Data Utilization Platform for Understanding, Utilizing, and Simply Analyzing Various Data of Business Systems in the Railway Field

Hidenori YAMAMOTO *, Takeshi HANDA *, Yuko KATO *, Kenji KAWASAKI *, and Takashi TSUNO **

Abstract: Recently in the railway field and the industry field, it becomes more necessary to use data effectively in order to store and succeed to fields’ know-how by information technology and in order to plan for further increase in business efficiency. However, there is a large amount of various data from plural different railway business systems, so it is difficult to utilize these data transversely. In order to solve the above problems, we propose a data utilization platform, especially including platform architecture, data relation generating/visualizing functions with data relation network model, and analysis components that can be reused in plural applications, for the railway field as the first instance.

Key Words: data utilization platform, data relation network, data relation visualizing function, analysis component, analysis trial and error.

1. Introduction

Recently in the railway field and the industry field, as many experts are retiring, it becomes more necessary to use data effectively in order to store and succeed to fields’ know-how by information technology (IT) and in order to plan for further increase in business efficiency. Especially in the railway field, for example, under the smart maintenance concept, they try to change from previous time based management (TBM) to condition based management (CBM). In CBM, by analyzing real-time equipment condition data, they make rational decisions based on individual equipment conditions and future forecast values; they optimize maintenance work and formulate a medium to long term maintenance strategy [1],[2]. In addition, new services such as yield management that maximizes profits by forecasting passenger demand and selling train seats in the optimal price range and service content in ticketing are in progress in the development of new services that lead to further improvement in the profitability of railway operators.

As described above, in order to realize business improvement and new services, the need for utilizing various data held by various business systems in railway companies is expanding [3],[4]. In order to solve the above, there is a growing expectation that the use of IoT technology and big data, which has been increasing in recent years, has been particularly active [5],[6]. However, because each railway company’s business systems such as transportation, maintenance, and sales have been individually developed, it is difficult to utilize various data transversely. Furthermore, the amount of work to perform data processing and analyzing data by hand is very large, and it becomes a problem that the costs for processing and analyzing data are getting higher, and the quality of analysis result is dependent on an analyzer’s skill level and learning level to the field.

In order to solve the above problems, for the railway field as the first instance, we propose a data utilization platform that can utilize various data from plural business systems in the railway field and can support people who do not have enough knowledge about each business system or data analysis in understanding those data and easily starting analyses. As the primary users of our platform, we specifically anticipate two types of analysts: IT analysts and field analysts. IT analysts analyze various data across departments in various ways in order to discover potential problems and develop solutions across the company. They do not have enough knowledge about field data of each business system but have high IT literacy. Field analysts, on the other hand, analyze mainly their own data in order to develop solutions to problems that have emerged in their own department. They have enough knowledge about field data of each business system but do not have high IT literacy.

In order to realize the above platform, we must solve the following issues.

(1) Acquire various data from plural different business systems, manage them centrally, and integrate and provide them for various applications.

(2) Support analysts in understanding a large amount of various data and support them in selecting in data necessary for required analyses, even if without referring many specifications.

(3) Support analysts in quick trial and error of simple analyses. Facilitate the addition of new data source, addition and utilization of new data, addition and utilization of new analysis methods, and addition and utilization of new technologies.

In this paper, in Section 2, we describe previous work related to the above platform. In Section 3, we describe our proposed technique, the data utilization platform for utilizing a large amount of various data of business systems in the railway field. In Section 4, we describe examples of applying our
platform to use cases in a railway field. In Section 5, we conclude our technique described in this paper.

This paper is modified and added to our regular paper of SICE Annual Conference 2019 [7]. Especially in Section 2, we add details about problems in data utilization and the problems that cannot be solved by previous works, and in Section 4, we add use case details and evaluation results.

2. Previous Work, Problems, and Approaches

As part of data utilization, Jagadish [8] described the big data analysis pipeline: data acquisition, information extraction and retrieval, data integration and aggregation and representation, modeling and analysis, and interpretation. Figure 1 shows problems of data utilization in the railway field. Usually, as data preparation, we select data from a data lake that stores the data from business systems, cleanse it, and build a data warehouse (DWH). And we try analyses using data from DWH. Here we target the data from the railway business system, so (a) it is necessary to understand large amounts of various data for selecting data from the data lake, (b) it takes time to select data suitable for analysis and select a data processing method for building a DWH, (c) it takes time to understand trends and meanings of data and find analysis perspective before analyzing.

In relation to the above problem (a), through the above-described big data analysis pipeline, Aitoh-Okine [9] described big data challenges in the railway field: heterogeneity, inconsistency and incompleteness, merging data, and so on. As heterogeneity, most of the data collected from different databases in railway business systems are not homogeneous. As inconsistency and incompleteness, it is very clear that there are uncertainty, errors, and missing values for most of the data collected from databases in railway business systems. And as merging data, analyzing big data in the railway field is not only about large databases but also the merging of different databases to extract information for further analysis. Therefore, it is very difficult to understand various data and to select suitable data for analysis in the railway field.

In relation to the above problem (b), in recent years, self-service business intelligence (BI) has been introduced as a tool that enables end-users who do not have special IT skills to easily design and execute data analysis and reporting within the framework of one tool [10],[11]. In conventional BI, it takes time and cost to define data requirements and to become able to visualize and analyze data. And it is not possible to add or change data quickly. On the other hand, in self-service BI, data preparation, data modeling, fast aggregating, visualizing, and analyzing are performed in one tool and completed with user initiative. Self-service BI supports interactive analysis, non-standard reports, and supports trial and error with speed, flexibility, and relatively easy operability. Self-service BI is pursuing user-driven analysis with speed, flexibility, and relatively easy operability. However, when it is utilized for data analysis using various data from plural different business systems in the railway field targeted in this paper, it is required to design after the user sufficiently understands those businesses and target data for data preparation and data analysis. And in order to deal with a large amount of various data from plural different business systems, it should be necessary to support user work reduction for managing and analyzing those data.

Also, regarding data preparation, a platform technology that can apply different data cleansing rules automatically is being considered [12]. However, when it is utilized for data analysis using a large amount of various data from plural different business systems in the railway field targeted in this paper, although cleansing rules are easy to re-use, the data preparation is very difficult for beginners about analysis and those unfamiliar with data if they do not know what data to analyze and how to analyze it.

Lastly, in relation to the above problem (c), Li [13] described a multi-source information management system on the railway geological environment. In this system, to integrate data from various sources and to process various analyses, spatial analysis, 3D visualization, and so on, it has some special function modules for each purpose. In the above, these functions are sufficient if the purpose of the analysis is clear, but if that is not clear, they are not suitable for analysis trial and error.

In order to solve the above problems, in this paper, we take the following approaches. Mainly approach (1) is for the whole, approach (2) is for solving problems (a) and (b), and approach (3) is for solving problem (c).

1. Platform architecture consisting of data acquisition, data extract, transform, and load (ETL), data store, data analysis, and feedback, with common data model, common application programming interface (API) and so on.

2. Data relation generation/visualization over business systems with data relation network model.

3. Analysis component as software program component of analysis logics that can be reused in plural applications.

3. Data Utilization Platform

Following the approaches described in Section 2 in this paper, we propose a data utilization platform for the railway fields.

3.1 Concept

Figure 2 shows the concept of the data utilization platform for the railway field [14]. IT analysts or field analysts who are users of our platform analyze using various data of business systems for the purpose of solving business problems and improving business work. As a pre-stage to conduct a detailed analysis specialized in the field and work contents, by using the platform, the users, especially IT analysts, can select combinations of data to be analyzed from various viewpoints, and they can find out some relationships or tendencies among the data. After finding out any relationships or tendencies, the users, es-
especially field analysts in each field and each business, will conduct a detailed analysis (for example, planning of CBM, and so on).

Figure 2 shows that the users, especially IT analysts, can uniformly deal with data from plural different business systems in conducting analyses. They can understand relationships of data across systems and can easily select combinations of data that are likely to be useful for analysis from a large amount of data, even if it is not necessary to know the specific professional knowledge or overall understanding of a particular business system. In addition, the developer who develops an analysis logic incorporates analysis logic as a software component (analysis component) into the platform. The users can select analysis components and easily construct, change, and add analysis processed by their combinations.

From the above, the users, especially IT analysts can easily execute analysis by trial and error for various data combinations. In other words, they repeat the following procedures (a) to (c) until they find some relationships or tendencies among the data that may be useful for analysis: (a) Designate key items and browse related data. (b) Select data items to be analyzed simply from the results in (a) and set them as target data in the platform, and determine analysis contents, and select and set the analysis components. (c) Execute simple analysis with the settings in (b) and confirm the analysis results.

After finding some relationships or tendencies among the data by that trial and error, the users, especially field analysts, try a detailed analysis in the domain.

3.2 Platform Architecture

Figure 3 shows the architecture that realizes the data utilization platform for the railway field we described in Section 3.1. As shown in Fig. 3, steps of “Data Acquisition” from business systems, “Data ETL,” “Data Store,” “Data Analysis,” and “Feedback” to business systems, are executed sequentially. Table 1 shows the main functions in the above steps of our proposed platform.

Table 1 Main functions of data utilization platform.

| Step       | Function                  | Explanation                                                                 |
|------------|---------------------------|-----------------------------------------------------------------------------|
| Data       | Online acquiring          | Acquire data online from business systems                                    |
| Acquisition| Offline acquiring         | Acquire batch data offline from business systems                             |
| Data ETL   | Data cleansing            | Cleanse and process acquired data so that it can be utilized easily          |
| Data       | Data converting           | Convert acquired data to a common data model                                 |
| Analysis   | Data relation generating  | Generate data relation network over systems about a variety of data acquired from several different business systems |
| Data Store | Data store                | Store structure-converted data and analysis results to a database            |
| Data       | Analysis component        | Execute analysis logic that are composed commonly as a software component    |
| Analysis   | Common API                | API provided to applications for executing analysis component and accessing data |
| Feedback   | Data relation visualizing | Visualize generated data relation network                                    |

3.3 Data Relation Generating/Visualizing Functions

In this section, we describe a data relation generating function and a data relation visualizing function [15] as one of the functions to realize the data utilization platform for the railway field described in Section 3.2. These functions are designed to support analysts in understanding of data from plural different business systems [16],[17].

The data relation generating function extracts reference relationships among data across systems for various data collected from plural different business systems. That is, this function aims to generate something like an entity relationship (ER) diagram showing reference relationships among data, using only target data without using the data specification. As shown in Fig. 4 and Table 2, this function generates and visualizes the data relation network consisting of four layers. The data rela-
3.4 Analysis Component

In this section, we describe the analysis component as one of the functions to realize the data utilization platform for the railway field described in Section 3.2. This function is designed to facilitate the addition and utilization of new data, the addition and utilization of new analysis methods, and the addition and utilization of new technology.

The analysis component is a software program component of analysis logic that can be reused in plural applications. As shown in Fig. 5, we define two types of analysis components: a common analysis component and a domain analysis component. The common analysis component can be used in common by all applications, regardless of the category (industry type, domain, business, and so on). On the other hand, the domain analysis component is field-dependent and can be commonly used only in the domain. In addition, analysis processes that are unique to the application and cannot be standardized shall be generated individually for each application as application logic.

As shown in Fig. 6, the analysis component is executed by being called from an application via the API and performs analysis processing using data stored in a database (DB). In addition to the facilitation of flexible combinations of analysis logic and replacement, we also take into consideration that analysis for a large volume of data takes a lot of time and that the analysis result data itself can be used for further analysis. We design an analysis component according to the following concept.

A complete process (analysis logic) is executed with a certain set of data as an input, and execution result data of the process is output. This is the basic unit of the analysis component. It shall be executed asynchronously, triggered by an application call (API). As shown in Fig. 6, it is also possible to combine one or more analysis components and execute them as a series of processes. The analysis results are stored in a DB and can be referenced and reused from their own and other applications.

| Layer                  | Relation               | Explanation                                                                 |
|------------------------|------------------------|-----------------------------------------------------------------------------|
| System Layer           | Specification          | - Link among systems that contain tables linked by the table layer          |
|                        | Data Relation          | - Link weight is the total number of links between tables                    |
| Table Layer            | Specification          | - Link among tables that contain the same data items                        |
|                        | Data Relation          | - Link weight is the total number of identical data items                   |
| Data Item Layer        | Specification          | - Link among data items (columns) contained in the same table              |
|                        | Data Relation          | - Link among data items from different tables via common data items         |
|                        |                        | - Link weight is the total number of tables that contain linked data items  |
| Correlation Relation   |                        | - Link among joinable data items based on specification data relation       |
|                        |                        | - Link weight is the strength of the correlation between linked data items  |
| Data Value Layer       | Appearance             | - Link among data values that appear simultaneously in the records of the   |
|                        | Frequency Relation     | same table                                                                   |
|                        |                        | - Link among data values of data items in different tables via common data |
|                        |                        | items and data values                                                       |
|                        |                        | - Link weight is the frequency of appearance of linked data values          |
4. Use Case and Evaluation

In order to evaluate our proposed platform and technique, we describe examples of applying our platform, especially data relation generating/visualizing functions and the analysis component to use cases in the railway field. This time, in this paper, we evaluate especially the functional usefulness of our proposed platform, and evaluations of non-functional requirements such as performance are future work.

We have the following two use cases that correspond to the problems of data utilization in the railway field shown in Fig. 1. In each use case, we evaluate whether (data relation generating/visualizing and analysis component) functions of our proposed platform can be used. These details will be described later.

- Use case 1: Understanding data relationships from plural different business systems (corresponding to problems (a) and (b) in Fig. 1)
- Use case 2: Simple analysis (corresponding to problem (c) in Fig. 1)

4.1 Use Case 1: Understanding Data Relations

In this use case, in order to easily understand data from plural different business systems for utilizing these data, we will grasp the whole picture and find some relationships among data across different systems without referring to specifications.

The data used this time is generated by the simulator on the assumption of the data of a total of seven systems where open data is added to six railway business systems shown in Table 3. The data is prepared as a comma-separated values (CSV) format file. Table 3 shows the number (table, data item, and data value) of data created by the simulator and open data. This time, in order to focus on the effectiveness of the function rather than performance, we focus on only representative tables and data items compared to the actual systems.

4.1.1 Data relation network (system, table, data item)

The data relation network data is generated using data of these business systems as input. The data relation visualizing component receives the above-generated data relation network data, generates a view in response to a request from the web browser, and displays the view on the browser. Figure 7 shows the system layer, table layer, and data item layer of the results of visualizing data relation network generated by operating the data relation generating function using the data shown in Table 3 as input.

The nodes in each layer of the data relation network (each representing a system, a table, and a data item) are color-coded for each system, and links among nodes are drawn.

4.1.2 Data relation network (data value)

Figure 8 shows the data value layer of the results of visualizing the data relation network generated by operating the data relation generating function using the above data as input.

Here, referring to the data relationship network shown in Fig. 8, for example, (a) there is a claim (“Type: Claim”) on a certain date “Date: 2017/8/25,” and when following the links, there is a tweet “Hot!!” (“Tweet: Hot!!”), and when following another link, (b)(c)(d) in a certain train (“Train ID: 722M”), the preset temperature of air conditioning is 25°C (“Preset temp.: 25°C”) in two vehicles (“Vehicle ID: Tc01-4,” “Vehicle ID: M02-4”). Although (e) the room temperature is 25°C to 26°C (“room temp: 25°C to 26°C”) in the one vehicle (“Vehicle ID: Tc01-4”), (f) the room temperature is not only 25°C to 26°C (“room temp: 25°C to 26°C”) but also 26°C to 28°C (“room temp: 26°C to 28°C”).

Table 3 Number of data of railway business systems.

| System             | Table | Data Item | Data Value |
|--------------------|-------|-----------|------------|
| Transportation     | 2     | 24        | 3349       |
| Vehicle            | 3     | 23        | 1889       |
| Vehicle maintenance| 1     | 5         | 2409       |
| Power management   | 1     | 5         | 2366       |
| Human flow monitoring| 1   | 2         | 200        |
| SNS monitoring     | 1     | 9         | 300        |
| Open data          | 4     | 21        | 613        |

Fig. 7 Data relation network (system layer, table layer, data item layer).

Fig. 8 Data relation network (data value layer).
temp: 26°C to 28°C") in another vehicle ("Vehicle ID: M02-
4"). From the above, it is inferred that the cause of the “hot”
claim is the above second vehicle ("Vehicle ID: M02-4").

4.1.3 Results

By referring the data relation network (system layer, table
layer, and data item layer) described in Section 4.1.1, we can
confirm relationships among data across different systems at
one time through views in different layers. This helped to grasp
the whole picture of the target data across different systems.

About the data relation network (data value layer) described
in Section 4.1.2, as we described in Section 3.3, the data value
layer shows relationships among data values that appear simul-
taneously across different systems. Therefore, as we described
in Section 4.1.2, we can visualize events that appear to occur
simultaneously across different systems. This is useful for un-
derstanding the event data before a detailed analysis.

In this way, our proposed technique helps to get a quick hit
before entering a detailed analysis.

4.2 Use Case 2: Simple Analysis

In this use case, in order to narrow down the data for an-
alyzing and find some analysis perspectives, we will perform
trial and error of selecting a few pieces of data and finding their
trends, correlations, and so on.

4.2.1 Analysis component

In our proposed platform, we implement analysis compo-
nents to carry out a simple analysis. Table 4 shows examples of
analysis components. Any of these can be used in order to eas-
ily check data and prepare data in advance, for analyses. Here,
the analysis component executes a complete process for visu-
alizing the calculated results with the input of a set of data and
outputs process execution result data. They are executed asyn-
chronously by using an application call (API) as a trigger.

4.2.2 Simple analysis using analysis components

Figure 9 shows an example of an implementation for using
analysis components. As shown in Fig. 9, we provide a screen
(for example, a graphic user interface by Web) for selecting
analysis components for easy use for users. (1) Select one or
more arbitrary data items in the data item layer (showing rela-
tionships among data items included in tables of each system)
of the data relation network. (2) Select available analysis com-
ponents for the selected data items in the selection screen. (3)
Execute selected analysis components (click the selected but-
ton), and transit to the corresponding screen (in this case, a
graph output screen, or a spreadsheet tool starts). We check the
results. If the results are not satisfactory, we change data and
analysis components and try steps (1) to (3) repeatedly. After
we get satisfactory results (for example, we get clear character-
istics, tendencies, and singularities of the target data, or we find
points to start a detailed analysis), we output selected data to
the data mart file for a detailed analysis that we may try after
that.

4.2.3 Results

As we described in Section 4.2.2, by preparing a certain num-
ber of analysis components for performing standard analysis
processing described in Table 4 and making them easy to use in
combination, we could perform trial and error easily.

4.3 Evaluation

Through two use cases described in Section 4.1 and 4.2, for
data utilization using data in the railway field, especially for
data analysis, we showed functional usefulness of our proposed
platform. Especially the data relation visualization is useful to
facilitate understanding of data from plural business systems.
And the analysis component is useful to perform trial and error
of simple analyses quickly.

As described above, this time, we used data generated by the
simulator and limited the number of data items and the amount
of data values compared to the actual systems. Therefore, our
future work is using actual data of railway business systems and
secure processing performance for large amounts of data.

| Table 4 | Examples of analysis component. |
|---------|---------------------------------|
| Type    | Explanation                      | For Analysis                   |
| Line Chart | Create a line chart from selected data | Check transitions and trends of data |
| Bar Chart | Create a bar chart from selected data | Check distribution of data for each unit, group |
| Stacked Bar Chart | Create a stacked bar chart from selected data | Check distribution of data for each unit, group including items |
| Scatter Chart | Create a scatter chart from selected data | Check existence of correlation between data |
| Correlation Analysis | Calculate correlation coefficient from selected data | Check strength of relationships of data |
| Statistics | Calculate statistics from selected data | Check characteristics of data |
| Data Mart Output | Output selected data to mart file | Prepare data for other detailed analysis |

Fig. 9 Simple analysis application.
5. Conclusion

In this paper, we propose, for the railway field as the first instance, the data utilization platform that can utilize various data from plural business systems in the railway field and can support people who do not have enough knowledge about business systems or data analysis in understanding those data and easily starting analyses. The concept of our platform is selecting combinations of data to be analyzed from various viewpoints prior to conducting a detailed analysis that specializes in a field or work contexts, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily. We applied our platform and technique to two assumption use cases: understanding data relations and tendencies, and executing simple analysis quickly for starting the work contents, and finding out certain data relations and tendencies, and executing simple analysis quickly for starting the detailed analysis easily.

References

[1] A. Yokoyama: Innovation in railways: Platform for maintenance innovation using ICT, JR EAST Technical Review, No. 29, Summer 2014, pp. 1–4, 2014.
[2] I. Koshiishi: Maintenance issues for railway facilities and future prospects in maintenance, JR EAST Technical Review, No. 29, Summer 2014, pp. 5–8, 2014.
[3] A. Yokoyama: Arrival of a great turning point in maintenance: Smart maintenance initiative, JR EAST Technical Review, No. 22, Spring 2012, pp. 5–6, 2012.
[4] Y. Saito, H. Yahagi, and T. Onodera: Construction of a decision-making support system for proposing track maintenance plans, JR EAST Technical Review, No. 29, Summer 2014, pp. 5–8, 2014.
[5] J.A. Stankovic: Research directions for the internet of things, *IEEE Internet of Things Journal*, Vol. 1, No. 1, pp. 3–9, 2014.
[6] X. Xu, Q.Z. Sheng, L. Zhang, Y. Fan, and S. Dustdar: From big data to big service, *IEEE Computer*, Vol. 48, No. 7, pp. 80–83, 2015.
[7] H. Yamamoto, T. Handa, Y. Kato, K. Kawasaki, and T. Tsuno: Development of data utilization platform for utilizing a large amount of various data of business systems in the railway field, *Proc. SICE Ann. Conf.* 2019, pp. 607–612, 2019.
[8] H.V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J.M. Patel, R. Ramakrishnan, and C. Shahabi: Big data and its technical challenges, *Communications of the ACM*, Vol. 57, No. 7, pp. 86–94, 2014.
[9] N. Attob-Okinke: Big data challenges in railway engineering, *Proc. 2014 IEEE Int. Conf. Big Data*, pp. 7–9, 2014.
[10] Gartner IT Glossary: Self-service business intelligence, https://www.gartner.com/en/information-technology/glossary/self-service-business-intelligence
[11] Gartner: Market guide for self-service data preparation, https://www.gartner.com/en/documents/3906957/market-guide-for-data-preparation-tools, 2016.
[12] M. Dallachiesa, A. Ebaid, A. Eldawy, A. Elmagarmid, I.F. Ilyas, M. Ouzzani, and N. Tang: NADEEF: A commodity data cleaning system, *SIGMOD ’13 Proc. 2013 ACM SIGMOD Int. Conf. Manag. Data*, pp. 541–552, 2013.
[13] W. Li, C. Tang, and R. Huang: Multi-source information management system of railway geological environment based on GIS technology, *Proc. Int. Conf. E-Product E-Service E-

Entertain., pp. 1312–1315, 2010.
[14] T. Handa, Y. Kato, H. Yamamoto, K. Kawasaki, and T. Tsuno: Traffic information service platform for supporting cross-sectional analysis of railway operations, *Proc. 54th Symp. Congr. Jpn. Railw. Cybernet.*, p. 6, 2017 (in Japanese).
[15] T. Yamada, Y. Kato, Y. Maekawa, and T. Tomiyama: Interactive service for visualizing data association using a self-organizing structure of schemas, *Proc. 2017 IEEE 10th Conf. Service-Oriented Comput. Appl. (SOCA)*, pp. 230–233, 2017.
[16] T. Tsuno, Y. Takamura, and T. Takamura: Two products for the Hitachi Data Science Platform Big Data Discovery and NX context-based data management system, *Hitachi Review*, 2018, Vol. 67, No. 7, pp. 114–120, 2018.
[17] Hitachi, Ltd.: Hitachi Data Science Platform/Big Data Discovery, http://www.hitachi.co.jp/products/it/society/product_solution/mobility/dsp_bdd/, 2018.

Hidenori YAMAMOTO (Member)

He received his B.S., M.S., and Ph.D. degrees from the University of Tokyo, Japan, in 1999, 2001, and 2017, respectively. In 2001, he joined Research & Development Group, Hitachi, Ltd. His research interests include information and control systems in the fields of railway, power, and industry. He won the Encouragement for Invention Prize of the Kanto Area Invention Award in 2014, and the SICE Technical Award in 2016.

Takeshi HANDA

He received his B.S. and M.S. degrees from Tohoku University, Japan, in 2011, and 2013, respectively. In 2013, he joined Research & Development Group, Hitachi, Ltd. His research interests include information systems in the fields of railway. He won the SICE Academic Encouragement Award in 2013. He is a member of the Institute of Electrical Engineers of Japan.

Yuko KATO

She received her B.S. and M.S. degrees from Keio University, Japan, in 2006 and 2008, respectively. In 2008, she joined Research & Development Group, Hitachi, Ltd. Her research interests include information systems in the fields of railway. She won the Encouragement for Invention Prize of the Kanto Area Invention Award in 2015.

Kenji KAWASAKI

In 1993, he joined Hitachi, Ltd. His research interests include information service systems in the fields of railway. He is a member of Information Processing Society of Japan.

Takashi TSUNO

In 2001, he joined Hitachi, Ltd. After experiencing software development, he is engaged as a system engineer.