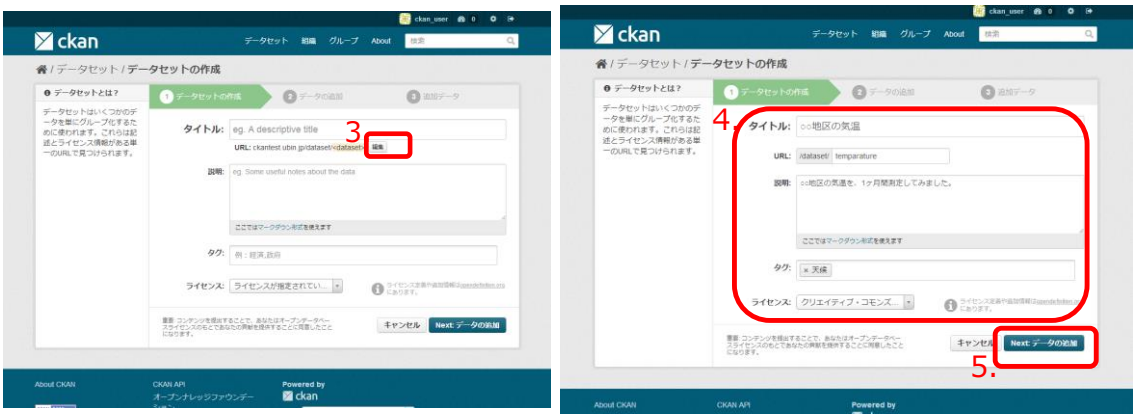


Figure 11-5 Creating a Dataset

11.3.5 Adding Data

After adding a dataset, add data belonging to that dataset.

In this example, "aed.geojson" and "ade.kml" are uploaded as data for "AED Locations." Its procedure is shown below following the preceding section.

1. Add data to be added. (Figure 11-6, Left)
   - Choose a file or link for the data by selecting “Link to a file,” “Link to an API,” or “Upload a file.” In this instance, "Upload a file" is selected, so that "aed.geojson," the file to be uploaded, is selected.
     - ➢ In order to "upload a file," it is necessary to configure datastore on the server.
2. Fill in the "description" field and enter "GeoJSON" into the "Format" field.
3. Click the "Save & add another" button and repeat steps 1 and 2 to upload "aed.kml." Enter "KML" into the "Format" field this time.
4. Click the "Next: Additional Info" button.
5. Select "Public" in the "Visibility" field (Figure 11 6, right).
   - If you select "Private," only users belonging to the organization owning the dataset can see it. This function can be used for confirmation purposes before publishing open data.
6. Enter additional information as required including the author producing the data, the author's e-mail address, the maintainer, the group they belong to, etc. All of these items are optional.
7. Click the "Finish" button. This completes adding data.

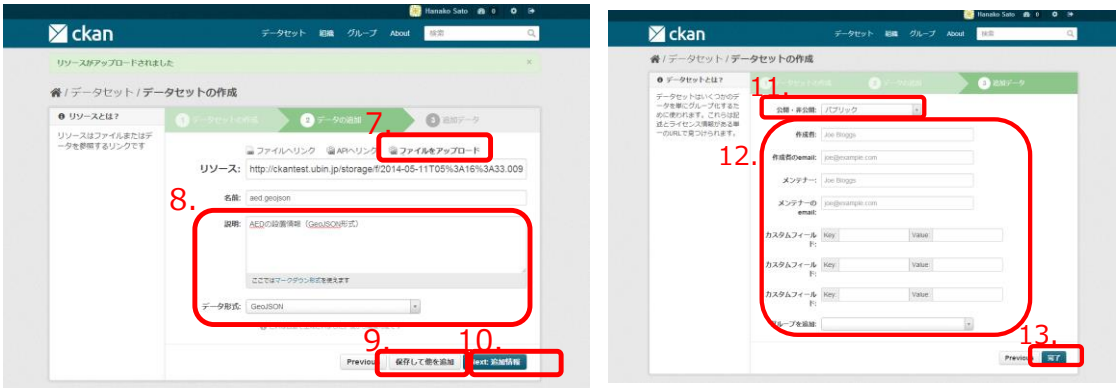


Figure 11-6Adding Data

When adding data is finished, a screen like one shown in Figure 11-7 appears.

Figure 11-7 Screen after Adding Data

Add data managed by Division B and C in accordance with the same procedure as indicated in this and preceding sections.

After those data are added, the list of datasets will be as shown in Figure 11-8.

Figure 11-8 Screen Showing the List of Datasets after Adding Datasets

11.3.6 Administration Page of CKAN

CKAN's sysadmin users can access the administration page of this data catalog system[63].

The sysadmin user can view, edit, and delete all users, datasets, and data in the data catalog.

The following functions are available on the web page.

- Changing the "look and feel" of the CKAN site (Figure 11-9, upper left)
- Changing organizations to which a dataset belongs (Figure 11-9, upper right)
- Deleting a dataset
- Managing users（Figure 11-9Bottom)

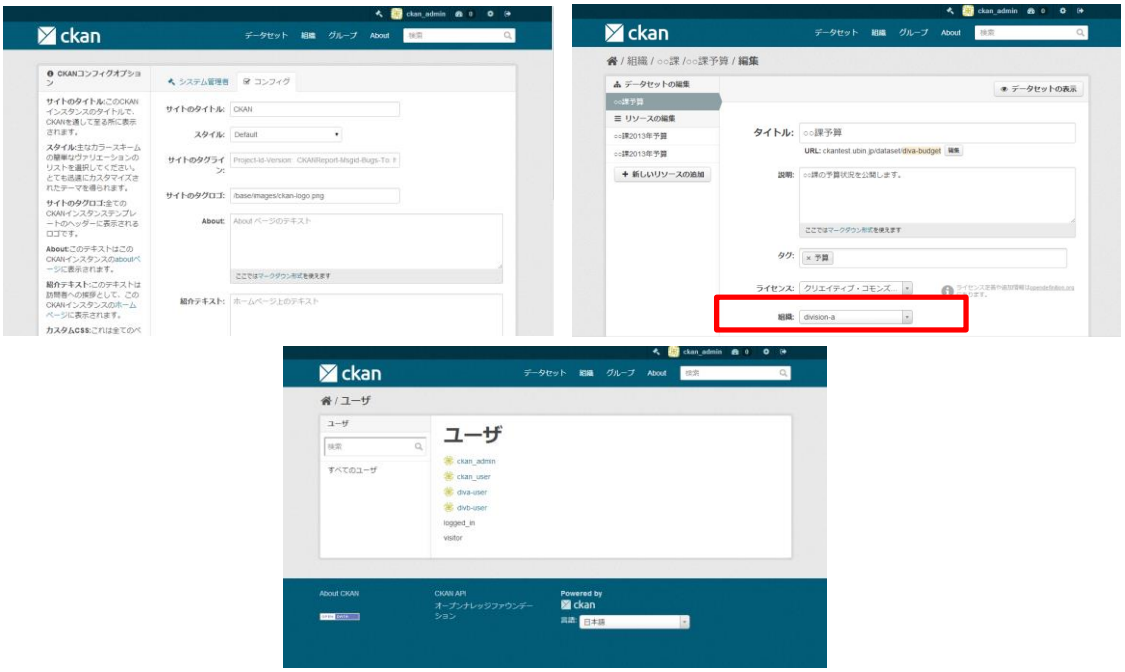


Figure 11-9 Administration Screen of CKAN

[63] Creating the first CKAN sysadmin user, however, requires issuing commands from the command line interface on the server that CKAN is installed on.

References

1. **Shafranovich, Y.** Common Format and MIME Type for Comma-Separated Values (CSV) Files. [Online] October 2005. RFC 4180. http://www.ietf.org/rfc/rfc4180.txt.

2. **Obama, Barack.** Transparency and Open Government. [Online] 2009. http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment.

3. —. Building a 21st Century Digital Government. [Online] 2012. http://www.whitehouse.gov/the-press-office/2012/05/23/presidential-memorandum-building-21st-century-digital-government.

4. **Moussa, Feras and Yoshino, Takeshi.** *Streams API.* [Online] November 5, 2013. W3C Working Draft. http://www.w3.org/TR/streams-api/.

5. **Gandon, Fabien, Schreiber, Guus and Beckett, Dave.** *RDF 1.1 XML Syntax.* [Online] February 25, 2014. W3C Recommendation. http://www.w3.org/TR/rdf-syntax-grammar/.

6. **Feigenbaum, Lee, et al., et al.** *SPARQL 1.1 Protocol.* [Online] May 21, 2013. W3C Recommendation. http://www.w3.org/TR/sparql11-protocol/.

7. **Bray, Tim, et al., et al.** Extensible Markup Language (XML) 1.1 (Second Edition). [Online] August 16, 2006. W3C Recommendation. http://www.w3.org/TR/xml11/.

Japan is Back. [Online] June 14, 2013. Cabinet Decision. http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf.

Ministry of Internal Affairs and Communications Advancement of Open Data in Statistics. [Online] May 28, 2013. http://www.soumu.go.jp/menu_news/s-news/01toukei01_02000024.html.

10. Declaration to be the World's Most Advanced IT Nation. [Online] Cabinet Decision. http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf.

Ministry of Internal Affairs and Communications Release of White Paper on Information and Communications in Japan in Open Data, April 19, 2013. http://www.soumu.go.jp/menu_news/s-news/01tsushin02_02000053.html.

12. Simple Data Format. [Online] March 16, 2014. http://dataprotocols.org/simple-data-format/.

13. **Open Knowledge Foundation.** Open Data Handbook. *What is Open Data?* [Online] http://opendatahandbook.org/en/what-is-open-data/.

https://www.gov.uk/government/publications/open-data-charter, [Japanese translation] http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html.

15. **Google.** *GTFS-realtime.* [Online] July 26, 2012. https://developers.google.com/transit/gtfs-realtime/.

16. **European Commission.** Digital Agenda for Europe. [Online] 2011. http://ec.europa.eu/digital-agenda/digital-agenda-europe.

17. 5 ★ Open Data. [Online] http://5stardata.info/, [Japanese translation] http://5stardata.info/ja.

Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society Open Government Data Strategy. [Online] July 14, 2012. http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html.

Road-map for Promoting Open Government Data. [Online] June 14, 2013. http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou3.pdf.

Liaison Conference of Chief Information Officer (CIO) of Ministries and Agencies Basic Ideas (Guidelines) for Data Publication of Ministries and Government Offices to Promote Secondary Use. [Online] June 14, 2013. http://www.kantei.go.jp/jp/singi/it2/densi/.

Japan Open Data Charter Action Plan. [Online] October 29, 2013. http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf.

Revision History of the Open Data Guide

Edition	Date of publication	Major revisions
Version 1 (Draft)	June 5, 2014	Appeal for Opinions on the Version 1 (Draft) Started (opinions accepted until June 30, 2014)
Version 1	July 31, 2014	Version 1 is available

Open Data Guide

Rules and Technical Guide for Open Data

Version 1

2014

Published By
Open Data Consortium