Open Data as an Infrastructure
—Impact of Availability of Government Data as Open Data on the Japanese Economy—

Hideyuki Tanaka
Professor, Interfaculty Initiative in Information Studies/ Graduate School of Interdisciplinary Information Studies, The University of Tokyo

Soichiro Takagi
Associate Professor, Center for Global Communications, International University of Japan
Visiting Associate Professor, Interfaculty Initiative in Information Studies, The University of Tokyo

Abstract
This study regards the data owned by national and local governments (public data) as infrastructure that will support the future development of Japanese society and economy. This article estimated the volume of public and private database assets and the impact of the provision of public sector data as open data on Japanese macro-economy.

The study estimated the value of data stocks owned by administrative bodies as well as private corporations by regarding them as “database assets”, and simulated the impact of the provision of public data as open data on Japanese macro-economy. The estimated value of data assets comes to about 2.7 trillion yen for private sector data and about 3.7 trillion yen for public sector data. The provision of public data as open data was estimated to boost up GDP by 158.6 billion yen to 701 billion yen, depending on the assumed parameters. The study clarified the potential of public data to be used as a new type of infrastructure and showed the effect of such use quantitatively by estimates based on published data, thereby could contribute to the related fields of study.

Keywords: Open data, public data, infrastructure
JEL Classification: E65, H49, O38

I. Introduction

In recent years, a trend of offering a new services by incorporating open data has become prevalent. In the previous data usage, a database was constructed and held within an individual organization without disclosure. For this reason, using data and providing a service from a third party has been difficult. However, by disclosing a data in a format available for a secondary usage, such data becomes open data which can possibly broaden new horizons.

Let’s see an example of searching book collection at a public library. If a book collection data is not disclosed in an available format for external use, it is difficult to search a book across several local governments’ libraries at once from single portal website. In this case,
one can search only from libraries of the same local government. However, there is a need for searching a book across local governments’ boundaries. If local governments’ data on book collections are disclosed and externally accessible, a new service can be provided. “Calil,” the cross-library search service, that is listed in “the Data Catalog Site” (data.go.jp) by the Japanese government, is one example of such services. This service is provided by a private company instead of a government agency. Making data open and publically available could possibly bring a positive economic impact as it may encourage private sectors to start new business. For example in Europe and the U.S., large businesses with open data have been developed. Well-known cases may include real estate related advanced information services such as MRIS and Locatable, as well as an insurance service with a weather information¹ called Total Weather Insurance.

Data access enables an innovational development at an individual or an organizational level, as the G8 Open Data Charter of 2013 suggests. Open data could possibly serve as an infrastructure supporting a future social/economic development. Regarding the economic impact of open data, as explained later in this paper, some of the prior studies have already conducted estimations. They, however, are mainly based on a flow-based and market-scale estimation. And as far as the authors know, no work has treated open data as a stock-based infrastructure, or estimates impacts to macro economy based on economic models. Additionally, many of the prior studies tend to have been conducted with data based on individual investigations. Thus, it is difficult to trace a similar estimation where objects (i.e. points in time, countries) are different. On the other hand, by positioning open data as a new infrastructure, this study aims to identify impacts the infrastructure provides on the Japanese macro economy based on disclosed data, with applicable methods for the third parties.

Although there are a variety of definitions, the study basically depends on the definition of the “Open Data Guide” (Open Data Promotion Consortium 2014), “the data in a machine readable format, disclosed under the rule that allows a secondary usage including commercial use.” Although information providers of open data under the definition includes private business operators besides government agencies, this study conducts an analysis by focusing on the disclosure of public data held by government agencies such as the national and local governments. This is because the public data can be directly manipulated as policy variables. In other words, the authors assume that disclosing public data accumulated by the government agencies functions as a supporting infrastructure of economic activities. Based on this assumption, the authors will estimate an economic value of such public data and simulate impacts from partial disclosure of the data on the Japanese macro economy.

The remainder of the paper is structured as follows: In Chapter II, the prior studies regarding the economic effects of open data are reviewed. In the course of the review, the authors show significance of the present study that deal with a database as an infrastructure. Chapter III shows the overview of the methodology of this study. The analyses are divided

in two parts: Chapter IV estimates database assets. Chapter V simulates the impact of open data on the Japanese macro economy based on the estimation of Chapter IV. In Chapter VI, the authors discuss the results of these estimations and simulations and shows the conclusion of this study.

II. Related Literature

The estimations with regard to the economic effects of public data or open data so far have been conducted in three major subject areas. Specifically, they are: (1) Estimation regarding an economic value of public data, (2) Estimation of a market scale related to the use of public data or open data, and (3) Estimation of a ripple effect for a macro economy due to the use of public data or open data.

First, the prior studies related to the economic value estimation of public data are explained. The work of Pira International (2000) is known as an early study. This is to estimate the economic value of data held by the public sectors of the 15 EU nations. Based on the expenses required annually for refining/building public data, it is calculated by multiplying a certain ratio between expenses and revenues. Specifically, the estimated value of the whole 15 countries indicated within the range from 27.7 billion to 134 billion euros, with a median at 68.5 billion euros.

Also, the work of DotEcon (2006) is known for an estimation focusing on the UK. In this case, it is estimated based on the survey conducted by the Office of Fair Trade of the UK toward the persons in charge of the government’s public data. The survey reveals the amount of income (GBP 394 million) gained from providing, selling and licensing public data through a questionnaire. Further, by calculating consumer surplus based on the income, it also estimates the total economic value of public data as being worth GBP 590 million. Additionally, if such usage is available for further promotion, the economic value is assumed to be up to GBP 1.11 billion.

These prior studies have already identified some economic value of the government-held public data. However, such calculations are based on the flow-based numerical value such as an annual expense/income. Since this value is not necessarily handled as a stock, there is a limitation to employing these methods to analyze open data as an infrastructure. Additionally, it should be noted that these prior studies do not verify a ripple effect of the public data on the macro economy from the view point of public data as an infrastructure.

As to the second point, preceding studies estimate the market scale. In this field of study, a market scale created through using public data by private sectors is estimated. It should be noted that an estimated object is neither an economic value nor a market value of the public data per se. For example, Dekkers et al. (2006) estimated the market scale of 25 EU countries and Norway based on the gross sales values of enterprises that use public data information such as geography, weather, and traffic. The result may vary between 10 billion and 48 billion euros, whose median is set to be 27 billion euros.

In addition to information such as geography, weather, and traffic, Vickery (2011) target-
ed a broad range of public data including cultural contents, and estimated a market scale to be 28 billion euros. However, the estimation method is the application of other estimations in comparison to GDP, including the ripple effect on the Australian economy by ACIL Tasman (2008) which is discussed later in this paper. As Jitsuzumi et al (2013) pointed out, application as such has its limitation in reflecting on a difference of industrial structure per country and region.

The work of Hitachi Consulting (2012) estimates the private sector’s market scale by incorporating Japanese public data. Instead of simply applying the prior studies’ estimated value in Europe to Japan based on the GDP ratio, the work estimates the Japanese market scale incorporating data related to the Japanese domestic market scale in 9 fields such as geography, weather, traffic, and healthcare. As an estimated value, it indicates a JPY 513.9 billion market scale.

As to the third point, the authors mention the prior studies regarding the ripple effect of open data on the macro economy. ACIL Tasman (2008) estimated the economic effect of Australian geographic information. By reflecting the benefit of employing geographical information, which improves the related sectors’ productivity and enables the use of new natural resources, the ripple effect on the macro economy is estimated to be in a value range of 6.43 million to 10.26 million Australian dollars. However, this estimation requires noting three major points: First, the data is limited to geographical information. As the other prior studies have discussed, besides geographic information, there are fields where open data is exploited in economic activities. However, ACIL Tasman (2008) does not deal with such fields. Second, the data is not limited to public data only. If it covers the entire geographical information, it should include private sectors’ data. The third point is the fact that a sufficient explanation is not necessarily provided on an estimation of the ripple effect on the macro economy. Although an explanation for the structural element of the ripple effect estimation model is provided, it does not include the structural formula or the specific numerical value of parameters in the report, which causes a limitation because it is difficult for third parties to verify the estimation.

Vickery (2011) is the other work with regard to the third point. It estimates an economic effect on 27 European countries. The estimated value is shown as 140 billion euros, however does not mean that the author structures an economic ripple effect model for the estimation. Vickery (2011) simply applies the prior estimation such as that of ACIL Tasman (2008) based on a comparison with the GDP ratio. This simple application cannot reflect different industrial structures among countries. There is also another limitation other than the reflection of industrial structures, because Vickery (2008) includes cultural contents as data even though ACIL Tasman (2008) covers only the geographical information.

As mentioned previously, a certain amount of prior studies exists with regard to the economic value or the market scale of public data or open data. As far as the authors know, however, there are three limitations in the prior studies. Firstly, no estimation of public data as a stock has been performed. This study treats public data as an infrastructure of economic activities. Furthermore, the data needs to be estimated as a stock in order to estimate its ef-
fect based on a production function. Secondly, regarding the estimation of the ripple effect of the Japanese public data on the macro economy, there are only a few estimations based on Europe’s prior studies which are the comparisons based on the GDP ratio. Therefore, performing an estimation by reflecting the Japanese economic characteristics as much as possible is anticipated. Thirdly, it is difficult to apply the methodology of the prior studies in different countries or times because they are estimated based on their own research data. This may cause issues such as a difficulty in comparing the estimated values among multiple nations. In order to complement mainly these three limitations, this study estimates the effects on the macro economy by the Japanese public data as follows.

Table 1 summarizes the estimations of prior studies discussed in this chapter and indicates the values in the case of applying them to Japan based on the GDP ratio. Although the value applied to Japan is a simplified estimation value, it will be referred to upon validating the estimated result of this study in the later part.

### III. Basic Framework of the Analysis

This study estimates an economic effect of open data through the following 2-step-approach. In Step 1, the study grasps quantitatively how much data is available with a potential for value creation within the public and private sectors. In this study, the data which can

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Estimations of prior studies and simple applications to Japanese economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Source</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>Economic Value, flow base</td>
<td>Pira International (2008)</td>
</tr>
<tr>
<td></td>
<td>DotEcon (2006)</td>
</tr>
<tr>
<td></td>
<td>Dekkers et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Vickery (2011)</td>
</tr>
<tr>
<td></td>
<td>Hitachi Consulting (2012)</td>
</tr>
<tr>
<td></td>
<td>ACIL Tasman (2008)</td>
</tr>
<tr>
<td></td>
<td>Vickery (2011)</td>
</tr>
</tbody>
</table>

Source: The authors configured the chart based on the prior studies
produce this value is regarded as a part of intangibles and called a “DB (database) asset”.

At the first stage, brand value and other parts, which cannot be categorized as DB value are removed from intangibles. This is performed by identifying the relationship of RC (Relational Capital) such as brand value and advertising expense. Subsequently, the relationship between the DB asset and IT capital/IT labor is assessed based on information service enterprises whose intangibles are positive. By applying the relationship obtained here, the private sector’s DB asset that is activated by accelerating open data of public sector will be estimated.

In the second step, a part of the public sector’s DB asset is assumed to be duplicated/transferred to the private sector by exploiting open data. By positioning the transfer from the public sector to the private sector as an increase of the DB asset of the private sector, impacts are analyzed through simulation with the DSGE (Dynamic Stochastic General Equilibrium) model. For conducting analysis, several patterns are assumed at different ratios of transfer from the public sector to the private sector, and different levels of frequency of usage and application in the private sector are assumed. Overall economic impacts on GDP are simulated based on these assumptions. In the following Chapter IV, DB asset’s estimation in the first stage is explained, and the detailed method and results regarding the open data impact on the macro economy in the second step is provided in Chapter V.

IV. Estimation of DB asset

IV-1. Introduction of the concept of DB asset

This study regards the DB asset as a part of intangibles. According to Bontis (1999), intangibles are classified into Human Capital (HC), Structural Capital (SC), and Relational Capital (RC). RC is regarded as being deducted from intangibles as it includes brand values estimated from advertising expenses (Sydler et al., 2014), which is different from the database property discussed in this paper. SC is explicit knowledge that is estimated by research and development expense, which comprises the central part of the DB asset. HC is considered as tacit knowledge that is estimated by labor expense. The authors assume that input of labor generates the DB asset when the estimation is limited in the DB enterprises as mentioned later and that the DB asset is consist of SC and HC in this study (see Figure 1).

IV-2. Estimation of DB asset

This study estimates the DB asset based on the data of information service enterprises. Information service enterprises in this paper consist of listed companies among the three business types of “Software (business support)”, “Professional information site”, and “Financial information services”, of an enterprise information service called “SPEEDA” (by UZABASE, Inc.) and other enterprises whose database services can be confirmed by the explanation of the Annual Securities Report (91 enterprises in total). It is assumed for the
above enterprises that their DB assets might increase by the government’s open data policy.

The authors separate the enterprises into two groups. The first group is “DB Enterprises”, whose intangibles are basically positive with the DB properties available for confirmation\(^2\). In the following analysis, DB enterprises were selected in order to minimize errors of the aforementioned components of intangibles, such as SC, as much as possible, upon clarifying the relationship between IT investment (IT capital and IT related labor) and the DB Asset. There are 54 DB enterprises for the analysis.

The second group is “Non-DB Enterprises” comprising those which do not fulfill the DB enterprises conditions above. There are 37 “Non-DB” enterprises for the analysis. The company data of the analysis are of the panel data between 2001 and 2013.

The method of the analysis takes the following steps: First, by focusing on the Non-DB enterprises, in order to exclude RC (Relational Capital) among intangibles, the relationship between RC and advertising expense is identified. For all cases of enterprises, intangibles were calculated based on Formula (1).

\[
I_{i,t} = MV_{i,t} + DEPT_{i,t} - ASSET_{i,t}
\]

\(I\) means intangibles, \(MV\) is aggregated market value (after treasury stock adjustment), \(DEPT\) is interest-bearing debt balance, and \(ASSET\) indicates total asset. Additionally, \(i\) indicates enterprise and \(t\) the respective year. Data in Section IV-2 is converted to real term with the GDP deflator.

In order to identify the relationship between RC and advertising expense, panel data re-

---

\(^2\) Precisely, the DB enterprises are selected if their value of intangibles measured based on Annual Securities Reports are positive for more than two-thirds of the whole measuring periods.
gression analysis was performed for Non-DB enterprises with Formula (2)³.

\[ I_{i,t} = \alpha + \beta_1 K_{i,t} + \beta_2 L_{i,t} + \beta_3 AD_{i,t} + e \] (2)

\( K \) means the sum of tangible and intangible fixed assets excluding goodwill (hereinafter referred to as tangible fixed assets), \( L \) means the total labor expense, and \( AD \) means the advertisement expense. Those items for which \( AD \) data were not stated on Annual Securities Reports were replaced with zeros. In the respective variables, those over a value of average value ± eightfold standard deviations are removed as outliers (refer to Table 2a for the statistical summary). As a result of the Hausman Test, a fixed-effects model is selected.

Result of the analysis is as per stated in Table 2b. A statistically signified value was shown as a result in any coefficients of \( K \), \( L \), and \( AD \). Amongst these results, by focusing on \( AD \), \( RC \) can be calculated with Formula (3) by applying the coefficient (16.267).

\[ RC_{i,t} = 16.267 AD_{i,t} \] (3)

The coefficient between \( K \) and \( L \) also needs to be discussed here. Basically, for those

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_i )</td>
<td>435</td>
<td>-1,437</td>
<td>15,736</td>
<td>-65,416</td>
<td>96,975</td>
</tr>
<tr>
<td>( K_i )</td>
<td>435</td>
<td>6,297</td>
<td>14,505</td>
<td>0</td>
<td>89,675</td>
</tr>
<tr>
<td>( L_i )</td>
<td>435</td>
<td>2,709</td>
<td>4,169</td>
<td>0</td>
<td>19,851</td>
</tr>
<tr>
<td>( AD_i )</td>
<td>435</td>
<td>230</td>
<td>576</td>
<td>0</td>
<td>4,809</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_i )</td>
<td>-1.244***</td>
<td>0.246</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>( L_i )</td>
<td>-2.901***</td>
<td>0.495</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>( AD_i )</td>
<td>16.267***</td>
<td>2.209</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>10514.64***</td>
<td>1535.222</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: N = 435 (unbalanced), Adj. R²: .381, Fvalue: 7.847 (p < .001)
Hausman test \( \chi^2 = 83.82 \), Prob < 0.001
* p < 0.10; ** p < 0.05; *** p < 0.01 (the same hereinafter).

³ The variables are not logarithmically transformed since the estimated values of the respective variables calculated on an enterprise basis could be applied to a macro economic base (the same shall apply hereinafter).
non-DB enterprises with negative intangibles, tangible fixed assets and labor have a negative correlation with the intangibles. Japanese enterprises are said to have resolved the three excess capacities (facility, employment, and liability) by the mid 2000’s (Cabinet Office, 2006, p. 133). Nevertheless, in the Japanese services business, which is pointed out for its low productivity on an international basis, their excess capacity of facilities and employments are concerned. The excess capacity of tangible fixed assets and labor may have been negatively evaluated by the market for the non-DB enterprises. Regarding the intangible amount of the DB enterprises, it is assumed to be compensated with the DB asset value, although the tangible fixed assets and labor were negatively evaluated. As a result, the DB enterprises’ DB asset is calculated with Formula (3b) by excluding $RC$ from $I$ and adding the effects of $K$ and $L$ in order to compensate their negative effects. The statistical summary of $RC$ of whole samples and other variables that are mentioned in the next subsection is shown in Table 3.

$$DB_{i,t} = I_{i,t} - RC_{i,t} + 1.244K_{i,t} + 2.901L_{i,t}$$

(3b)

### Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RC$</td>
<td>384</td>
<td>3,387</td>
<td>10,563</td>
<td>0</td>
<td>79,951</td>
</tr>
<tr>
<td>$DB$ asset</td>
<td>384</td>
<td>12,431</td>
<td>29,417</td>
<td>-45,056</td>
<td>262,461</td>
</tr>
<tr>
<td>$K(IT)$</td>
<td>384</td>
<td>497</td>
<td>2,443</td>
<td>0</td>
<td>39,552</td>
</tr>
<tr>
<td>$L(IT)$</td>
<td>384</td>
<td>111</td>
<td>160</td>
<td>0</td>
<td>1294</td>
</tr>
</tbody>
</table>

### IV-3. Relationship between DB asset and IT investment

Subsequently, the relationship between the above calculated DB asset and IT investment/IT labor (IT stock, IT-related labor) is identified. IT stock is calculated by multiplying the industry-classified IT stock ratio of the JIP Database 2014 version provided by Research Institute of Economy, Trade and Industry, by the total assets of applicable enterprises. The IT investment stock ratio within the capital stock is calculated based on the real IT capital stock and the real net sectoral capital stock (JPY 1 million, price as of 2000) of the JIP Database. As the industrial categorization differs, “91 Information Services Business (Internet-related services)” in the JIP Database is evenly applied. Since data of 2012 and after is unavailable, the ratio of 2011 is applied.

IT labor expense is calculated by multiplying the total labor cost of applicable enterprises by the ratio of internal and external employees shares of information system personnel
within the entire employees of information service sectors based on the result of 2013 Actual Conditions Survey of Information Technology. Due to the limitation of categorization, the ratio of “Information Services” is applied here too.

Panel data analysis is performed based on DB enterprises whose major intangibles might be a database. The random-effects model is selected as a result of the Hausman Test\(^4\). The result of panel data regression analysis is shown in Table 4. Due to the obtained coefficient, the relationship between DB asset \((DB, K\ (IT), \text{and } L\ (IT))\) are as shown in Formula (4).

\[
DB_{i,t} = 5646.157 + 2.623 K\ (IT)_{i,t} + 44.623 L\ (IT)_{i,t}
\] (4)

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result of panel data analysis of DB enterprise</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent valuable : (DB\ asset)</th>
<th>coefficient</th>
<th>Standard error</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K(\text{IT}))</td>
<td>2.623***</td>
<td>0.434</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(L(\text{IT}))</td>
<td>44.623***</td>
<td>9.426</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>5646.157*</td>
<td>3363.846</td>
<td>0.0941</td>
</tr>
</tbody>
</table>

Note: \(N = 384\) (unbalanced)  
Adj. \(R^2 = .135\)  
Hausman test \(\chi^2 = .129, \text{ Prob}=0.9374\)

\(\text{IV-4. Estimation of DB asset in private and public sectors}\)

In this section, the authors estimate the Japanese macro level DB asset in the private and public sectors separately. Specifically, estimations are conducted by applying the relationship of \(DB, K\ (IT), \text{and } L\ (IT)\), which was identified with the enterprise level data in the preceding section. Also, in this case, the constant term shown in Formula (4) of the preceding section is not particularly reflected, and only the DB assets which are produced by IT stock and IT labor are subject to the estimation. The estimation in this section is as of 2011 in nominal terms.

Firstly, an estimation of the private sector’s DB asset is explained. It is supposed that a certain ratio of the data processing and information services’ sector (Japan Standard Industry Classification 392), that the majority of aforementioned enterprises for the analysis be-

\(^4\) While the analysis of the Non-DB enterprise’s relationship between intangibles and advertisement expense shown in Table 2 is conducted in a fixed effect model, the analysis shown in Table 4 is conducted in a random-effects model. One possible factor that causes such difference between them may be that because the latter dependent variable subtracts RC, which is affected by AD, from the former dependent variable. Residuals might not be fixed because AD easily fluctuates due to the influence of enterprises’ performances.
long to, exploits DB assets.

The private sector’s $K_{\text{IT}}$ is calculated based on the JIP data’s IT capital stock (information services business) value (JPY 3,277,883 trillion, in nominal term). IT capital stock of data processing and information services’ sector is estimated by proportional division based on the gross sales of the domestic production amount of the Current Survey of Selected Service Industries: Information Service Business (software business, data processing and information services) and domestic output of the Input Output Table (108 sections) in 2011. Subsequently, the estimated IT capital stock of the data processing and information services’ sector is multiplied by the ratio of DB enterprises’ total assets against all sample enterprises (8.6%). The calculated value, JPY 50,085 billion, is set as the entire Japanese private IT capital stock, $K_{\text{IT}}$.

Secondly, for the private sector’s $L_{\text{IT}}$, the authors calculate compensation of employees of data processing and information services’ sector based on the value of compensation of employees: JPY 6,331,156.5 billion as obtained from the 2011 Input Output Table (information system) and data of the Current Survey of Selected Service Industries (software business, data processing and information services). In addition, after multiplying the ratio of information system personnel (16.6%) based on the Actual Conditions Survey of Information Technology, the IT labor value ($L_{\text{IT}}$) of the entire Japanese DB enterprises is determined as the value JPY 58,295 billion, by multiplying the DB enterprises’ labor expense ratio (22.4%) among the 2011 data of the enterprises of the analysis. By substituting the value of $K_{\text{IT}}$ and $L_{\text{IT}}$ calculated as above to Formula (4b), an estimated value of the private sector’s DB asset is calculated as follows\(^5\).

\[
DB = 2.623K_{\text{IT}} + 44.623L_{\text{IT}}
\]

Estimated value of the private sector’s DB asset: JPY 2.7327 trillion

The public sector’s DB Asset estimation is explained subsequently. For the estimation of public sectors, IT capital stock of the JIP data and compensation of employees on the Input Output Table are used.

Public sector’s $K_{\text{IT}}$ calculation is based on the value of JIP data’s IT capital stock (of other governments, JPY 3,693,213 trillion in nominal term). We should note that IT capital stock needs to be limited because IT capital stock includes mission-critical tasks of an internal administration (i.e. basic resident register system of a local government), and not all of which is used for a database. Therefore, the authors calculate database-related IT capital stock by focusing on the value of per contracted partner (public service) of the Current Survey of Selected Service Industries: Information Service Business. Thus, the IT capital stock amount ($K_{\text{IT}}$) used for estimating the public sector’s DB asset is calculated to be JPY

\(^{\text{An interval for this estimation cannot be estimated since coefficients of results based on enterprises (excluding constant terms) are applied to the macro economy. Also, with a 95% confidence interval, the coefficient between } K_{\text{IT}} \text{ and } L_{\text{IT}} \text{ shown in Table 4 can be shown as follows: } K_{\text{IT}} \text{ lower limit } = 1.770; K_{\text{IT}} \text{ upper limit } = 3.447; L_{\text{IT}} \text{ lower limit } = 26.090; L_{\text{IT}} \text{ upper limit } = 63.156.\)
1,205.965 billion.

For public sector’s L (IT), the value of compensation of employees (JPY 14.501379 billion) obtained from the 2011 Input Output Table (public service) is set as a starting point. The public sectors in this analysis are limited to general sectors. Also, based on the materials of the National Personnel Authority and the Ministry of Internal Affairs and Communications, the general sector’s ratio out of entire government officials and local government employees is calculated to be 0.2402. In addition, by multiplying the ratio of the IT labor ratio\(^6\) (0.0112) and the corresponding ratio (0.3265) of the data processing and information services’ sector in the preceding paragraph, IT labor (L (IT)) for public sector’s DB asset estimation is calculated as JPY 12.737 billion.

In the same manner as the private sector, by substituting the value of K (IT) and L (IT) calculated above to Formula (4b), an estimated value of the public sector’s DB asset is calculated as follows.

Estimated value of the public sector’s DB asset: JPY 3.7314 trillion

\(^6\) The ratio is calculated based on the IT personnel data of IPA (2014).
V. Estimating the Economic Impact of Open Data

V-1. Incorporation of database assets in production function

In this section, it is assumed that a certain portion of the DB asset in the public sector (3,731 billion yen) which is estimated in the previous section is incorporated in the DB asset in the private sector (2,733 billion yen), and its impact on the whole economy is estimated by DSGE (Dynamic Stochastic General Equilibrium) analysis. Because the amount of the public DB asset that can be opened and utilized in the private sector depends on various factors such as the utility of the data, privacy protection, and costs regarding data provision, four variations of analysis are set in terms of the percentage of the public DB asset which is transferred to the private DB asset: 10%, 25%, 33%, and 50%\(^7\). Additionally, the transfer and incorporation of the data cannot be completed instantaneously, therefore it is assumed that the transfer of data of the above extent takes 10 years in each variation. The increase of the private DB asset is shown according to the four variations in the Table 5.

Table 5

<table>
<thead>
<tr>
<th>Percentage of public DB asset which is transferred to private sector</th>
<th>Increase of private DB asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>13.655%</td>
</tr>
<tr>
<td>25%</td>
<td>34.137%</td>
</tr>
<tr>
<td>33%</td>
<td>45.060%</td>
</tr>
<tr>
<td>50%</td>
<td>68.273%</td>
</tr>
</tbody>
</table>

In terms of incorporating the DB asset into the macroeconomic model of DSGE analysis, the present study refers to the production function in “Relation analysis between the amount of data flow and economic growth” of MIC (2014, p. 113). (eq. 5)

---

\(^7\) There is no reliable evidence on which areas of public DB asset are used as open data, but according to the Open Data Institute (2015) which has surveyed the use of public data in UK, the data which is used by more than 25% firms of respondents covers various sectors such as geography/maps, transportation, environment, population/society, education, health/medical, economy, energy, climate, and residents. It is inferred that open data can be used in various sectors, therefore, the present study adopts a range from 10% to 50% as the use rate of the public DB asset in the private sector.
In MIC (2014), $K$ is the stock of information asset, Data is the amount of data flow. This model assumes constant return in terms of capital and labor, and increasing return in terms of the stock of information asset, because of the network externality. Taking into account the concept of this model, the present study uses the production function of the eq. 6.

$$Y = AK^{\alpha}_{all}L^{\beta}(K,Data)^{\gamma}, \alpha + \beta = 1$$

$Y$ is GDP, $A$ is technological standard, $K_{all}$ is capital, $L$ is labor, $DB$ is database asset. Parameter $\gamma$ defines the scale of impact of DB asset on economy. Based on the results of estimation in MIC (2014), three variations are adopted as the value of $\gamma$ as shown in Table 6. There are four variations on the transfer from public to private DB asset, and three variations on $\gamma$, therefore twelve variations in total are being estimated.

### Table 6
Calibration of parameter $\gamma$

<table>
<thead>
<tr>
<th>Value of $\gamma$</th>
<th>Description based on MIC (2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>Coefficient of GPS data, and minimum significant value</td>
</tr>
<tr>
<td>0.03</td>
<td>Coefficient of weather data, one of the representative fields of public data re-use</td>
</tr>
<tr>
<td>0.06</td>
<td>Coefficient of customer data, accounting data, e-sales log and maximum value among all fields</td>
</tr>
</tbody>
</table>

### V-2. Base model

Based on the production function which is introduced in the previous section, this section constructs the models which incorporate the effect of the re-use of open data. The base model follows Griffoli (2011) which is a real business cycle model with monopolistic competition. First, representative household follows the utility function in eq. 7.

$$E^t \beta(\delta)y^t + E^{t+1} \beta(\delta)y^{t+1} = \frac{1}{1+E} [c^t + E(1-E)\beta(1-\delta)] + w^t l^t + r^t i^t + Z^t \psi^t l^t + \frac{1}{2} \delta k^t i^t$$

$E$ is expectation, $\beta$ is discount rate, $\psi$ is weight for labor, $\delta$ is depreciation rate of capital, $c$ is consumption, $l$ is labor, $k$ is capital, $w$ is wage, $r$ is interest rate, $y$ is output, $i$ is investment flow, $\psi$ is technological level.

---

\* Considering that most of the re-use of public data is centered in the field of geography, maps, and weather, it is possible to use 0.02 or 0.03 as the value of $\gamma$. However, for the firms which use public data for their core revenue, it is possible that the effect of the reuse of data can be boosted to the level of customer data. Therefore, 0.06 is also used as the value of $\gamma$.

\* $E$ is expectation, $\beta$ is discount rate, $\psi$ is weight for labor, $\delta$ is depreciation rate of capital, $c$ is consumption, $l$ is labor, $k$ is capital, $w$ is wage, $r$ is interest rate, $y$ is output, $i$ is investment flow, $\psi$ is technological level.
\[ E_t = \sum_{t=0}^{\infty} \beta \left[ \log C_t + \psi \log(1 - l_t) \right] \]  

(7)

\( C \) is consumption, \( l \) is labor, therefore \( 1 - l \) represents leisure. Households maximize eq. 7 under the following budget constraint.

\[ c_t + k_{t+1} = w_t l_t + r_t k_t + (1 - \delta) k_t \]  

(8)

\( k \) is capital, \( w \) is wage, \( r \) stands for interest rate, \( \delta \) is depletion rate of capital. \( w_t l_t + r_t k_t = y_t \), where \( y \) is the output. Additionally, investment flow follows eq. 9 and 10.

\[ i_t = k_{t+1} - (1 - \delta) k_t \]  

(9)

\[ i_t = y_t - c_t \]  

(10)

From the first order condition under the budget constraint (eq.8), the following Euler equation is obtained.

\[ \frac{1}{c_t} = \beta E_t \left[ \frac{1}{c_{t+1}} (1 + r_{t+1} - \delta) \right] \]  

(11)

First order condition on \( w \) is as follows.

\[ \psi \frac{c_t}{1 - l_t} = w_t \]  

(12)

In terms of the firm section, each firm \( i \) produces output following the Cobb-Douglas function with Harrod-Neutral technological progress:

\[ y_{it} = k_{it}^\alpha (e^{Z_r l_{it}})^{1-\alpha} (DB)^{\gamma} \]  

(13)

where \( z \) defines the level of technology. The profit of firm is described as follows:

\[ k_{it}^\alpha (e^{Z_r l_{it}})^{1-\alpha} (DB)^{\gamma} - w_t l_t - r_t k_t \]  

(14)

First order condition for \( k \) and \( l \):

\[ k: \alpha k_{it}^{\alpha-1} (e^{Z_r l_{it}})^{1-\alpha} (DB)^{\gamma} = r_i \]  

(15)

\[ l: k_{it}^{\alpha} (1 - \alpha)(e^{Z_r l_{it}})^{-\alpha} (DB)^{\gamma} = w_i \]  

(16)

Dividing (15) by (16) yields the optimal capital to labor ratio:

\[ k_{it} r_i = \frac{\alpha}{1 - \alpha} w_i l_{it} \]  

(17)

Under monopolistic competition, price is determined by:

\[ p_{it} = \frac{e}{e - 1} mc_i p_t \]  

(18)
where \( p_i \) is the firm-specific price, \( mc_i \) is marginal cost, and \( \varepsilon \) is the elasticity of substitution.

For simplification, individual firms take market price \( p_i \), therefore \( mc_i = \frac{\varepsilon - 1}{\varepsilon} \). Combining the marginal cost and production function, the following conditions are obtained:

\[
\begin{align*}
\omega_i &= (1 - \alpha) \frac{1}{l_{it}} \frac{(\varepsilon - 1)}{\varepsilon} (DB)^{\gamma} \\
\alpha_i &= \frac{1}{k_{it}} \frac{(\varepsilon - 1)}{\varepsilon} (DB)^{\gamma}
\end{align*}
\]

\( V-3 \). \textit{Incorporating the impact of open data policy}

As discussed previously, this study assumes four patterns as the ratios of the public DB assets which are utilized by private sector: 10\%, 25\%, 33\%, and 50\%. Additionally, it is supposed to take 10 years to reach these ratios. Accordingly, the DB asset in private sector increases depending on the eq.21, on and after which “DB” is expressed as “\( kdb \)”.

\[
kdb_t = \omega_2 kdb_{t-1} + kdb_{\text{add}}
\]

where \( kdb_{\text{add}} \) is the increase in each year, and \( \omega_2 \) is set as 0.999. \( kdb_{\text{add}} \) depends on eq. 22,

\[
kdb_{\text{add}, t} = \omega kdb_{\text{add}, t-1} + ekdb_{\text{add}}
\]

where \( ekdb_{\text{add}} \) is the temporary shock to determine the ultimate volume of the increase of the DB asset. Speed of the increase follows AR(1): Auto Regressive One and \( \omega \) is set as 0.9. Because \( kdb \) is defined as the asset which does not duplicate with \( k \), \( kdb + 1 \) is incorporated in the equations in order to avoid affecting the output when \( kdb \) is zero. Therefore, if \( kdb \) is in its initial value of 0.00566, and \( \gamma = 0.02 \), \( (DB)^{\gamma} \) becomes 1.000113.

In order to control the comparative scale of values in the DSGE analysis incorporating the abovementioned equations, initial values of the variables are set as in Table 7. Initial values are set based on the 2013 national accounts as comparative values when \( y \) (GDP) is 1. However, it should be noted that the actual steady state is calculated with the balance between endogenous variables and parameters, therefore it could diverge from initial values to some extent. Steady state values are shown in Table 8. Based on the above initial values, temporary shocks, which is \( ekdb_{\text{add}} \) and the corresponding four patterns on the ratio of transferring DB asset is as shown in Table 9.

As this study focuses on the Japanese economy, calibration of parameters is based on the estimations on Japan as much as possible if available, and for the parameters without Japanese estimation, calibration is taken from Griffoli (2011). As seen in Table 10, \( \alpha, \beta, \delta \) are based on the estimates on Japan by Sugo and Ueda (2008).
Table 7
Initial values of endogenous variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Initial value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>1</td>
<td>Gross Domestic Products (Expense)</td>
</tr>
<tr>
<td>( c )</td>
<td>0.8</td>
<td>Final expense of private and public sector</td>
</tr>
<tr>
<td>( k )</td>
<td>6.3</td>
<td>Net asset</td>
</tr>
<tr>
<td>( i )</td>
<td>0.2</td>
<td>Total fixed asset formation</td>
</tr>
<tr>
<td>( l )</td>
<td>0.3</td>
<td>7.2 hours in a day</td>
</tr>
<tr>
<td>( w )</td>
<td>1.71</td>
<td>Calculated from ( w )( \times l )=income of labor</td>
</tr>
<tr>
<td>( r )</td>
<td>0.008</td>
<td>Bank of Japan, average rate for domestic banks, February 2015</td>
</tr>
<tr>
<td>( z )</td>
<td>0.67</td>
<td>Set from labor productivity which is calculated from labor income</td>
</tr>
<tr>
<td>( kdb_add )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( ekdb_add )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( kdb )</td>
<td>0.00566</td>
<td>Ratio of private database asset to ( y )</td>
</tr>
</tbody>
</table>

Table 8
Steady state values

<table>
<thead>
<tr>
<th>Variables</th>
<th>Parameter ( \gamma )</th>
<th>0.02</th>
<th>0.03</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td></td>
<td>0.83196</td>
<td>0.832108</td>
<td>0.832553</td>
</tr>
<tr>
<td>( c )</td>
<td></td>
<td>0.576298</td>
<td>0.576387</td>
<td>0.576652</td>
</tr>
<tr>
<td>( k )</td>
<td></td>
<td>4.26103</td>
<td>4.26203</td>
<td>4.26503</td>
</tr>
<tr>
<td>( i )</td>
<td></td>
<td>0.255662</td>
<td>0.255722</td>
<td>0.255902</td>
</tr>
<tr>
<td>( l )</td>
<td></td>
<td>0.318703</td>
<td>0.31872</td>
<td>0.318773</td>
</tr>
<tr>
<td>( w )</td>
<td></td>
<td>1.4803</td>
<td>1.48056</td>
<td>1.48136</td>
</tr>
<tr>
<td>( r )</td>
<td></td>
<td>0.065025</td>
<td>0.065025</td>
<td>0.065025</td>
</tr>
<tr>
<td>( z )</td>
<td></td>
<td>-1.11E-15</td>
<td>-1.11E-15</td>
<td>-1.11E-15</td>
</tr>
<tr>
<td>( kdb_add )</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( kdb )</td>
<td></td>
<td>0.00566</td>
<td>0.00566</td>
<td>0.00566</td>
</tr>
</tbody>
</table>
Results

Under the setting in previous sections, impulse response analysis is conducted. Analysis is conducted on 12 patterns depending on parameter $\gamma$ which defines the impact of the DB asset on GDP, and the ratio of transferring DB asset from the public to the private sector. In all 12 patterns, the increase of GDP at the time 10 years later is calculated by specifying the difference of $y$ from the steady state, and this difference and the real GDP of Japan in 2014 (527.6 trillion yen). As seen in Table 11, it ranges from a minimum of 158.6 billion yen to a maximum of 701 billion yen at the point 10 years later.

Dynamic paths of major endogenous variables in the case of minimum and maximum impact are shown in Figure 2 and 3. The vertical axis is the difference from the steady state, and the horizontal axis is year. In detail, in Figure 2, output (GDP: $y$) increases following eq. 13, with the increase of the DB asset ($k_{db}$). On the other hand, as seen in $i = y_t - c_t$ in eq. 10, consumption ($c$) and investment ($i$) also increase when $y$ increases. Capital ($k$) also increases by the increase of investment ($i$). Wage ($w$) and interest rate ($r$) also increase following eq.

\[ y_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ c_t = y_t - i_t \]

\[ i_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ w_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ r_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ e_{k_{db} \text{ add}} \]

<table>
<thead>
<tr>
<th>Ratio of transferring database asset</th>
<th>$ek_{db _add}$</th>
<th>Value of $k_{db}$ 10 years later</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.000993</td>
<td>0.00643</td>
</tr>
<tr>
<td>25%</td>
<td>0.001172</td>
<td>0.00759</td>
</tr>
<tr>
<td>33%</td>
<td>0.001267</td>
<td>0.00821</td>
</tr>
<tr>
<td>50%</td>
<td>0.00147</td>
<td>0.00952</td>
</tr>
</tbody>
</table>

\[ \text{Table 9} \]

\[ \text{Values of } ek_{db \_add} \]

\[ \text{Table 10} \]

\[ \text{Calibration of parameters} \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\delta$</th>
<th>$\psi$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.37</td>
<td>0.995</td>
<td>0.06</td>
<td>1.75</td>
<td>10</td>
</tr>
<tr>
<td>Reference</td>
<td>Sugo and</td>
<td>Sugo and</td>
<td>Sugo and</td>
<td>Griffoli</td>
<td>Griffoli</td>
</tr>
</tbody>
</table>

V-4. Results

Under the setting in previous sections, impulse response analysis is conducted. Analysis is conducted on 12 patterns depending on parameter $\gamma$ which defines the impact of the DB asset on GDP, and the ratio of transferring DB asset from the public to the private sector. In all 12 patterns, the increase of GDP at the time 10 years later is calculated by specifying the difference of $y$ from the steady state, and this difference and the real GDP of Japan in 2014 (527.6 trillion yen). As seen in Table 11, it ranges from a minimum of 158.6 billion yen to a maximum of 701 billion yen at the point 10 years later.

Dynamic paths of major endogenous variables in the case of minimum and maximum impact are shown in Figure 2 and 3. The vertical axis is the difference from the steady state, and the horizontal axis is year. In detail, in Figure 2, output (GDP: $y$) increases following eq. 13, with the increase of the DB asset ($k_{db}$). On the other hand, as seen in $i = y_t - c_t$ in eq. 10, consumption ($c$) and investment ($i$) also increase when $y$ increases. Capital ($k$) also increases by the increase of investment ($i$). Wage ($w$) and interest rate ($r$) also increase following eq.

\[ y_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ c_t = y_t - i_t \]

\[ i_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ w_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

\[ r_t = \frac{1}{\delta} \left( \alpha \cdot k_{db} + \beta \cdot c_t + \delta \cdot k_{db} + \psi \cdot i_t + \epsilon \right) \]

10 The analysis is conducted with Matlab R2014b and Dynare4.2.0.
Table 11
Impact on GDP by open data policy (ratio, 100 million yen)

<table>
<thead>
<tr>
<th>Transfer rate</th>
<th>Value of $\gamma$</th>
<th>Ratio</th>
<th>Amount (100 mill.)</th>
<th>Ratio</th>
<th>Amount (100 mill.)</th>
<th>Ratio</th>
<th>Amount (100 mill.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.030%</td>
<td>1,586</td>
<td>0.045%</td>
<td>2,380</td>
<td>0.090%</td>
<td>4,769</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.035%</td>
<td>1,866</td>
<td>0.053%</td>
<td>2,801</td>
<td>0.106%</td>
<td>5,614</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.038%</td>
<td>2,014</td>
<td>0.057%</td>
<td>3,023</td>
<td>0.115%</td>
<td>6,060</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.044%</td>
<td>2,328</td>
<td>0.066%</td>
<td>3,495</td>
<td>0.133%</td>
<td>7,010</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2
Dynamic paths of variables ($\gamma = 0.02$, transfer rate 10%)
Comparing the wage and interest rate, whereas the wage rises for a longer term under the influence of consumption and labor supply (eq. 12), the interest rate starts to fall at 10 years towards steady state, lacking such additional factors. As labor supply \( l \) is determined by the balance of eq. 12 and 19 which defines the relations between consumption, wage, and labor supply, \( l \) continues to grow until the middle of the term before gradually shifting. The dynamic path of \( z \) is not displayed because there is no variance on \( z \).

Comparing Figure 2 on minimum impact and Figure 3 on maximum impact, there is not much difference on the paths of variables between these two results. The dynamics of variables is larger in Figure 3 corresponding to the difference of the increase of the DB asset \( kdb \).

VI. Conclusion

The present study considers public data (data owned by central and local governments) new infrastructure, and discussed how the release of those data affects the Japanese economy at the macro-level. For this purpose, this study (1) estimated the value of the stock of data considering it to be a DB asset, and (2) conducted simulation analysis on the impact of
the release of public data on the whole Japanese economy. In the course of the estimation, this study proposed the method based on the publicly available information so that following studies can apply a similar approach.

In terms of (1) estimation of the DB asset, this paper proposed the method to estimate the DB asset based on the publicly available information such as financial reports, stock prices, and input-output tables. Using this method, the specific results of the estimation concerning the DB asset have been presented, dividing the private and public sector.

On the DB asset in private sector, the estimated value is 2.732 trillion JPY, based on the data from 2001 to 2013. Compared with the data which is simply applied from previous studies based on the ratio of GDP (Table 1), the estimated value of this study is not far from the market size of open data in previous studies. However, it is several times larger than the estimation of Hitachi consulting (2012), 510 billion JPY, which summed up the results of a survey in the Japanese market. This difference might be because the present study conducted the estimation on the all data processing and information services, whereas the estimate of Hitachi consulting (2012) covers 9 sub-sectors.

On the other hand, the DB asset in the public sector is estimated to be 3.731 trillion JPY. This value is a little smaller than the median of a prior study (4.02 trillion JPY) which is applied from Pira International (2000) on the economic value of public data in the EU. However, it is between the maximum and minimum of this prior study.

Comparing the results of the present study to prior studies, the DB asset in the private sector is a little larger, but they do not diverge too far from the results which are obtained by applying prior studies. In this regard, the results have a certain consistency with the results of prior studies.

In terms of (2) simulation analysis on the impact on the Japanese economy, the impact of open data policy in which a part of the DB asset in the public sector is released and incorporated in the private sector, is estimated by DSGE model. The results suggest GDP is boosted by a range from a minimum of 158.6 billion JPY to a maximum of 701 billion JPY, depending on the calibration of parameters. These values suggest the difference of GDP in a single year, which is ten years later from the latest year, therefore, the impact could be much larger if the impacts in each year are aggregated through the period.

Compared to the results which simply transpose the prior studies on economic impact of open data to the Japanese economy based on the GDP ratio as shown in Table 1, the results of the prior study exceeds 3 trillion JPY even as a minimum effect, therefore larger than the result of the present study, 701 billion JPY. This disparity might be due to the difference of the scope of analysis because the present study limits the increase of value added which is produced by IT capital and IT labor, whereas the prior study includes the entire market size.

Taking the abovementioned discussions into consideration, this study provides the following two academic contributions. First, this study presents the methodology to estimate the DB asset as infrastructure, so that third-party scholars can adopt a similar method with publicly available information. This enables comparative analysis across countries and periods. Second, this study applied this newly developed methodology to data on the Japanese
economy and obtained the estimation which has no wide discrepancy with prior studies. This suggests the validity of the methodology to some extent, and enables the quantitative estimation on how much the provision of public data affects the Japanese economy.

On the other hand, there are several limitations in the estimation of the present study. The first half of the analysis, which is the estimation of the DB asset, focuses on data processing and information services. However, it is widely supposed that new businesses could be generated by using public data in a wide range of business sectors such as transportation and health care. Therefore, the scope of the present study can be expanded to include more sectors. Additionally, it is noted that the estimation of the DB asset in the public sector is conducted with IT related capital: K (IT) and IT related labor: L (IT). Estimation concerning the public sector can not necessarily utilize these elements as efficiently as in private sector.

In addition, it should also be noted that the elements which affect the DB asset are limited to two factors such as IT related capital (K (IT)) and IT related labor (L (IT)). There is also a limitation on applying the relation between these elements and the DB asset in the private sector to the macroeconomic estimation. There are possibly a variety of factors which affect DB asset formation, therefore, future exploration is desirable concerning a more suitable method for macro-level estimation such as focusing on total factor productivity, for instance. This paper presents a certain method for estimating the DB asset, but further improvement is possible.

In terms of the second half of the paper, the simulation on the effect on the economy on the macro-level, the extent of the positive effect on the GDP widely depends on the value of parameter γ. The question of what type of data is the most valuable has been discussed as one of the most important topics to promote open data policy, but currently there is no clear agreement on this matter. Although various events such as Ideathon and Hackathon are conducted frequently, the contribution of open data on the creation of big business is still not evident: at least in Japan. Therefore, there are still challenges to be met until the increase of the DB asset actually results in products or services which can fulfill latent demand and boosts output as expected. As shown in MIC (2014), data in the fields such as weather, accounting, customer, and sales logs are supposed to have higher economic impact. On the other hand, personal data is not generally the target of open data, and it is required to be cautious in the personal data protection.

This paper focuses on opened public data as an infrastructure that supports the future development of society and economy, and discusses the scale of the data used as the DB asset and how the provision of publicly-owned data as an infrastructure affects the Japanese economy on the macro-level. This approach has not been conducted in prior studies on the estimation on open data, thus it contributes to the related research field. However, there are several challenges in this approach as discussed in this section, therefore, there could be further progresses in this research field.
References


