Development and Evaluation of a New Platform for Accelerating Cross-Domain Data Exchange and Cooperation

Daiji Iwasa¹ · Teruaki Hayashi¹ · Yukio Ohsawa¹

Received: 23 July 2019 / Accepted: 20 November 2019 / Published online: 29 November 2019
© The Author(s) 2019

Abstract

The technologies for collecting and analyzing data are developing significantly, enabling us to gain new knowledge from various types of data. Recently, the considerably increasing expectation for cross-domain data exchange and cooperation has attracted attention from data markets. However, as we lack data-utilization knowledge, an accurate evaluation of data is difficult; a non-price indicator for deciding the exchange of data is needed. Thus, in this study, we propose a new online platform for stakeholders to communicate regarding data utilization. Our online platform, called Web Innovators Marketplace on Data Jackets (Web IMDJ), is built with reference to the process of IMDJ workshops. Web IMDJ is superior to the conventional paper-based IMDJ (hereafter, Table IMDJ) in terms of reducing the burden on conducting workshops. Notably, Table IMDJ and Web IMDJ have different communication media, and this may affect the data-utilization knowledge proposed in the workshop. Therefore, we conducted workshops on both platforms under a controlled experimental environment to compare the proposed data-utilization knowledge. Consequently, the knowledge proposed in Web IMDJ gained equal or higher ratings by third-party evaluators (those who did not join in the experimental workshop). By contrast, subjects themselves evaluated the knowledge proposed in Table IMDJ as superior to Web IMDJ. These results revealed that both workshops have advantages as data-utilization platforms. Furthermore, we derived the best practices to utilize data effectively from a detailed analysis of the data obtained from the experimental workshops.

Keywords Data Jacket · Face-to-face versus online · Market of data · Online data marketplace for innovators

✉ Daiji Iwasa
ganzaa.0707@gmail.com

¹ Department of Systems Innovation, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan

Introduction

Situations and Issues Surrounding Data Utilization

Recently, terms such as Big Data, artificial intelligence (AI), Internet of Things (IoT), and cloud computing, have become popular in various fields. These terms sometimes, called buzzwords, express vague concepts regarding the technology of the future, and indeed indicate the development of technology [1–4]. For example, regarding Big Data and IoT, sensors are improving and could provide timely and high-fidelity data [5]. Smart city projects utilize city resources more efficiently with data acquired from sensors and are, thus, advancing worldwide [6]. AI involves various technologies including the improvement and versatility of data analysis methods, such as regression analysis [7], classification by trees [8] or vectors [9], probabilistic classifiers [10], and artificial neural networks [11]. Cloud computing represents the improvement of technology and service forms relating to resource distribution [12]. Additionally, the growth of computers' calculation speeds and environments conducive to learning data analysis increase our expectation of gaining new knowledge through data utilization.

Thus, why do these terms describe only vague concepts? One of the biggest reasons could be that we cannot utilize data as per our expectations, despite the improvement of various technologies. The challenge in fully utilizing data surrounds the difficulty in exchanging data. Data holders, especially companies, cannot publish their data for free because data are their asset. Consequently, data users experience difficulty obtaining the kind of data a company owns and may, therefore, be unable to acquire the data they actually need. Moreover, when we exchange data via a purchase, it is still challenging to set a reasonable price because of the lack of knowledge surrounding data utilization [13–15].

A data market, a platform to buy/sell data, works as a method for accelerating data exchange. There are various online markets such as Every Sense¹ [16], Japan Data Exchange Inc.,² Azure Marketplace³ [17], CKAN⁴ [18], Qlik DataMarket⁵ [19], and Dawex⁶ [20]. Additionally, some research groups have surveyed the existing data markets and classified data marketplaces based on various aspects [21–23]. In these platforms, data holders provide the outlines and prices of their data based on which data users decide whether to buy the data. These pieces of information are useful in that they make it possible to search for existing data and the company that owns them. In Japan, some experiments utilizing such platforms have been

¹ <https://every-sense.com/eng/>. Accessed June 2019.

² <http://j-dex.co.jp/en/index.html>. Accessed June 2019.

³ <https://azuremarketplace.microsoft.com/en-us>. Accessed June 2019.

⁴ <https://ckan.org/>. Accessed June 2019.

⁵ <https://www.qlik.com/us/>. Accessed June 2019.

⁶ <https://www.dawex.com/en/>. Accessed June 2019.

Original Dataset

Date	Customer ID	Customer name	Purchase items	Purchase amount
2017/10/01	123456	AAA	EEE	123
2017/10/02	234567	BBB	FFF	4,567
2017/10/30	345678	CCC	GGG	890
2017/10/31	456789	DDD	HHH	12,345

Data Jacket (DJ)

DJ No. XX [Title: Supermarket point of sales (POS)]	
Outline	Monthly POS data of XXX supermarket
Variable label	Date, Customer ID, Customer name, Purchase items, Purchase amount
Format	CSV

Fig. 1 Sample of Data Jacket (DJ)

conducted to verify whether they perform as expected.,⁷Japanese only. Accessed June 2019.⁸

However, the issue of data pricing still remains. Liang et al. [14] classified the existing data pricing models according to various aspects: economical [24, 25], game theory [26, 27], and auction [28, 29] based. More recently, various types of improved data pricing and trading models have been proposed [30–32], showing the high expectation of data exchange. Nonetheless, as discussed in [14], the wide diversity of data prevents us from optimal trading.

Indeed, we do not need to price data to exchange data. Here, we focus on the communication of stakeholders and invent data-utilization scenarios that describe the purpose of data exchange. Additionally, when data are traded based on the data-utilization scenarios, the risk of business opportunity loss can be reduced by devising a contract. Therefore, the devised scenarios assist not only in exchanging data at a reasonable price, but also in facilitating the collaborations needed to achieve a given scenario. Furthermore, by accumulating such scenarios and their prices as data-utilization knowledge, predicting the outcome of data utilization and determining the price of data itself in the future can be made possible.

Communication Platform for Data Utilization

Innovators Marketplace on Data Jackets (IMDJ) [33] is a workshop-based method to facilitate communication about data utilization. Data Jacket (DJ) [34] is a framework to hold the data summary as structured information. Figure 1 illustrates a DJ created

⁷ https://www.iajapan.org/press/iot-field-trial_20181204.html.

⁸ <https://www.sakura.ad.jp/information/pressreleases/2017/12/05/90200/>. Japanese only. Accessed June 2019.

for the sample point of a sales dataset. The original dataset contains variable labels such as date, customer ID, customer name, purchase items, and purchase amount as well as their corresponding values. In contrast, the DJ contains the title, outline, and some other contents of the dataset, but does not include their corresponding values. As a result, data holders can publish their data, reducing the risks by describing the data summary in the DJ format. Further, the data summary is useful for discussion because those who are not familiar with the data can easily participate in discussions. Therefore, this study discusses the utilization of data using DJ in an IMDJ workshop. “[Face-to-Face Versus Online Communication](#)” section provides detailed procedures of the IMDJ workshop.

Several data-utilization projects with IMDJ are in progress in various domains. For example, in Tokyo Marunouchi, a demonstration experiment is being conducted since May 2018, aiming at creating a new town by utilizing data from multiple industries.⁹ Therein, several projects have begun based on data-utilization scenarios proposed by IMDJ [35]. Additionally, Kyodo Printing Co., Ltd. is working on data-distribution support projects with IMDJ.¹⁰ Thus, IMDJ has contributed to the promotion of data utilization.

However, IMDJ has a major problem in that it involves a large operational burden (see “[Comparison of the Whole Process of Table IMDJ and Web IMDJ](#)” section for details) to prepare a workshop and save the proposed knowledge. This not only results in disregarding the introduction of IMDJ to projects but also complicates the continuous discussion of data utilization with IMDJ. Indeed, we often conduct a workshop only once in IMDJ-introduced projects. This prevents the acceleration of data utilization and the accumulation of data-utilization knowledge. Moreover, from the viewpoint of innovation, we cannot expect much impact from a single workshop. To address this limitation, we propose a new online platform to reduce the burden of workshops. Also, by comparing IMDJ with our platform, we show how our platform contributes to promoting data utilization.

The outline of this paper is as follows. “[Related Studies](#)” section details the origin and process of IMDJ. Furthermore, to compare the conventional paper-based IMDJ (hereafter, Table IMDJ) with Web IMDJ, we introduce studies that compare face-to-face and online communications. “[Web IMDJ](#)” section introduces Web IMDJ, our online platform for conducting IMDJ workshops with less burden. In “[Comparison Experiment of Table and Web IMDJs](#)” section, we describe an experimental workshop for discussing the influence of different communication media, i.e., face-to-face and online communications in Table IMDJ and Web IMDJ, respectively. Moreover, we propose the most effective IMDJ operation method for promoting data utilization. Finally, we conclude this paper in “[Conclusion](#)” section.

⁹ Press release, Japanese only. Accessed June 2019. <https://pr.fujitsu.com/jp/news/2018/05/14-2.html>.

¹⁰ Press release, Japanese only. Accessed June 2019.
<https://www.kyodoprinting.co.jp/release/2018/20181026-2389.html>.

Related Studies

IMDJ

IMDJ [33] is a workshop method for discussing data utilization. In this section, we detail the process of an IMDJ workshop, which is classified into the following three phases:

- I. Preparation for the workshop;
- II. Implementation of the workshop;
- III. Preservation of knowledge proposed.

First, in the preparation phase, we determined the purpose and theme of the workshop, and the DJs to be used in the workshop. DJs can be obtained after they are registered by participants from a DJ Platform,¹¹ as well as by reusing DJs from past workshops. As the DJ Platform allows us to search from among the existing DJs in a natural language, we can easily select and use DJs matching the theme of IMDJ. The search algorithm of the DJ Platform inherits that of DJ Store [36]. The number of DJs used in a workshop typically ranges from 25 to 30. After determining the DJs to use, we created a map, in which the relationship among DJs was visualized to activate the discussion in the workshop. The relationship was visualized by extracting keywords from the content of the DJs via the KeyGraph algorithm [37]. The DJs and included keywords are extracted as nodes, and connected by an edge when the relationship is detected by KeyGraph. Figure 2 shows a visualized map on which ideas are placed using sticky notes. The black and red nodes in the figure represent the keywords and the DJs, respectively. After completing the visualized map, we printed it on a large-size paper (generally, A0 size: $841 \times 1189 \text{ mm}^2$ or B0 size: $1000 \times 1414 \text{ mm}^2$) for a smooth discussion. In addition, we created DJ cards in which the necessary DJ contents, typically title, outline, and variable label, are described. The DJ cards are the green notes in Fig. 2.

Subsequently, in the workshop phase, we discussed data utilization based on the visualized map. A general workshop was organized with 7–8 people per group. This phase comprises the following three steps:

- II-A. Present requirements;
- II-B. Propose solution scenarios;
- II-C. Evaluate the solution scenarios.

In step II-A, participants present requirements related to the contents of DJs and keywords on the map. As such, it becomes easier to conduct discussions based on the DJ when devising solutions. The position of the participants is not determined

¹¹ <https://datajacket.org/?lang=english>. Accessed June 2019.

Fig. 2 Illustration of the visualized map and ideas after IMDJ workshop

to allow them to present any requirements. Requirements are represented by yellow sticky notes and placed near the related DJ.

In step II-B, participants create solution scenarios that represent a series of actions to meet the presented requirements based on the DJs. Solutions, i.e., items which seem useful to achieve the solution, are represented by blue sticky notes with the DJ IDs. As the number of DJs on the map is between 25 and 30, other data are often necessary for devising solutions. Therefore, we describe not only the existing data but also the data assumed to be acquired in the future by red sticky notes, as additional DJs.

In step II-C, participants evaluate the proposed solution scenarios. The participants can evaluate any solution except for their own. Here, the monkey bill, a virtual bill valid only in the workshop, is used for evaluation. Ten monkey bills are distributed to each participant at the beginning of the IMDJ workshop, and the participants ‘pay bills’ to evaluate the solutions. The number of paid bills is represented by a small paper that is attached to the evaluated solution.

An essential characteristic of the evaluation in IMDJ is that negotiation of payment amount is permitted. For example, in case an evaluator pays one monkey bill when the solution proposer thinks that his solution is worth two monkey bills, negotiation starts. The evaluator either accepts the negotiation or pays only one monkey bill. In the latter case, the proposer of the solution tries to improve his solution to gain two monkey bills. Therefore, permission for negotiation activates discussion and leads to the improvement of solutions. Figure 2 illustrates the visualized map after the IMDJ workshop, and Table 1 summarizes the paper color and written content.

Table 1 Color of sticky notes and written content

Color of papers	Written content
Green	Data Jacket card (DJ card)
Yellow	Data user's requests or social issues (Requirement)
Blue	Solution scenarios for requirements (Solutions)
Red (large)	Additional Data Jackets
Red (small)	Number of paid monkey bills

Finally, in the knowledge preservation phase, we transcribed all the content described in the papers and saved them to the database. As the data-utilization knowledge proposed in the workshop has a hierarchical structure of requirement, solution, and DJ, the knowledge is stored in the RDF/XML format [38]. The hierarchical structure represents the network of solutions created by a combination of different types of DJs and requirements satisfied by multiple solutions. RDF/XML is a syntax, defined by the W3C, to express a resource description framework (RDF) graph as an extensible markup language (XML) document [39]. The knowledge saved here is used in the demonstration experiments, as introduced in “[Communication Platform for Data Utilization](#)” section, and it is utilized to improve accuracy when searching for DJs.

Face-to-Face Versus Online Communication

The comparison of face-to-face and online communications has been widely studied in the fields of education and medical care. Especially, in the education field, the educational gap is such a critical issue that online education has attracted attention as a solution. The high expectations for online education are well expressed by the Babson Survey Research Group, which has been publishing reports on American online education every year since 2003; it shows that the number of students receiving online education is increasing every year [40]. This report also reveals that the percentage of academic leaders who rate the learning outcomes from online education similar or superior to that from face-to-face instruction is increasing: 71.4% and 57.2% in the surveys of 2016 and 2003, respectively. Thus, online education has not only been improving, but also has already achieved a high level. By contrast, Bowers et al. [41] pointed out that in online education, students scarcely communicate with other students and often quit learning along the way because of feelings of isolation. As a result, blended learning, which combines conventional face-to-face and online education systems, has attracted attention; various studies have stated that blended learning improves student performance [42]. The aforementioned studies reveal that online education has been improved to a level equal to or higher than that face-to-face education; however, further promotion of blended education is necessary to improve student performance.

A similar trend exists in medical treatment. Olthuis et al. [43] compared the extent to which treatment is impacted by online cognitive behavior therapy (CBT) with therapist support, online CBT without any support, face-to-face CBT, and no treatment. They determined that online and face-to-face CBTs showed no significant differences in clinically meaningful improvement in anxiety. This study suggests that online CBT has improved to a level equal to that of face-to-face CBT. Wentzel et al. [44] argued that it is essential for patients and doctors to mutually understand treatment, especially in blended (online and face-to-face) mental health care; they proposed a framework that promotes mutual understanding, showing that the environment for blended mental care has been developing towards practical implementation.

Although the aforementioned studies focus on the one-to-one relationship between teachers and students/doctors and patients, some studies focused on team communication. Salter et al. [45] compared face-to-face and online communications in a reflective discussion that encouraged students to consider and talk about what they had observed. They found that face-to-face communication tends to cause divergent arguments, whereas online communication leads to exhaustive arguments. They also argued that threading arguments in face-to-face communication and continuing to discuss online communication promoted collective understanding. Christina et al. [46] focused on the relationship between team trust and team effectiveness and reported it to be stronger in virtual teams than in face-to-face teams. Thus, team trust is more critical in virtual teams; however, when dialogs from virtual team interactions were documented, the relationship weakened. They, thus, concluded that documenting team interactions is a viable complement to trust-building activities, particularly in virtual teams. Abrams et al. [47] compared data richness trade-offs between face-to-face and online focus groups; compared to face-to-face focus groups, online groups did not facilitate rich data.

Comparisons between face-to-face and online communications have also been studied in workshop methods derived from chance discovery. Some researchers have focused on the innovators marketplace (IM), a predecessor workshop method of IMDJ. They reported that both the feasibility and novelty of the business proposals put forward in online workshops received high evaluation [48]. It should be noted that subjects were already in a relationship of trust in advance; this seemed to affect the result. Wang et al. [49, 50] built an online system called iChance based on IM and reported that online discussion invented more ideas than paper-based IM.

These various studies overviewed in this section illustrate some of the differences in communication media methods. Furthermore, the studies conducted on this topic in various fields show that the utility of online communication is recognized, and that the expectation for common implementation is high. Moreover, instead of the shifting trend toward online communication, blended communication, which can successfully combine the advantages of both face-to-face and online communications, has attracted attention. This indicates that face-to-face and online communications have their respective advantages; this can also be applied to the discussion on data utilization. Therefore, our mission is not only to promote data utilization by implementing Web IMDJ but also to ascertain the optimal operation method by understanding the advantages of both Table (face-to-face) and Web (online) IMDJ.

Fig. 3 Structure of setting necessary information of workshop and the way to participate in a particular workshop

Web IMDJ

As discussed in “[Situations and Issues Surrounding Data Utilization](#)” section, Web IMDJ focuses on discussions surrounding data utilization among stakeholders. In other words, data exchange or collaborations can occur based not on the data price, but on the discussion surrounding data utilization. Additionally, as Web IMDJ is an online platform that reduces the burden of a workshop, data utilization can continually be easily discussed.

In this chapter, we introduce the detailed structure and implementation of Web IMDJ to show that we can conduct a Web IMDJ workshop in the same sense as a Table IMDJ workshop. Additionally, we compared the processes of Table and Web IMDJ to show that Web IMDJ reduces the burden of workshops.

Structure of Web IMDJ

First, for Phase I as outlined in “[IMDJ](#)” section we input the date, theme, username, password, and admin password, all of which are stored in the database. Simultaneously, the server allocates a unique ID for each workshop. The ID and password are necessary for participating in the workshop. The admin password is necessary for the organizer, who may need to download the various data for the workshop. Figure 3 shows the structure of this process.

Once all information has been successfully stored, the browser automatically moves to a page for creating a map, which activates the discussion in the workshop. Here, we decide which DJs to use in the workshop by having participants input new DJs, or by selecting existing DJs. Then, to create a map, we input and submit the DJ IDs, which are used to obtain all the contents of the corresponding DJs through an application programming interface (API), and keywords are extracted from the obtained information via KeyGraph. The keywords and DJs are recognized as nodes

Fig. 4 Structure of the process from deciding on the DJ to visualizing the map

and are visualized as a network diagram. Additionally, DJ cards are created by the necessary DJ contents. As there are cases in which labels of nodes overlap or some unnecessary words may remain in the visualized map, it is necessary to clean the map manually. After creating the visualized map, the “Start IMDJ” button is clicked to save the map and allow people to begin participating in the workshop. Figure 4 shows the structure of the process from deciding the DJ to visualizing the map.

In Phase II (as outlined in “**IMDJ**” section), people participate in a specific workshop by submitting the ID and password of the workshop. Next, each participant enters a username to identify the participants. Parts of the operations of participants during the workshop are defined as events; the server is notified about the events firing in real time every time an event occurs. When the server is notified of the firing of the event, it performs its corresponding processes (see Table 2) and sequentially saves the changes. Real-time communication is established here through the Web Socket protocol which allows the server to identify clients at the first connection so that the server can actively send the response to every client [51, 52]. Therefore, ideas can be proposed and discussions conducted smoothly in Web IMDJ, as in Table IMDJ. Additionally, as the data-utilization knowledge proposed in the workshop has a hierarchical structure for requirement, solution, and DJ, MongoDB [53], a database in the JavaScript object notation (JSON) format, was selected as the database of Web IMDJ. Figure 5 shows the structure of the workshop phase.

In Phase III (as outlined in “**IMDJ**” section), it is not necessary for Web IMDJ to actively save knowledge, because the database is updated at every occurrence of an event during the workshop. However, whereas Web IMDJ stores data in the JSON format, the data of the past usage of DJs and communications in IMDJ are stored in the RDF/XML format [39]; thus, data must be integrated. Therefore, we implemented a function to download data in the RDF/XML format, enabling us to incorporate data immediately after checking it. Technically, we can automatically integrate these databases; however, to extract noisy data and to avoid security issues, we incorporate data after reviewing them. Figure 6 shows the structure of the workshop phase.

Table 2 Events defined in the workshop page, with the process that they trigger, and information stored in the database for that event

Event	Triggered process	Stored data
Participate in IMDJ	Show new participant's username on the page	Participant's username, ID
Click papers on the upper right area	Create text area with info as ID, proposer, etc.	Given attributes to the text area
Drag and drop the text area on map	Show the text area on every participant's map	Sharing status
Rewrite text area	Reflect changes in all participant's page	Update the content
Associate ideas and DJs	Reflect the association in case ideas are shared	The association
Comment by chat	Reflect the comments	The comment
Evaluate with monkey bills	Reflect the transaction and attach the number of paid bills to the solution	Updated information
Exit IMDJ page	Delete the participant's username from the page	Participation status

Fig. 5 Structure of the workshop phase **Fig. 6** Structure of the preservation phase

Fig. 7 Screen transition diagram of Web IMDJ

Implementation of Web IMDJ

In this section, we describe the implementation of Web IMDJ as of June 2019. Figure 7 shows the screen transition diagram. From the top page,¹² we can move to the pages for a trial workshop, preparation, participating in a specific workshop, and the detailed explanation of IMDJ.

When we log into a workshop, the workshop page appears (Fig. 8). Suppose that in Table IMDJ, the workshop proceeds as follows; the map is browsed to invent ideas, sticky notes are used to write ideas by hand and placed on the map to share, and discussion is conducted based on the ideas. Using the same procedure, we can conduct a workshop in Web IMDJ. To pick up the sticky notes, we click the upper right images, and the text area of the corresponding color appears in the left gray area. Ideas in this gray area are not shared, to allow the participants to conceptualize in a calm manner. When the participants are ready to share the ideas, they can drag and drop them on the center map, and the ideas appear on all participants' map. The chat is available in the right-center area for discussions. Moreover, these activities are defined as events, and processed in real-time (Table 2). Thus, Web IMDJ allows users to conduct the workshop in a similar way as Table IMDJ.

However, Table and Web IMDJ have fundamental differences in face-to-face and online communication media; this may affect the number and quality of the proposed data-utilization knowledge. For example, if the number and quality of data-utilization knowledge proposed in Web IMDJ are inferior to those of Table IMDJ for one workshop, the effect of Web IMDJ becomes questionable. In other words, it is essential to guarantee the number and quality of data-utilization knowledge in Web IMDJ. “Comparison Experiment of Table and Web IMDJs” section presents a comparison of Table and Web IMDJ under a controlled situation with regard to the

¹² <https://imdj.datajacket.org/>. Accessed June 2019.

Fig. 8 Workshop page

number and quality of the proposed solutions as well as in the knowledge-creation process.

Comparison of the Whole Process of Table IMDJ and Web IMDJ

As described, although it is essential to repeat the workshop to improve the scenarios, the workshop is often conducted only once per project because of the heavy burden of the workshop. In this chapter, we compare the processes of Table and Web IMDJ for each phase of the workshop to prove that Web IMDJ reduces this burden dramatically.

In Phase I (“**IMDJ**” section), we determined the workshop’s purpose, theme, and which DJs to use first. After determining this required information, we visualized the relationship of DJs as a network diagram. In Table IMDJ, the following six steps must be followed: (A) disassemble contents of DJ into sentence elements and output in comma-separated values (CSV) format; (B) import the CSV file into a visualization tool; (C) visualize a map using the KeyGraph algorithm [37]; (D) arrange the visualized map for smooth discussion; (E) print the visualized map on large-sized sheets (typically, size A0 or B0). In addition to these steps, DJ cards must be created.

On the other hand, in Web IMDJ, owing to automation, we only need to input DJ IDs and set several parameters to visualize a map. Additionally, as Web IMDJ allows for conducting online workshops, the map does not need to be printed. As a result, Web IMDJ reduces the time required for preparation by approximately 1–2 h.

In Phase II (“**IMDJ**” section), the process of the workshop is the same in both IMDJs; however, in Table IMDJ, we need to arrange the date and place. In contrast, Web IMDJ allows users to participate from remote locations if they have a computer connected to the network.

In Phase III (“**IMDJ**” section), all knowledge proposed in Table IMDJ must be converted into digital data; whereas in Web IMDJ, it can all be download at once. Qualitatively, Web IMDJ reduces the time required for preservation by approximately 30 min–1 h.

Table 3 summarizes the aforementioned arguments. Especially for the preparation and preservation phases, Web IMDJ shortens the steps and reduces the burden on the workshop. Moreover, although Table IMDJ requires a large printer and various tools to visualize the relationship among DJs, neither is necessary in Web IMDJ. Therefore, our contribution is not only to reduce the burden of workshops, but also to allow anyone to participate in the discussion about data utilization.

Comparison Experiment of Table and Web IMDJs

Experimental design

The purpose of the experiment was to compare Table and Web IMDJs with respect to the number and quality of the proposed solutions as well as the knowledge-creation process.

Table 3 Comparison of the whole process of Table and Web IMDJs

	Table IMDJ process	Web IMDJ process
Preparation for workshop phase	Determine the purpose and theme of the workshop Decide DJs to use Disassemble contents of DJ into sentence elements and output in CSV format Import CSV file into a visualization tool Visualize a map by using KeyGraph algorithm Clean the visualized map for smooth discussion Print the visualized map on large size sheets Create DJ cards	Input necessary information of the workshop Decide DJs to use Visualize a map automatically Clean the visualized map for smooth discussion
Workshop phase	Arrange the date and place for the workshop Set up the prepared implements	Inform about the ID and password Prepare computers
Preservation phase	Convert all knowledge to digital data	Click the download button

Table 4 Grouping to avoid affecting the workshop's order and theme

	First workshop	Second workshop
Group 1	Table IMDJ theme: city and transportation	Web IMDJ theme: tourism and consumption
Group 2	Table IMDJ theme: tourism and consumption	Web IMDJ theme: city and transportation
Group 3	Web IMDJ theme: city and transportation	Table IMDJ theme: tourism and consumption
Group 4	Web IMDJ theme: tourism and consumption	Table IMDJ theme: city and transportation

The subjects included 32 Japanese industrial workers interested in data utilization. We allowed subjects to conduct both Table IMDJ and Web IMDJ workshops on different themes. We include working people because they have a strong inclination to solve social issues with their data. Note that a regular IMDJ workshop uses 25–30 DJs and takes approximately 90 min with 7 or 8 people; however, these conditions put a heavy burden on subjects. Furthermore, the communication among 7–8 people is too complicated to address as a knowledge-creation process. Therefore, in this experiment, we used 10 DJs and allow 5 min for raising requirements, 10 min for proposing solutions, and 3 min for evaluating solutions with two people per workshop. As a result, we could finish the whole experiment within 1 h.

However, the orders of Table and Web IMDJs could affect the proposed data-utilization knowledge, because the subjects already have the experience of one of the workshop types when they participate in the second workshop. Additionally, the workshop theme and DJs used would affect the proposed knowledge. Thus, we prepared two themes and classified all 16 teams into four groups which exchange both the order and theme of the workshop. Table 4 summarizes the groups. Note that the themes are initially described in Japanese and translated in English.

The workshop themes were determined as “city and transportation” and “tourism and consumption” based on the words often included in search queries entered in the DJ Store [36], a platform for searching DJs. The DJ Store was considered because those who visit DJ Store are interested in data utilization, and search queries entered by them represent the fields of their interest. After determining the theme, we selected the 10 DJs to use in the workshop from the DJs searched for in the DJ Store using the themes as search queries.

On the day of the experiment, we delivered lectures to the subjects about IMDJ if they were not familiar with it. Thus, subjects understood IMDJ as a workshop method for accelerating data utilization before attending the workshop. When experimenting, we prepared a manual, which the practitioner used in proceeding with the experiment.

The procedure is as follows: consent was acquired to use proposed ideas for research, subjects were paired up, and details were provided about Table or Web IMDJs. When conducting a Table IMDJ workshop, we first distributed the necessary tools such as a visualized map, sticky notes, and DJ cards, and then the subjects placed DJ cards on the corresponding nodes. Then, subjects raised requirements for 5 min and proposed solutions for 10 min; this was followed by the evaluation of

Table 5 Evaluation index of each solution with detailed explanation

Evaluation indicators	Detailed explanation
Novelty	The extent of newness of a solution
Marketability	Can we expect a sufficient market size?
Feasibility	Degree of possibility to realize the solution
Utility	Degree of usefulness of the solutions

solutions. Although we used monkey bills for the evaluation in a standard workshop, we asked subjects to perform a five-grade subjective evaluation on the four indicators of novelty, marketability, feasibility, and utility, for a more detailed evaluation. A detailed explanation of each indicator is shown in Table 5. Although we assumed the evaluation time to be 3 min, it was often not possible to evaluate all of the solutions within 3 min. Accordingly, we prioritized the acquisition of the evaluation data of the subjects and extended the evaluation time. Further, we recorded conversations between subjects with voice recorders in Table IMDJ. When conducting a Web IMDJ workshop, first we explained how to use Web IMDJ in the trial workshop page where subjects could get used to operations by practicing. After the explanation, subjects logged into the workshop page of each team and experimented. The subsequent procedures are common to those used in Table IMDJ.

Comparison of the Number and Quality of Proposed Solutions

As described earlier, 16 teams participated in the experiment; however, we did not use the data of 5 teams because of operational mistakes on our end and the fact that some subjects performed other work not relating to the experiment during the experiment (it is necessary to prevent external influences to the order and theme of the workshop). Thus, we used the data of 8 teams after excluding the 3 teams who performed the experiments last. In Web IMDJ, we determined some incorrect operations derived from the interface, for example, some ideas were almost the same or not associated with other ideas. In this study, we corrected such obvious mistakes in a way that the subjects originally intended. However, as a matter of course, it is essential to improve the interface so as not to allow such preventable mistakes to occur.

Moreover, as mentioned earlier, we acquired the evaluation of solutions from the subjects; however, this evaluation is insufficient as subjects are likely to be biased. Therefore, we asked students and researchers (hereafter, referred to as the third-party) who did not participate in the workshop and are familiar with the process of IMDJ to evaluate the solutions. We selected those familiar with IMDJ as evaluators because it seemed difficult to ask individuals to correctly evaluate the solutions if they did not know the invention process. The evaluation method was the same as the subject's evaluation, i.e., a five-grade subjective evaluation with respect to the four indicators of novelty, marketability, feasibility, and utility. These four indicators were derived from past research which evaluates the scenarios based on solutions in novelty, feasibility, and utility [48]. Here, we added marketability as an indicator.

Table 6 Total amount of the proposed knowledge

	Requirement	Solution	Additional DJ
Table IMDJ	38	29	15
Web IMDJ	35	30	14

Table 7 Comparison of subject-based evaluation scores

Subjects	Novelty	Marketability	Feasibility	Utility
Table IMDJ	3.30	3.93	3.64	3.96
Web IMDJ	3.42	3.51	3.64	3.66
<i>p</i> value	n.s. (0.66)	** (0.008)	n.s. (0.95)	* (0.03)

n.s. no significant difference

*Significant difference in 95% of confidence interval

**Significant difference in 99% of confidence interval

Table 8 Comparison of third-party-based evaluation scores

Third party	Novelty	Marketability	Feasibility	Utility
Table IMDJ	3.33	3.41	3.46	3.67
Web IMDJ	3.06	3.55	3.46	3.72
<i>p</i> value	* (0.05)	n.s. (0.26)	n.s. (0.98)	n.s. (0.56)

n.s. no significant difference

*Significant difference in 95% of confidence interval

**Significant difference in 99% of confidence interval

The evaluators included 14 people, each of whom evaluated 19 or 20 solutions. The number of evaluations is 109 in the subject-based evaluation and 276 in the third-party-based evaluation. Note that the solutions by the subjects and third-party differed in that the former evaluate solutions by considering the discussion during the workshop while the latter evaluated solutions only according to the written pieces of contents.

Table 6 shows the total number of proposed data-utilization knowledge in both sessions of Table and Web IMDJ. No significant difference was observed in the number of solutions.

Tables 7 and 8 show the average subject and third-party-based evaluation values, respectively. As normality was rejected by the Shapiro–Wilk test, which tests the null hypothesis that data comes from the normal population [54], we used Mann–Whitney’s *U* test [55] to investigate whether there were significant differences. The values in parentheses in Tables 7 and 8 represent the *p* value, which is small if the means of the ranks in the two groups differ considerably. As shown in Tables 7 and 8, the marketability and utility of Table IMDJ are significantly higher in the subject’s evaluation; while in Web IMDJ, these indicators are higher in the third-party-based evaluation, although there is no significant difference. The novelty of Web IMDJ is higher in the subject’s evaluation, while the novelty of Table IMDJ

Fig. 9 Knowledge-creation process in an IMDJ workshop

is higher in the third-party-based evaluation. Regarding the feasibility, the value is almost the same in both.

Comparison of the Knowledge-Creation Process

Figure 9 shows the knowledge-creation process in the IMDJ workshop; that is, the workshop proceeds with discussion and idea invention based on the visualized map created from DJs. In the figure, *req*, *sol*, and *add* represent requirements, solutions, and additional DJs, respectively, and *remarks* represents utterances during the workshop. This process is common between Table and Web IMDJs. Moreover, in this experiment, it can be assumed that the workshop proceeds based on the same visualized map as we removed the influence of theme by dividing teams into four groups, as shown in Table 4. Therefore, by comparing the richness of *req*, *sol*, *add*, and discussion as well as the relationships among them, we can consider the feature of the knowledge-creation process. Here, the relationships among these indicators are discussed based on the number of common elements. The definition of elements is nouns obtained by decomposing each sentence into morphemes excluding meaningless nouns.

To eliminate ambiguity about elements, we explain this definition using an example. The title of DJ107 is railway congestion data that contain railway, congestion, and data as nouns. Here, “data” is the meaningless word because all DJs are about data. Then, let us suppose that the requirement “I want to avoid traffic congestion” is proposed in the workshop. As the DJ107 and the requirement have “congestion” in common, we add one to the shared element score between the DJ and requirements.

Table 9 shows the total number of elements included in the discussion and the average of the number of elements included in *req*, *sol*, and *add*. The total number of elements in remarks is much larger in Table IMDJ, showing that subjects had livelier discussions in the Table IMDJ. In contrast, the number of elements

Table 9 Total number of elements included in the discussion and the average of the number of elements included in the requirements (*req*), solutions (*sol*), and additional (*add*) DJs, per paper

	Remarks	<i>req</i>	<i>sol</i>	<i>add</i>
Table IMDJ	774	4.5	6.3	4.7
Web IMDJ	118	4.7	8.3	3.9

Table 10 Number of shared elements among requirements (*req*), solutions (*sol*), additional (*add*) DJs, remarks, and DJ in Table IMDJ

Table	<i>Req</i>	<i>Sol</i>	<i>Add</i>	Remarks	DJ
<i>Req</i>	2.1	3.6	1.2	2.3	1.9
<i>Sol</i>	4.8	2.6	1.8	2.2	3.4
<i>Add</i>	3.1	3.4	0.9	1.8	1.3
Remarks					159

Table 11 Number of shared elements among requirements (*req*), solutions (*sol*), additional (*add*) DJs, remarks, and DJ in Web IMDJ

Web	<i>Req</i>	<i>Sol</i>	<i>Add</i>	Remarks	DJ
<i>Req</i>	1.9	6.5	2.1	0.2	1.2
<i>Sol</i>	7.6	3.3	2.3	0.2	1.8
<i>Add</i>	5.1	5	1.5	0	1.3
Remarks					8

included in the solution is significantly larger in Web IMDJ, indicating that solutions contain more information in Web IMDJ.

Tables 10 and 11 show the number of shared elements among the requirements, solutions, additional DJs, remarks, and DJ, respectively. As shown, the number of shared elements between requirements, solutions, and additional DJs is more abundant for all except Req–Req in Web IMDJ. This suggests that when creating data-utilization knowledge in Web IMDJ, there is a tendency to refer to the other proposed knowledge. By contrast, the number of shared elements between knowledge (*req*, *sol*, *add*), *remarks*, and DJ is substantially larger for all in Table IMDJ. Additionally, the number of shared elements between remarks and DJ is more abundant for Table IMDJ, suggesting that subjects hold discussions based on the visualized map and the proposed ideas reflect the discussion. Note that we did not divide the number of shared elements by the number of elements because we want to make comparisons per workshop. For example, in Table IMDJ, the number of shared elements between remarks and DJ is 159 out of 774 elements of the whole discussion. The focus here is not on the proportion of contents related to the DJ among the discussion, but on the extent of the discussion related to the DJ. Thus, we determine the value of the shared elements to be 159.

Discussion

In this section, we discuss the results of the experiment described in “[Comparison of the number and quality of proposed solutions](#)” and “[Comparison of the knowledge-creation process](#)” sections detail.

Tables 7 and 8 show the subject- and third-party-based evaluations, respectively. As described earlier, a difference exists because the subjects evaluate solutions by considering discussion during the workshop, whereas third-party members evaluate solutions only according to the written contents. In other words, the effect of discussions during the workshop will appear as differences in evaluations. Note that as shown in Table 9, the number of elements in remarks is much more significant in Table IMDJ. Thus, the influence of the discussion appears to be notable in the Table IMDJ. In Tables 7 and 8, marketability and utility scores of the Table IMDJ are much higher in subject-based evaluation than those in the third-party-based evaluation; however, these scores in Web IMDJ are only slightly different. Therefore, active discussion in Table IMDJ seems to be a critical factor of evaluation, especially with regard to marketability and utility.

At the same time, in Table 8, the scores of marketability, feasibility, and utility in Web IMDJ are equal to or higher than those in Table IMDJ. This result suggests that for written contents stored as data-utilization knowledge, Web IMDJ achieves a similar or higher score compared to that of Table IMDJ.

However, the score of novelty in Tables 7 and 8 cannot be explained by the difference in the number of remarks. Thus, we need other evaluation methods to compare the novelty.

Concerning the knowledge-creation process, as argued in “[Comparison of the knowledge-creation process](#)” section, subjects actively held discussions in the Table IMDJ. Additionally, subjects hold discussions based on the contents and extracted keywords of DJs. These results are similar to the study by Abrams et al. [47], in which the authors argued that discussions by chat not only contain less information than face-to-face discussions but also often have nothing to do with the agenda.

In contrast, for IM, the predecessor of IMDJ, ideas proposed in the online workshop are superior in novelty and feasibility [48]. Further, subjects invented more ideas in an online workshop than in a face-to-face workshop [49]. Except for the novelty indicator, which we cannot evaluate effectively, these trends did not appear in our results.

These gaps result from the difference in the way subjects worked in IM and our research. In IM, subjects discussed to evaluate the proposed ideas. By contrast, the time for evaluation was too short in our experiment, such that subjects did not hold discussions during the evaluation. These differences in experimental procedures were considered to have a greater impact on the quality of solutions in Web IMDJ because there is a tendency for the creation and the evaluation of solutions to be separated due to restrictions in screen size. If we had taken more time for evaluation, not only the quality of solutions, but also the amount of conversation in Web IMDJ, could be larger.

From the viewpoint of divergence and convergence of discussion, the discussion in Table IMDJ is based on the DJ. Namely, subjects hold discussions without much divergence from the content of the DJ. However, according to Salter et al. [45], face-to-face discussion was so wide-ranging and loosely structured that divergent aspects of the topic were uncovered, whereas topics were explored more thoroughly in the online discussion. The reason for the resulting gap is that IMDJ focuses on inventing data-utilization knowledge, whereas reflective discussion allows for free discussion. In other words, IMDJ requires the convergence of the discussion as solutions, whereas reflective discussion focuses on practicing discussions.

To summarize, this experiment yields that the solutions in Table IMDJ were well considered based on active discussions, such that the data-utilization knowledge gained higher quality. However, the storage of all the discussions as data is cumbersome, and it is difficult to save them as reusable knowledge. On the contrary, in Web IMDJ, the proposed knowledge and discussions by chat can automatically be stored. Additionally, regarding the written contents on the sticky notes that are preserved as knowledge, Web IMDJ is equal to or better than Table IMDJ.

Regarding the knowledge-creation process, subjects held discussions based on the DJs, and the proposed solutions reflected the discussion in Table IMDJ. In contrast, in Web IMDJ, subjects invented solutions referring to other proposed ideas. Considering the comparison of these outcomes with regard to face-to-face and online communications, some earlier studies reported similar results; whereas, others reported different results. Contrastingly, we argue that the factor lies in the process or purpose of the communication.

Here, referring to the results and considerations, how should we proceed with data utilization in an IMDJ workshop? Although the number and quality of solutions in Web IMDJ are equal or superior to those in Table IMDJ with regard to the written contents, subjects actively conduct discussions in Table IMDJ. Thus, we should not conclude that it is better to only use Web IMDJ. Therefore, in the next section, we discuss in detail the most effective IMDJ operation method for promoting data utilization by analyzing the acquired data.

Most effective IMDJ operation method for promoting data utilization

In the previous section, we compared Table and Web IMDJs in terms of the number and quality of the solutions, and the knowledge-creation process. In this section, by analyzing the data acquired in the workshop in detail, we propose the most effective IMDJ operation method for promoting data utilization. Specifically, we consider the relationship between the number of combined DJs and the evaluation score, the influence of workshop order, and subjects' opinions.

Relationship between the number of combined DJs and evaluation score

Figures 10 and 11 show the total number of DJs and additional DJs associated with the proposed solutions in Table and Web IMDJs, respectively. These figures

Fig. 10 Total number of DJs and additional DJs associated with the proposed solutions in Table IMDJ

Fig. 11 Total number of DJs and additional DJs associated with the proposed solutions in Web IMDJ

show that the number of combined DJs is more abundant in Table IMDJ. In general, when a solution is associated with many DJs, it can achieve a high score, especially for feasibility, because many data are accessible. Table 12 shows Spearman's rank correlation coefficient [56] for the number of DJs related to the solution and evaluation scores. We used Spearman's rank correlation coefficient because the null hypothesis, that each evaluation value is derived from a normal population, was rejected and it was necessary to use a nonparametric method. The discussion has been presented in "Discussion" section.

Table 12 Spearman's rank correlation coefficient for the number of DJs related to the solution and evaluation scores

	Novelty	Marketability	Feasibility	Utility
The subjects	− 0.10	0.13	0.27*	0.29*
The third party	0.02	0.21	0.21	0.34**

*Significant difference in 95% of confidence interval

**Significant difference in 99% of confidence interval

Table 13 Number of included elements in each team (classified by groups of Table 4)

Elements in discussion	Table IMDJ	Web IMDJ
Group 1	127	65
Group 1	143	13
Group 2	116	0
Group 2	190	18
Group 3	44	0
Group 3	20	0
Group 4	91	22
Group 4	43	0

Influence of the workshop order

As shown in Table 4, we experimented with four groups to prevent the influences of the order and theme. Did every group have different performance scores? Table 13 shows the number of elements included in the discussion of all teams classified for each group. Groups 1 and 2 first participated in the Table IMDJ workshop, whereas Groups 3 and 4 participated in the Web IMDJ workshop. The number of elements included in the discussion significantly differs between the two. Specifically, for Groups 1 and 2, the number of elements included in the discussion in Table IMDJ exceeds 100, whereas that in Web IMDJ is also relatively large. By contrast, Groups 3 and 4 held less discussions in both Table and Web IMDJs.

Here, to analyze the relationship between the activeness of the discussion and the quality of the solution, we focus on the three teams with the largest amount of discussion and the three teams with the least amount of discussion. Tables 14 and 15 show the evaluation by the third party for the proposed solutions in Table and Web IMDJs. We used the third-party's evaluation here because the bias can significantly affect the subject's evaluation in the comparison of teams. We also performed a significant difference test using Mann–Whitney's *U* test [55]. As in the analysis in “[Comparison of the Number and Quality of Proposed Solutions](#)” section, the values in parentheses in Tables 14 and 15 represent the *p* value. Table 16 shows the number of proposed solutions among teams with the

Table 14 Comparison of evaluation scores of the teams holding heavy and least amount of discussions in Table IMDJ

Table IMDJ	Novelty	Marketability	Feasibility	Utility
Rich discussion (Table → Web)	3.07	3.34	3.61	3.43
Small discussion (Web → Table)	3.38	3.38	3.49	3.70
<i>p</i> value	n.s. (0.14)	n.s. (0.73)	n.s. (0.42)	n.s. (0.17)

*Significant difference in 95% of confidence interval

**Significant difference in 99% of confidence interval

Table 15 Comparison of evaluation scores of the teams holding heavy and least amount of discussions in Web IMDJ

Web IMDJ	Novelty	Marketability	Feasibility	Utility
Rich discussion (Table → Web)	2.67	3.69	3.62	3.90
Small discussion (Web → Table)	3.28	3.42	3.37	3.68
<i>p</i> value	** (0.009)	n.s. (0.19)	n.s. (0.33)	n.s. (0.17)

*Significant difference in 95% of confidence interval

**Significant difference in 99% of confidence interval

Table 16 Comparison of the number of proposed solutions of the teams holding heavy and least amount of discussions

	Table IMDJ	Web IMDJ
Rich discussion (Table → Web)	9	9
Lean discussion (Table → Web)	14	14

heaviest and least discussions. A discussion has been provided in “[Discussion](#)” section.

Subject opinion

After the experiment, we voluntarily obtained an unsigned questionnaire from the subjects. Table 17 shows the questions and answers. Seven subjects responded, and all comments that lead to personal information are eliminated.

We have two purposes for acquiring this questionnaire. One is to argue the difference between Table and Web IMDJs from the viewpoint of the subject, which corresponds to the first question. For example, we received answers “easy

Table 17 Subject opinion questionnaire

Questions	Answers
Please describe what you noticed and felt about the differences between Table IMDJ and Web IMDJ	<p>Comments for Table IMDJ</p> <p>Easy to overlook and to lead letters thoroughly</p> <p>Easy to get excited by sharing non-verbal information</p> <p>Unlikely to miss the points thanks to oral communication</p> <p>Comments for Web IMDJ</p> <p>Easy to put out many ideas</p> <p>Visualized map and sticky notes are easy to see, but it takes time to get used to</p> <p>Feel free to link a solution with DJs</p> <p>I can concentrate on my thoughts, but the sense of making solutions by discussion is weak. Feel like doing separate work</p> <p>I concentrated on my thoughts</p>
Please describe what you feel is convenient or inconvenient for Web IMDJ	<p>Convenient</p> <p>Able to link requirement, solution, DJ with drag and drop operation</p> <p>Accommodate the flexibility from the full bird's eye view to detailed browsing by scaling</p> <p>Fundamentally useful as a tool</p> <p>Multiple people can collaborate simultaneously at a distance</p> <p>Remain as data</p> <p>Inconvenient</p> <p>Linking operation is hard, and the map is confusing to see (color scheme, overlapping of characters and lines, etc.)</p> <p>Hard to start without explanation. Need materials for immediate operation</p> <p>Hard to operate when the screen is small</p> <p>Once expanded, the characters of the map remained large even if I shrink it</p> <p>Hard to grasp the information when many solutions appear</p> <p>Hard to notice that others stuck the ideas</p> <p>Even if the requirements and solutions are linked, it does not visualize on the map</p> <p>Need to change the font size manually</p> <p>I could not talk until I went to a specific chat room and I was at a loss where to talk the idea noticed on the spot. I felt they affected the progress</p>

Table 17 (continued)

Questions	Answers
Please describe the desired or required functions for Web IMDJ in promoting data utilization	3D network diagram Visualize the connection between DJs and solutions Auxiliary function to reduce the burden of creating scenarios such as scenario planning framework In addition to chat, need voice chat as used in a web meeting Function to automatically convert the input voice into text and display it Stepless zoom function Change chat notification from displaying red circle to changing the shape of tabs Notification when a sticky note appears Function to check unidentified sticky note Always show the chat. Mention function No chat function required Function for quickly reaching the sticky notes I like, such as “Favorites” or “Watch”

to get excited by sharing nonverbal information” in Table IMDJ and “I concentrate on my thoughts” in Web IMDJ. These opinions seem to represent contributing factors for the rich discussion in Table IMDJ versus the small discussion in Web IMDJ. However, as these opinions were much affected by individual differences, we only considered qualitative evaluation.

Another purpose is to examine the method of improving the interface of Web IMDJ from the opinion of subjects who experienced Web IMDJ; this purpose, answers the second and third questions. Among various answers, there is a high demand for improvements on the interface related to the association between DJs and solutions, and communication.

Discussion

Table 12 shows a correlation between the number of combined DJs in the solution and the feasibility and utility. This is because a combination of multiple DJs makes it easier to propose cross-domain solutions, resulting in a high score in utility. In other words, the small number of combined DJs in Web IMDJ tends to lower the solution quality because sufficient information cannot be reflected in the solution. Referring to the subject’s opinion, the small number of combined DJs can be attributed to the problem in the interface of associating DJs and solutions. Thus, by improving the interface, the quality of the proposed solution increases in Web IMDJ.

Considering the workshop order, the performance of the second workshop would be superior because of the experience gained in the first workshop, as shown in Table 15. In other words, for the solution proposed in Web IMDJ, the evaluation scores of marketability, feasibility, and utility are superior to the team that first used Table IMDJ. By contrast, Table 14 shows that the marketability score in Table IMDJ

is almost the same, and the score of feasibility is higher for the team who first used Table IMDJ. This result could be because of the vibrant discussion in Table IMDJ.

However, regarding the novelty, the teams that attended the Web IMDJ workshop first gave high evaluations for both Table and Web IMDJs. This result suggests that small discussion leads to the invention of new ideas. Additionally, Table 16 shows that the number of proposed solutions is higher for teams with lean discussions. Thus, although teams that first conducted Web IMDJ hold fewer discussions, they propose more solutions.

To summarize, the analysis results show that when we start from Table IMDJ, we can actively discuss and invent solutions superior in marketability and feasibility, as discussed in “[Discussion](#)” section. Here, the indicators in which differences appear are different because the number of targeted teams is as small as three. However, what is essential here is that the activeness of discussion affects the quality of the proposed solution, and thus it is necessary to improve the interface of Web IMDJ by referring to the studies to activate online discussions [57].

In contrast, when users start from Web IMDJ, they have fewer discussions, but propose more solutions, which are superior in terms of the score of novelty. This suggests that when we hold fewer discussions, we have less resistance to proposing idea-based solutions and can invent new solutions.

How is the influence of order of face-to-face and online communications argued in other fields? According to some subjects in the field of data research, participants are asked to communicate online before they perform group work face-to-face, for smooth progress of experiments. Additionally, in the field of medical care, there is a recognition that online communication makes it easier to distinguish the opinions of participants [58]. These results conflict with the result that subjects discuss actively when they first use Table IMDJ. This conflict is attributed to the difference in the purpose of the disciplines; that is, in the field of data research and medical care, communication itself is essential; whereas, inventing data-utilization scenarios is essential in IMDJ.

Based on the abovementioned discussions, we can propose an effective IMDJ operation method according to the given purpose: when we want to hold more discussions and acquire superior solutions in marketability, feasibility, and utility, we use Table IMDJ first; in contrast, when we want to acquire brand-new solutions in a casual atmosphere, we use Web IMDJ first. However, in general, IMDJ workshop essentially requires discussion among stakeholders, and thus, the activation of discussion is desired for overcoming the contextual gap among the different domains of stakeholders.

Therefore, it is better to start with Table IMDJ. It is also true that conducting a Table IMDJ workshop is cumbersome. Moreover, as argued in “[Discussion](#)” section, Web IMDJ allows inventing solutions equal or superior to those of Table IMDJ with regard to the written content stored as data-utilization knowledge. Consequently, by continuously holding discussions in Web IMDJ after the first Table IMDJ workshop, we can effectively progress data utilization.

An important opinion of the subjects is that Web IMDJ still has room for improvement. For example, a subject suggested adding voice chat, which is easy to achieve by integrating other apps. Here, recall that Abrams et al. [47] reported that the data richness of online audiovisual focus groups is the same as that of face-to-face focus groups. This indicates that enabling voice chat activates discussions in Web IMDJ.

Therefore, Web IMDJ has the potential to further contribute to the promotion of data utilization by improving the interface.

Overall, here we have outlined the most efficient IMDJ operation method. That is, after improving the critical interface issues as soon as possible, a Table IMDJ workshop must first be conducted, followed by continuing discussions on Web IMDJ.

Conclusion

To promote data utilization, we implemented Web IMDJ, a platform for discussing the invention of data-utilization scenarios. To examine the influence of different communication media, namely face-to-face or online methods, we compared both Table and Web IMDJs in terms of various aspects. Experimental results showed that the number and quality of solutions in Web IMDJ are equal or superior to those in Table IMDJ, in terms of the written contents; however, subjects actively discussed more in Table IMDJ. Further, we proposed the most efficient IMDJ operation method by combining the features of Table and Web IMDJs.

The contribution of this paper can be divided into two main aspects. First, we promote data utilization. Our platform, Web IMDJ, enables us to repeat the workshop in which we invent the data-utilization scenarios to be as useful as those in Table IMDJ, with lesser complications. Second, we compared the difference in the communication media with discussions on data utilization. As IMDJ is a complicated workshop method for inventing new knowledge, our study can provide suggestions into this research field.

In the actual discussions for data utilization, there are cases in which Web or Table IMDJ can not only be used independently, but also together, for better discussions. In such cases, it is necessary to expand the settings of our study to assess the quality of facilitation and communication.

We have been developing a platform integrated with the DJ site, DJ Store, and Web IMDJ, all of which are currently separate tools. Once completed, the platform will improve user experience and accelerate the accumulation of data-utilization knowledge. In the future, utilizing the accumulated data-utilization knowledge can inform determinations of proper data prices.

Acknowledgements This work has been supported by industrial collaborators of Ohsawa Laboratory, including Kozo Keikaku Engineering Inc. The authors would like to thank all the staff members of Kozo Keikaku Engineering Inc., and FUJITSU LIMITED for supporting the research; and Editage (www.editage.com) for English language editing.

Funding This study was partially supported by JST CREST [No. JPMJCR1304], JSPS KAKENHI [JP16H01836 and JP16K12428].

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval Not applicable.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- De Mauro, A., Greco, M., Grimaldi, M.: What is big data? A consensual definition and a review of key research topics. *AIP Conf. Proc.* **1644**, 97–104 (2015)
- King, B.: Creativity and artificial intelligence: a digital art perspective (2018). <https://doi.org/10.2139/ssrn.3225323>
- Madakam, S., Ramaswamy, R., Tripathi, S.: Internet of things (IoT): a literature review. *J. Comput. Commun.* **25**, 164 (2015)
- Rittinghouse, J.W., Ransome, J.F.: *Cloud Computing: Implementation, Management, and Security*. CRC Press, Boca Raton (2016)
- Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of things (IoT): a vision, architectural elements, and future directions. *Future Gen. Comput. Syst.* **29**, 1645–1660 (2013)
- Misbahuddin, S., Zubairi, J.A., Saggaf, A., Basuni, J., A-Wadany, S., Al-Sofi, A.: IoT based dynamic road traffic management for smart cities. In: 2015 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET), Islamabad, pp. 1–5 (2015)
- Draper, N.R., Smith, H.: *Applied Regression Analysis*. Wiley, New York (2014)
- Breiman, L.: *Classification and Regression Trees*. Routledge, New York (1984)
- Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. *Neural Proc. Lett.* **9**, 293–300 (1999)
- Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv.* **31**, 264–323 (1999)
- LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436 (2015)
- JoSEP, A.D., Katz, R., KonWinSKI, A., Gunho, L.E.E., Patterson, D., Rabkin, A.: A view of cloud computing. *Commun. ACM* **53**, 50–58 (2010)
- Stahl, F., Schomm, F., Vomfell, L., Vossen, G.: Marketplaces for digital data: Quo vadis? (No. 24). Working Papers, ERCIS-European Research Center for Information Systems (2015)
- Liang, F., Yu, W., An, D., Yang, Q., Fu, X., Zhao, W.: A survey on big data market: pricing, trading and protection. *IEEE Access* **6**, 15132–15154 (2018)
- Zhang, M., Arafa, A., Huang, J., Poor, H. V.: How to price fresh data. (2019). arXiv preprint [arXiv:1904.06899](https://arxiv.org/abs/1904.06899)
- Mano, H.: EverySense: An end-to-end IoT market platform. In: Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services (MOBIQUITOUS 2016). ACM, New York, NY, USA, pp. 1–5 (2016)
- Chappell, D.: Introducing the windows azure platform. David Chappell & Associates White Paper (2010)
- Winn, J.: Open data and the academy: an evaluation of CKAN for research data management. In: IASSIST 2013, Cologne (2013)
- Ponte, D.: Enabling an Open Data Ecosystem. ECIS 2015 Research-in-Progress Papers, Paper 55 (2015)
- Richter, H., Slowinski, P.R.: The data sharing economy: on the emergence of new intermediaries. *IIC-Int. Rev. Intellect. Prop. Compet. Law* **50**, 4–29 (2019)
- Stahl, F., Schomm, F., Vossen, G., Vomfell, L.: A classification framework for data marketplaces. *Vietnam J. Comput. Sci.* **3**, 137–143 (2016)
- Panarello, A., Tapas, N., Merlino, G., Longo, F., Puliafito, A.: Blockchain and IoT integration: a systematic survey. *Sensors* **18**, 2575 (2018)
- Perera, C., Liu, C.H., Jayawardena, S.: The emerging internet of things marketplace from an industrial perspective: a survey. *IEEE Trans. Emerg. Top. Comput.* **3**, 585–598 (2015)
- Shen, Y., Guo, B., Shen, Y., Duan, X., Dong, X., Zhang, H.: A pricing model for big personal data. *Tsinghua Sci. Technol.* **21**, 482–490 (2016)

25. Zheng, Z., Peng, Y., Wu, F., Tang, S., Chen, G.: An online pricing mechanism for mobile crowd-sensing data markets. In: Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, p. 26 (2017)
26. Niyato, D., Alsheikh, M. A., Wang, P., Kim, D. I., Han, Z.: Market model and optimal pricing scheme of big data and Internet of Things (IoT). In: 2016 IEEE International Conference on Communication (ICC), Kuala Lumpur, pp. 1–6 (2016)
27. Sen, S., Joe-Wong, C., Ha, S., Chiang, M.: Smart data pricing (SDP): Economic solutions to network congestion. *Rec. Adv. Netw.* **1**, 221–274 (2013)
28. Cao, X., Chen, Y., Liu, K.J.R.: An iterative auction mechanism for data trading. In: IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP), New Orleans, LA, pp. 5850–5854 (2017)
29. Susanto, H., Zhang, H., Ho, S., Liu, B.: Effective mobile data trading in secondary ad hoc market with heterogeneous and dynamic environment. In: IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, pp. 645–655. IEEE (2017)
30. Yang, J., Zhao, C., Xing, C.: Big data market optimization pricing model based on data quality. *Complexity* **2019**, 5964068 (2019). <https://doi.org/10.1155/2019/5964068>
31. Zheng, Z., Peng, Y., Wu, F., Tang, S., Chen, G.: ARETE: on designing joint online pricing and reward sharing mechanisms for mobile data markets. *IEEE Trans. Mobile Comput.* (2019). <https://doi.org/10.1109/TMC.2019.2900243>
32. Oh, H., Park, S., Lee, G.M., Heo, H., Choi, J.K.: Personal data trading scheme for data brokers in IoT data marketplaces. *IEEE Access* **7**, 40120–40132 (2019)
33. Ohsawa, Y., Kido, H., Hayashi, T., Liu, C., Komoda, K.: Innovators marketplace on data jackets, for valuating, sharing, and synthesizing data. In: Knowledge-Based Information Systems in Practice, pp. 83–97. Springer, Cham (2015)
34. Ohsawa, Y., Kido, H., Hayashi, T., Liu, C.: Data jackets for synthesizing values in the market of data. *Procedia Comput. Sci.* **22**, 709–716 (2013)
35. Ejiri Y., Ikeda E, Sasaki H.: Realization of data exchange and utilization society by Blockchain and Data Jacket: merit of consortium to accelerate co-creation. In: IEEE International Conference on Data Mining Workshops (ICDMW), pp. 180–182. IEEE (2018)
36. Hayashi, T., Ohsawa, Y.: Retrieval system for data utilization knowledge integrating stakeholders' interests. In: AAAI Spring Symposium 2018 Beyond Machine Intelligence: Understanding Cognitive Bias and Humanity for Well-being AI (2018)
37. Ohsawa, Y., Benson, N. E., Yachida, M.: KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor. In: Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries (ADL'98), pp. 12–18. IEEE (1998)
38. Hayashi, T., Ohsawa, Y.: Knowledge structuring and reuse system design using RDF for creating a market of data. In: 2nd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 607–612. IEEE (2015)
39. Beckett, D., McBride, B.: RDF/XML syntax specification (revised). W3C Recommendation 10.2.3 (2004). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.3972&rep=rep1&type=pdf>
40. Allen, I.E., Seaman, J.: Online Report Card: Tracking Online Education on the United States. Babson Survey Research Group; Quahog Research Group, LLC Alfred P. Sloan Foundation (ERIC Document Reproduction Service No. ED572777) (2016)
41. Bowers, J., Kumar, P.: Students' perceptions of teaching and social presence: a comparative analysis of face-to-face and online learning environments. *Int. J. Web-Based Learn. Teach. Technol.* **10**, 27–44 (2015)
42. Powell, A., Watson, J., Staley, P., Patrick, S., Horn, M., Fetzer, L., Hibbard, L., Oglesby, J., Verma, S.: Blending Learning: The Evolution of Online and Face-to-Face Education from 2008–2015. Promising Practices in Blended and Online Learning Series. Michigan; Nevada; New York; North Carolina; Pennsylvania; Utah; Washington (ERIC Document Reproduction Service No. ED560788) (2015)
43. Olthuis, J.V., Watt, M.C., Bailey, K., Hayden, J.A., Stewart, S.H.: Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. *Cochrane Database Syst. Rev.* **3**, CD011565 (2016)
44. Wentzel, J., van der Vaart, R., Bohlmeijer, E.T., van Gemert-Pijnen, J.E.: Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. *JMIR Mental Health* **3**, e9 (2016)

45. Salter, S., Douglas, T., Kember, D.: Comparing face-to-face and asynchronous online communication as mechanisms for critical reflective dialogue. *Educ. Action Res.* **25**, 790–805 (2017)
46. Breuer, C., Hüffmeier, J., Hertel, G.: Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators. *J. Appl. Psychol.* **101**, 1151–1177 (2016)
47. Abrams, K.M., Wang, Z., Song, Y.J., Galindo-Gonzalez, S.: Data richness trade-offs between face-to-face, online audiovisual, and online text-only focus groups. *Soc. Sci. Comput. Rev.* **33**, 80–96 (2015)
48. Ohsawa, Y., Okamoto, K., Takahashi, Y., Nishihara, Y.: Innovators marketplace as game on the table versus board on the web. In: 2010 IEEE International Conference on Data Mining Workshops, pp. 816–821. IEEE (2010)
49. Wang, H., Ohsawa, Y.: iChance: a web-based innovation support system for business intelligence. *Int. J. Organ. Collect. Intell.* **2**, 48–61 (2011)
50. Wang, H., Ohsawa, Y.: Idea discovery: a scenario-based systematic approach for decision making in market innovation. *Expert Syst. Appl.* **40**, 429–438 (2013)
51. Fette, I., Melnikov, A.: The WebSocket Protocol. RFC 6455, December 2011. <https://www.rfc-editor.org/info/rfc6455>
52. Pimentel, V., Nickerson, B.G.: Communicating and displaying real-time data with websocket. *IEEE Internet Comput.* **16**, 45–53 (2012)
53. Chodorow, K., Mongo, D.B.: *The Definitive Guide: Powerful and Scalable Data Storage*. O'Reilly Media, Newton (2013)
54. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965)
55. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947)
56. Spearman, C.: The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904)
57. Ito, T., Shibata, D., Suzuki, S., Yamaguchi, N., Nishida, T., Hiraishi, K., Yoshino K.: Agent that facilitates crowd discussion. In: 7th ACM Collective Intelligence 2019, Carnegie Mellon University, Pittsburgh, USA, June 13–14 (2019)
58. Owens, C., Sharkey, S., Smithson, J., Hewis, E., Emmens, T., Ford, T., Jones, R.: Building an online community to promote communication and collaborative learning between health professionals and young people who self-harm: an exploratory study. *Health Expect.* **18**, 81–94 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.