2020/3/31
国立大学法人東京大学

戦略的イノベーション創造プログラム
Cross-ministerial Strategic Innovation Promotion Program

開発項目「データジャケットの国際標準化」平成３１年度～令和１年度成果報告書

目　次

１．研究開発の成果と達成状況

(1) 研究開発の成果

①　要約

データ市場における参加者間コミュニケーションの前提条件として、データジャケットを中心として、参加者による記述内容、提出後のコミュニケーション過程とそのレギュレーション（信頼性・創造性を高めるための制約）、その他約束事、データジャケットの利用事例、関連して有効となる技術を中心とする標準化文書を提示した。

②　本文（共同研究、再委託研究による成果を含む。）

内容梗概：本報告書は、データ市場における参加者間コミュニケーションの前提条件

(a)DJ が含む要素やその論理的位置づけ：コミュニケーションの前提知識となる

(b)コミュニケーションのプロセスと参加者同士の約束事

(c)その他の前提条件

を標準化対象とする。その具体的な情報の単位として参加者が提示しあう情報がデータジャケット（Data Jackets：以下 DJ と略）である。DJ とは、データの使途に関する主観的期待を込めた要約を人が記したテキスト情報である。既に Innovators Marketplace on Data Jackets なるデータ連携プロセスによりビジネスイノベーションの場において初期に準備する情報として大澤らが提案し、日本国内の企業群や経済産業省等の省庁への提供を中心としつつ台湾、インド、中国などの大学にも提供してきた。一連の手法は、データ連携のみならずデータを介した企業間の連携や学術的に新規性と社会貢献性を兼ねた成果を生み出してきた。

一方､DJ の記載方法を十分に理解しないユーザからは不十分な内容が DJ となって登録されたり、コミュニケーションのプロセスにおいて本質的なステップが欠落したりすることによって成果を生み出さない場合も散見されていた。かつ、海外の企業への提供がほぼ未踏に近い状況であった。本プロジェクトでは、これらの状況を克服し、DJ を世界の産学界において広く利用してもらうため、

① DJ の記述内容（一階述語論理によって、変数セット V, 述語セット P, 関数セット F の意味を含む情報をテキストで書いたもの）

② DJ 提出後のコミュニケーション過程のガイドライン（信頼性・創造性を高めるためのプロセスモデルで、一階述語論理でのアブダクションに沿う推論過程をなし、結果として pictogram の一種で表せるデータ取引のアーキテクチュアも生成する）

③ 上記を実現するためのレギュレーション（提示された内容の利用権、IMDJ の内外からみた信頼性を確保するための CROP ルール等）

④ DJ の使用事例

⑤ IMDJ におけるコミュニケーションと思考の支援技術

を中心とする標準化文書を提示した。また、ここで特に重大な点は、本件は国際標準化を目的としていることである。このため、成果文書は以下の制約を守るものである。

(1) 使用言語は、データ連携を行う者が出身国によらず読解可能な言語（英語）とする

(2) 一般的なユーザが用いることが前提となるため､標準化の内容は国内外のユーザにとって用いることのできるものでなければならない

以下､成果を英文にて付す(editage 社による英文校正済み)。本文書は、途中の報告会で合意を得たとおり大澤の責任において自由に利用し、国際的に広める活動を行ってゆく。

Data Jackets as a Global Technology for Data Free Flow with Trust

Yukio Ohsawa, Teruaki Hayashi, Gensei Ishimura (The Univ. of Tokyo)

1. Abstract

This document presents the prerequisites for the globalization of communication toward the creative marketplace of data. Data jackets (DJs) used here are human-made metadata for datasets, reflecting people's subjective or potential interests. DJs are used for creating proposals and knowledge to satisfy latent requirements in society, via the coupling of potentially connective datasets. Thus, DJs are the basis of data-mediated coupling places for participants from multiple domains in businesses/sciences. If the owner of a dataset cannot open the data to the public, he/she presents its DJ to the Innovators Marketplace on Data Jackets (IMDJ), because a DJ does not contain the corresponding data. By visualizing the relevance among DJs, participants in the market of data are enabled to discuss why, how, and which datasets should be coupled. Here, participants communicate to create ideas to couple/use/reuse data, or to plan future collaborators. The IMDJ can be extended by marrying living labs, where participants enhance their own sensitivity, to contexts in the open society. The required data can be searched for by using the tools developed on an for DJs (e.g. DJ store, or Variable Quest). These tools enable higher-order creativity, such as analogical inventions of data analysis methods. Here, we show the logics for the recent documentation of the strategic globalization of DJs.

2. The Goal of Data Jackets

2.1 The Effects of Data Jackets for Data Management with Co-creation

Exchanging data: DJs link the requirements and the data of the participants of the data market, according to the matching between requirements, solutions, and DJs stored in the DJ storage, via the sessions of the IMDJ.

Connecting data jackets: DJs are connected via common variables and common words, where their commonality is defined on computed similarity (e.g., inverse co-occurrence frequency, distance in the embedment space). These connections should be visualized to aid humans' creativity in data coupling scenarios.

Correcting data jackets: the owner of the dataset can correct the data by adding useful variables to the DJs, via creating data-use scenarios in IMDJ.

Collecting data and data jackets: if the data do not yet exist, it is possible to design and collect data corresponding to the collected DJs used. Also, it is possible to keep collecting DJs from participants in IMDJ sessions.

Discussions for using data: regarding the connections of the DJs above, participants can propose, evaluate, and improve ideas on how to use each dataset and its coupled data-metasets.

2.2 The merits of stakeholders

A. Data controller (control-right owner)
- is enabled to find someone (in B, D, or E) who may discover the utility of A's data
- is enabled to find someone (in B, C, D, or E) who may buy A's data

B. Data provider (not always identical to A)
- is enabled to find someone (in A) who may have B deal with their own data
- is enabled to find someone (in C) who may deal with the data that B deals with
- is enabled to find someone (in D or E) who may discover the data utility
- is enabled to find someone (in C, D, or E) who may buy and/or use the data
- is enabled to price the data via the discovery described above

C. Data enablers (including platform organizers)
- are enabled to find some (in A or B) who may have C deal with their data
- are enabled to find some (in D or E) who may discover data utilities or use data
- are enabled to find some who may buy data (E)

D. Business data scientists (including statistics/AI technicians)
- are enabled to find someone (in A or B) who may provide data
- are enabled to find someone (in E) who may show requirements for data analysis
- are enabled to find someone who (E) may buy the results of D's data analysis

E. Intermediaries (data-using businesses)
- are enabled to find someone (in A or B) who may provide data
- are enabled to find someone (in D) who may provide methods for data analysis
- are enabled to find someone (F) who may buy E's service/products

F. Final consumers (the customers of D)
- are enabled to learn about products that are useful in one's own life

For the stakeholder names above, see Deloitte LLP's "Open growth: Stimulating demand for open data in the UK" in 2012.

3. The Definitions of DJ and IMDJ

3.1 The definition of data jacket

The Intuitive explanation of a DJ, first introduced in [1], and extended in later publications and social activities, is as follows: a piece of digest information of a dataset, that does not open the content of the data, but includes the title, the abstract, and variables, which may represent the subjective expectation of the data owner or potential data users regarding the utility of the data.

The relationships between the variables should be logically modeled to enable knowledge generation by reasoning about data coupling. By this, we can clarify the boundary between knowledge to be derived by resolution, and by hypotheses or facts that cannot be.

The (formal) definition of DJ (appeared in [2], refined in [3]): a DJ_i for a dataset d, suffixed by i , is: by DJ_i (d) := $\{F_i(d), P_i(d), V_i(d), U_i(d),\}$ where

F_i (d): the set of functions defined on the variables in V_i(d)

P_i (d): the set of predicates that express relations among variables in V_i(d)

V_i (d): the set of variables in DJ_i(d)

U_i (d): the set of use cases of (d) based on the information in DJ_i(d).

In U_i(d), the hypothesis to be validated can be expressed by logical formulae, using the elements of F_i(d), P_i(d), and V_i(d), or corresponding natural language text. The meanings of each element of F_i(d), P_i(d), and V_i(d) are explained by human(s).

The basic idea of a DJ comes from what we and salesclerks do in the shopping stores of media such as movie DVDs, where only quite superficial pieces of information are presented on DVD jackets on shelves, for customers to read. The contents should be hidden to reduce the risk that the details may be copied and used freely by anyone, without them having to pay, or by rivals. Such a policy may seem to be in the spirit of data closure, but is a useful idea for the market of data, where each data owner takes part by filling in and disclosing a DJ showing the digest information (the title, the summary, and variables in the data) of existing data, or data to be collected in the future.

In a DJ, one writes out "variables" as long as they can be written and made public. This set of publicly available variables, and other information that the DJ provider wants to advertise, is called a “jacket”. For example, if the dataset is about nutrition in food, “protein” or “fat” may be a variable. If it is consumption data, “date and time,” etc., may be variable. Among the variables contained in a single data set, the set of variables that the data controller thinks can be disclosed becomes a public variable list. The DJ also includes a summary of the data, reflecting not only the objectively true composition of the data, but also the potential expectations of people in the market regarding the utilities of the data (see http://www.datajacket.org for entering DJs).

Over 4000 DJs have been collected so far by private companies, researchers, and governments. In contrast to data contents, DJs are easy to collect, and can describe links between data easily for ordinary users. As a result, the participants of the IMDJ, which are stated later, are communicating for/by buying/selling/sharing/exchanging data easier than if they were forced to open their data.

3.2 The Attributes of DJs

In order to include all of the elements in list 3.2, a DJ must be composed of the following attributes. These elements, however, can be fixed to fit the case of each user, as far as the elements listed in section 3.1 are included. The bracket before the colon shows the elements in that 3.1 correspond to each attribute, with * being used for exceptions.

The title of the dataset (*P, F*): “healthcare data including blood test, vital data such as weight, blood pressure, etc.” can be put as a predicate because, for example, **care(personal_ID, blood_pressure)** means the patient of the **personal_ID** takes care of his/her blood pressure. Furthermore, **blood_pressure** can be also represented as a function of (body’s) **weight** because blood pressure can be a result of weight care, i.e., **blood_pressure = care(weight)**.

Name and contact information of the DJ’s provider (*): these are saved confidentially

The provenance of the dataset (*): the name of the owner/URL /title of papers/sources of the dataset

The outline of the dataset (*P, F*, and possibly *V*): what the dataset is about, e.g., “human’s health data collected for daily healthcare regardless of whether the person has a disease or lives a normal

life.” This is also a predicate (*P*) and/or a function (*F*), in a similar sense to the title. The outline can be written subjectively, reflecting the expectations, which can be stated in the later attributes. As a result, predicates and functions are enriched e.g., a predicate **sell_low-fat_food_to(personal ID)** may be learned from this dataset.

The type of data (*V*, *): table/figure/text/real value/graph/image/movie/sound/other. “Time series” has been provided as a data type in DJ entry systems so far, because this is a frequent type of data, but it should be expressed by either time or date in the variable labels. Other types are not covered as variables, but are useful in the combinatorial use of datasets.

The variable labels of a dataset (*F*, *V*): the names of the variables i.e., attributes, are entered as variables separated by "|" e.g., date|time|temperature|place. Some variables here represent functions over other variables, e.g., **temperature** may mean **temperature(time, place)**.

A proposal(s) or use scenario (*U*, on *P*, *F*, and/or *V*): the representation of the process for inferring the expected outcome/typical analysis example(s) with data. This possibly includes tools for analysis, simulation, or other methods for using the data.

The expected outcome of the process and other anticipations for using (*P*, *F*, *U*): obtained knowledge, analysis results, the output of tools, and hypothetical causality between variables e.g., “find a word of high fitness to the user’s interest” as **fitness(word) > const**.

Other supplements (*): data or tools that can be used with the dataset. These are not covered by *F*, *P*, or *V*, but are useful for learning, validating, or inventing *F* and *P* on the values of *V*.

The method and/or cost for data collection or creation (*): e.g. traffic sensor (1k $ per car) /online articles (free).

The data sharing policy (*): With anyone/with specific people/under particular conditions/undecided/shareable for research purpose/not shareable/after purchasing the data for suitable prices, which may be preset or may be set via negotiation. Among these choices, the “specific people” can refer to the DJ provider if he/she is the controller of the data. In cases where the dataset does not contain personal data (not necessarily data that can be used to directly identify the data subjects), the data controller is responsible to ensure that the rights of the data subject are not violated by sharing the DJ or the corresponding dataset, in the way described in the data sharing policy described here. The controller (the data subject, or others such as the information bank, if trusted) or the processor (data enabler, data scientists, intermediaries, etc.) of the data should be the specific people declared here. Also, they are supposed to manage the data, in which should not be given to anyone disallowed by any condition according to the descriptions here. The description here is used only for planning the scenario to use the data in IMDJ (Section 4). The conditions for trading data should be written elsewhere in the data catalog.

Comments (*): Other impressions and requests about DJs, etc.

4. Innovators Marketplace on Data Jackets

4.1 The definition of Innovators Marketplace on Data Jackets

The IMDJ is a process to create and revise models of businesses and data-use scenarios for the businesses. There are three aspects: (A) data users' requirements, (B) DJs, and (C) proposals of uses for data. (A), (B), and (C) can be respectively regarded as the global, local, and the glocal models, defined as follows.

A global model: The representation of the desired causality. A participant playing the role of the user in the IMDJ expresses this as a requirement without restriction by knowledge specific to a domain represented by terms, variables, or predicates, as described in Section 3.

A local model: A set of the elements described in Section 3 in a domain, which may be connected via causal relations in the domain that are believed to be true by the provider of the DJ. In other words, a local model is a model included in a DJ.

A glocal model: The representation of a connection between local and global models, using elements in local models. Such a model should be created as a result of reasoning. In other words, a glocal model is a solution for satisfying the requirements of the global model.

4.2 The procedure of IMDJ

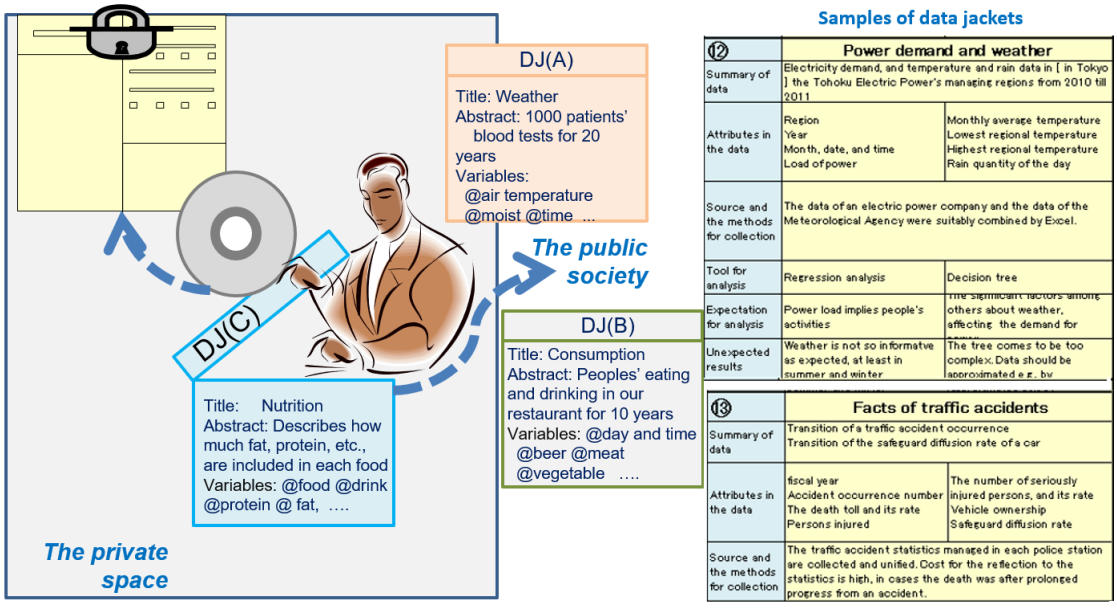


Figure 1. The owner of data may hide and lock one's own data somewhere. The digest information including the abstract and (not the values) of variables in the data are put on DJs and published. The provider of a DJ can be the controller of the corresponding data, but others can also provide DJs for datasets they do not hold.

IMDJ has been executed, including the three-step process below, to create and revise models of

businesses and data-use scenarios.

Step 1) The sharing of DJs for datasets.

The owner of a dataset may hide and lock their own data somewhere. Only the digest, i.e., the abstract and (not the values) of variables in the dataset are put on a DJ and published, as shown in Figure 1.

Step 2) The shared DJs are linked via features (variables, words, etc).

One may also explicitly declare links to other data (using Resource Description Framework, i.e. RDF). Both links should be visualized in a graph by KeyGraph® in the implementations so far, as shown in Figure 2. KeyGraph here visualizes the names of variables and words that form connections between DJs (the cards in Figures 1 and 2). This allows the participants to intuitively understand the roles of words (i.e., variables or others) from their positions in the map.

Note that DJs should not be shown as in a sheer data catalog, where data and metadata are registered and shown in a list, because the IMDJ should be an environment where stakeholders can discuss the expected uses of combined datasets (as represented by DJs). In Step 2 here, a map can be created by linking the DJs through words that represent concepts and variables. In this map, DJs can be linked via the same variables or variables of the same meaning (e.g., “date” and “yy yy/mm/dd”), or via words that are common to multiple DJs. Thus, the connections among DJs represent the commonly expected utilities of various datasets among the expectations of DJ providers, or the possibility to connect the data represented by the DJs. This is different from data catalogs, where metadata and the RDF declaration of the links among data in Linked Open Data (LOD) are registered as objective truths.

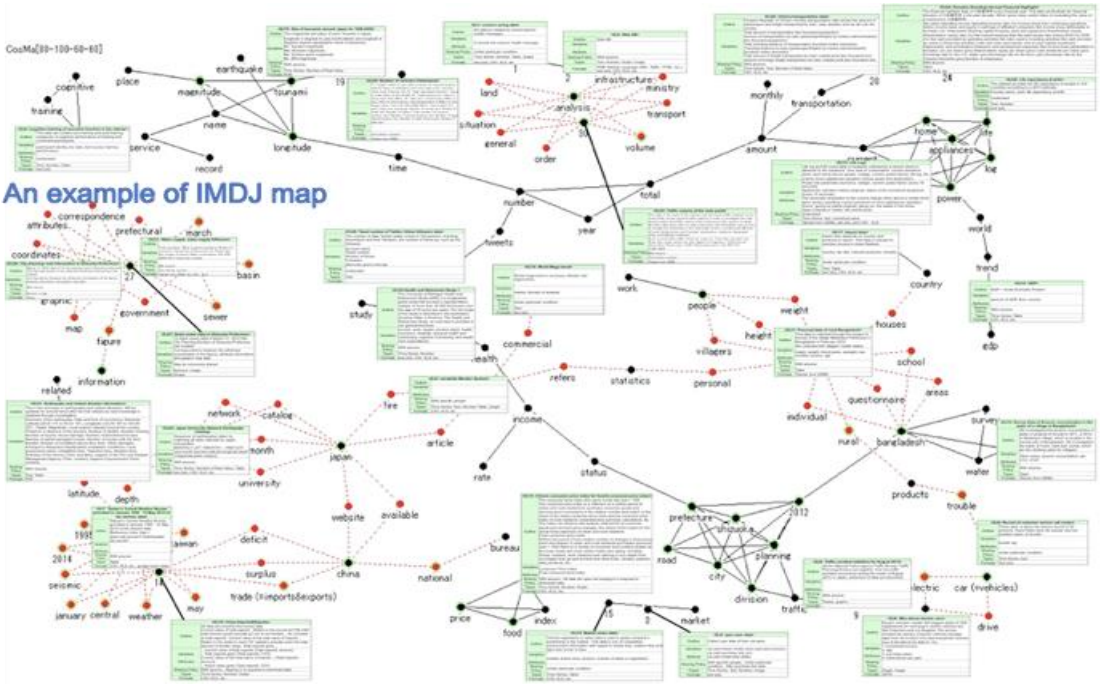


Figure 2. The visualized connections of DJs, representing the potential connectivities of data

Step 3) Communication for trading or pricing data.
The participants' businesses are potentially (because participants are just planning the businesses) connected via DJs, more potentially (because data may be confidential) via data. Once they are connected via DJs, they can plan businesses with more concrete connections (because now data can be exchanged) via data. See Figures 3 (a) and (b).

In Step 3, requirements—what they need in daily lives or businesses—are shown by the potential users of data, or by the final consumers of the services or products to be created by using data. Ideas for using data (for analysis, machine learning, simulation, or just to be read) that satisfy these requirements are called solutions. Solutions can be proposed by anyone in the market of data by connecting DJs.

In Figure 3, the doctor in the right-hand makes a request for knowledge that would be useful for predicting sudden changes in health conditions. This doctor is a potential user who is showing their requirements. The negotiation between a data provider, an analyst who plays the role of data scientist here, and a user, as the final consumer here, may reach an agreement to share data, based on the analyst's solution. The analyst here may be regarded as a solution provider, as well as the user of data. The solution is here to combine datasets, one about weather, another about liquor consumption, an other about blood tests that is owned by the doctor herself, to discover knowledge such as "one tends to drink beer if the air temperature is higher by three degrees than the average of the 10 days before, and gets one's g-GTP increased." If the idea matches a user's (i.e., the doctor's) requirement, the idea and the data represented by the DJs used in the idea will be priced on the negotiation. Here, a person who wants to add links or causalities coming from one's own thoughts may do so. Providers can also explicitly declare links between data, as in LOD. However, this is not required before attending the IMDJ, because the point of the IMDJ is to discover latent links that are not easy to declare in advance.

4.3 The Four Layers in IMDJ

As a quick summary, the four layers composing the market of data using IMDJ are described here:

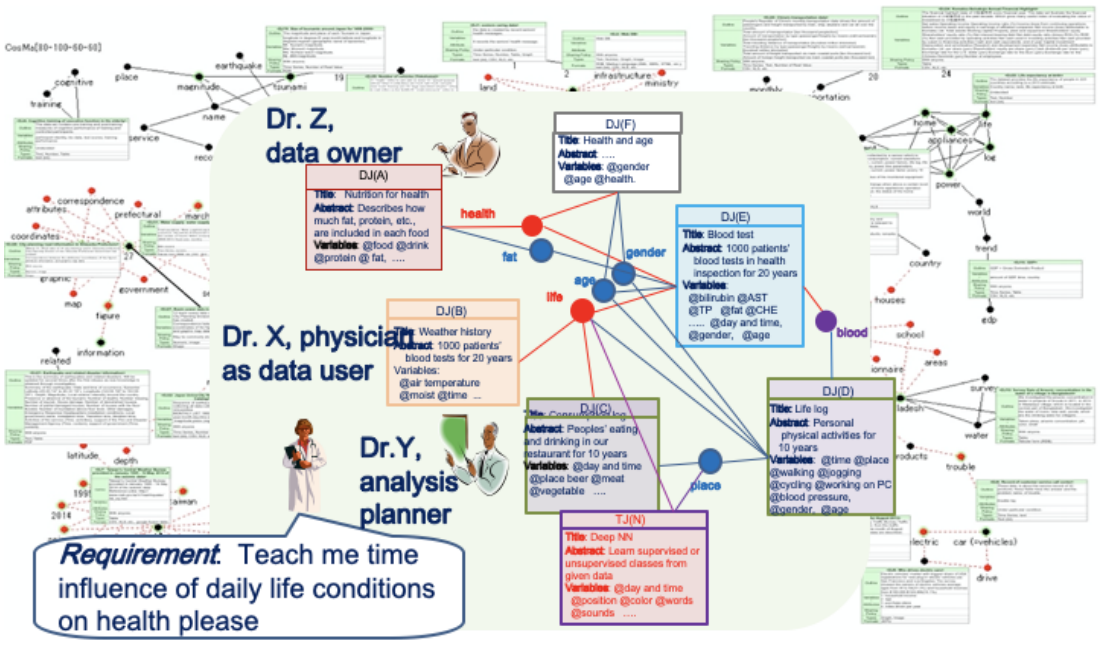
Layer 4: Requirements given by potential users of final consumers of data, who may not have the vocabulary by which to talk or think about the data.

Layer 3: Solutions given by data analysts or planner(s) of data analysis. These participants can be called data scientists, who should consider satisfying the requirements of the users or the final consumers. Layers 3 and 4 meet in the communications detailed in Step 3.

Layer 2: DJs provided by any participant in the market of data. The summaries of the datasets and the names of variables should be stated subjectively, reflexing the DJ providers' expectations about the utility of the data.

Layer 1: The datasets. They do not have to really exist, because one can first exchange DJs and then create and collect the content of a new dataset corresponding to the DJ to which a high price has been assigned, or via which a solution has been highly-priced (i.e., for which a high utility value has been found in the IMDJ).

(a)



(b)

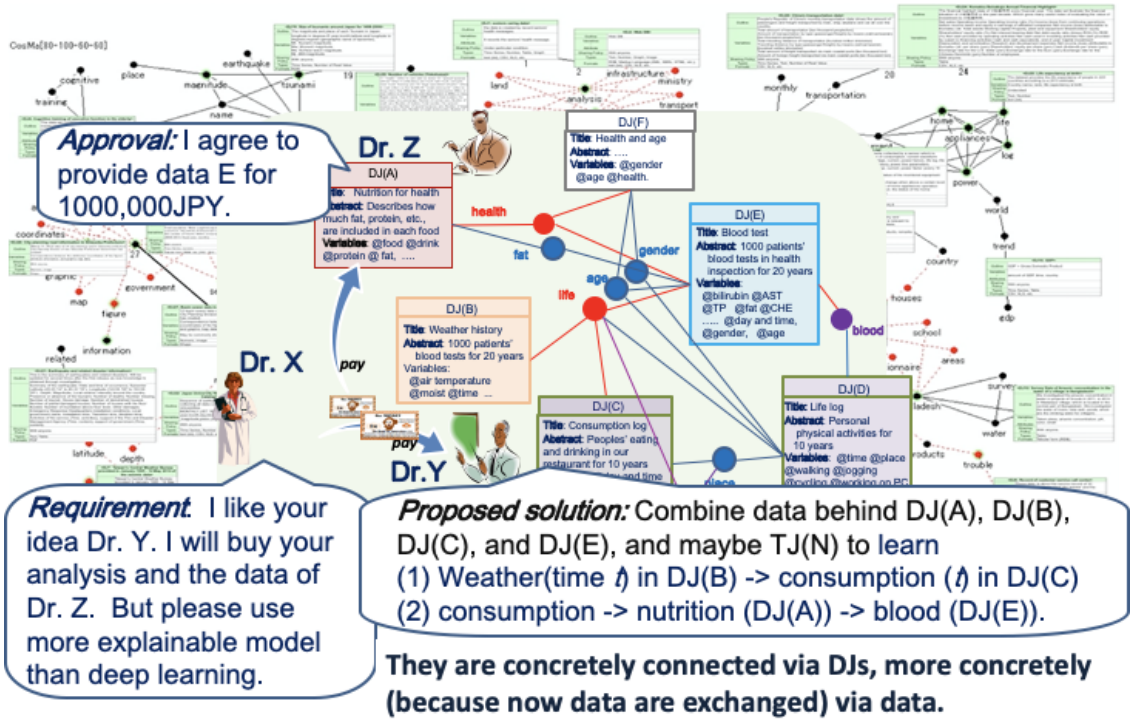


Figure 3. An example of communication in IMDJ. A requirement of the user, who is a final consumer of data in Section 2.2, is shown in (a), and a solution is presented by the data scientist in (b) that is then evaluated by the user. The user may also price the data. The data scientist can do this if he buys the data for satisfying the requirement.

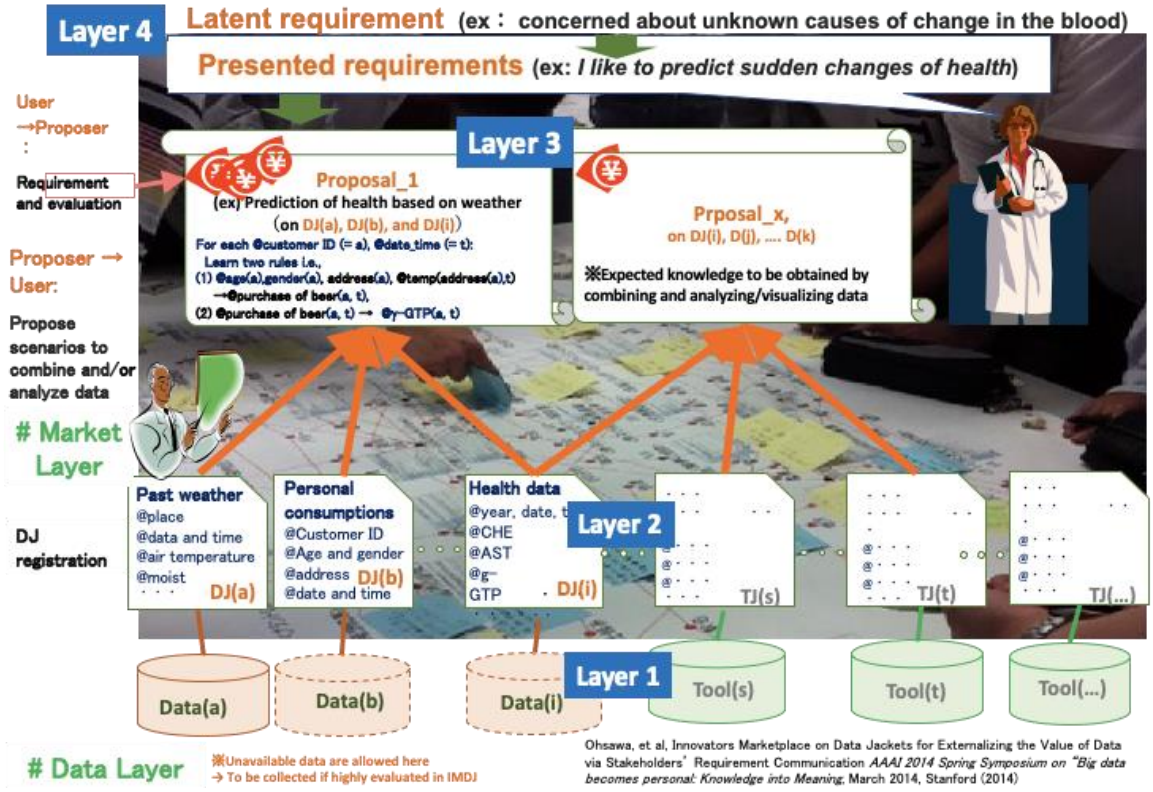


Figure 4. The four layers of IMDJ

4.4 The connection of DJs in IMDJ, corresponding to First Order Logic

As shown in Figure 5, the requirement-driven communication in the IMDJ is processed by connecting DJs. For example, suppose the global model *G* is the requirement to know the influence of weather on health, represented as "**health** ← **weather**." By relating **health** to **g-GTP_high(person ID, date)** and weather to **hot(date)**, *G* corresponds to *G'* in clause (1). A variable **person_ID** bound by an existential quantifier (∃) is ignored in the illustration in Figure 5.

$$G'\text{: } \exists \text{person\_ID } \{\text{up\_g-GTP(person\_ID, date)} \leftarrow \text{hot(date)}\} \quad (1)$$

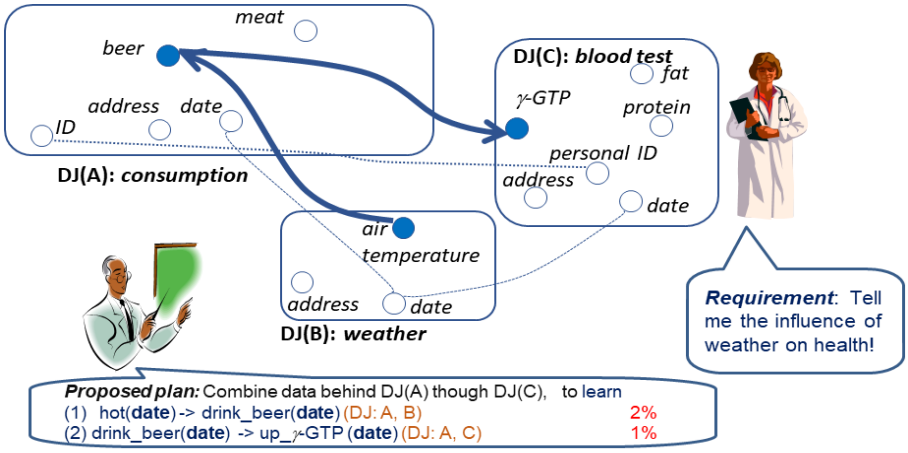
G' can be derived by the combination of clauses (2) and (3), which form a solution.

$$\text{up\_g-GTP(person ID, date)} \leftarrow \text{drink\_beer(person ID, date)} \quad (2)$$

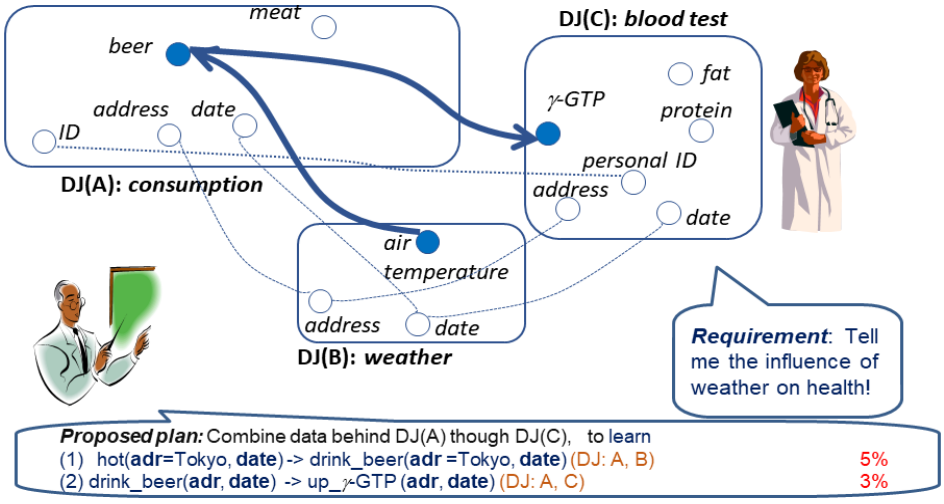
$$\exists \text{person ID } \{\text{drink\_beer(person ID, date)} \leftarrow \text{hot(date)}\} \quad (3)$$

Here, **hot(date)** and **up_g-GTP(person ID, date)** mean **air_tempr(date) - air_ tempr (date-10) > *a*** [deg] and **gGTP(date) - gGTP(date-10) > *b*** [u/l] respectively, for constants ***a*** and ***b***. The values of ***a*** and ***b*** are obtained from datasets represented by DJs. For example, ***a*** is obtained from data B, represented by DJ(B) for the variables in question {date, address, air temperature, etc}. A function such as **air_tempr(date)** is supported by this dataset, and a predicate such as **hot** is defined over **air_tempr** and **date**. Thus the logical part of the reasoning in and between the local models i.e., DJs, can be expressed in the first order logic (FOL).

(a)



(b)



(c)

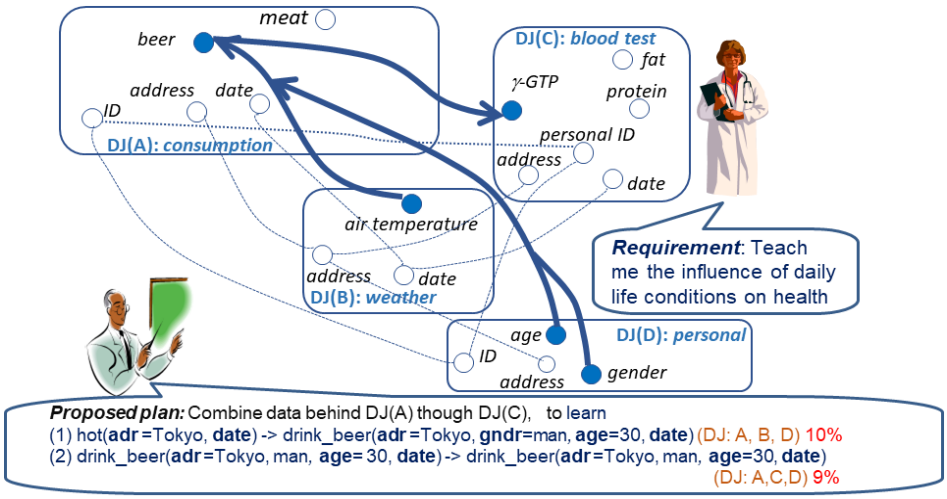


Figure 5. The abductive reasoning steps in the communication in IMDJ

If the obtained theory, *T*, i.e., the solution as a glocal model, is not satisfactory (here the low confidence percentage in the colored letters) for explaining the goal (i.e. it does not satisfy the

requirement), other variables such as address (**adr** in Figure 5) can be additionally used, as shown in Figure 5(b). Furthermore, new DJs may be added to obtain a satisfactory *T* and evaluate it, using data corresponding to the DJs referred to in *T*. This enables participants in the IMDJ to design new datasets by proposing the addition of new DJs (by filling and attaching a red post-it).

As in this figure, the negotiation between a provider, an analyst, and a user may reach an agreement to share data, based on the analyst's idea in the IMDJ. The idea is here to combine datasets, one about (A) food consumption including beer, (B) weather, (C) health, and (D) personal information, to discover knowledge such as "one tends to drink beer too much and get an increased g-GTP if the air temperature is higher by three degrees than the average of 10 days before." If this knowledge may be useful for the doctor who requested the knowledge to predict sudden changes in health conditions, then this idea and the datasets to be used to validate this hypothesis will be imported, according to the conditions (the price of data, the contract of collaboration, etc) agreed in the negotiation.

The FOL-based structure of global (goal), glocal (solution), and local (DJs), once illustrated, can be used to draw a structure of interactions among stakeholders, as shown in Figure 6. The exchange of data (solid lines in Figure 6), and of money (dotted) in the opposite direction to the data, are shown by arrows. The personal datasets (frames with red shadows) and others (frames with blue shadows) are not explicitly classified in advance. They can be distinguished in this figure however because personal variables (the set of IDs, address, name, etc.) are in some of the DJs. The timelines of the changes in arrows (shown by *t* here) were obtained in the abductive reasoning process illustrated in Figure 5. Thus, the obtained diagram can be used as a basis for the pictogram created in this project (SIP2 for DFFT).

The data sharing policy can be also reconsidered with this visualization, as stated at the end of Section 3.2. The types of participants in the IMDJ in Section 3.2 are here used for planning the scenario to use the data in the IMDJ, including the reconsideration of the policy. To enable such reconsideration, which tends to be hard to be explicitly computed because of latent human relationships, a network such as that in Figure 6 should be drawn manually, rather than by relying on tools for automatic visualization. To date (2014-2019), This drawing has been included explicitly as a step in Action Planning in the next section. It is applied in the public and the private sections of data marketplaces, but essentially it is the product of all of the steps of abductive reasoning in the communication in the IMDJ, as seen from the structural similarity between Figure 5(c) and Figure 6.

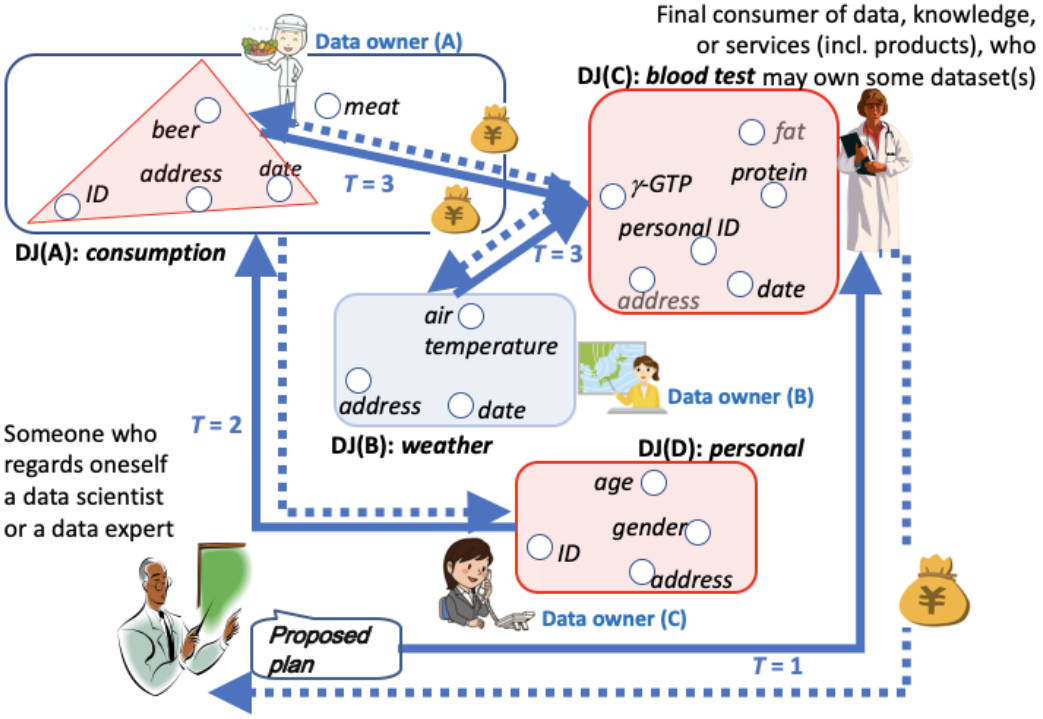


Figure 6. The connection of DJs via the trading actions among participants in IMDJ

One important lesson from this section is that the communication in IMDJ, which is organized through frank and easy dialogue, should ideally follow logical abductive reasoning. For this reason, it is clear from this section that each DJ should include logically connected contents (*V*, *P*, and *F* in Section 3.3). Thus, the participants must check if the DJ they provide includes the variables (corresponding to *V*), the expected relationships between variables (corresponding to *P*), and the expected effects of variables on businesses or human lives (corresponding to *F*). This does not mean that a DJ should be entered using FOL, like a code for an automatic inference machine. Therefore, a DJ can be composed of easy natural language, as long as the above condition is satisfied.

4.5 Action planning: a post-process of IMDJ

Action planning (AP [4]) should be used as a post-process of IMDJ for realizing the solutions proposed in IMDJ, as shown in Figure 7. This should be followed by real actions where real resources (including data) can also be collected.

Definition: The method to support creating (1) analysis scenarios and (2) strategic business scenarios by externalizing and relating elements for the realization of ideas.

Goal: To refine the logical relationship among the elements in Section 3.1, in the realization of solutions. As subgoals, the identifications of risks (including human factors such as people with interest conflicts), new opportunities (incl. human factors such as potential customers and resource suppliers), and the elimination of blind spots that arise when making decisions, are aimed to create practical scenarios.

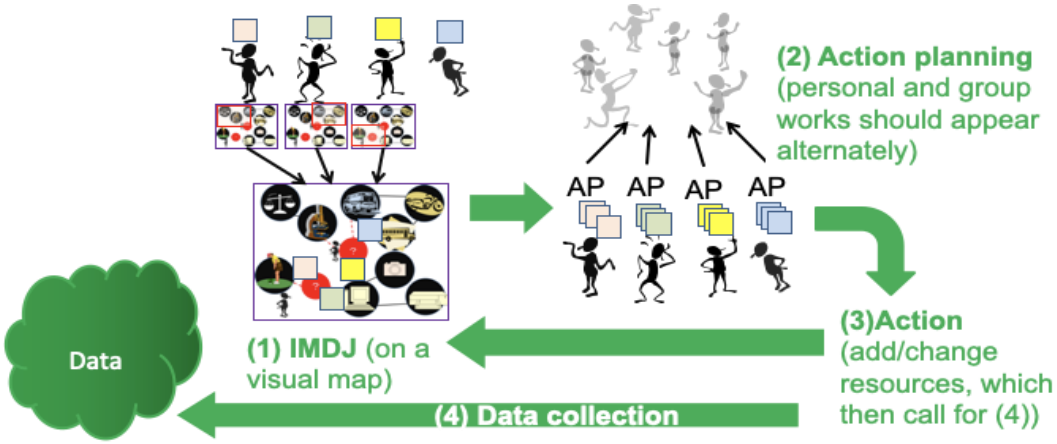


Figure 7. The process of IMDJ: connection to the post-process

AP is composed of steps for refining the ideas created in the IMDJ on a map into a more concrete and feasible scenario. This is achieved by completing an AP sheet, as shown in Figure 8.

Step 1. Requirements Analysis (AP1): From the solution created in the IMDJ, by considering the background factors of the externalized requirement, the potential requirement that is the requirement that should be truly solved can be obtained. Inventing a solution that meets the potential requirements leads to an essential problem-solving solution.

Step 2. Element externalization (AP2): The stakeholders (targets, etc.), resources (data, time, etc.), and 4P necessary to guide the solutions studied in AP1 are externalized here.

Step 3. Element and action scenarization: By grouping the relationships between the elements examined in AP2 according to certain rules (realization process, revenue model, system flow, etc.), the contradiction between elements is resolved, and the existence of missing elements is clarified and reconsidered when scenarizing the idea.

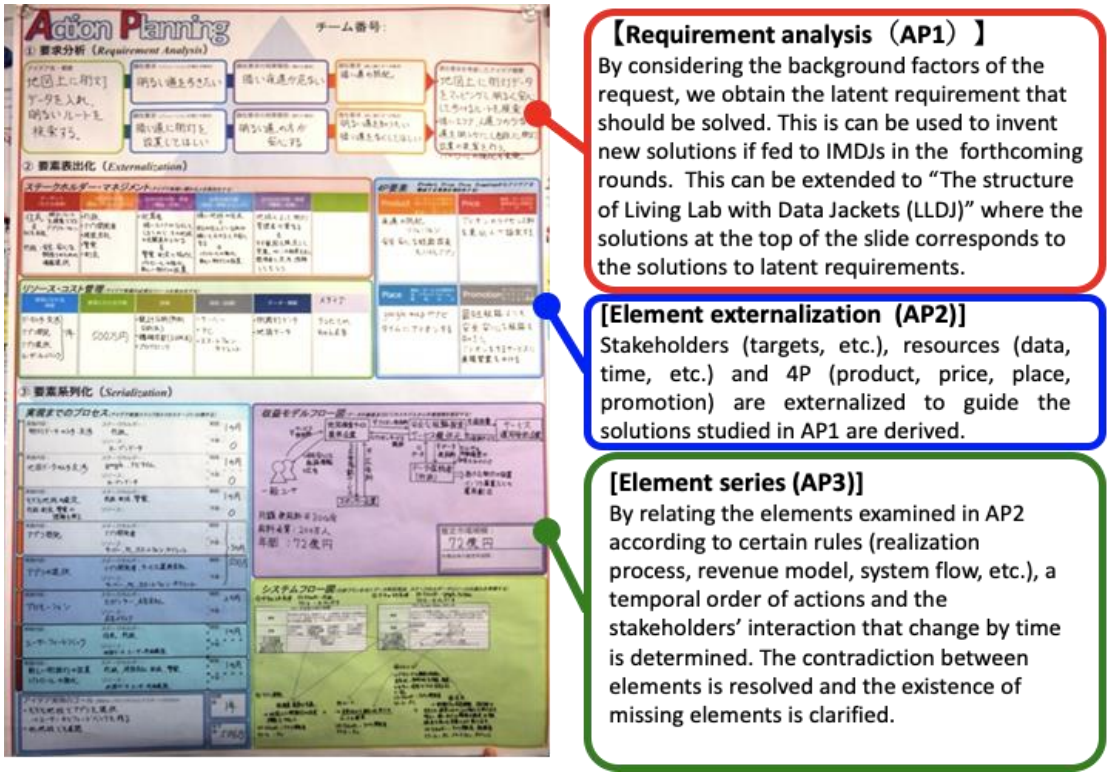


Figure 8. Action Planning (AP) by completing an AP sheet

By completing an AP sheet, participants can delve more deeply into the latent requests of other participants. This allows them to create a scenario of analysis to business realization, whereby those requests can be satisfied. As a result, it becomes possible to examine the stakeholders, resources, costs, and time allocation, which had not been previously considered. Moreover, a fundamental revision of the original ideas is enabled based on the results of noticing these latent requests anew.

- Analysis flow: Drawing a flow chart of the system for achieving the solution, and composing an overall picture of the process and system to derive the expected analysis results from the data use.
- Business Flow: Drawing a flow chart of the earnings model of solution achievement, and estimating the market scale of the services to be provided, while considering the budget and the stakeholders.

AP can invite new participants who were not involved in the previous step of the IMDJ, to execute in a way like a living lab. On the way, an action plan is visually represented in the scenario chart, as shown in Figure 6 (or as in the manually drawn example in Figure 9 and 10, preceded by its left-hand frame, which shows the analysis plan). This reflects the FOL-based abductive reasoning.

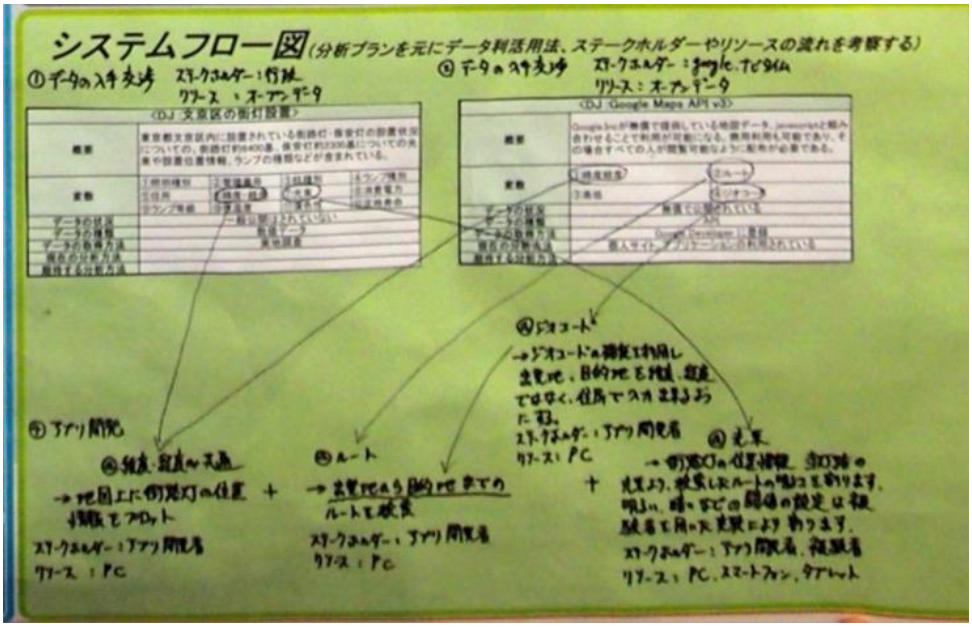


Figure 9. An example of data-use scenario (use case) charts: the analysis flow char

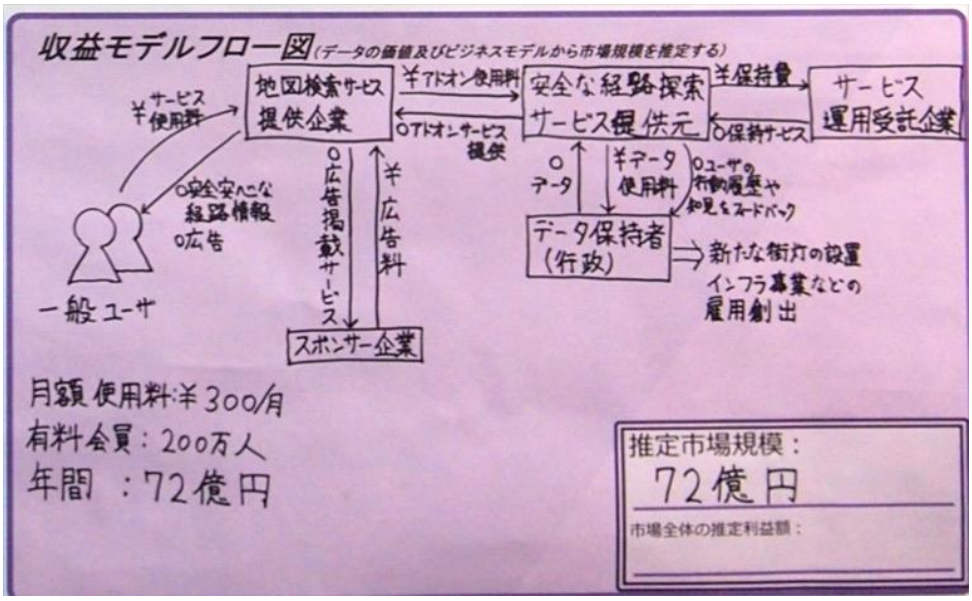


Figure 10. An example of data-use scenario (use case) charts: the business flow char

4.6 The structure of the Living Lab with Data Jackets (LLDJ)

We are integrating the IMDJ with Living Lab (LL [5, 6, 7]) to both deepen and widen the requirements, and also to tighten the trust of the information and the participants in the IMDJ. Furthermore, we can establish trust in activities from external society by integrating the IMDJ with LL. Here we focus on the effects of LL that contribute to Organizational Citizenship Behavior (OCB). OCB has been defined as individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization [8]. Contextual Performance (non-task related work behaviors and activities contributing to the social and psychological aspects of the organization [9]) and Extra-role

Behavior (behavior attempting to benefit the organization beyond existing role expectations [10]), are linked to OCB and are in the target of LL pointed out by Sae Kondo [3]. The mediating effect of political skills includes the ability to sense the influence of individuals on others and the intentions of others [11]. By introducing LL with this new focus, we expect to raise the sensitivity of participants to deep and wide social requirements, to widen and deepen their trust in the IMDJ.

The obtained process follows four simple steps:

Step 0) Set up the topic *Z*.

Step 1) Open the LL relevant to ***Z***. This entails communicating requirements and solutions for the requirements, considering the real situations about ***Z***, by visiting real sites and/or inviting real workers or habitants.

Step 2) Make the set **RLL(*Z*)** of the requirements obtained in Step 1 (KeyGraph can be used here to show the structure of the requirements).

Step 3) Search for DJs, using the words in the requirements in **RLL(*Z*)** as the query, and start with IMDJ. Collect the participants in IMDJ from real sites relevant to these DJs.

Step 4) The solution(s) are executed as real actions at the working site of the IMDJ participants. They are also added together with the requirements presented in Step 3, and then returned to Step 1. The participants relevant to these added elements are called.

For externalizing deeper requirements, the communication is regulated by setting a rule that each solution in Step 1 and also in Step 3 must be proposed, after asking a deep reasoning question i.e., “why is do you require it?” on the limit-handling (Stanford group, etc [12]). This process is shown in Figure 11. The LLDJ contributes to solutions for more general social issues than the IMDJ.

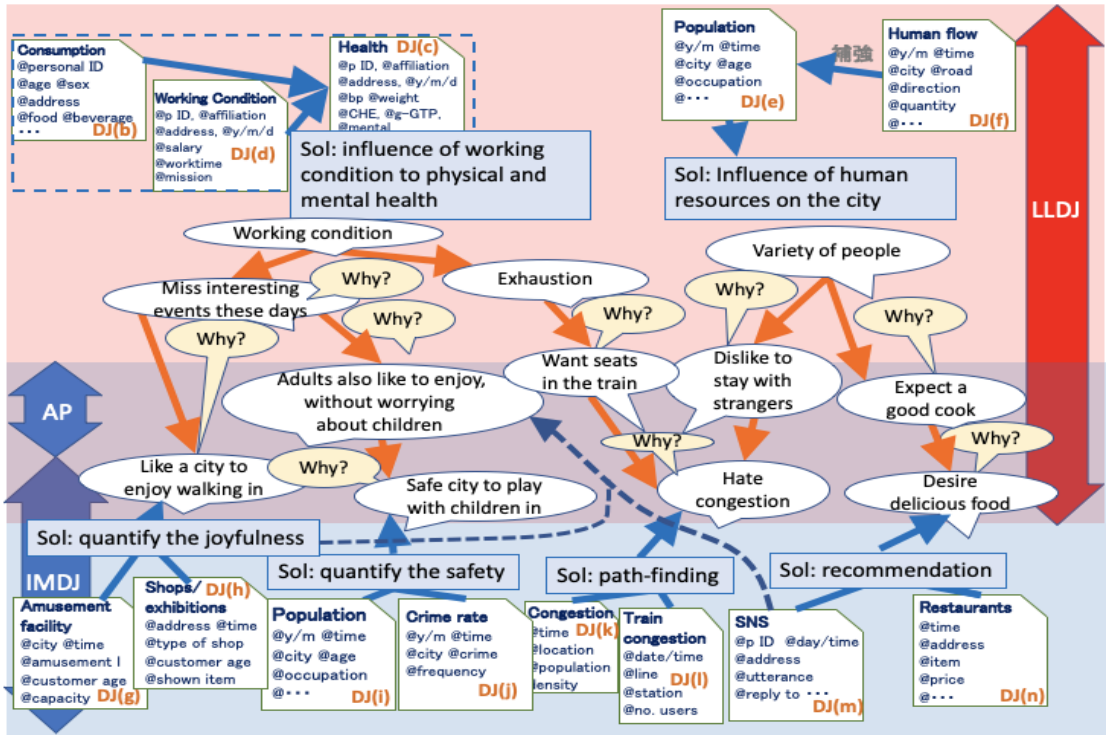


Figure 11. Living labs with deep reasoning questions combined with DJs

5. Trust and Rights

5.1 Basic principles regarding rights and reliability, as the foundation of "trust"

The coupling of data creates new utility for said data. This may cause a change in the dimensions of the information, as in the case above. As a result, new participants are always expected to join with the progress of the IMDJ. Here, trust becomes a noteworthy checkpoint. We assume that trust will be built in(and of) the IMDJ if the following conditions hold. By this, society will ensure that the IMDJ follows laws, social norms, and ethics.

The above establishment can be hardly made without visualizing or understanding the data flow. For example, who can be stakeholders involved in business, and whose opinions should be considered for executing a real business? The IMDJ starts without inviting all of the stakeholders, but instead starts from a few members. This number may be too small to answer a given question. One advantage of the IMDJ is that more members can be added afterwards via interaction with society. In this sense, the social acknowledgement of the trustworthiness of the IMDJ will play an essential role.

1. Rights of DJ's providers in the IMDJ (conditions promised to providers)
   - ○ DJ's providers are assured that they can control how their DJ should be used.
     i. The information contained in the DJ is only shared among the defined stakeholders. Participants in the IMDJ are restricted to the defined stakeholders. This definition is explained to the providers in advance. If no definition is made, this means that the DJ is open to the public. Thus, DJs are open by default, but can be restricted by the provider if desired.
     ii. The provider can request the withdrawal of their DJ and/or their datasets. They can also restrict their circulation and/or use whenever they want, after an IMDJ session has been held. Their request is reported immediately to the related stakeholders, and is executed as appropriately as reasonably practicable.
2. Right of participants in the IMDJ (conditions promised to participants)
   - ○ The participants in the IMDJ, and only they, will be able to use any ideas of solutions for given problems proposed in the IMDJ, under the CROP principle ("Controlled Reach of Presentation" as explained in the next page).
   - ○ The participants in the IMDJ can express their opinions freely in the IMDJ, except for illegal or unethical opinions.
3. Reliability of DJs as treated in the IMDJ (conditions promised to DJ's users)
   - ○ DJ's users are assured that the DJ that they are using is sufficiently reliable.
     i. Data sets represented by DJs are authentic (not falsified, fabricated, nor stolen).

ii. Any unethical intention, such as the exposure of confidential information, is prohibited.
iii. "Reliability" here does not necessarily mean "accuracy," nor "objectivity," because a DJ can also be described subjectively. It instead refers to a broader concept containing a reliable ecosystem among various actors such as owners, providers, users, and analysts.

4. Reliability of the IMDJ in society (conditions promised to the society)
   - The society is assured that the IMDJ definitely follows laws, social norms, and ethics.
   - The above concept is named "ELSI (Ethical, Legal, and Social Implications)." This is almost always referred to in the area of research and development within science and technology today [13]. At first, in 1988, the concept was proposed by Nobel Laureate Watson, J. at the press conference regarding the Human Genome Project (HGP) [14].

5.2 The CROP principle in the IMDJ

Based on all the conditions for trust in (and of) the IMDJ (see 5.1), all of the rights to use and realize ideas created in the IMDJ are right-protected, under the principle of the "Controlled Reach of Presentation (CROP)".

1. Each participant, and each participant only, has the right to use any idea for any solutions, and to use related DJs for any problems proposed by the others in the IMDJ freely.
2. This right is assured only if it is exercised fairly and ethically, and it is allowed by all of the providers of the referenced DJ.
   - (see 5.1.2)
3. Any proposer of any idea in the IMDJ can request the restriction of the rights of the others (1) at any time later, and this should be realized as far as possible.
   - (see 5.1.1)
4. However, the proposers have to accept that the use and propagation of their idea might not be able to be interrupted by their ex-post facto requests.
   - (see 5.1.1)

5.3 "Data-Trust Ecosystem" of the IMDJ in society

The IMDJ does not stand alone, but is embedded in society as a "***Data-Trust Ecosystem***," as follows.

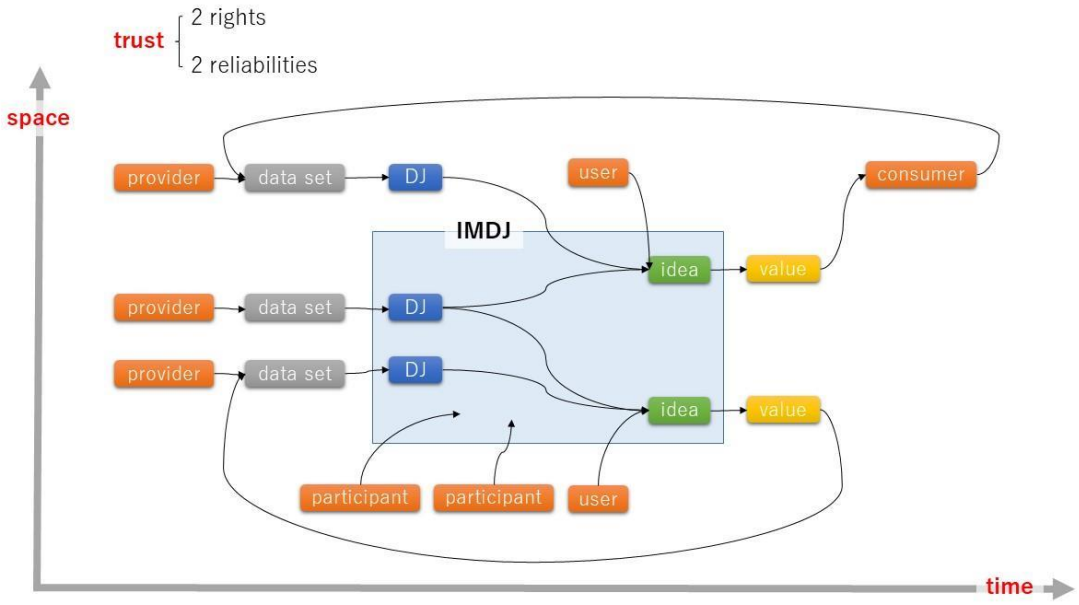


Figure 12. "Data-Trust Ecosystem" of the IMDJ in society described in the Spatio-temporal chart

- Ecosystem
  - The IMDJ is not independent, nor is it isolated from society. It is deeply embedded in the context of the Data-Trust Ecosystem.
- Trust
  - "Trust" is the gameboard of any game. It is a platform for any kind of communication around data. This means that if the trust is broken or lost, we can never continue any game, however valuable it is.
  - Trust in the IMDJ is developed by the accumulation of IMDJ sessions. The process is strongly recursive, and is of self-reference.
  - Trust is not only granted to IMDJ sessions, but also through other factors, such as ideas, consumers, a combination of several players and/or processes as subsystems, or the whole system.
- Spatial aspect
  - The axis "space" expresses the diversity and interaction of players involved in the data-related activities. This implies that there exists an ecosystem, which is composed of data, data sets, DJ, IMDJ, problems, solutions, ideas, values, various players, various processes, and trust.
  - Not only inside the IMDJ sessions but also outside of them, there are many players that interact mutually. There are many "implicit" DJs, participants, users, problems, solutions, and ideas, outside of IMDJ sessions.

- ○ A player does not always have the same role. A consumer can also be a provider at the same time. A provider can also be a participant of the IMDJ, or a user of ideas from the IMDJ.

- Temporal aspect
  - ○ The axis “time” expresses the temporal development of the data-trust ecosystem. It describes the history (hysteresis) of data communication (transaction), which itself provides the basis of trust in society regarding the data-trust ecosystem.
- Spatio-Temporal aspect - all things are in a state of flux
  - ○ Consumers who use values created by the IMDJ produce new data sets via their consumption, which again becomes a DJ for a future IMDJ.
  - ○ In addition, values created by the IMDJ themselves also produce new data sets, to prepare the DJ for a future IMDJ.

Such a comprehensive system can be envisioned as the Data-Trust Ecosystem.

6. Tools for empowering the IMDJ

The IMDJ is a powerful method for empowering the data marketplace, by activating the creativity of participants. This is the very central function of the marketplace, which tends to be discounted if the data marketplace is confused with the concept of a “data trading marketplace.” In the latter, only the data themselves are monetized, or in the case of “data free flow,” all data are provided free to anyone without evaluating the utility values of data. However, creativity is not always easy to activate, if we only regulate the process of communication using the regulated DJs, following the standard presented in this report. In this section, we demonstrate tools for aiding this activation of creativity.

6.1 DJ Store

The DJ Store is a retrieval system for data. It is used not only to list the information in a DJ, but also to reuse discussions on past IMDJ workshops as knowledge for data utilization. The DJ Store can find potential data related to users’ interests, by bridging the knowledge gaps between stakeholders (data owners, providers, or consumers) [15].

In the experiments, we found that the data retrieved from queries related to the requirements or the solutions were larger than data from the queries related to the data. This may potentially be useful for considering the data usage.

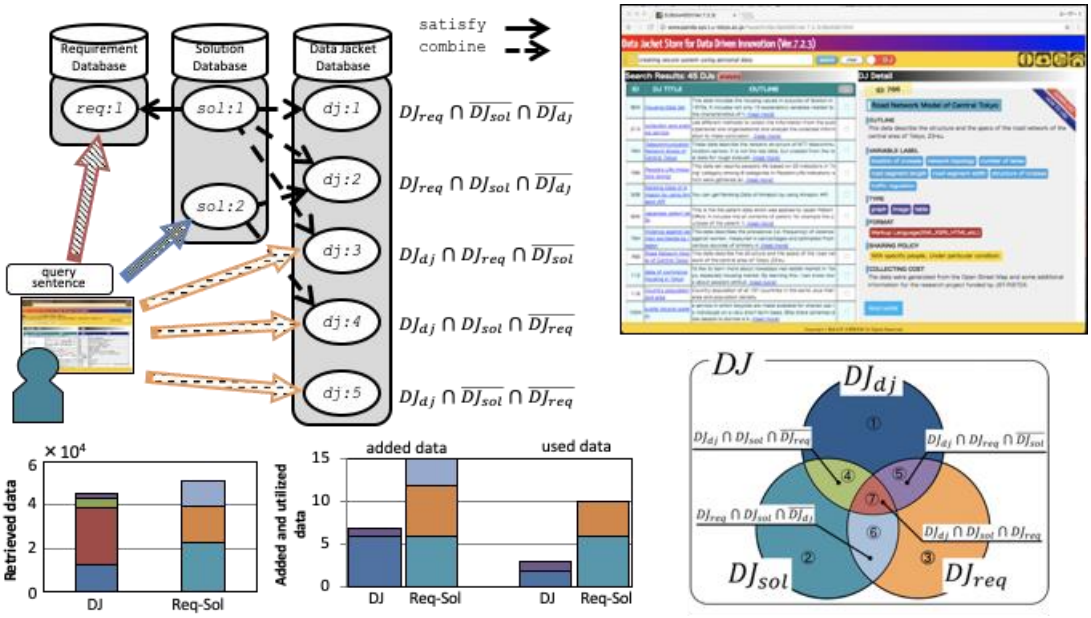
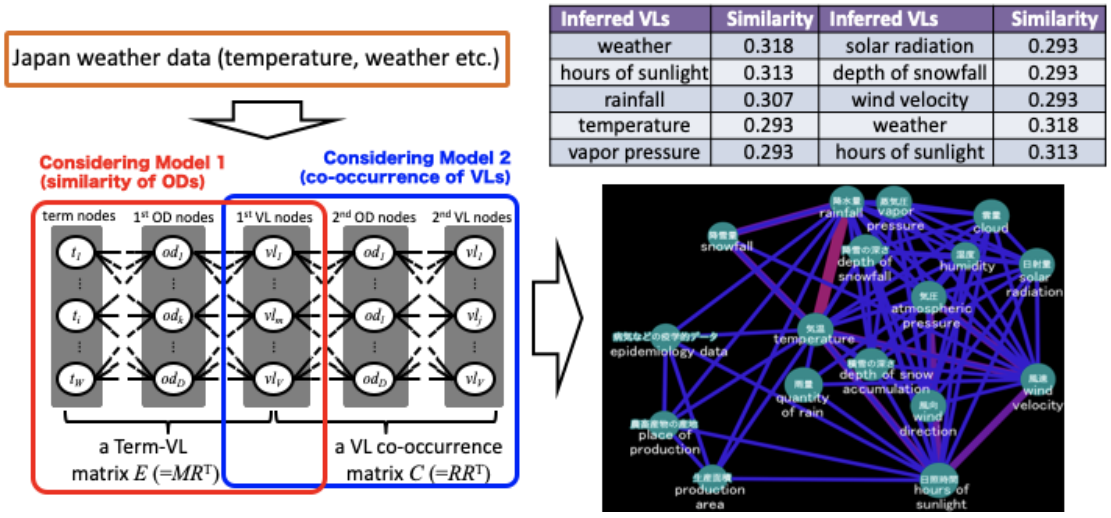


Figure 13. The knowledge base of data utilization, and the retrieval process.

6.2 Variable Quest

VARIABLE QUEST (VQ) is a matrix-based inference method for variable labels (VLs), which are the names/meanings of variables in DJs [16]. VQ infers related VLs from the free text queries whose VLs are missing or unknown. It focuses on the similarity of outlines of data, and the co-occurrence of VLs.



Inferred VLs	Similarity	Inferred VLs	Similarity
weather	0.318	solar radiation	0.293
hours of sunlight	0.313	depth of snowfall	0.293
rainfall	0.307	wind velocity	0.293
temperature	0.293	weather	0.318
vapor pressure	0.293	hours of sunlight	0.313

Figure 14. The inferring process of variable labels using VQ and its result

6.3 Data Matching System: Treasuring Every Encounter of Data Affairs

A matching system that externalizes the needs of the data users and visualizes the potential linkages

with the data of the data holders is required. This matching system is Treasuring Every Encounter of Data Affairs (TEEDA). Following improvements in the web platforms used for data exchange and trading, there are increasing opportunities for users to obtain data from data holders and providers. However, access is limited to unilateral information provision from the data holders, and there are few means for learning what kinds of data the users wish, or for what purpose. TEEDA is a matching system that externalizes the needs of the data users and visualizes the potential linkages with the data of the data holders.

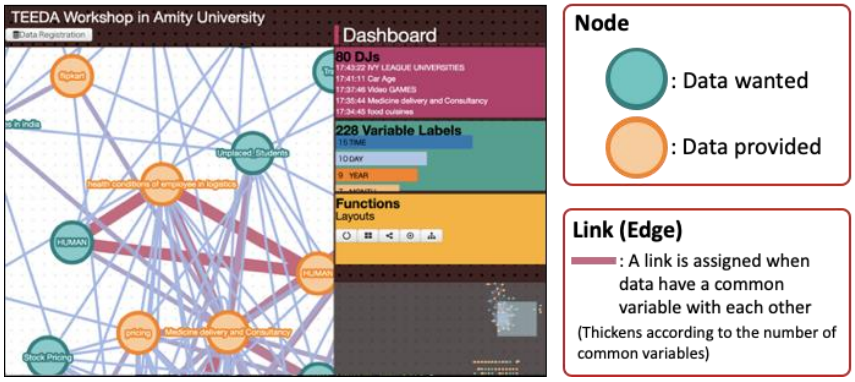


Figure 15. The process to collect and match the data requests and the providable data, and the interface of TEEDA.

6.4 Human Resource Finder

Even when considering the same stakeholders, their relationships with a business might change, depending on the context of the business. Human Resource Finder (HRF) is the recommender system for stakeholders and their relationships with a data business scenario [17]. The HRF is a system that recommends stakeholders and their relationships with a data business scenario. Using the databases of scenarios created in AP and DBpedia (information from Wikipedia that has been structured with RDF), HRF can calculate the relationships between stakeholders and their relationship likelihoods, from a scenario whose stakeholders are unknown.

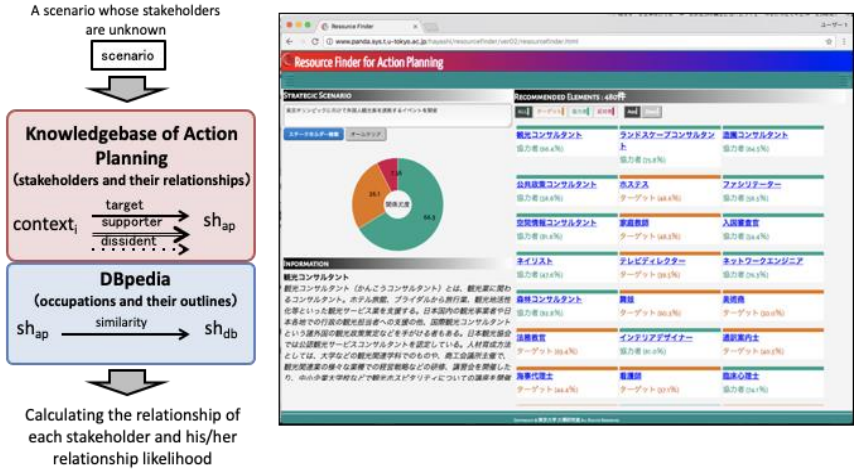


Figure 16. The process to estimate the potential stakeholders with the relationship likelihoods.

6.5 Web-based IMDJ

Web-based IMDJ supports communication among stakeholders in the market of data. By

connecting with the database of the DJ Store, Web-based IMDJ reduces the burden of organizing the discussion for cross-disciplinary data collaborations. Furthermore, Web-based IMDJ supports communication among stakeholders in the market of data [18].

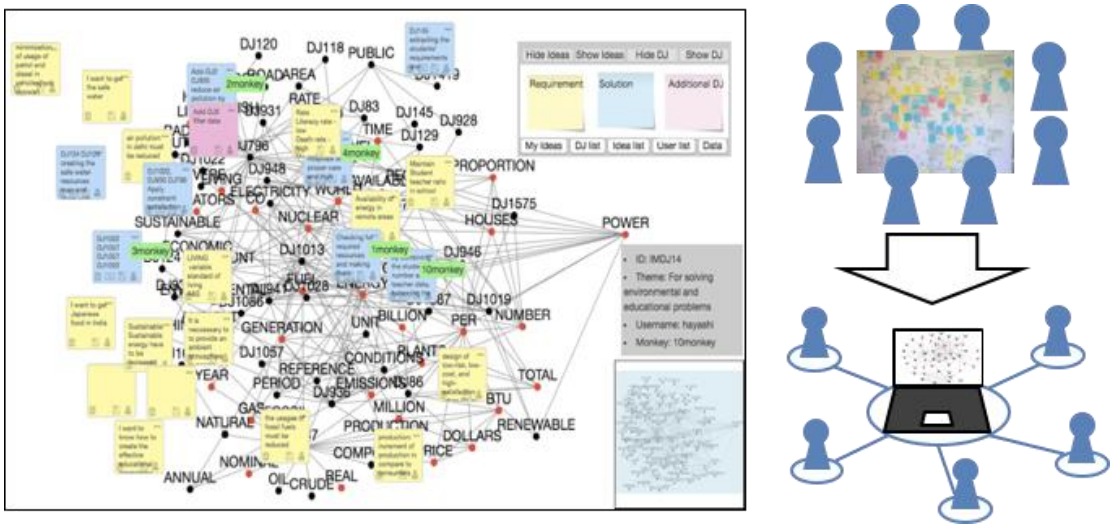


Figure 17. The interface of Web-based IMDJ

6.6 Virtuora DX

The digital transformation system Virtuora DX introduces parts of DJs and DJ Store functions. Fujitsu deploys IMDJ/AP as a consulting service, based on Virtuora DX. Virtuora DX has been used in the Data Co-creation Project in Tokyo Marunouchi Area, and Marunouchi Data Consortium, among others [19].

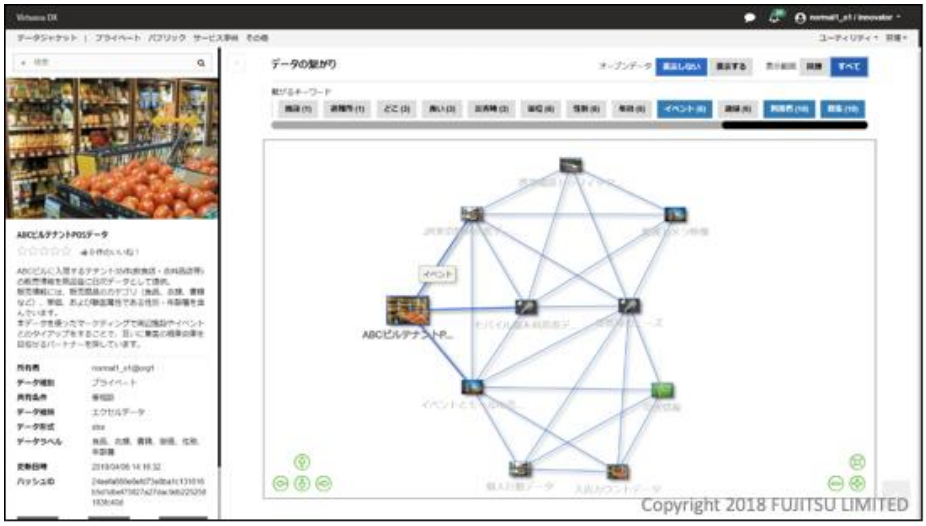


Figure 18. The interface of Virtuora DX

7. Case studies of the IMDJ

7.1 Safety of roads for pedestrians

In this case, the datasets represented by the following DJs were combined as the basis of the solution. The idea was to compute a safe path between two arbitrary points for walkers at night, responding to the requirements for pedestrians' safety on dark roads (published only domestically in Japan [20]).

DJ1: Locations and types of city lights

DJ2: Road map contained in the data of Google Maps.

Although dataset 1, corresponding to DJ1, was confidentially owned by the local government, the government allowed the solution provider to use this dataset for this restricted purpose. This enabled the realization of the safe-path finder (in the right), which was positively received by female students.

Requirement R: Pedestrians' safety on dark roads

Solution S: Find a safe walking path by mapping city lights on the road map

Data for realizing S: {**DJ1**: Location of city lights, **DJ2**: map data}

Figure 19. The case to make a safe path for walkers at night.

7.2 Improvement of working conditions

The sets of {requirement **R**, solution **S** using data represented by data jackets} in the following examples was obtained as a result of the IMDJ. In this case, the datasets represented by the following DJs were combined as the basis of a solution to investigation of the influence of workers' resting on the organizational business performance of the firm. This was responding to the requirement to evaluate workers' satisfaction (published only domestically in Japan).

DJ825: Paid holidays in each industrial domain (factor *X*)

DJ675: Stock price in each industrial domain (factor *Y*)

Although dataset 1, corresponding to DJ825, was owned by a business consultant, we used this dataset for the restricted purpose to satisfy his own requirement above. This enabled us to discover the positive correlation between the two factors. This was a surprise, because participants normally believed that a company's business performance is the better if members work harder and for longer.

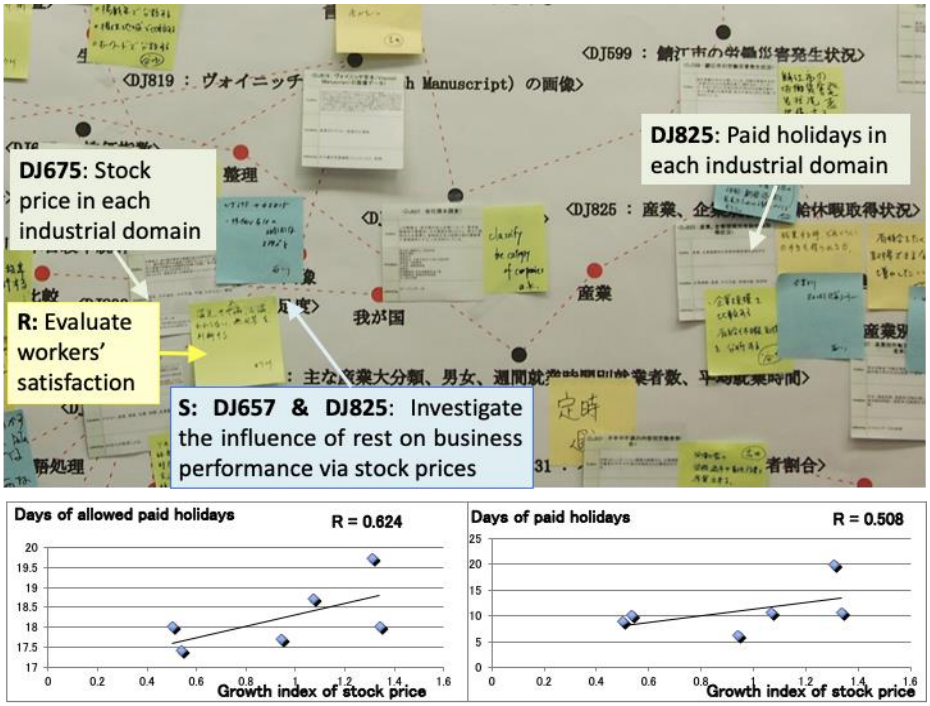


Figure 20. The case of IMDJ for improvement of working conditions

7.3 Soccer

Here, he requirement was to evaluate and improve the defensive capability of a soccer team. The proposed solution was to detect "lines" of teammates for passing the ball from a full-view video, and from data on the players' body directions. Whether the teammates are conscious of such a line is an essential factor, because players' eyes should be focused on the ball in the front, and they should also pass the ball quickly to teammates along these lines [21].

A tool for coaching players was developed, which as a result improved the team's performance (from below the 32nd place in a local league to 8th).

Requirement R: Evaluate and improve the defense quality

Solution S: Detect "lines" from full-view video, using

{**DJa**: full-view video, **DJb**: body direction}

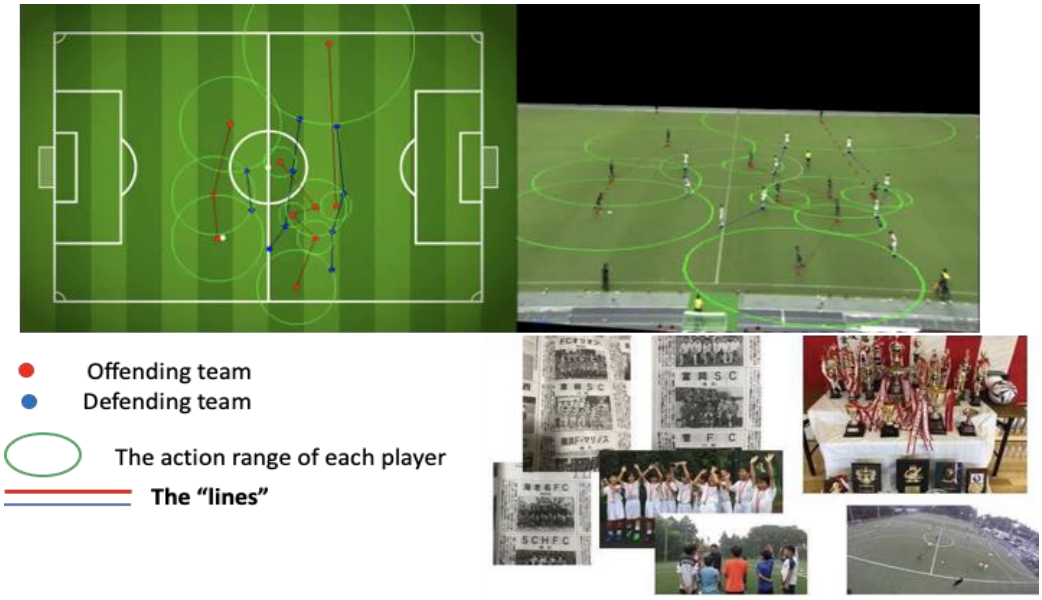


Figure 21. Top: the software showing the lines and their movements in the game. Bottom: The trophies and news articles as evidence of the improvement of the team performance after introducing the tool for coaching.

7.4 “Tangled String” diverted to not only predicting but also explaining changes in stock prices

Tangled String (TS) was originally invented to detect the message with the highest social impact from a log of human communications (DJ1), to realize the solution S1 below for satisfying the requirement R1 [22]. To satisfy R1 After detecting such a message, additional information (DJ2) is needed to validate the credibility of the message corresponding to the chosen messages.

Requirement R1: collect information that is useful for decision making

Solution S1: obtain a high-impact event in a sequence and relate it to external information

Data for realizing S1: {**DJ1**: log text of communication, **DJ 2**: information about disasters}

Simply put, TS obtains a set of consecutive periods with dynamic changing topics, within a certain periods (called pills). It also obtains the switching points (called wires) between such periods. After TS was invented, it came to be required in the following set.

Requirement R2: explain events in the tipping points of consumer behaviors in the market

Solution S2: detect a high-impact event and relate it with external information to explain the causality

Data for realizing S2: {**DJ3**: consumptions or purchase history, **DJ4**: social events and news}

For details, see [23] The encouraging results include (1) the connections from/to trends in the market came to be visualized. This aided in explaining the changes in the market at the desired time scale, which was associated with middle-term changes caused by political decisions, and long-term changes due to innovations in the industry. (2) The change points that were found as entrances to pills by TS coincided with high precision with real increases in each stock price.

TS is also strong in that the trend shifts, in both the overall stock market and the price of each stock, can be explained with one string. This has been analyzed or discussed separately in previous work. As a result, we can relate the changes in a log in DJ3 with other data representing the social trends in DJ4.

Requirement R3: Detect tipping points in customers/investors' behaviors
Solution S3: Extract timings of changes, and link them with external information
Data for realizing S3: {DJ3: log of consumers' buying behaviors, DJ4: TS}

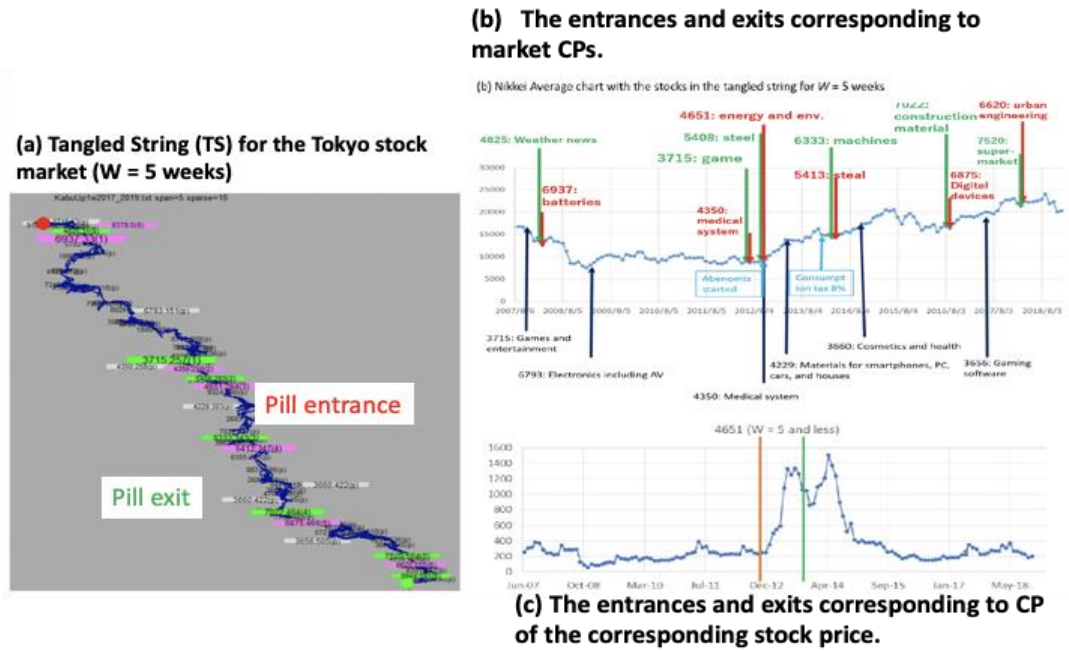


Figure 22. Tangled String diverted to predicting and/or explaining changes in stock prices.

7.5 Not only predict, but explain changes in sales:

To the same requirement-solution set as in the last, i.e.,

Requirement R: explain events at the tipping points of consumer behavior in the market
Solution S: detect a high-impact event and relate it with external information to explain the causality
Data for realizing S: {**DJ3**: log of consumption/purchase history, **DJ4**: social events/news}

Graph-based entropy (GBE) is an index of the diversity of events in their distribution to parts of a co-occurrence graph. It is computed to detect the signs of structural changes in the data that are informative in explaining the latent dynamics of consumers' behavior. For the data on the position of sale (POS), a change in GBE is regarded as a sign of the appearance, the separation, the disappearance, or the uniting of consumers' interests. Experiments show that GBE outperforms baseline methods that can be used to detect change, and that it is useful in explaining substantial changes and their signs in consumers' preference of items in supermarket stores [23].

In the case below, the bridging edge between the two clusters in the graph is cut in the 10th week, which the marketers interpreted as independent growth of the lower cluster, which corresponds to spices for cooking stew. This was surprising, because the 10th week in the data was August, but the result of Google Trend Search supported this, as shown in the curve in Figure 23.

We used the DJ Store (Section 6.1), which is a search engine that explores DJs fitting a query that may not include any of the words contained in a DJ. This was enabled by the collected relationships between requirements, solutions, and DJs in the previously executed IMDJs. A query tends not to include words in DJs, but tends to include words in requirements or solutions, because participants in the IMDJ communicate through informal communications, as well as in the query.

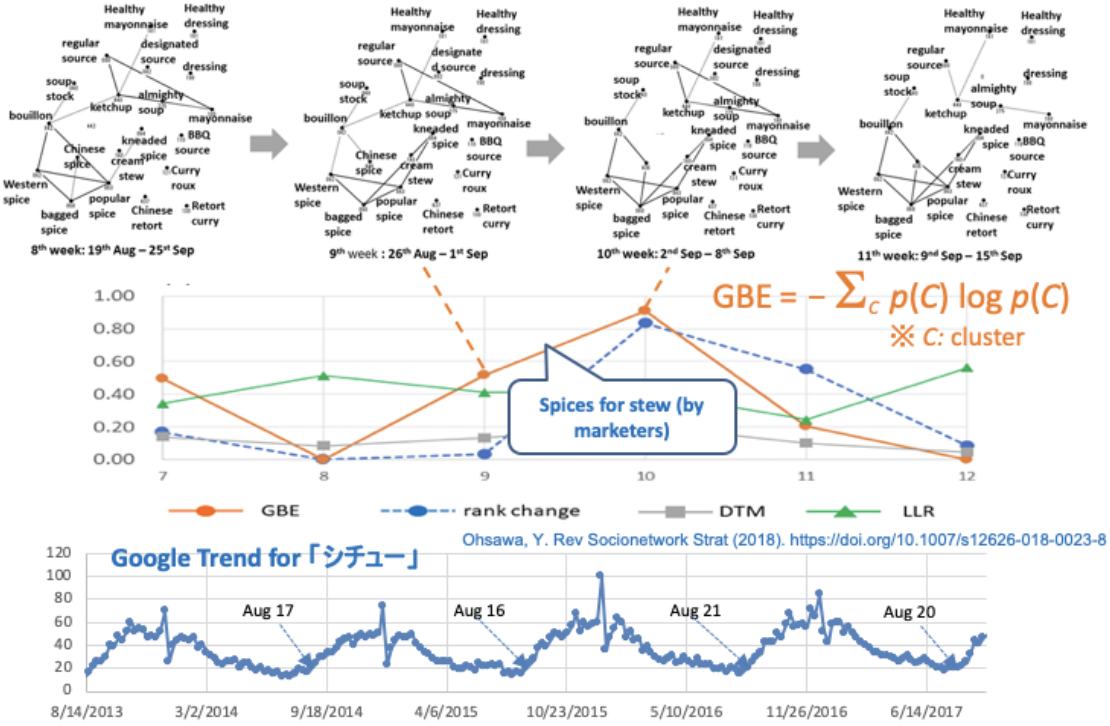


Figure 23. The changes in graph-based entropy (GBE) correspond to the structural change in the market, which corresponds to the change in the consumers' motivation (bottom).

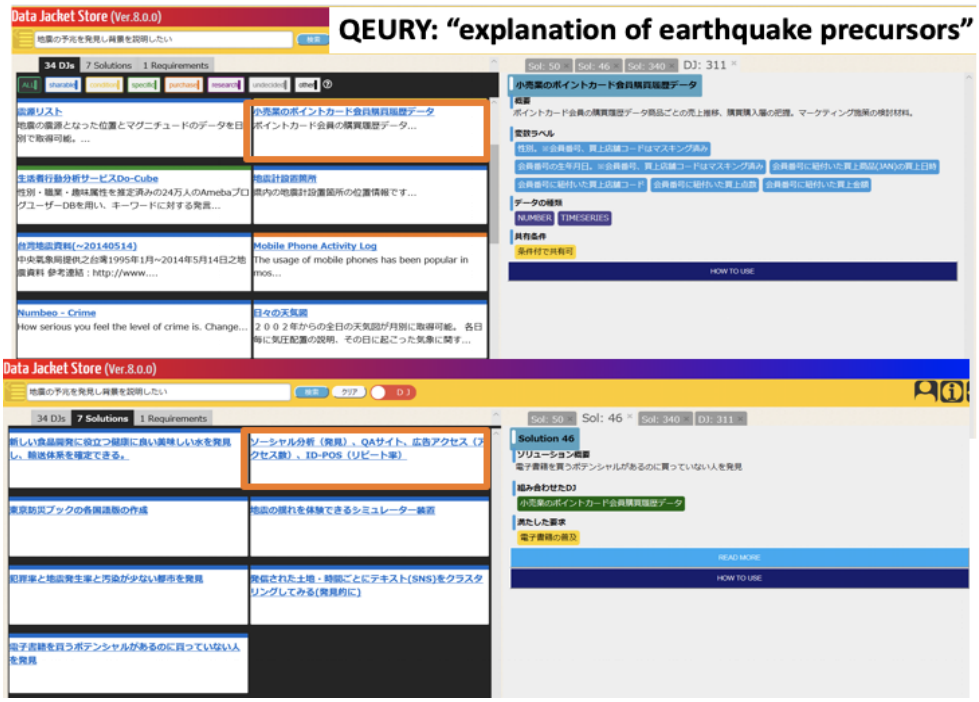


Figure 24. The use of a DJ store for a new interest in the scientific purpose

For example, the user entered this query "explanation of earthquake precursors." An unexpected result was POS data (in the orange-colored frames), which did not appear to have any relation to the

query. However, this result was due to the previous requirement to use POS data to explain the precursor of changes in the market. This came to be a meaningful and helpful result, as explained below. Here, the user introduced a simple model of land crust dynamics to explain the precursory process of the activation of earthquakes. The model was composed of the two phases described below:

Phase 1: The diversity of clusters for which the foci of earthquakes distribute increases from state (a) or (b) to state (c). Here, a cluster is separated to create a seismic gap.

Phase 2: Earthquakes converge to a smaller number of clusters (Figures 25 (c) to (e)) possibly via state (d). Here, the clusters of foci are combined, because earthquakes will occur in the seismic gap in (c). In reaching state (e), earthquakes may occur in the seismic gap.

When reaching (c), ((e) or (d) if the epicenters on the bridge occur in a sufficiently large density), the value of entropy called RESI in this study increases (decreases). Thus, the saturation in the increase of RESI implies the state is in between (c) and (d) or (c) and (e). This represents a precursory condition of a large earthquake. Thus, on the similarity to the market dynamics shown in Fig.23, RESI, which is similar to GBE, has been adopted to detect and explain signs of earthquakes.

From 1993–1995, new clusters emerged until 1994. This corresponded to the increase in RESI depicted in region (a) until 1994 (Figure 26). This increase saturated in 1995, before the magnitude (M) 6.9 earthquake occurred in October 1996. Furthermore, there was an increase in clusters until 1999 in region (c), when the saturation of RESI occurred. Then, the M6.1 earthquake at the red cross in the upper figure of Figure 26 occurred in August 2000, which reduced the RESI substantially [25].

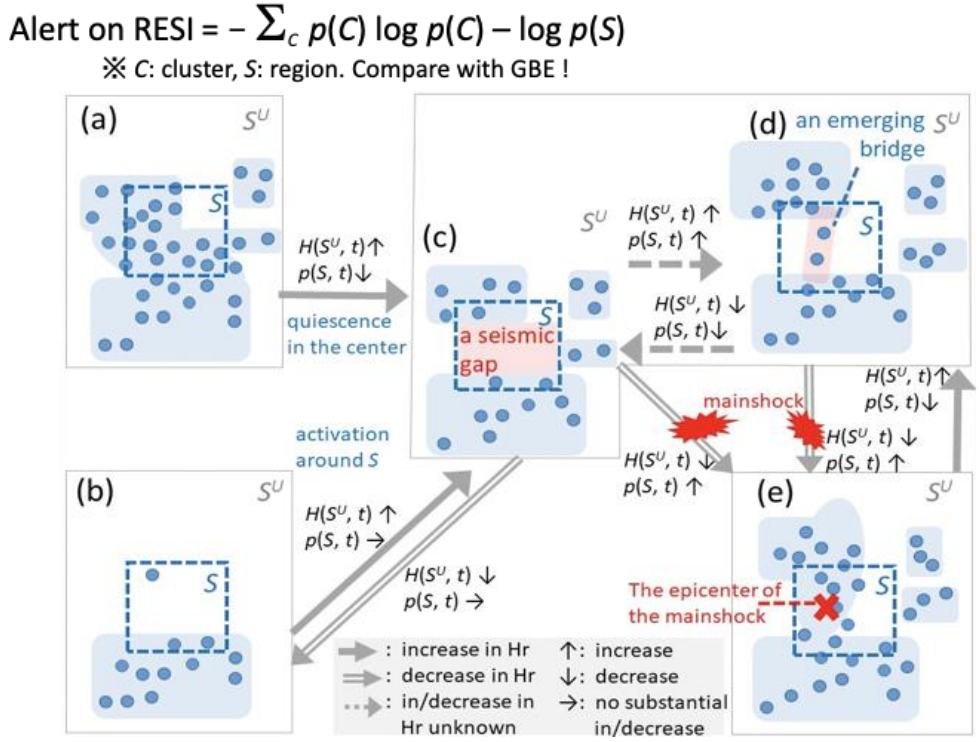


Figure 25. Transitions of earthquake activation. In each state from (a) through (e), the appearance of seismic gaps and bridges are regarded as precursor candidates for earthquake activation. Each earthquake in a sparse region is illustrated to form one cluster in (b) and (d).

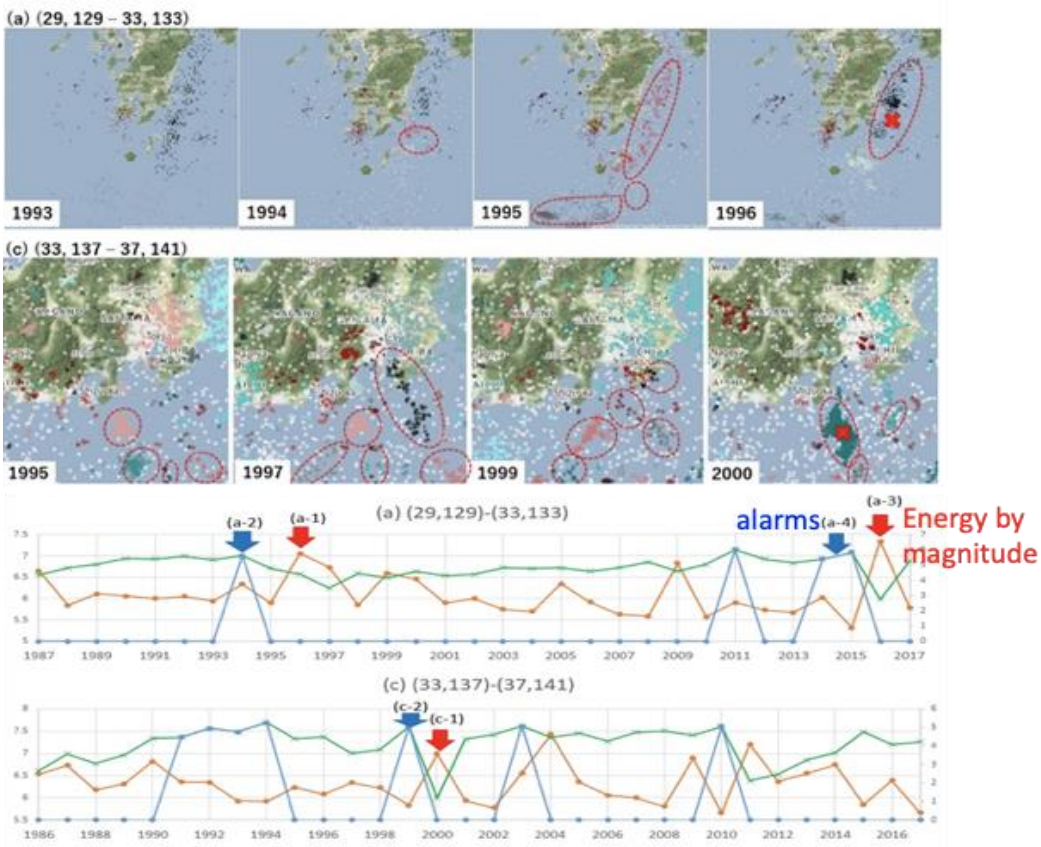


Figure 26. The structural change in the epicenter distribution

7.6 Data Co-creation Project in Yokohama City (2017-2018)

IMDJ/AP have been used to generate scenarios to solve public issues in Yokohama City (Fig. 27). In the project, staff members used the DJ Store based on generated scenarios in IMDJ/AP (https://djp.iais.or.jp/s/djplatform). Figure 28 shows an example of the created scenario for providing support for raising children in Yokohama City, from various perspectives such as daily life, movement, and playgrounds [26].

Figure 27 Innovators Marketplace on Data Jackets and Action Planning

Figure 28. The example of the created scenario

7.7 Data Co-creation Project in Tokyo Marunouchi Area (2018-2019)

Daimaruyu refers to the area around Tokyo Station, which is made up of the first characters (in Japanese kanji) of the neighborhoods of Otemachi, Marunouchi, and Yurakucho. This area is the core of the Japanese business world. An experimental data utilization project in the Daimaruyu area started in May of 2018. The project, which was started by four organizations (Mitsubishi Estate Company, Fujitsu, Softbank, and the Ohsawa Laboratory of the University of Tokyo) ended up comprising twelve organizations, including data providers, analysts, and users. It resulted in three major outputs in just six months, creating DJs and related technologies for data exchange and utilization platforms [27].

Figure 29. Proof-of-concept of Daimaruyu Project and the members

7.8 DJs in the Data Trading Alliance

In the Data Trading Alliance (DTA), DJs have been introduced as the part of the data catalog (see details in https://data-trading.org/public-release/).

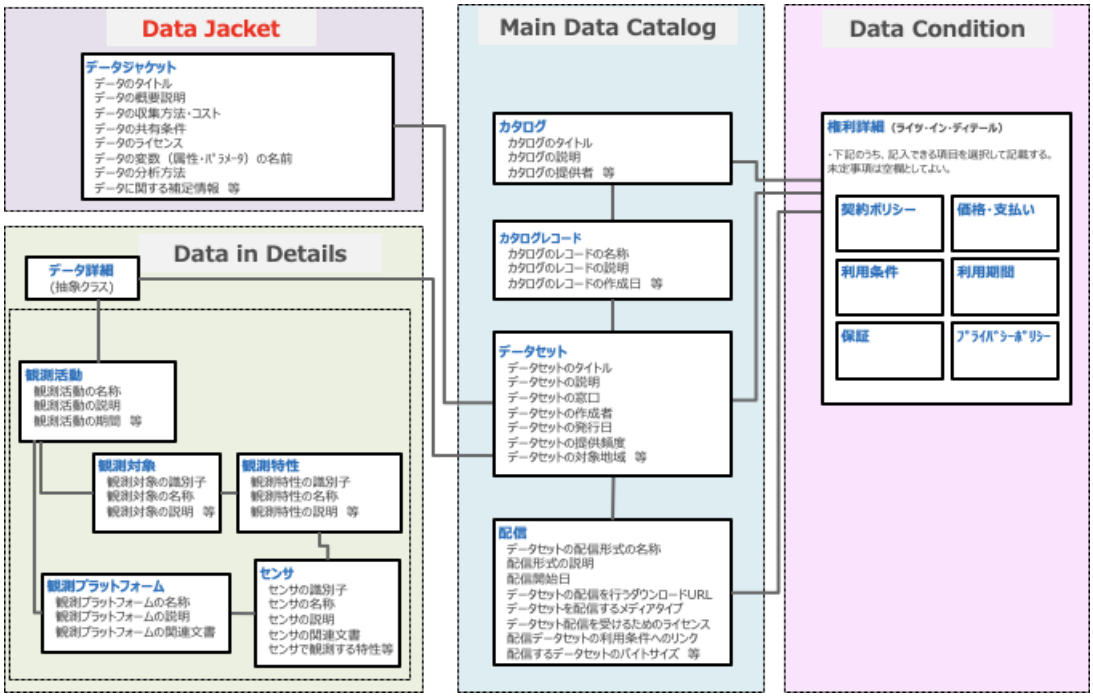


Figure 30. The data catalog using data jackets

8. Conclusions

This document proposed the globalization of prerequisites for the communication in the marketplace of data, including (a) the elements included in each DJ and their logical positioning, (b) the IMDJ process and promises between participants, and (c) other prerequisites. Such prerequisites are desired to be standardized for consistent, efficient, and productive interaction of stakeholders to realize reasonable and trustworthy data trading. The statements about (b), such as CROP in Section 5, are coordinated to be consistent with an ordinary letter of assurances for industrial property rights. However, all these parts are dedicated to the creativity of stakeholders that is the core requirement in the market of data, because the value of datasets cannot be evaluated or priced without the externalization of their latent use scenarios. For this purpose, the formalization of the elements of DJ in (a) are given in the simple framework of FOL to obtain explainable solutions for the requirements by abductive reasoning in Section 4, rather than the use of RDF for representing the relevance of datasets or the use of DL-liteR for combining datasets or [28] or for obtaining search queries for data [29]. This document partially refers to statements in [30].

References

[1] Ohsawa, Y., Kido, H., Hayashi, T., Liu, C., Data Jackets for Synthesizing Values in the Market of Data, Procedia Computer Science, Vol.22, pp.709-716, 2013.
[2] Ohsawa, Y., Hayashi, T., Kido, H., Restructuring Incomplete Models in Innovators Marketplace on Data Jackets, in Magnani, L., Bertolotti, T. (eds), Springer Handbook of Model-Based Science, pp.1015-1031, 2017.
[3] Ohsawa, Y., Kondo, S., Hayashi, T., Data Jackets as Communicable Metadata for Potential Innovators – Toward Opening to Social Contexts, Proc. of the 19th Int'l Conf. Intelligent Systems Design and Applications (ISDA), 2019.
[4] Hayashi, T., Ohsawa, Y., Processing Combinatorial Thinking: Innovators Marketplace as Role-based Game plus Action Planning, International Journal of Knowledge and Systems Science, Vol.4(3), pp.14-38, 2013.
[5] Følstad, A., Living Labs for Innovation and Development of Information and Communication Technology: a Literature Review, EJ. Virtual Organizations and Networks, Vol.10, pp.99-131, 2008.
[6] Leminen, S., Westerlund, M., Nyström, A.G., Living Labs as Open-Innovation Networks, Technology Innovation Management Review, pp.6-12, 2012.
[7] Almirall, E., and Wareham, J., Living Labs: arbiters of mid- and ground-level innovation, Technology Analysis & Strategic Management, Vol.23, No.1, pp.87-102, 2011.
[8] Organ, D.W., A Restatement of the Satisfaction-Performance Hypothesis. J. Management, Vol.14, No.4, pp.547-557, 1988.
[9] Borman, W.C., Motowidlo, S.J., Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. C. Borman, & Associates eds., Personnel selection in organizations. San-Francisco: Jossey-Bass, pp.71-98, 1993.
[10] Dyne, V.L., Cummings, L. L., McLean Parks, J., Extra-role behaviors: in pursuit of construct and definitional clarity, Research in Organizational Behavior, Vol.17, pp.215-285, 1995.
[11] Ohshima, R., Miyazaki, G., Haga, S., The Mediating Effect of Political Skill in Influencing the Effect of the Big Five Personality Domains on Organizational Citizenship Behavior, Japan. Assoc. Industrial/Organizational Psychology J., Vol.32, No.1, pp.31-41, 2018.
[12] Eris, O., Effective Inquiry for Innovative Engineering Design, Kluwer Academic, 2004.
[13] Briggle, A., Mitcham, C. Ethics and Science: An Introduction, 2012.
[14] Cook-Deegan, R., The Gene Wars: Science, Politics, and the Human Genome, 1994.
[15] Hayashi, T., Ohsawa, Y., Retrieval System for Data Utilization Knowledge Integrating Stakeholders' Interests, AAAI Spring Symp., pp.221-222, 2018.
[16] Hayashi, T., Ohsawa, Y.: Inferring Variable Labels Using Outlines of Data in Data Jackets by Considering Similarity and Co-occurrence, International Journal of Data Science and Analytics, Vol.6, No.4, pp.351-361, 2018.
[17] Hayashi, T., Estimating Contextual Relationships of Stakeholders in Scenarios Using DBpedia, IEEE-ICDM Workshops 2015, pp.698-705, 2015.
[18] Iwasa, D., Hayashi, T., Ohsawa, Y.: Development and Evaluation of a New Platform for Accelerating Cross-Domain Data Exchange and Cooperation, New Generation Computing 2019.

[19] FUJITSU Co. Ltd, Intelligent Data service Virtuora DX, 2019. https://www.fujitsu.com/jp/products/network/carrier-router/dataexchange/virtuora-dx/saas/
[20] Ikegami, K., Hayashi, T., Ohsawa, Y. Creative Communication and Action Process in Utilization of Data, IEICE Tech. Rep., Vol.114, No.343, pp.45-50, 2014.
[21] Takemura, K., Hayashi, T., Ohsawa, Y., Aihara, D., Sugawa, A., Computational Coach Support Using Soccer Videos and Visualization, IEICE Tech. Rep., Vol.117, No.440, pp.93-98, 2018.
[22] Ohsawa, Y., Hayashi, T., Tangled string for sequence visualization as fruit of ideas in innovators marketplace on data jackets, Intelligent Decision Technologies, Vol.10, No.3, pp.235-247, 2016.
[23] Ohsawa, Y., Hayashi, T., Yoshino, Takaaki., Tangled String for Multi-timescale Explanation of Changes in Stock Market, Information, Vol.10, No.3, 118, pp.1-21, 2019.
[24] Ohsawa, Y., Graph-Based Entropy for Detecting Explanatory Signs of Changes in Market, Rev Socionetwork Strat, 2018.
[25] Ohsawa, Y., Regional Seismic Information Entropy for Detecting Earthquake Activation Precursors Entropy, Vol.20, No.11, 861, 2018.
[26] Kano, E., Onishi, K., Building a Framework of Resolving Public Issues by Utilizing Data –Public and Private Data Utilization Scenario Creation Platform Project–, IEICE Tech. Rep., Vol.117, No.440, pp.67-72, 2018.
[27] 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, IEEE International Conference on Data Mining Workshops, pp.180-182, 2018.
[28] Nikolaou, C., Grau, B.C., Kostylev, E.V., Kaminski, M., Horrocks, I., Satisfaction and Implication of Integrity Constraints in Ontology-based Data Access, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, pp.1829-1835, 2019.
[29] Cima, G., Lenzerini, M., Poggi, A., Semantic Characterization of Data Services through Ontologies, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, pp.1647-1653, 2019.
[30] Ohsawa, Y., Hayashi, T., Ishimura, G., Kondo, S., Shiromizu, T., Visualization of Data Value Coupling based on Data Jacket Logic, IEICE Technical Report Artificial Intelligence and Knowledge-Based Processing, Vol.119, No.413, pp.33-38, 2019.

(2) 目的に照らした達成状況（共同研究、再委託研究による成果を含む。）
～　国際的外部評価（Dawex 社における意見収集）～

国際化に資する事業としての達成状況の客観的評価を行うため、国際外部評価のための調査を実施した。評価者としたのは Dawex 社（仏）である。世界に 9000 社が参加するデータ市場プラットフォームをわずか開業から 5 年で生み出した社員４０名余の企業であり、フランス政府による強力な支援も得ている。2020 年 1 月 10 日に同社を終日訪れ、データジャケットとその国際標準化案について事業者の視点から評価をしてもらった。結果的に、本プロジェクトは全体的に成功しているため、失敗要因の分析は行わない。

午前中にまず Dawex 社側のデータ市場ビジネスについての説明を受け（後述）、次に IMDJ に関するレクチャー（大澤）を 1 時間実施した。既に示した英語資料の当時の版を示して DJ,とその利活用手法のほか、スーパーマーケットやシステム開発、不動産会社、SNS データを取扱う会社、丸の内データコンソーシアムなどとの実施例も紹介した。
午後にはまず IMDJ ワークショップを約 1 時間実践した。次に、Web-IMDJ、DJ ストア、Variable Quest、TEEDA など、IMDJ に関連する諸技術を紹介し、世界規模でのデータ市場に対する DJ の適用性を論じた。

論点１：データプラットフォームとデータ市場
AI ビジネスや研究ではデータを大規模データが不可欠とされる中、多種多様なデータが利用可能となることが重要となってくる。外部を交えてのデータ交換と利活用は困難である。参加するすべての企業をマネージする能力を持っている企業は稀であることなどを考えて、データマーケットプレイスは今後重要な役割を担うことが期待される。
日本では DDFT という提案で首相が牽引する中で DTA が様々な企業の連携で重要な役割をしており、日本での導入企業数はまだ少ないが日本とのデータ連携は強化することは世界規模のデータ市場プラットフォームとしても有益と考えられる。
一般に、データプラットフォームの使い方は 2 種類あり、企業や分野内でのプラットフォームの展開、もう一つはオープンなマーケットプレイスである。いずれでもデータの trust には①データ提供者に対する信用、②データそのものの信用がある。

論点２：データ市場プラットフォーム事業の現状（Dawex 社）
Dawex 社は、参加企業のニーズやデータをマネージするプラットフォームを展開していく Facilitating the Exchange of Data をミッションとし、サービスは次の 2 つを中心とする。
DEP：data exchange platform
GDM：global data marketplace
中央集権的なデータプラットフォームだけでなく、ブロックチェーンを使って分散的管理も行い、分野ごとに特化したプラットフォームも実現している。農業分野、行政、スマートシティなど、多様なデータを使う機関に環境の構築を支援するのがビジネスモデルである。control and security, traceability, licensing and regulatory compliance を達成することを重視している。システムは User Journey on the Data Marketplace モデルをとり、マッチング、検索エンジンなど、データ利活用など支援技術を提供する。メタデータは Data Theme と Data Offering で構成し、データに関する情報（タイトル、概要、ライセンス、カラム名）を記入する。License Contract は Sharing Policy と似た役割を持ち、入力すると Contract を自動出力でき、データに関する取り決めが理解可能になる。

論点３：データ市場技術としての DJ の導入意義

IMDJ について）データの追加、データの要求・ソリューションを書く、データを繋ぐコンセプトとなるキーワードを選ぶ手法等、全プロセスにおける人の思考を更新してゆくインタラクションフェーズに、事業に適用する観点から興味が寄せられた。さらに、得られる知識、データのネットワークの解釈の仕方と評価方法に興味が寄せられた。

支援技術についても Web-IMDJ、DJ ストア、Variable Quest、TEEDA について）データ利活用知識の構造化の方法、データ検索システムのインターフェース、データの可視化方法と見せ方など、今後利用する事業者の視点から技術的質問があった。

Dawex 側のシステムのようにデータ利活用の全プロセスで支援技術を提供し、メタデータはデータに関する情報を人に理解してもらうという方針はデータ市場において重要であり、DJ はこの観点と相性が良いことが明らかとなった。また創造性を支援する点で同社におけるような手法との相補的な役割も期待できるとのことであった。

論点４：将来の見通しと検討課題

今後の見通し : Data marketplace は萌芽的な市場であり、まだ発展途上である。globalization についてはデータ取引における満足度評価の方法を確立し、ビジネスユースを示すことが重要である。DJ を介してこれから共に育てていくような関係を日本とも維持していきたい。

技術の具体的な用い方としては、Dawex 社のメタデータ、特に Data Theme では主観的にデータについて複数記述することができるので、DJ と相互に連携できる。一組織内でもデータが散在する現状、データの整理と新たな知見を得るワークショップの実施は重要との視点から、DJ 技術とコラボレーションをめざす動きはぜひ前向きに進めたい由の希望を得た。データ取引における満足度評価の方法等の共通の課題である。

２．研究発表・講演、文献、特許等の状況（共同研究、再委託研究も含む。）

(1) 研究発表・講演

[2] Yukio Ohsawa, Data Jackets may Change with Your Drifting Speed in the Real Field、*2nd International Symposium on Socionetwork Strategies in the Market of Data - Intercultural Collaboration and Life Space Innovation-*, Shimane, Japan (2020)

[3] 大澤幸生・近藤早映・早矢仕晃章・石村源生・白水督久, データジャケット論理に基くデータ価値連成の可視化, 信学会人工知能研究会データ市場特集 VI : 異文化連携と生活空間イノベーション AI2019-48 (2020)

[4] 早矢仕晃章・石村源生・白水督久・大澤幸生, 創造的データバリューチェーン記述モデルの提案, 信学会人工知能研究会データ市場特集 VI : 異文化連携と生活空間イノベーション AI2019-49 (2020)

[5] 石村源生・早矢仕晃章・白水督久・大澤幸生, IMDJ(Innovators Marketplace on Data Jackets)を利用した新しい科学技術コミュミュニケーション手法の提案, 信学会人工知能研究会データ市場特集 VI : 異文化連携と生活空間イノベーション AI2019-50 (2020)

[6] Ohsawa, Y., Kondo, S., Hayashi, T., Data Jackets as Communicable Metadata for Potential Innovators – Toward Opening to Social Contexts, *Proc. of the 19th Int'l Conf. Intelligent Systems Design and Applications* (ISDA), 2019.

[7] Ohsawa, Y., Hayashi, T., Ishimura, G., Kondo, S., Shiromizu, T., Visualization of Data Value Coupling based on Data Jacket Logic, *IEICE Technical Report Artificial Intelligence and Knowledge-Based Processing*, Vol.119, No.413, pp.33-38, 2019.