

Featured Articles

Infrastructure for Multilayer Interoperability to Encourage Use of Heterogeneous Data and Information Sharing between Government Systems

Kazuki Adachi
Ryosuke Takazawa
Hiroshige Shibata
Akifumi Yato
Dan Yamamoto

OVERVIEW: In recent years, the use of heterogeneous data with various different meanings and structures has become essential for sharing information between government systems and for promoting open data, an area of growing activity. Infrastructure for Multilayer Interoperability enables the combination of different types of data, the cross-platform use of data, and seamless information sharing between government systems that use different data formats. To enable the implementation of Infrastructure for Multilayer Interoperability, Hitachi has developed techniques for acquiring the vocabulary data needed for standardizing data formats, and for designing databases that manage vocabulary data. By utilizing these techniques in tasks such as data integration within organizations, application development, and open data creation, Hitachi is contributing to organizational process improvement and the development of new services that impact society.

INTRODUCTION

THERE has been growing activity in recent years, particularly in Europe and the USA, involving moves such as making public data held by government institutions available to the private sector (open data) and enabling information sharing between government systems (data integration and conversion). Japan's "Declaration to be the World's Most Advanced IT Nation" (initial Cabinet decision on June 14, 2013 and subsequent revision on June 30, 2015)⁽¹⁾ stipulated work on the establishment of standardized "vocabulary platforms" that facilitate the combination of data and the use of data from different sources with the aims of developing new businesses and improving the performance of existing industries as well as achieving efficiency and transparency in government through progress on open data and providing highly convenient electronic government services. In accordance with this declaration, the Ministry of Economy, Trade and Industry and the Information-technology Promotion Agency, Japan have embarked on a project to develop the Infrastructure for Multilayer Interoperability (IMI)⁽²⁾ as a common vocabulary platform. Fig. 1 shows an overview of the IMI.

This article provides an overview of the IMI, describes what Hitachi is doing to implement such infrastructure, and considers the outlook for the future.

OVERVIEW OF THE IMI

Trends in Open Data and Information Sharing

Responding to the Open Data Charter agreement reached at a G8 summit, the Japanese government established the DATA.GO.JP⁽³⁾ data catalog*¹ as a web portal that enables searches to be performed across all of the open data provided by various government institutions. As of July 2015, it held more than 13,000 datasets.

In the USA, the implementation and use of the National Information Exchange Model (NIEM)⁽⁴⁾ as an information exchange platform has enabled information sharing between numerous government systems, particularly in the Department of Homeland Security, the Department of Justice, and the Department of Health and Human Services. In Japan, the social security and tax number system [identification (ID) number system] which entered use in January 2016

*1 Data catalog: A portal site that facilitates access to data with functions that include listing data, cross-agency searching of data items, and collation.

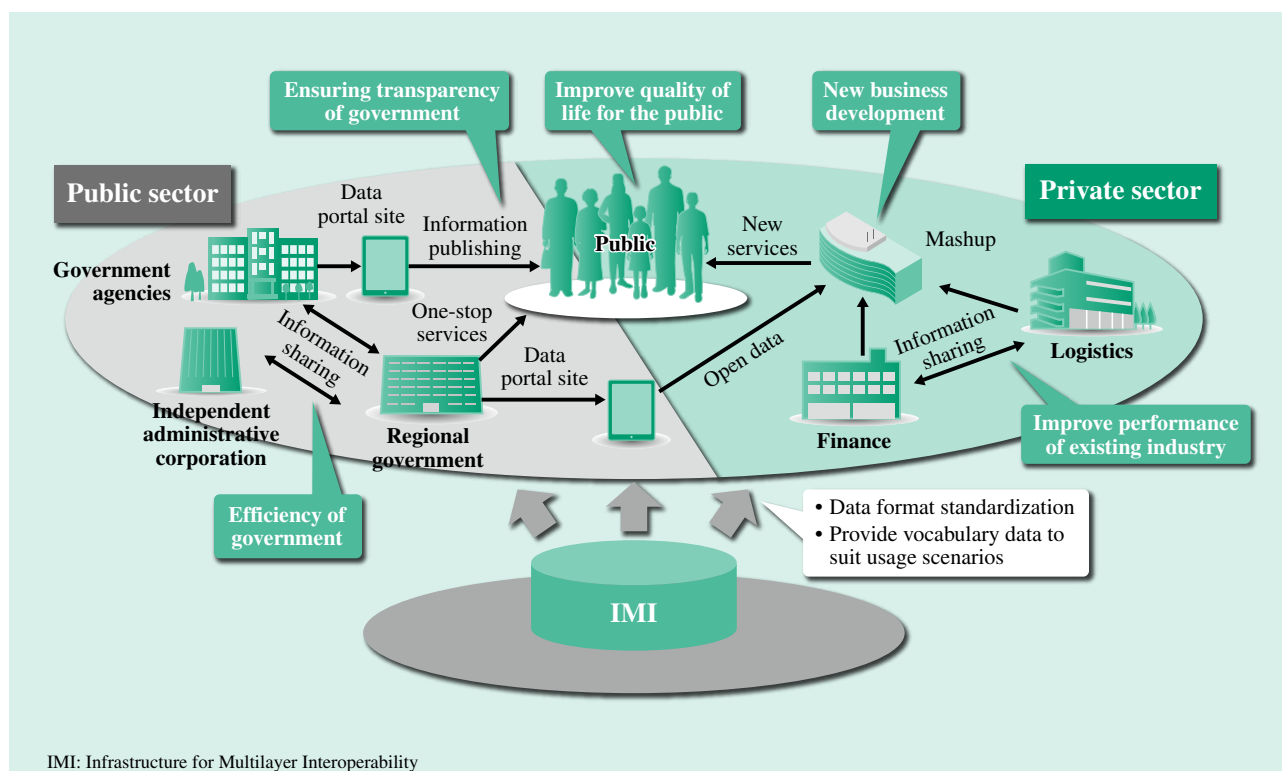


Fig. 1—Overview of the IMI.

The IMI enables the combination of different types of data, the cross-platform use of data, and seamless information sharing between government systems that use different data formats.

is expected to promote information sharing between government systems.

Difficulties of Using Data of Different Types

As government institutions that provide open data and government systems that share information become more numerous in these days, they are using data created in a wide variety of data formats (notation of terminology, data structures, etc.).

Using the automated external defibrillator (AED) data published by local governments as an example, different local governments list the names of facilities where AEDs are available under different titles, such as “facility name” or “name.”

Likewise with the addresses of facilities where AEDs are available: in some cases these are formatted as a single field containing everything from the postal code to the lot number, in others as different fields for the postal code, prefecture, municipality, and lot number.

In this way, even if data are conceptually the same, when data have different notations and structures, humans can understand that the meaning is the same, however computers interpret them as having different meanings. This complicates the integration of heterogeneous data (data of different types) published

by numerous different government institutions, and the achievement of seamless information sharing across government systems.

Using the IMI to Enable Use of Heterogeneous Data and Seamless Information Sharing across Government Systems

To overcome the difficulties described above, it is necessary to standardize data formats by concentrating on terminology and collating vocabulary data that includes the notation, structure, and meaning of each term, notation constraints, and the relationship between different terms. As a means of achieving this, Hitachi is working on establishing the IMI to provide a database for managing vocabulary data and an application programming interface (API) for using this vocabulary database.

In the case of the AED example described above, an AED search application has been developed that uses the vocabulary data provided by the IMI to standardize the format of the AED data published by each local government and enable searches to be performed for AEDs that cover the entire country⁽⁵⁾. Fig. 2 uses AEDs as an example to show how the standardization of data formats works.

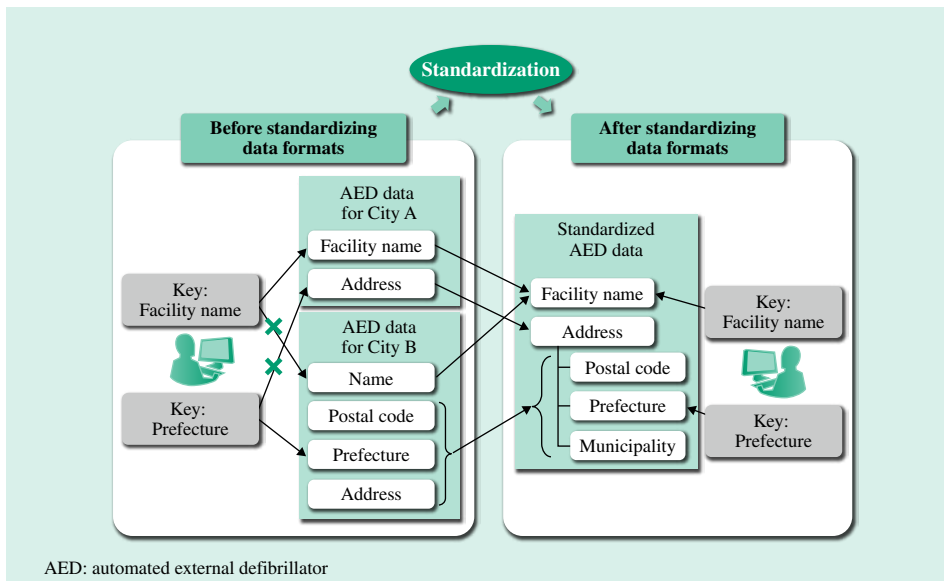


Fig. 2—Standardization of Data Formats.

Standardizing the notation and structure of data terms facilitates the use of different types of data (heterogeneous data).

Preparation for the use of heterogeneous data and seamless information sharing across government systems requires a phase in which the vocabulary data required for the standardization of data formats is collated and a phase in which data is produced in a standardized format from the data held by the data providers.

Fig. 3 shows how the use of heterogeneous data and seamless information sharing across government systems is achieved using the IMI.

WORK TOWARD IMPLEMENTING THE IMI

Challenges of Implementing the IMI

Hitachi participated in the Development of Conceptual Model of Vocabulary Database for Information Sharing and Implementation and Operation of Pilot System project of the Information-technology Promotion Agency, Japan, which involved the development of a conceptual model representing the form that the IMI should take and also identifying the

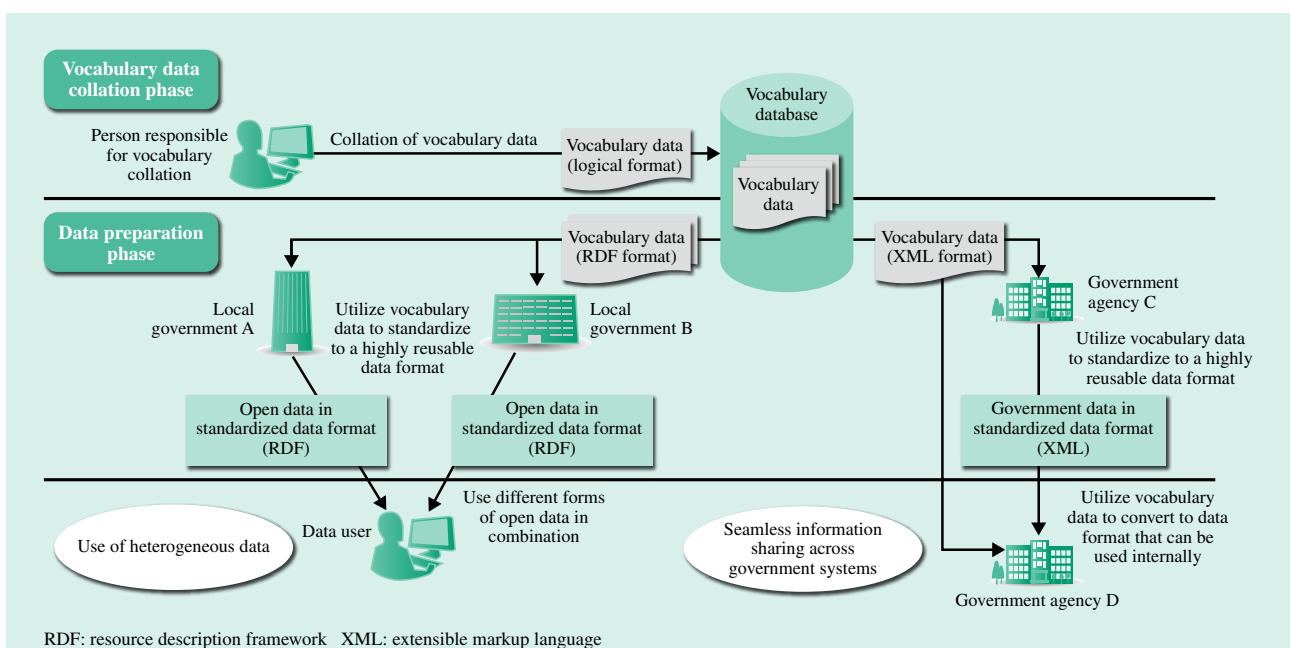


Fig. 3—How Use of Heterogeneous Data and Seamless Information Sharing across Government Systems is Achieved Using the IMI. Using the IMI for heterogeneous data and seamless information sharing across government systems facilitates the use of data by standardizing the formats of heterogeneous data.

challenges associated with its implementation⁽⁶⁾. The two major challenges are as follows.

The first is that the vocabulary data collation phase requires using usage scenarios for information publishing and sharing in a variety of fields (transportation, finance, organizations, etc.) as a basis for the collation of information such as the structure and meaning of terms, notation constraints, and the relationship between different terms. The difficulty with this is that collating information such as the scope of usage scenarios or the meaning of terms used in a particular field requires someone with expertise in that field.

The second challenge is that, in the subsequent data preparation phase, because data needs to be prepared in a variety of formats to suit different users and usage scenarios, including the use of heterogeneous data and seamless information sharing across government systems, the vocabulary data must support all of these different data formats.

Vocabulary Data Collation Technique for Publishing and Sharing of Information

To overcome the challenges posed by the vocabulary data collation phase, Hitachi has developed a

vocabulary data collation technique for publishing and sharing information.

This technique enables large quantities of high-quality vocabulary data to be collated efficiently for the purposes of information publishing and sharing, even by people who lack expertise, and ensures a certain level of quality even if adequate support from experts in the field is not available.

Hitachi participated in a project run by the Organization for Small & Medium Enterprises and Regional Innovation, Japan in which it used this technique to collate vocabulary data for information sharing between organizations over a period of approximately four months from November 2014 to February 2015.

Specifically, this identified three categories of usage scenarios (information about business systems at the Organization for Small & Medium Enterprises and Regional Innovation, Japan, application forms for small and medium-sized enterprises, and company data held by external company data sites and other sources), and collected and collated terms including corporate and joint-stock company terminology.

The following is a summary of the collated vocabulary data.

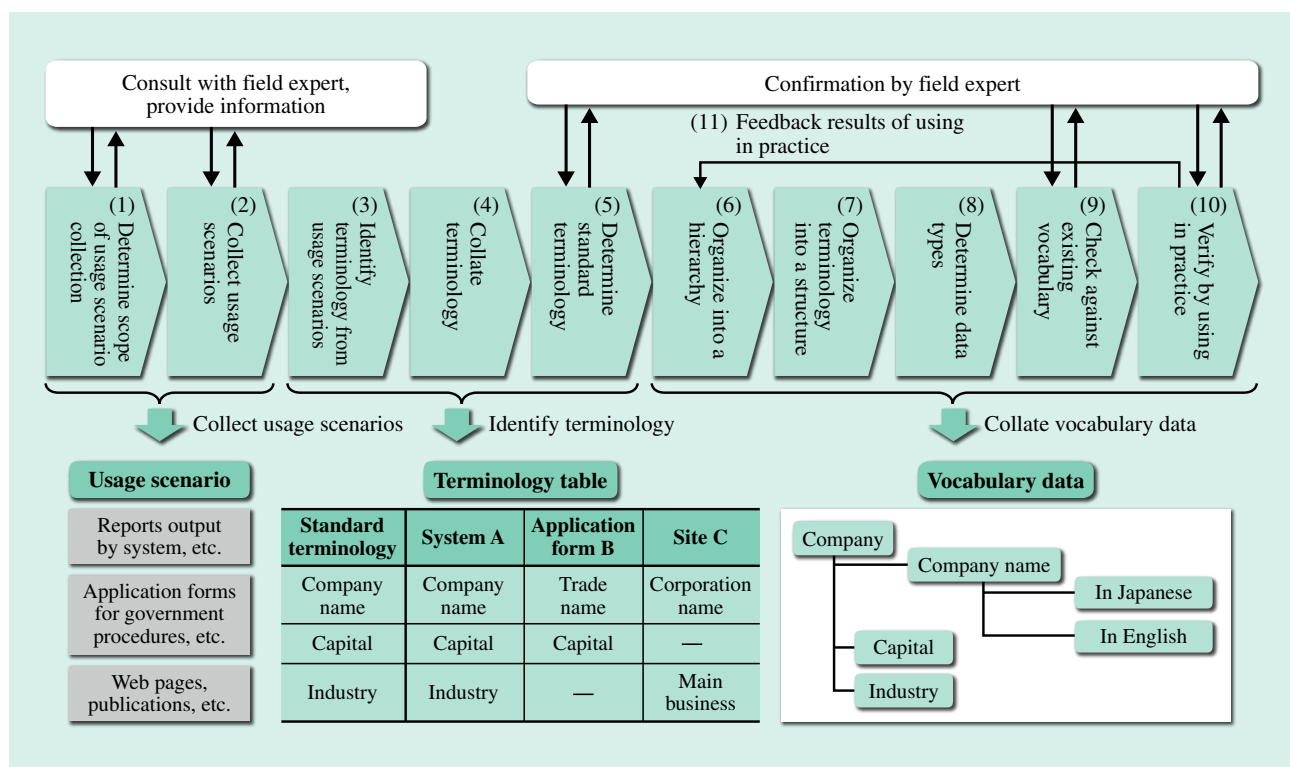


Fig. 4—Vocabulary Data Collation Process.

The steps in the process, which include collecting samples (usage scenarios) of actual data, such as web pages; identifying terminology; and collating vocabulary data, are performed with assistance from an expert in the field.

- 2,504 usage scenarios were collected.
- 61,337 terms were identified.
- 1,304 terms were included in vocabulary data.

Fig. 4 shows the process for using the vocabulary data collation technique for publishing and sharing information to collate vocabulary data for information sharing between organizations.

In collating vocabulary data, the core vocabulary⁽⁷⁾ collated by the Development of Conceptual Model of Vocabulary Database for Information Sharing and Implementation and Operation of Pilot System project of the Information-technology Promotion Agency, Japan was used to reference common terms such as organization, name, and address to conduct the collation work more efficiently and to ensure compatibility with other vocabulary data.

Vocabulary Database Design Technique

To overcome the challenges posed by the data preparation phase, Hitachi has developed a vocabulary database design technique.

This technique stores vocabulary data that, in the vocabulary data collation phase, was collated in logical formats that are independent of the data format and provides it in physical formats that are suitable for users and associated applications, such as the resource description framework (RDF)^{*2} format, which facilitates the combination of data on the web and searches across multiple data sets, and the extensible markup language (XML) format, which can easily be checked to verify that linked data is free from errors and inconsistencies.

To trial the technique, Hitachi implemented and operated a pilot vocabulary database system that stores vocabulary data collated in logical formats and provides it in physical formats via an API. This work was part of the Development of Conceptual Model of Vocabulary Database for Information Sharing and Implementation and Operation of Pilot System project of the Information-technology Promotion Agency, Japan.

The pilot system was used to store vocabulary data provisionally collated into five fields (facilities and services, roads, tourism, events, and systems), and it was operated in conjunction with prototype tools for each field to verify the vocabulary data and its suitability for providing the vocabulary data obtained via the API in a physical format.

CONCLUSIONS

This article has provided an overview of the IMI, which facilitates the use of heterogeneous data and seamless information sharing across government systems, and described what Hitachi is doing to implement such infrastructure.

Hitachi intends to encourage the use of heterogeneous data and information sharing across government systems by continuing to develop technologies for providing and using the IMI. Through the use of these technologies, Hitachi intends to contribute to achieving Social Innovation by facilitating data integration, application development, uses for big data that include open data and Internet of things (IoT) data, and other such activities that are likely to be undertaken by organizations, particularly government agencies, so that it can encourage organizational process improvement and the development of new services that impact society.

REFERENCES

- (1) “Declaration to be the World’s Most Advanced IT Nation” (Jun. 2015), <https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20150630/siryou1.pdf> in Japanese.
- (2) Information-technology Promotion Agency, Japan, “Overview of Infrastructure for Multilayer Interoperability,” http://imi.ipa.go.jp/doc/IMI_Overview_v2.pdf in Japanese.
- (3) DATA.GO.JP, <http://www.data.go.jp/> in Japanese.
- (4) National Information Exchange Model (NIEM), <https://www.niem.gov/Pages/default.aspx>
- (5) AED Open Data Platform, <http://hatsunejournal.jp/w8/AEDOpendata/> in Japanese.
- (6) Information-technology Promotion Agency, Japan, “Development of Conceptual Model of Vocabulary Database for Information Sharing and Implementation and Operation of Pilot System: Conceptual Model Document” (Feb. 2015), http://goikiban.ipa.go.jp/contents/20150203_1_1.pdf in Japanese.
- (7) Information-technology Promotion Agency, Japan, “Core Vocabulary of Infrastructure for Multilayer Interoperability,” <http://imi.ipa.go.jp/ns/core/2> in Japanese.

*2 RDF: A framework recommended by the World Wide Web Consortium (W3C) for using three elements (subject, predicate, and object), collectively called a “triple,” to represent the relationships between data.

ABOUT THE AUTHORS



Kazuki Adachi

Government & Public Systems Department 4, Government Solution Operation, Government & Public Corporation Information Systems Division, Information & Telecommunication Systems Company, Hitachi, Ltd. He is currently engaged in the design and implementation of open data systems such as the DATA.GO.JP data catalog.



Ryosuke Takazawa

Government & Public Systems Department 4, Government Solution Operation, Government & Public Corporation Information Systems Division, Information & Telecommunication Systems Company, Hitachi, Ltd. He is currently engaged in the design and implementation of public-sector systems.



Hiroshige Shibata

Government & Public Systems Department 4, Government Solution Operation, Government & Public Corporation Information Systems Division, Information & Telecommunication Systems Company, Hitachi, Ltd. He is currently engaged in the design and implementation of public-sector systems.



Akifumi Yato

Security Research Department, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd. He is currently engaged in the research and development of public-sector systems. Mr. Yato is a member of The Japanese Society for Artificial Intelligence (JSAI).



Dan Yamamoto

Security Research Department, Center for Technology Innovation – Systems Engineering, Research & Development Group, Hitachi, Ltd. He is currently engaged in the research and development of public-sector systems. Mr. Yamamoto is a member of The Institute of Electronics, Information and Communication Engineers (IEICE).