

# Data Jackets as a Global Technology for Data Trading with Trust

## データジャケットの国際標準化

Yukio Ohsawa, Teruaki Hayashi, Gensei Ishimura  
大澤幸生, 早矢仕晃章, 石村源生

The University of Tokyo

# The summary of this document

**This document proposes the globalization of prerequisites for the communication in the marketplace of data, including**

- (a) The elements included in a Data Jacket DJ and their logical positioning
- (b) The process of Innovators' Marketplace on DJs (IMDJ) and the promises between participants
- (c) The representation of a DJ (incl. the meaning of the variable set, the predicate set, and the function set written by first-order predicate logic)
- (d) The guidelines for the communication process in IMDJ (a process model for enhancing reliability and creativity, that is an abductive inference process in first-order predicate logic generating a data transaction architecture to be visualized as a pictogram)
- (e) Regulations to achieve the above (the right to use the presented content, the CROP rules to ensure reliability from inside and outside IMDJ, etc.)

**Also here we show**

- (a) DJ use cases
- (b) Support technology for communication and thinking in IMDJ

# Data jackets defined in the First Order Logic

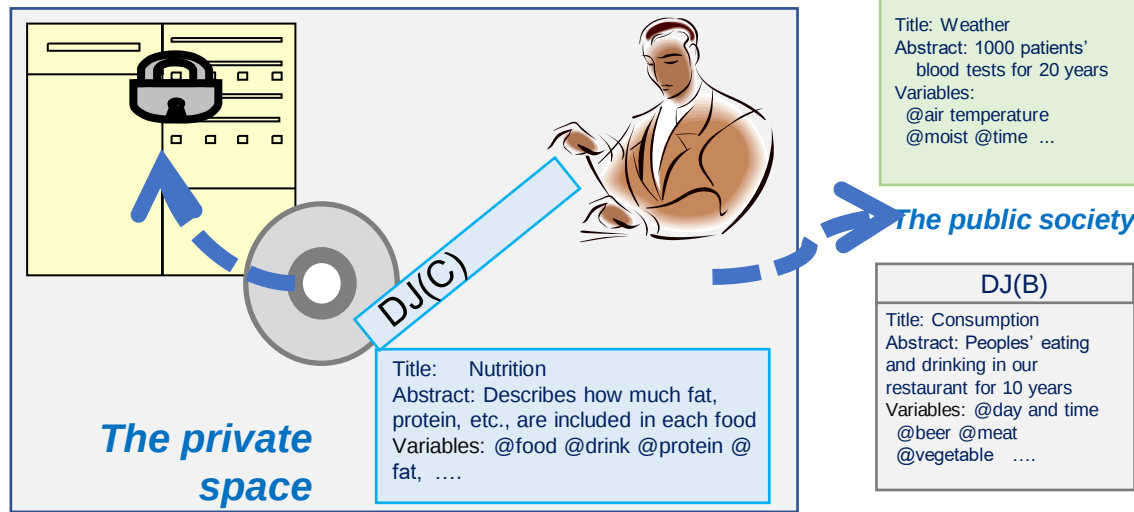
An extension of Ohsawa Y., Hayashi T., Kido H., Restructuring Incomplete Models in Innovators Marketplace on Data Jackets. Magnani L., Bertolotti T. (eds) *Handbook of Model-Based Science*, 1015-1031 Springer (2017)

**A Data Jacket (DJ)** is 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, that may represent the subjective expectation of data owner or potential data users about the utility of the data.

**Definition** a data jacket ( $DJ_i$ ) for a dataset  $d$ , suffixed by  $i$  is **defined by information corresponding to  $DJ_i$**   
 $(d) := \{F_i(d), P_i(d), V_i(d), U_i(d)\}$  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  $(d)$  and their values based on the information in  $DJ_i(d)$ .

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



| ⑫                                     | Power demand and weather                                                                                                           |                                                                              |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Summary of data                       | Electricity demand, and temperature and rain data in [ in Tokyo ] the Tohoku Electric Power's managing regions from 2010 till 2011 |                                                                              |
| Attributes in the data                | Region                                                                                                                             | Monthly average temperature                                                  |
|                                       | Year                                                                                                                               | Lowest regional temperature                                                  |
| Source and the methods for collection | Month, date, and time                                                                                                              | Highest regional temperature                                                 |
|                                       | Load of power                                                                                                                      | Rain quantity of the day                                                     |
| Tool for analysis                     | The data of an electric power company and the data of the Meteorological Agency were suitably combined by Excel.                   |                                                                              |
| Expectation for analysis              | Regression analysis                                                                                                                | Decision tree                                                                |
| Unexpected results                    | Power load implies people's activities                                                                                             | the significant factors among others about weather, affecting the demand for |
|                                       | Weather is not so informative as expected, at least in summer and winter                                                           | The tree comes to be too complex. Data should be approximated e.g. by        |

# The Effects of DJs for Data Management with Co-creation

***Collecting/correcting data and data jackets:*** The market can collect **data jackets** from participants in the market of data. Even if the data do not exist yet, one can design and provide data corresponding to highly evaluated **data jackets**. The provider can correct the data or data jackets adding useful variables in the data-use scenarios created in IMDJ .



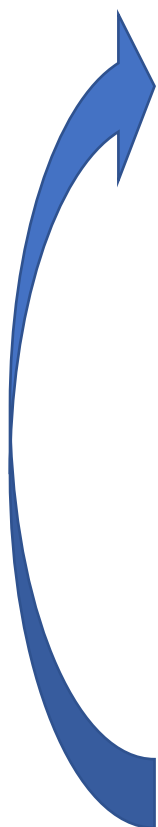
***Connecting data jackets:*** connect **data jackets** via common variables and common words, where the commonality is defined by the similarity in a certain distance space (e.g., inverse of co-occurrence frequency, distance in embedment space). The connections should be visualized to assist human(s'/'s) creativity about data use scenarios.



***Discussing for using data:*** on the connections of **data jackets** visualized above, participants propose, evaluate, and improve ideas about how to use each dataset and the combined data-metaset.



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



# The stakeholders' merits of DJs in the market of data

## A. Data owner

**enabled to** find someone (in **B, D, or E**) who may discover the utility of **A's** data

**enabled to** find someone (in **B, C, D, or E**) who may buy **A's** data

## B. Data provider (not always identical to A)

**enabled to** find someone (in **A**) who may have **B** deal with the own data

**enabled to** find someone (in **C**) who may deal with the data **B** deals with

**enabled to** find someone (in **D, or E**) who may discover the data utility or use the data

**enabled to** find someone (**C, D, or E**) who may buy and/or use the data

**enabled to** price the data via the discovery above

## C. Data enablers (including platform organizers)

**enabled to** find some (in **A or B**) who may have **C** deal with their data

**enabled to** find some (in **D or E**) who may discover data utilities or use data

**enabled to** find some who may buy data (**E**)

## D. Business data scientists (including statistics/AI technicians)

**enabled to** find someone (in **A or B**) who may provide data

**enabled to** find someone (in **E**) who may show the requirement for data analysis

**enabled to** find someone who (**E**) may buy **D's** results of data analysis

## E. Intermediaries (data-using businesses)

**enabled to** find someone (in **A or B**) who may provide data

**enabled to** find someone (in **D**) who may show/provide methods for data analysis

**enabled to** find someone who (**F**) may buy **E's** service/products

## F. Final consumers (the customers of D)

**enabled to** learn relations of data, technologies, services, and products to one's own life

# The Innovators' Marketplace on Data Jackets (IMDJ)

**Innovators Marketplace on Data Jackets (IMDJ)** is a process to create and revise models of businesses and data-use scenarios for the businesses in three aspects, i.e, (A) data users' requirements, (B) data jackets, and (C) proposals of use cases of data. (A), (B), and (C) can be respectively regarded as the *global*, *local*, and the *glocal* models defined as follows.

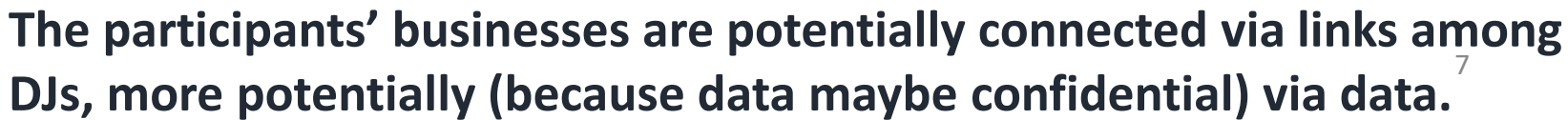
**A *global* model:** The representation of desired causality, that a participant playing the role of user in IMDJ expresses as a requirement without restriction by knowledge specific to a domain represented by terms, variables, or predicates.

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

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

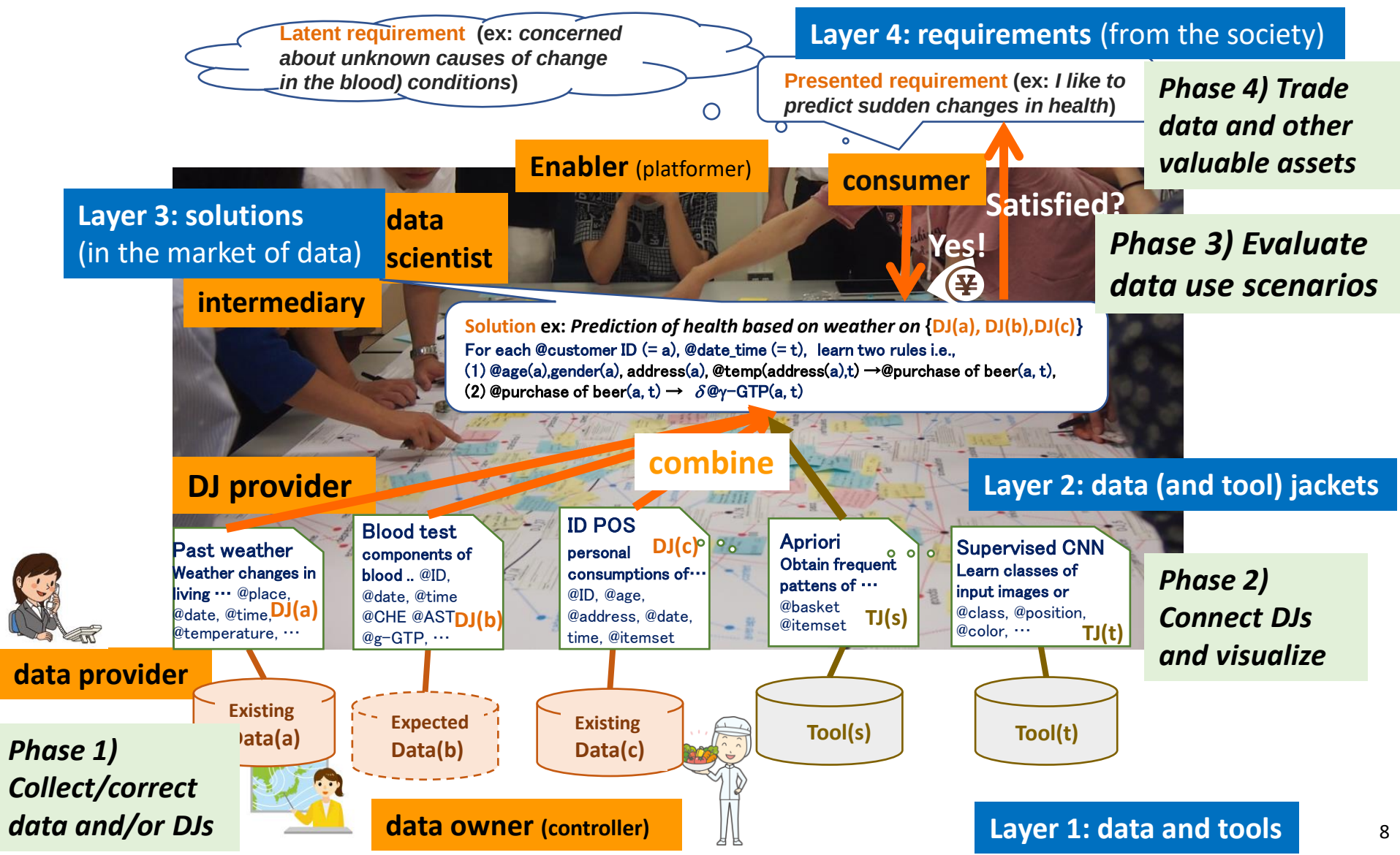


## CosMa[80-100-60-60]



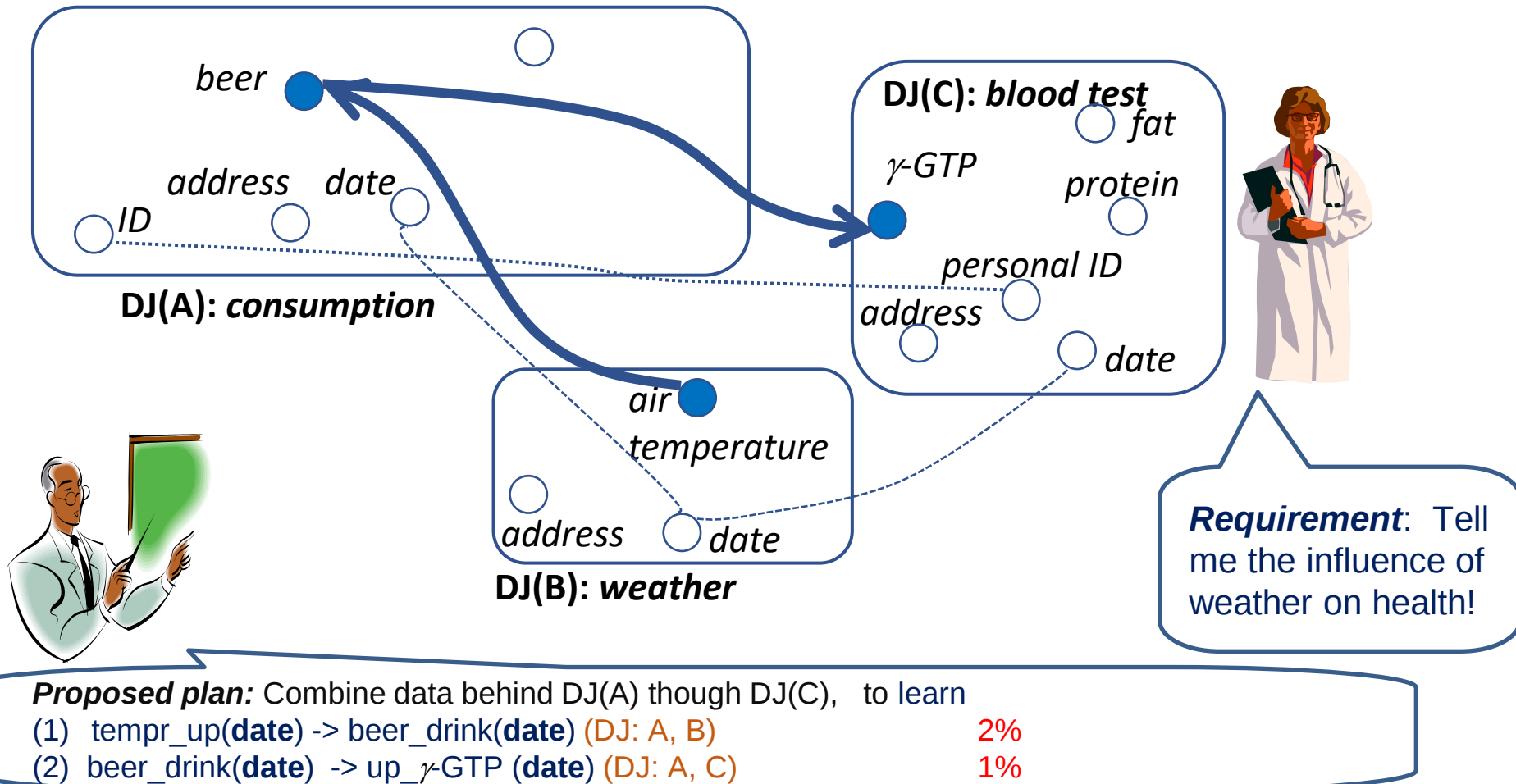
# The **Four Phases** crossing the **Four Layers** of IMDJ for the **Seven Stakeholders** in the market of data

Extending Ohsawa, et al, Innovators Marketplace on Data Jackets for Externalizing the Value of Data via Stakeholders' Requirement Communication AAAI 2014 Spring Symposium on "Big data becomes personal: Knowledge into Meaning, March 2014, Stanford (2014)



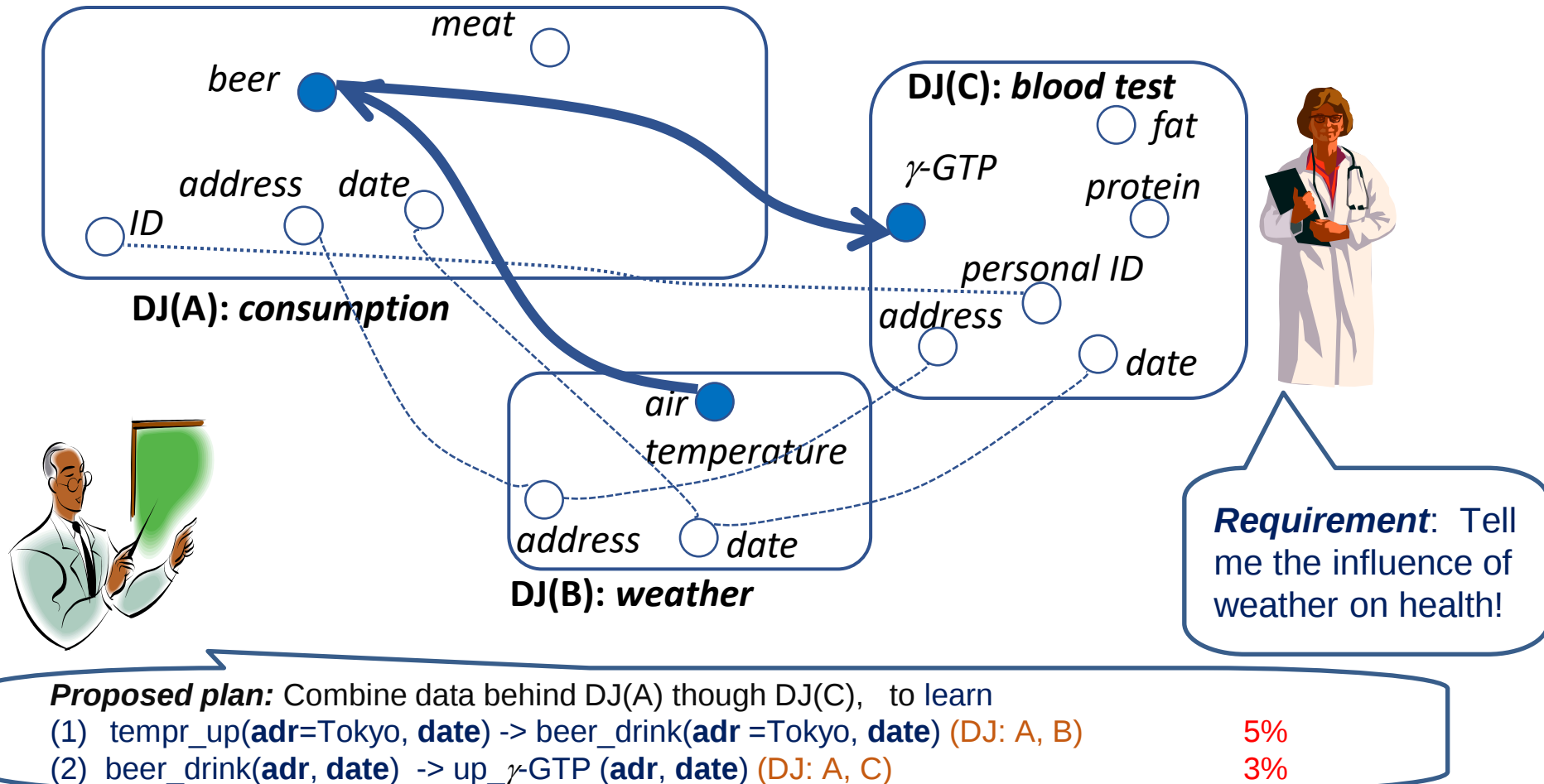
# The abductive communication in IMDJ to create use cases of datasets (1)

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



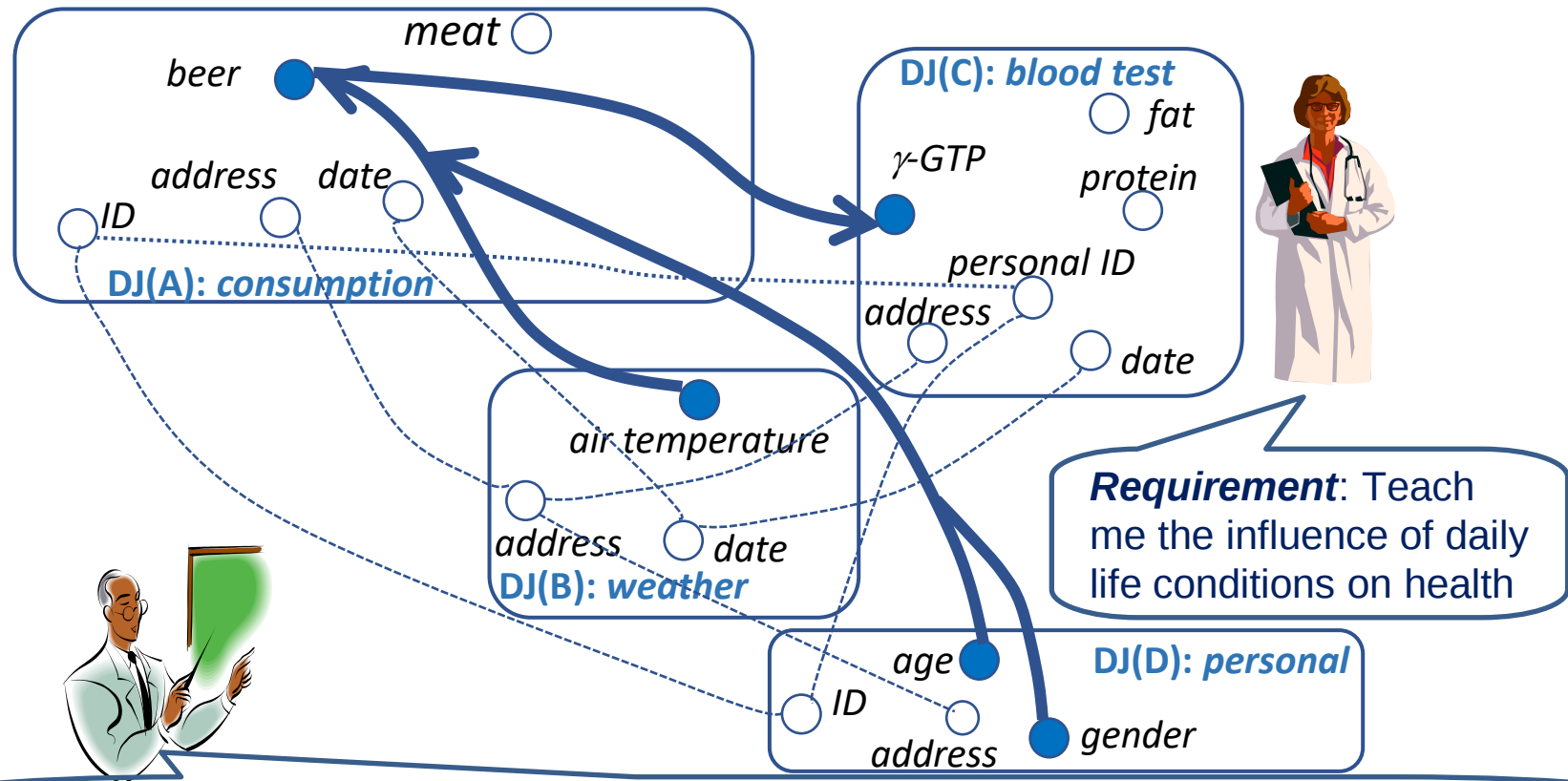
# The abductive communication in IMDJ to create use cases of datasets (2)

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



# The abductive communication in IMDJ to create use cases of datasets (3)

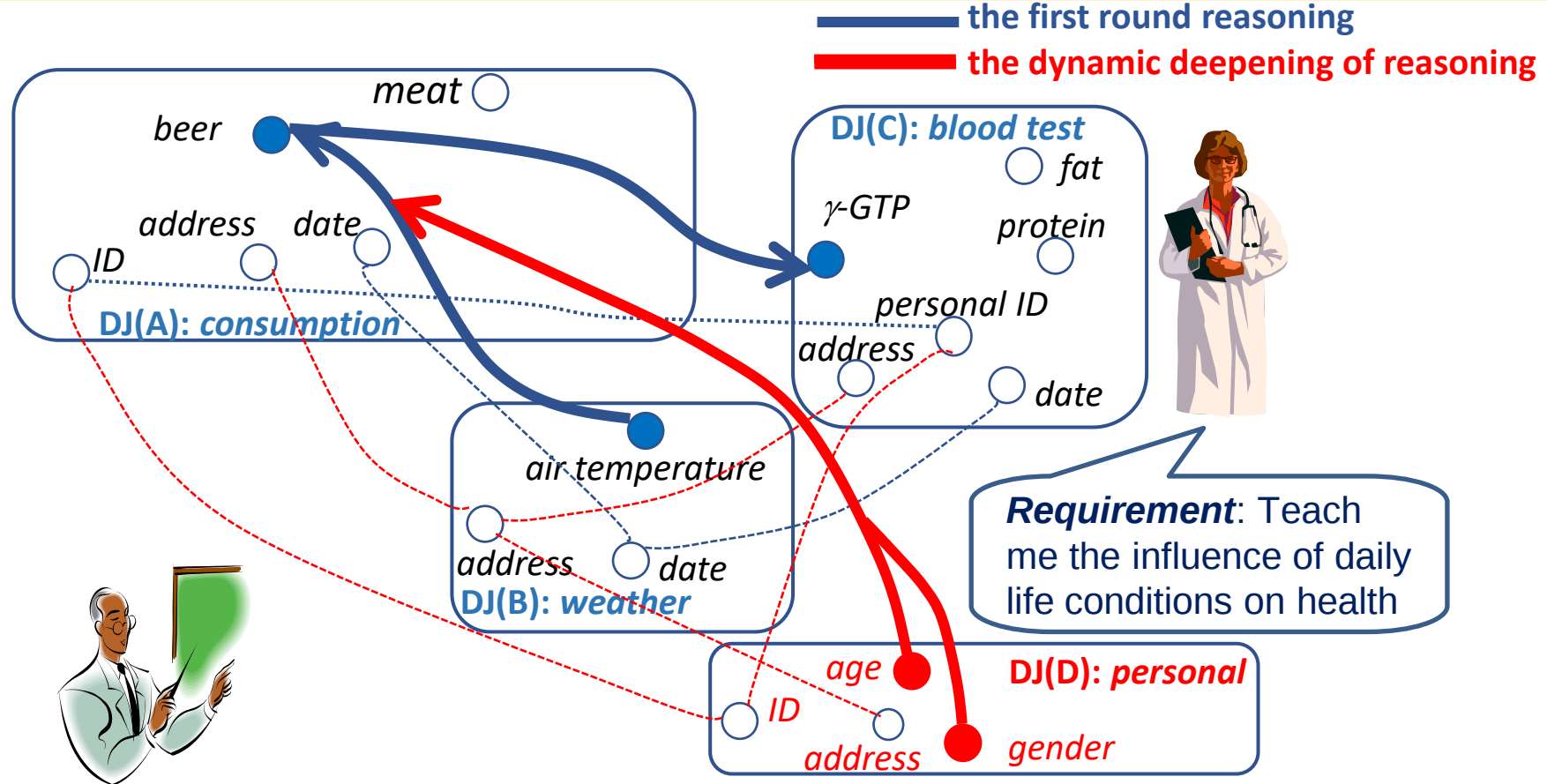
Ohsawa, Y., Kondo, S., Hayashi, T. (2019) Data Jackets as Communicable Metadata for Potential Innovators – Toward Opening to Social Contexts, 19th Int'l Conf. on Intelligent Systems Design and Applications (ISDA 2019)



**Proposed plan:** Combine data behind DJ(A) though DJ(C), to learn

- (1) `tempr_up(adr=Tokyo, date) -> beer_drink(adr=Tokyo, gndr=man, age=30, date)` (DJ: A, B, D) 10%
- (2) `beer_drink(Tokyo, man, 30, date) -> up_ $\gamma$ -GTP (Tokyo, man, 30, date)` (DJ: A, C, D) 9%

# The abductive communication in IMDJ to create use cases of datasets (summary)

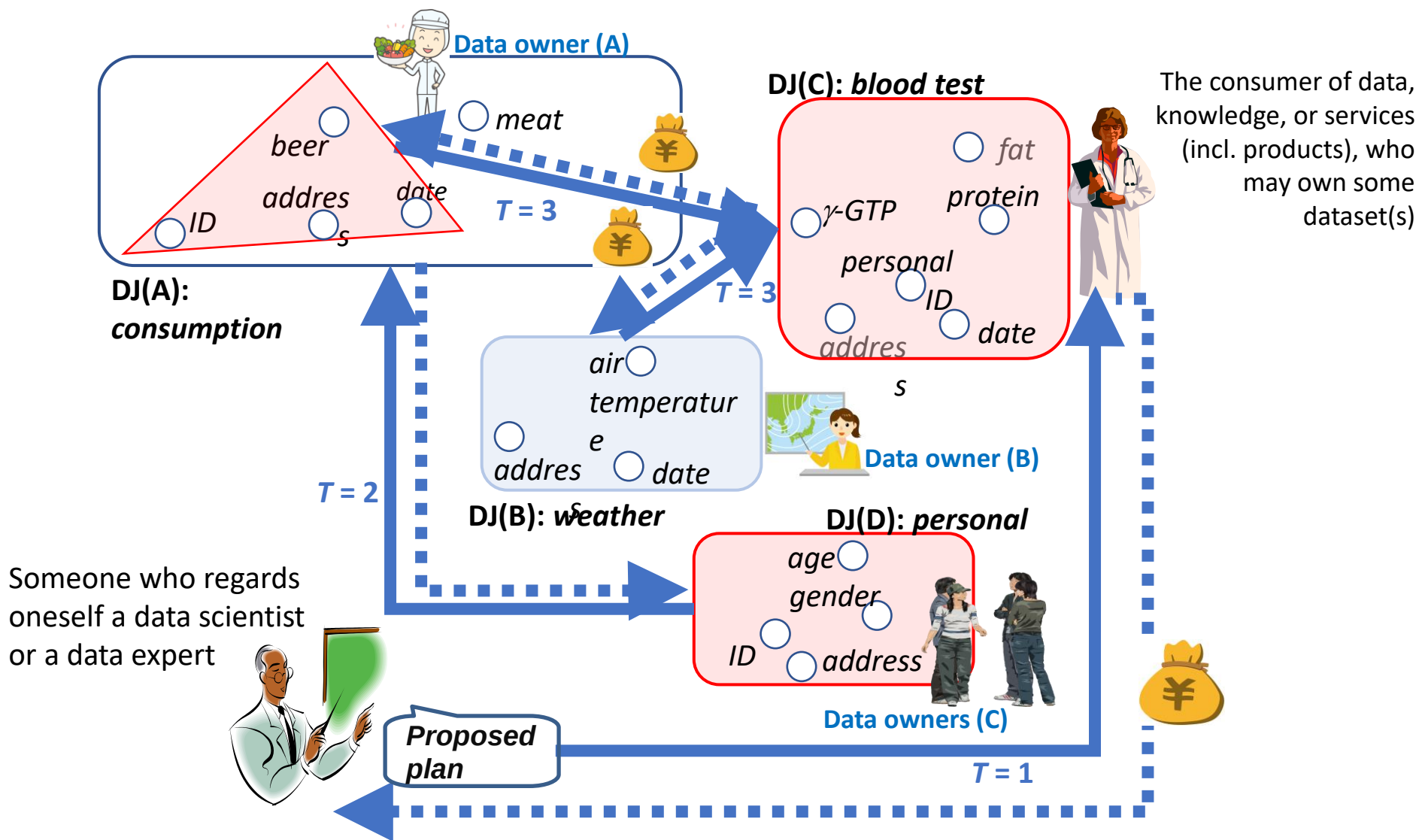


**Proposed plan:** Combine data behind DJ(A) though DJ(C), to learn

- (1) `tempr_up(adr=Tokyo, date) -> beer_drink(adr=Tokyo, gndr:man, age:30, date)` (DJ: A, B, **D**)
- (2) `beer_drink(Tokyo, man, 30, date) -> up_γ-GTP (Tokyo, man, 30, date)` (DJ: A, C, **D**)

Confidence increases (e.g. from 1% to 9%) by adding the red variables in the personal dataset **D**.

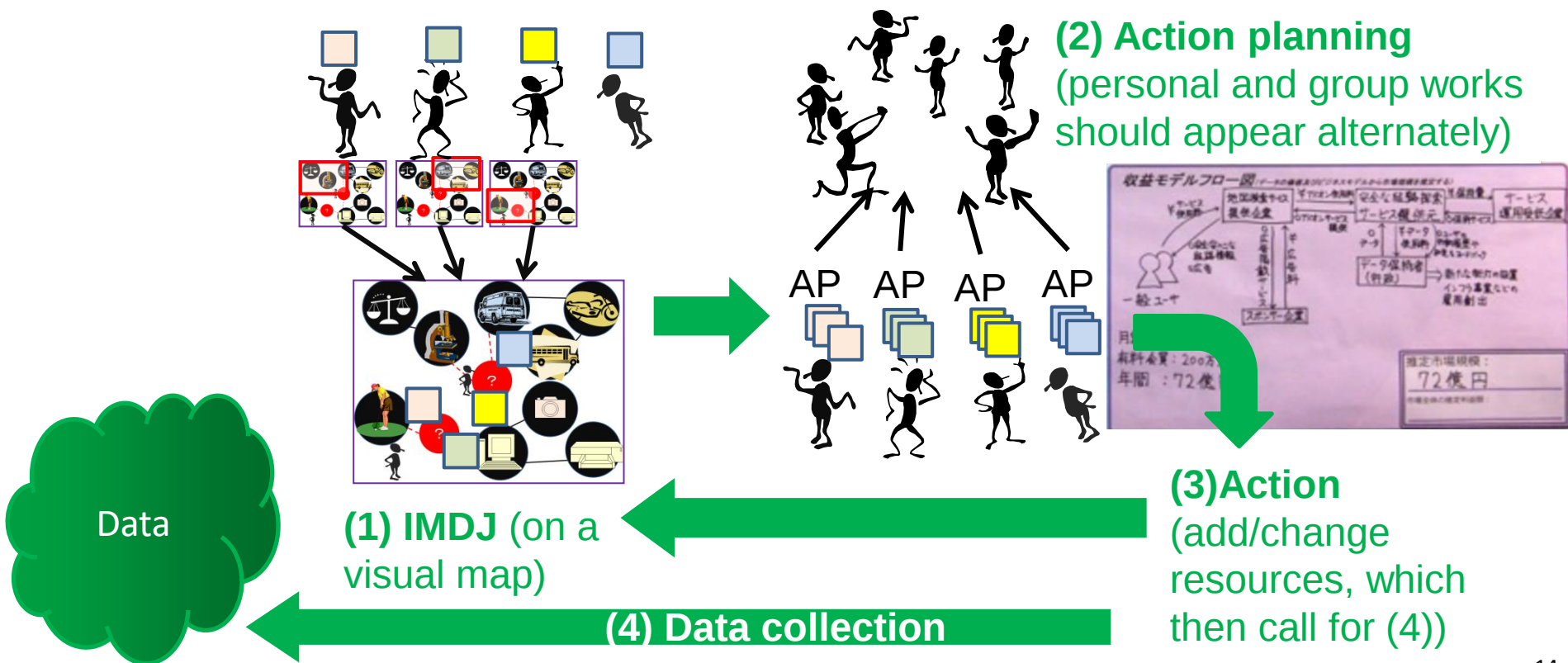
# The stakeholders' value exchange after the abductive communication



# Action planning as a post process of IMDJ

**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 as the post process of IMDJ:** To identify the logical relationship among the elements in 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 in creating practical scenarios.



# The definition of Trust in the context of IMDJ

The trust in/of IMDJ is established *iff* the following rights and reliabilities are ensured.

## 1. Right of DJ's providers in IMDJ (conditions promised to providers)

- DJ's providers are ensured that they can control how their DJ should be used.
  - The information of DJ is absolutely ***kept among defined stakeholders***. The definition is noticed to the providers in advance. If no definition is made, it means that DJ are open to the public. Participants in IMDJ are restricted to the defined stakeholders.
  - The provider can request ***withdrawal*** of their DJ and/or data sets and ***restriction*** of their circulation/use whenever they want after IMDJ. Their request is reported immediately to the related stakeholders, and is executed appropriately.

## 2. Right of participants in IMDJ (conditions promised to participants)

- Participants in IMDJ, ***and only they***, are ensured to be able to use any ideas of solutions for given problems proposed in the IMDJ, under the CROP principle (see the next page).
- They are ensured to express opinions freely in IMDJ except for illegal or unethical ones.

## 3. Reliability of DJ treated in IMDJ (conditions promised to DJ's users)

- DJ's users are ensured that DJ which they use are ***sufficiently*** reliable.
  - Data sets represented by DJ are authentic (not falsified, fabricated, nor stolen).
  - Any unethical intension such as exposure of confidential information is prohibited.
  - The reliability doesn't necessarily mean "accuracy" - DJ can be described subjectively.

## 4. Reliability of IMDJ in the society (conditions promised to the society)

- The society is ensured that IMDJ follows laws, social norm, and ethics (referring to "ELSI").

# The rules for the participants in IMDJ

## The CROP principles

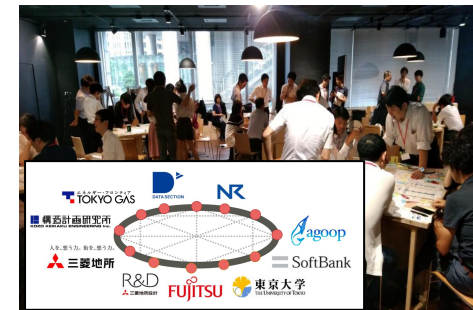
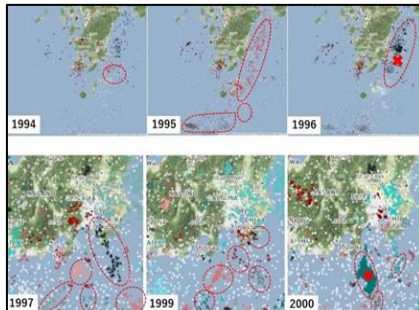
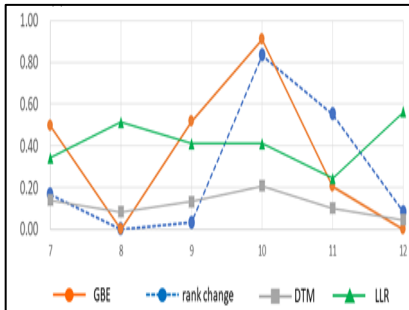
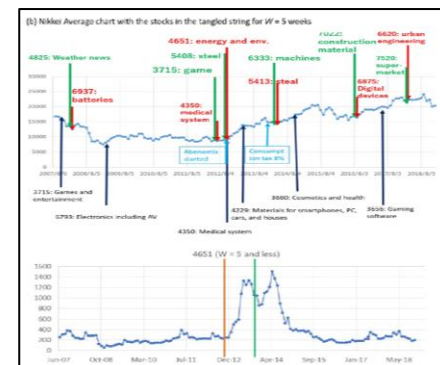
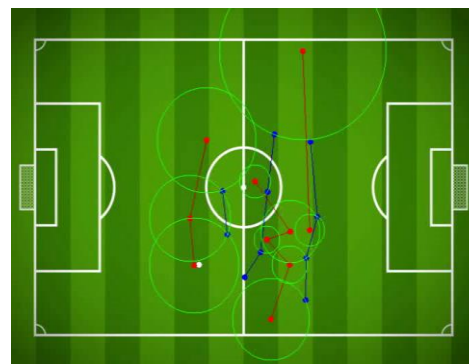
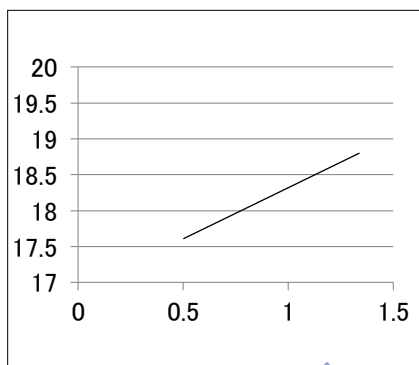
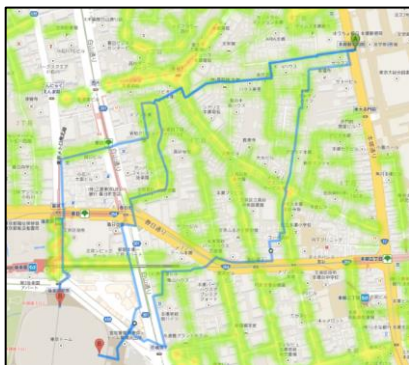
Based on all the definition for trust in and of IMDJ (see the last page), all the ideas created in IMDJ are right-protected under the boundary defined by the Controlled (in or after the session) Reach Of the Presentation.

- (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) The above right is assured only if it is exercised fairly and ethically, and it is allowed by all of the providers of the referenced DJ.
- (3) Any proposer of any idea in the IMDJ can request the restriction of the rights of the others at any time later, and this should be realized as far as possible.
- (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.

# The cases of IMDJ in businesses and sciences

8 cases

1. Streetlights mapping
2. Stock price and payed holidays
3. Football coaching system
4. Stock price analyzer
5. Graph-based Entropy for POS
6. GBE for change detections
7. Yokohama City Project 2017
8. Daimaruyu Project 2018



# Case 1. Safety of roads for pedestrians

Requirement  $R_2$ : Pedestrian's safety on dark roads

Solution  $S_2$ : Find a safe walking path by mapping city lights on the road map

Data for realizing  $S_2$ :  $\{DJ_1$ : map data,  $DJ_2$ : Location of city lights}

<DJ26: バスの運行状況データ>

|       |                                                                 |
|-------|-----------------------------------------------------------------|
| 概要    | 対象となるバス会社、およびバスのリアルタイムな運行状況データ。                                 |
| データ属性 | バスの号車、経路、バスの位置、バスが運行している区間、遅延回数、遅延時分、平均速度、運行状況(混み具合)、バスの向き、人の動き |
| 前提条件  | GPS、自社システムへの入力、バスの基本的な運行                                        |
| 分析の期待 | 道路状況の把握、到着時間の予測、バス停での待ち時間、ロケーションの案内                             |

< DJ26: Bus Movements>  
Based on road maps ( $DJ_1$ )



| 照明種別 | 管理No | 柱種    | 適合ランプ     | 町名 | 丁目 | 番地号 | 緯度 | 経度        | 光束         |       |
|------|------|-------|-----------|----|----|-----|----|-----------|------------|-------|
| 街路灯  | 1    | ボール大型 | CM220     | 湯島 | 1  | 13  | 2  | 354209902 | 1394603203 | 19800 |
| 街路灯  | 2    | ボール大型 | CM220     | 湯島 | 1  | 12  | 5  | 354209792 | 1394601561 | 19800 |
| 街路灯  | 3    | ボール大型 | CM190     | 湯島 | 1  | 12  | 3  | 354209675 | 1394559999 | 20800 |
| 街路灯  | 4    | ボール大型 | CM220     | 湯島 | 1  | 11  | 13 | 354209923 | 1394559225 | 19800 |
| 街路灯  | 5    | NTT共架 | LED33W    | 湯島 | 2  | 11  | 15 | 354215362 | 1394555127 | 2635  |
| 街路灯  | 6    | ボール大型 | CM220     | 湯島 | 1  | 11  | 10 | 354210382 | 1394557204 | 19800 |
| 街路灯  | 7    | ボール大型 | CM220     | 湯島 | 1  | 7   | 12 | 354210922 | 1394554851 | 19800 |
| 街路灯  | 8    | ボール大型 | CM220     | 湯島 | 1  | 7   | 01 | 354211217 | 1394553007 | 19800 |
| 街路灯  | 9    | ボール   | 1394HF125 |    |    |     |    |           |            | 8000  |
| 街路灯  | 10   | ボール大型 | CM220     |    |    |     |    |           |            | 9800  |
| 街路灯  | 11   | 東電共架  | LED33W    |    |    |     |    |           |            | 2635  |
| 街路灯  | 12   | 東電共架  | LED33W    |    |    |     |    |           |            | 2635  |
| 街路灯  | 13   | 東電共架  | LED33W    |    |    |     |    |           |            | 2635  |
| 街路灯  | 14   | 東電共架  | LED33W    |    |    |     |    |           |            | 2635  |
| 街路灯  | 15   | 東電共架  | HF125     |    |    |     |    |           |            | 8000  |
| 街路灯  | 16   | 東電共架  | HF125     |    |    |     |    |           |            | 8000  |
| 街路灯  | 17   | ボール   | 1394HF200 |    |    |     |    |           |            | 9900  |
| 街路灯  | 18   | 東電共架  | HF125     | 湯島 | 1  | 7   | 9  | 354209141 | 1394554870 | 8000  |
| 街路灯  | 19   | 東電共架  | LED33W    | 湯島 | 2  | 30  | 12 | 354223533 | 1394556237 | 2635  |
| 街路灯  | 20   | 東電共架  | HF125     | 湯島 | 1  | 7   | 01 | 354211217 | 1394553007 | 8000  |

DJ<sub>2</sub>

Locations and  
types of city lights,  
confidentially  
owned by the local  
government

$DJ_2$  Locations and types of city lights, confidentially owned by the local government

$S_2$  Developed tool:  
Visualize the optimized path for walkers at night  
Show  $p$  for  $\max_{p \text{ in path(A, B)}} \min\{\text{light}_{X \text{ in } p}(X) / \text{length}(p)\}$

## Evaluation (users' comments)

- I was feeling my daily path was too dark, but learned a light path from the map
- We can find a safe route to take after drinking
- Short paths turned out to be dangerous! etc.



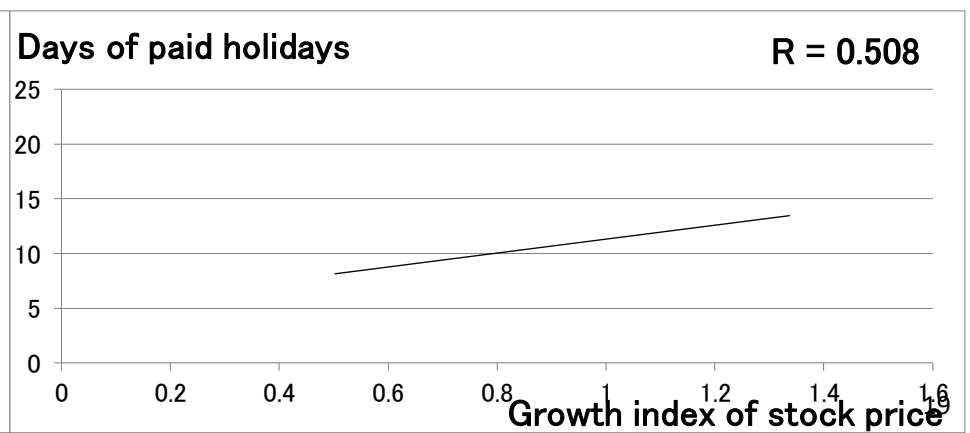
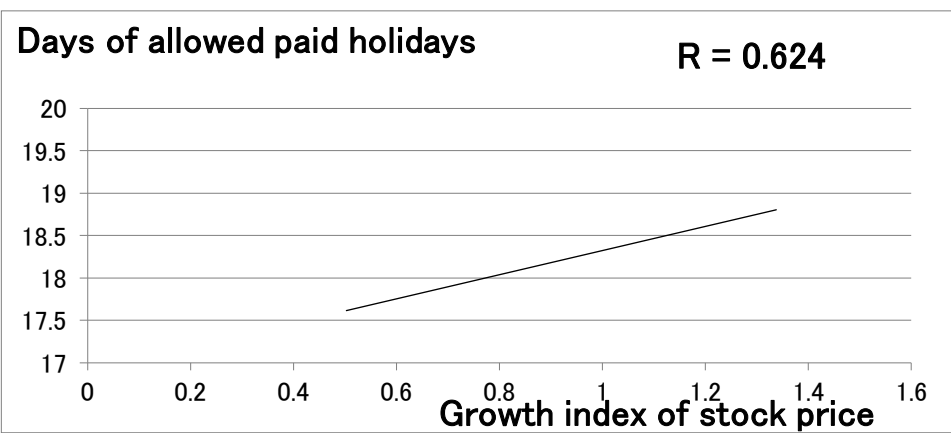
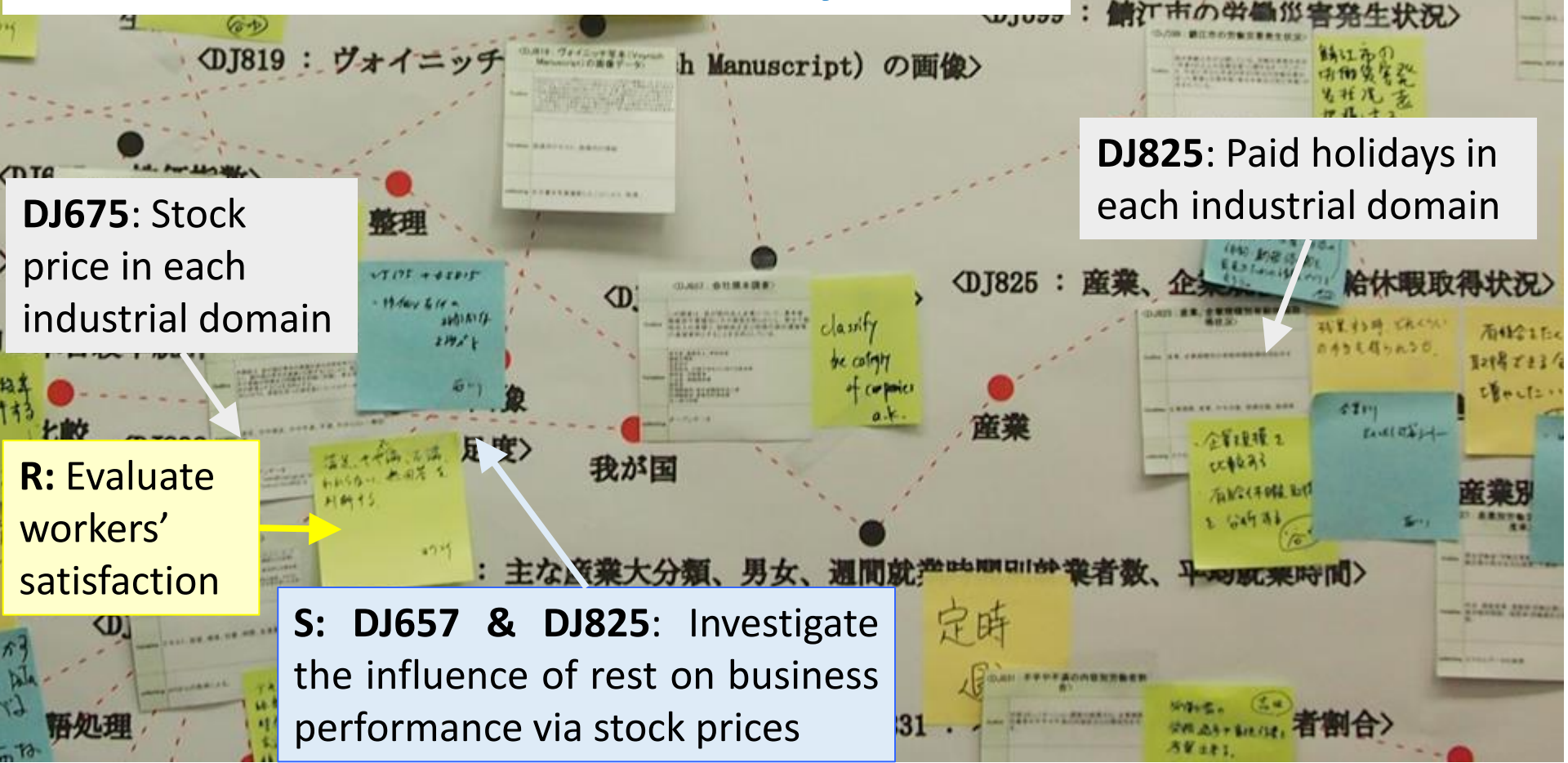
$R_2$

Traffic condition

Added  $DJ_2$

$S_2 = DJ_1 + DJ_2$

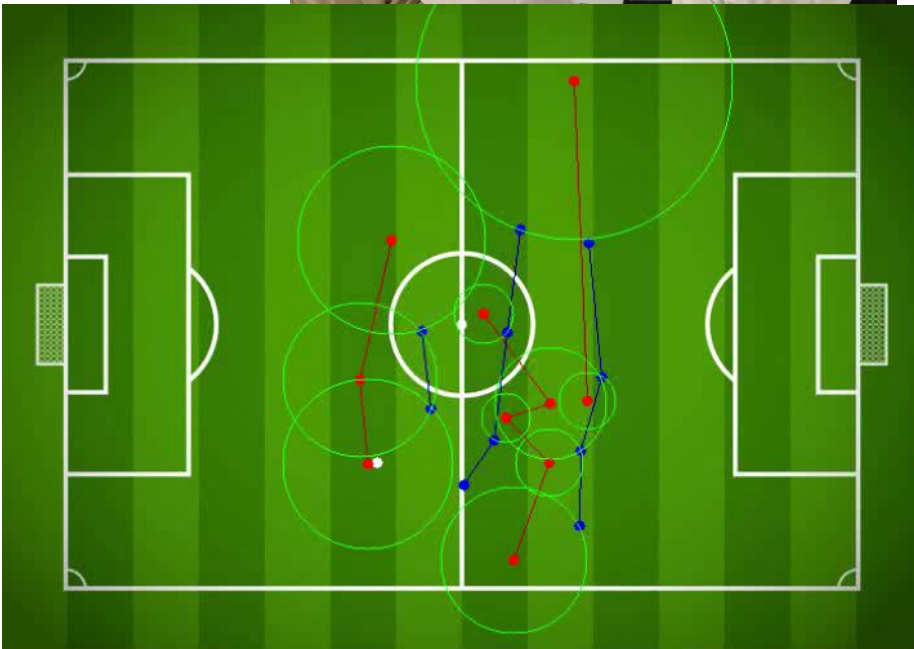
# Case 2. IMDJ 2 hours, AP 1 hour, analysis 1 hour



# Case 3: Soccer!



**R:** Evaluate and improve the defense quality  
**S:** Detect “lines” from full-view video, using  
 $\{DJ_a$ : full-view video,  $DJ_b$ : body direction $\}$



- Offending team
- Defending team
- The action range of each player
- The “lines”

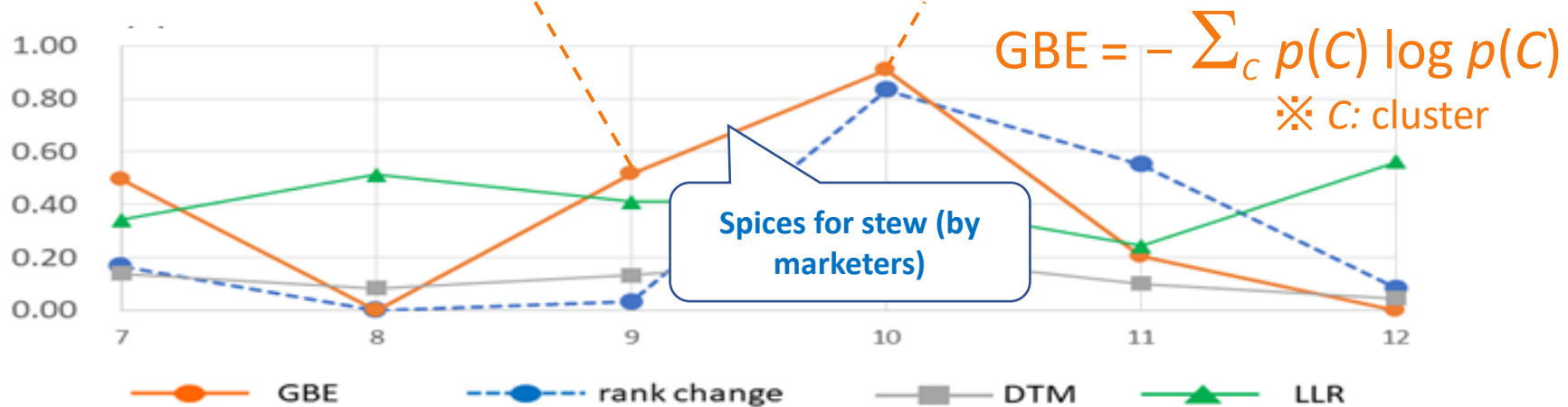
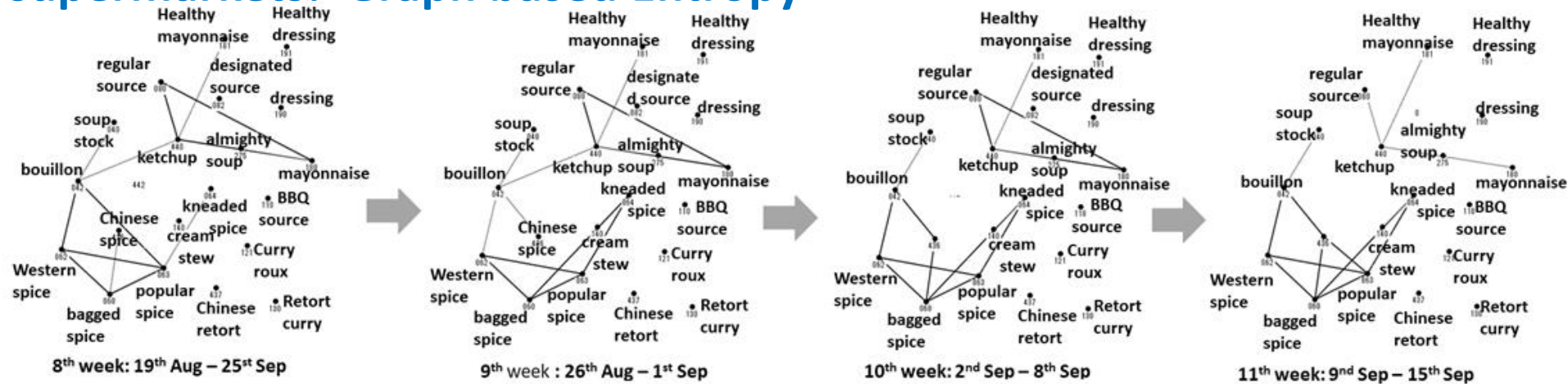
竹村・早矢仕・大澤(東大)・相原(レグロ)・須川(クロスハック):  
サッカー中継映像からの選手・ボールの座標抽出とその分析のた  
めの可視化手法, 信学会AI研究会・データ市場特集(2018.2.17)



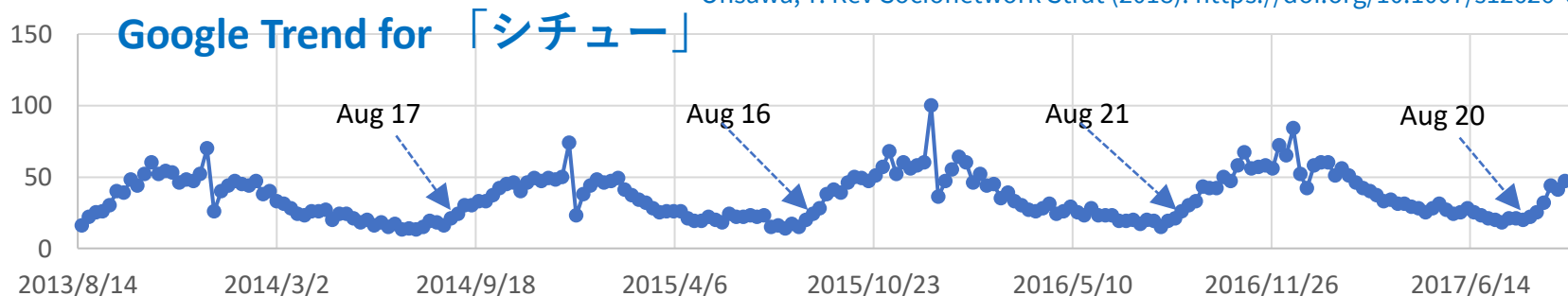
**(b) The entrances and exits corresponding to market CPs.**

21

# Case 5. Not only predict, but explain changes in the sales of supermarkets: Graph based Entropy



Ohsawa, Y. Rev Socionetwork Strat (2018). <https://doi.org/10.1007/s12626-018-0023-8>



# IMDJ on the Web

## (2) Join IMDJ from your own PC

### (1) Start IMDJ

**Info about IMDJ**

date Nov 11 2017

Outline (ex : IMDJ on India)

Theme (ex : Design safe and comfortable town)

User Name

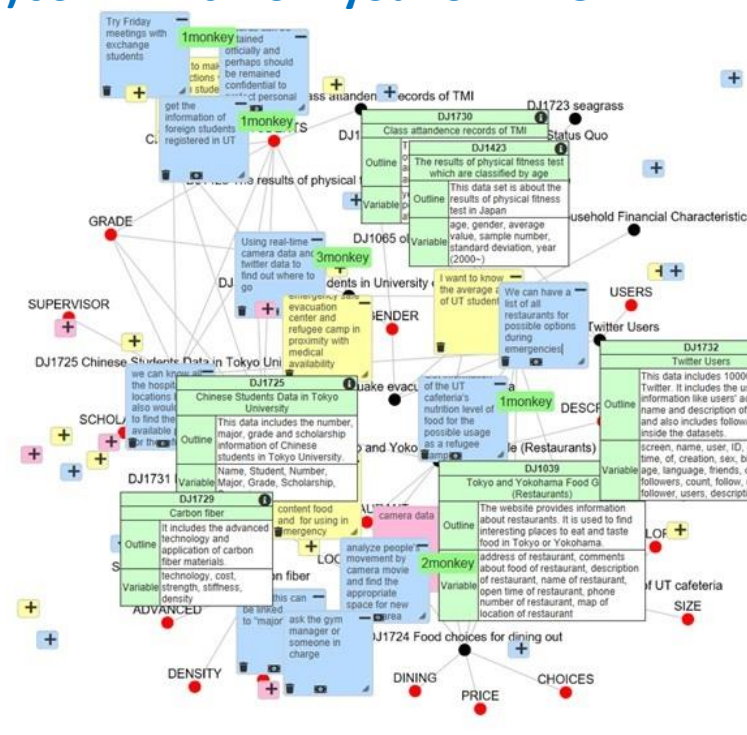
DJs to use (input DJ ID)

IMDJ ID

IMDJ password

admin password

Draw Keygraph



Hide Ideas | Show Ideas | Hide DJs

Requirement | Solution | Additional DJ

My Idea | DJ list | Idea list | Users

ID: IMDJ88

Theme: Resilience Engineering

全体 | sol1 | add11 | sol19 | sol8 | sol22

Rika Get information of the UT cafeteria's nutrition level of food for the possible usage as a refugee camp

link to req17 req1728 req12

Comments below

talk to others...

Hide Chat

### (4) DJ search enabled

Data Jacket Store (Ver.8.0.0)

地盤の予見を発見し再建を説明したい

34 DJs | 7 Solutions | 1 Requirements

小売業のポイントカード会員購買履歴データ

小売業のポイントカード会員購買履歴データ

ポイントカード会員の購買履歴データ...

購買履歴

性別、職業、趣味属性を指定済みの24万人のAmebaブログユーザーDBを用い、キーワードに対する発見...

台湾地震資料(2014.05.14)

Mobile Phone Activity Log

The usage of mobile phones has been popular in mos...

Numbeo - Crime

日々の天気

2002年からの全日の天気図が月別に取得可能。各日時に天気配置の説明、その日に起こった気象に関する...

地下水、河川の水質データ

災害情報

全国の生産量の4割を占めるミネラルウォーターの産地である山梨県の地下水、河川の水質データ...



Iwasa, D., Hayashi, T., Ohsawa, Y., Development and evaluation of a new platform for accelerating cross-domain data exchange and cooperation", New Generation Computing, Nature (in the process to appear in 2020)

地震の予兆を発見し背景を説明したい

検索

34 DJs 7 Solutions 1 Requirements

ALL

sharable

condition

specific

purchase

research

undecided

other

## 震源リスト

地震の震源となった位置とマグニチュードのデータを日別で取得可能。...

## 生活者行動分析サービスDo-Cube

性別・職業・趣味属性を推定済みの24万人のAmebaブログユーザーDBを用い、キーワードに対する発言...

## 台湾地震資料(～20140514)

中央気象局提供之台湾1995年1月～2014年5月14日之地震資料 参考連結 : http://www....

## Numbeo - Crime

How serious you feel the level of crime is. Change...

## 小売業のポイントカード会員購買履歴データ

ポイントカード会員の購買履歴データ...

## 地震計設置箇所

県内の地震計設置箇所の位置情報です...

## Mobile Phone Activity Log

The usage of mobile phones has been popular in mos...

## 日々の天気図

2002年からの全日の天気図が月別に取得可能。 各日毎に気圧配置の説明、その日に起こった気象に關す...

地震の予兆を発見し背景を説明したい

検索

クリア

DJ

34 DJs 7 Solutions 1 Requirements

新しい食品開発に役立つ健康に良い美味しい水を発見し、輸送体系を確定できる。

## 東京防災ブックの各国語版の作成

犯罪率と地震発生率と汚染が少ない都市を発見

電子書籍を買うポテンシャルがあるのに買っていない人を発見

ソーシャル分析(発見)、QAサイト、広告アクセス(アクセス数)、ID-POS(リピート率)

## 地震の揺れを体験できるシミュレーター装置

発信された土地・時間ごとにテキスト(SNS)をクラスタリングしてみる(発見的に)

## QEURY: "explanation of earthquake precursors"

Sol: 50 Sol: 46 Sol: 340 DJ: 311

## 小売業のポイントカード会員購買履歴データ

## 概要

ポイントカード会員の購買履歴データ商品ごとの売上推移、購買購入層の把握。マーケティング施策の検討材料。

## 変数ラベル

性別。※会員番号、買上店舗コードはマスキング済み

会員番号の生年月日。※会員番号、買上店舗コードはマスキング済み

会員番号に紐付いた買上商品(1AN)の買上日時

会員番号に紐付いた買上店舗コード

会員番号に紐付いた買上点数

会員番号に紐付いた買上金額

## データの種類

NUMBER

TIMESERIES

## 共有条件

条件付で共有可

HOW TO USE

Sol: 50 Sol: 46 Sol: 340 DJ: 311

## Solution 46

## ソリューション概要

電子書籍を買うポテンシャルがあるのに買っていない人を発見

## 組み合わせたDJ

小売業のポイントカード会員購買履歴データ

## 満たした要求

電子書籍の普及

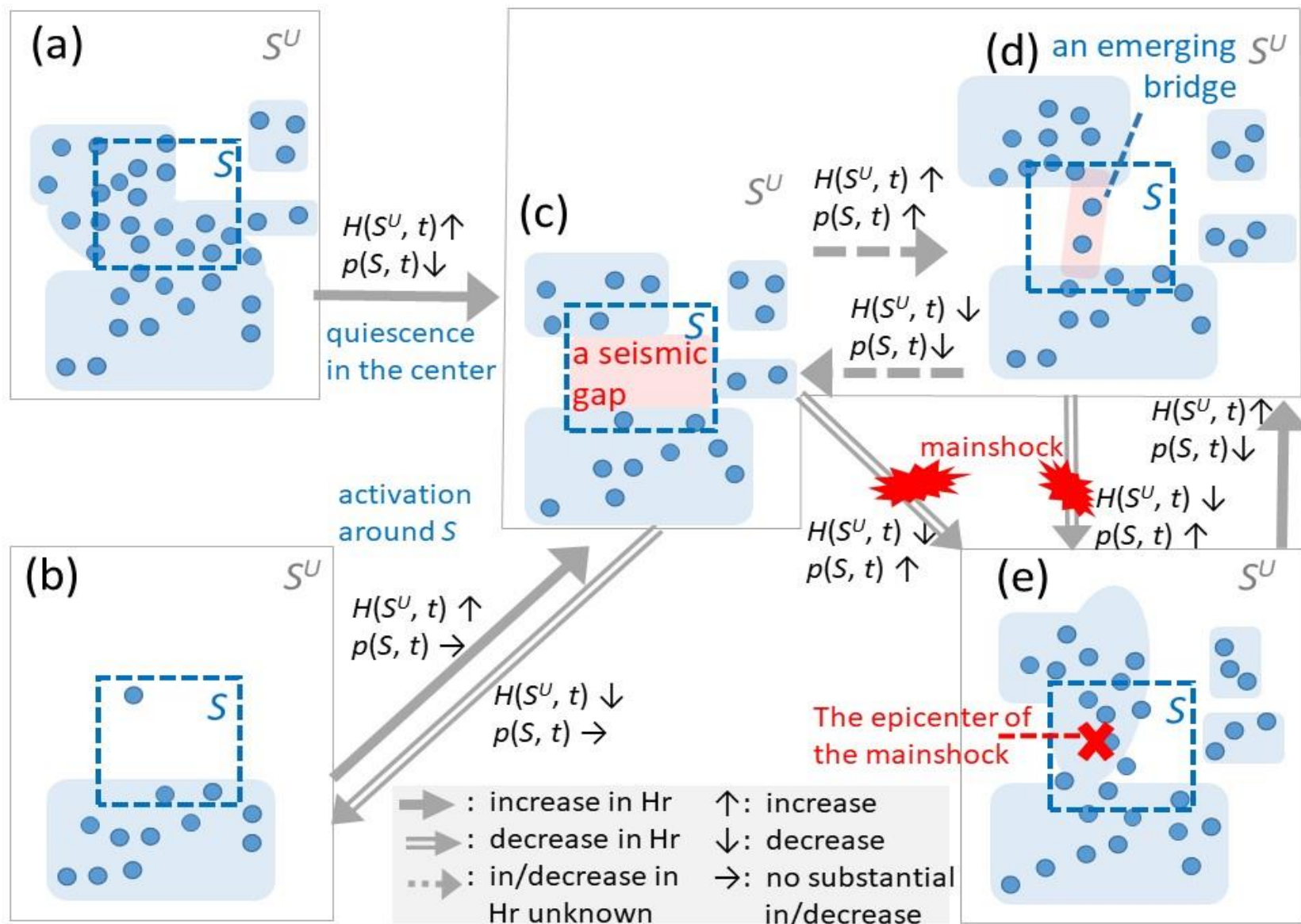
READ MORE

HOW TO USE



Alert on RESI =  $-\sum_c p(C) \log p(C) - \log p(S)$

※ C: cluster, S: region. Compare with GBE !

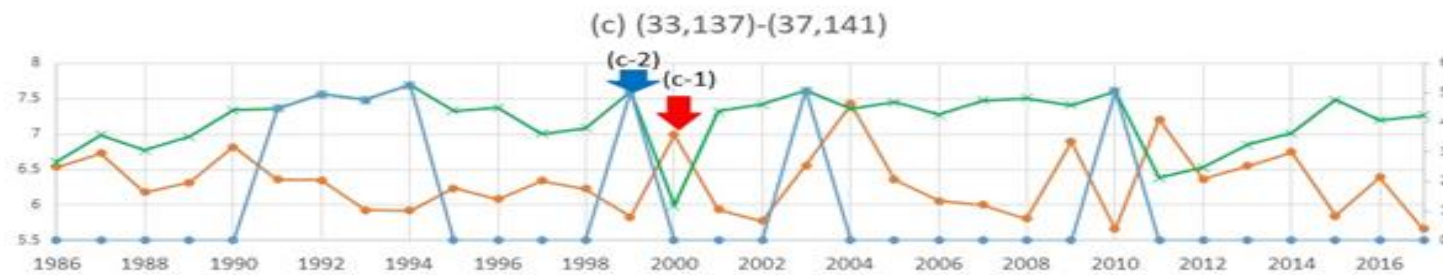
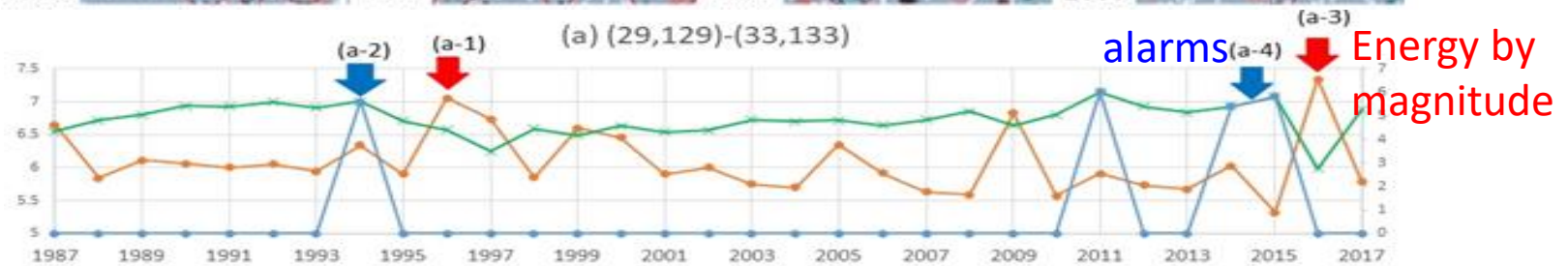
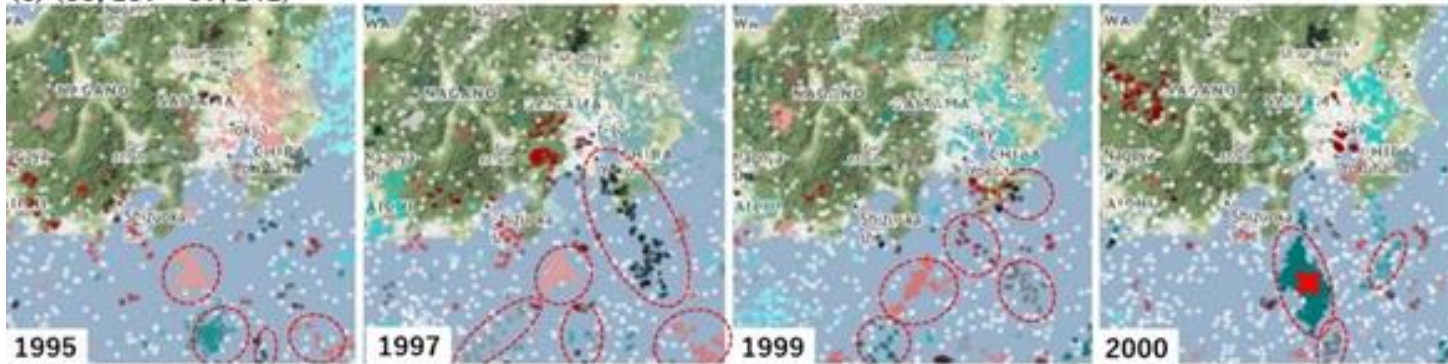


# Case 6. The structural change in the epicenter distribution

(a) (29, 129 – 33, 133)



(c) (33, 137 – 37, 141)



# Case 7. Data Co-creation Project in Yokohama City

- IMDJ/AP have been used for the generation of scenarios to solve public issues in Yokohama City.
- Implemented DJ Store based on the generated scenarios in IMDJ/AP (<https://djp.iais.or.jp/s/djplatform>)



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.

### 「子連れで楽しい街〜ヨコハマ」みなとみらい地区での移動支援に関する実証企画（案）

週末に市民及び観光客が集まるみなとみらい地区において、ランドマークスポット間の移動手段として自動運転実証実験を実施。各乗降場所にベビーカーやレンタル授乳室の貸出事業者と連携し、それぞれのステーションを設置。子育て世代のモビリティをサポート。自動運転車の車内には利用可能な写真データを用意しておき、後部座席からSNSへ感想などをアップしてもらい拡散。

コース全長2.2km

みなとみらいでは馬車なども実施中

車内にSNS映える利用可能な写真（QRコード等）を準備し、利用者に「自動運転+ベビーカー」を拡散

NTT DATA

DADWAY

各乗降場所ごとにベビーカー貸出事業者によるステーションを設置することで乗り捨ても可能。

# Case 8. Data Co-creation Project in Tokyo Marunouchi Area



人を、想う力。街を、想う力。

 三菱地所

 FUJITSU

 SoftBank

 東京大学  
The University of Tokyo

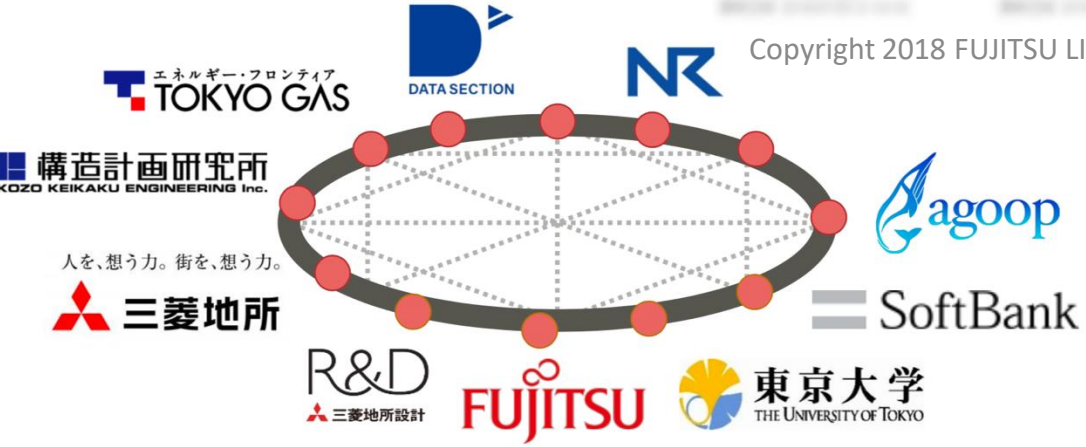
2018年5月14日

報道関係各位

三菱地所株式会社  
富士通株式会社  
ソフトバンク株式会社  
国立大学法人東京大学

**業種を超えたデータ活用で新たな街づくりを目指す実証実験を  
東京・丸の内エリアで開始**

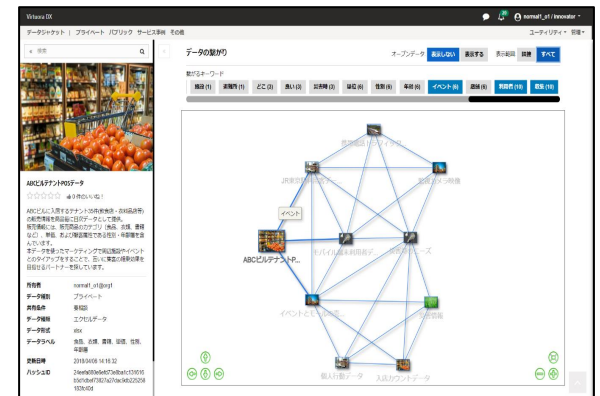
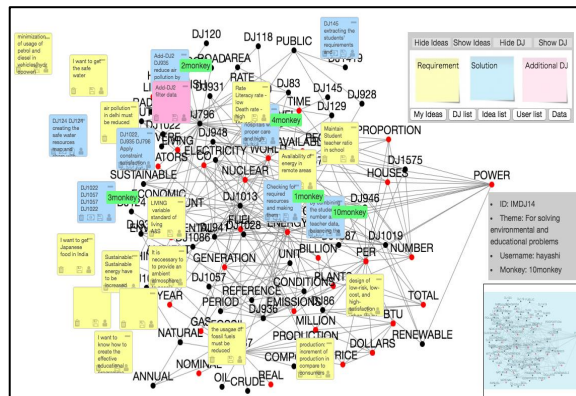
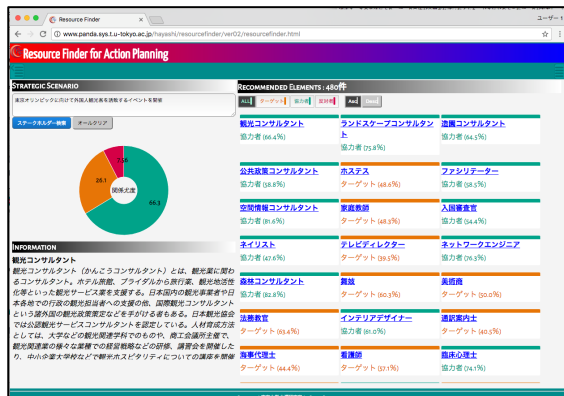
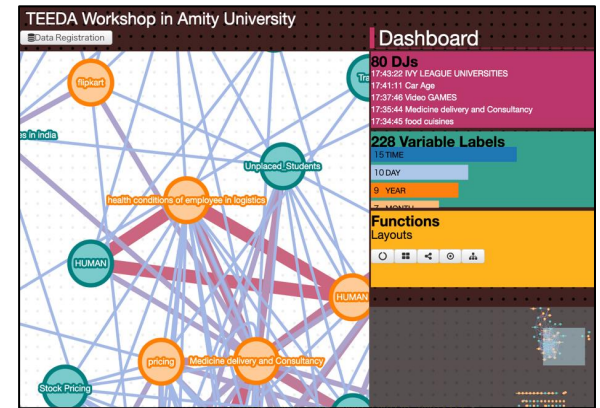
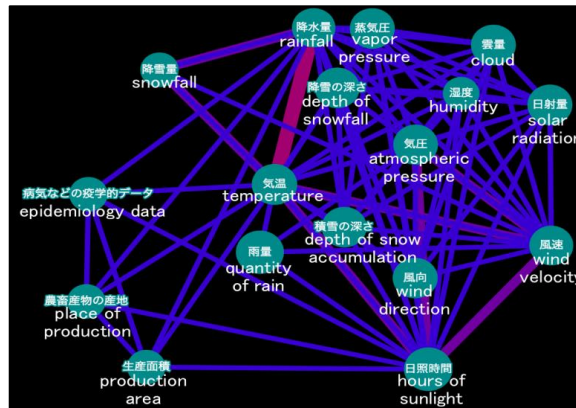
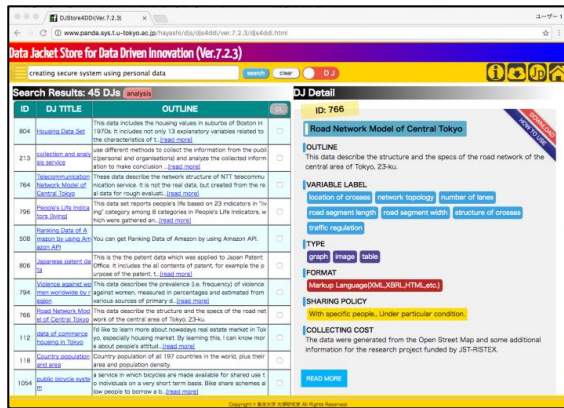
～産学連携でデータ活用による新たな街づくりへ～



- ❑ Create new business by industry-academia collaboration
- ❑ Expand Co-creation with data centered in activities
- ❑ Utilize data jackets for data exchange and utilization platform

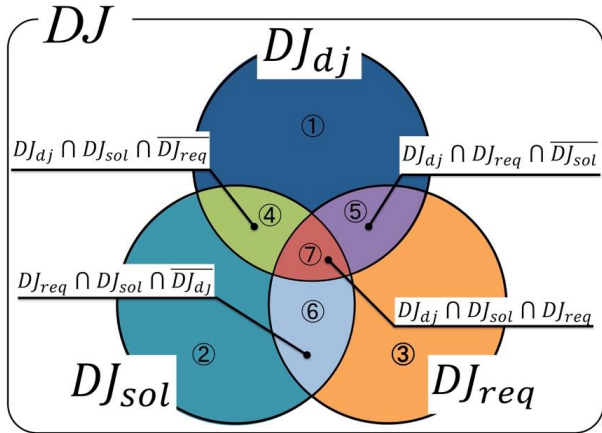
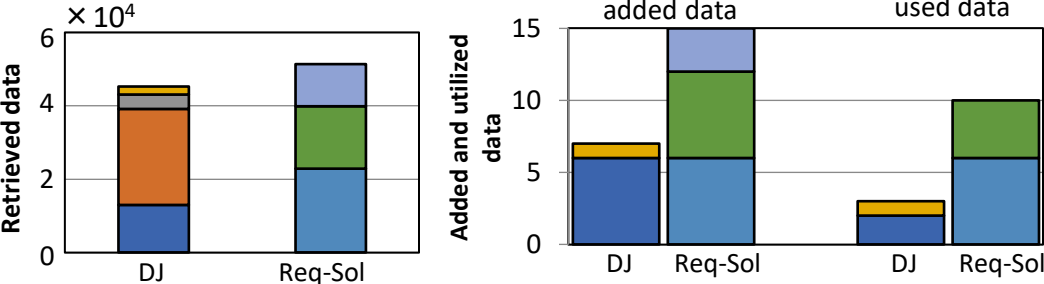
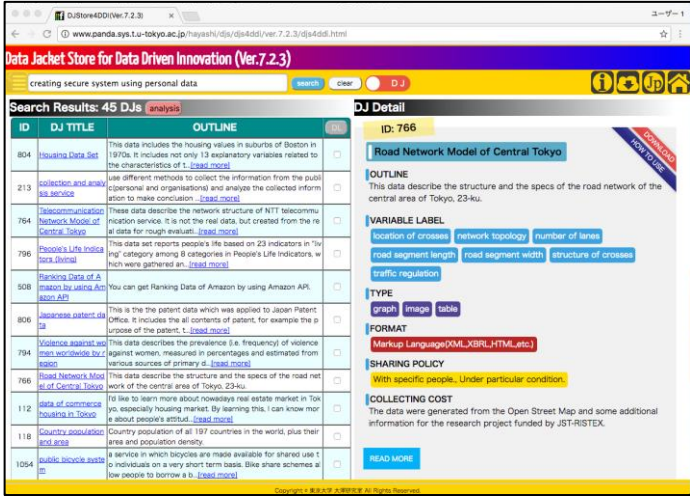
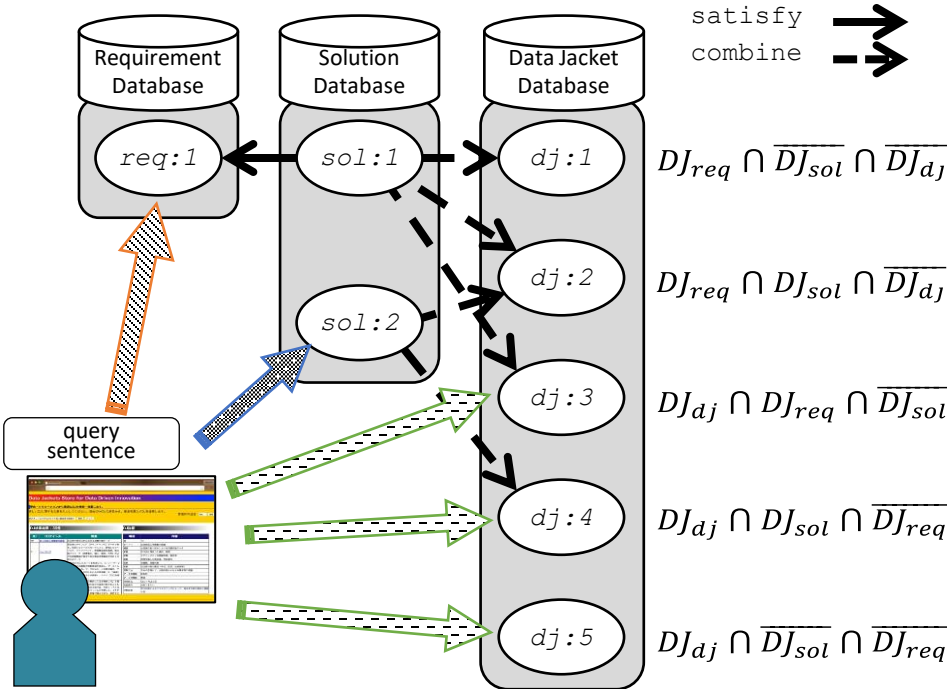
# Tools for aiding the creative thoughts in IMDJ

1. Data Jacket Store
2. Variable Quest
3. Data Matching System: TEEDA
4. Human Resource Finder
5. Web-based IMDJ
6. Virtuora DX



# Tool 1. Data Jacket Store

Developing a data retrieval system "DJ Store" taken into account the knowledge gap between stakeholders (data owners, providers, or consumers) involving structured knowledge for data utilization.



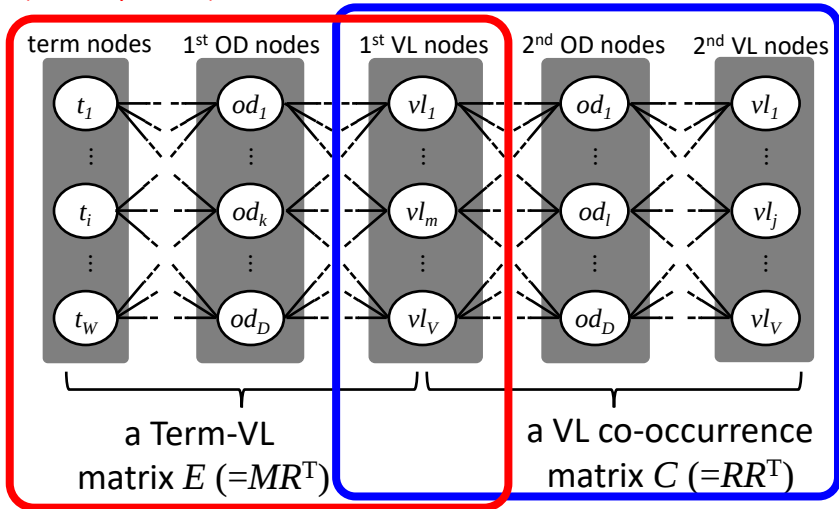
- Hayashi, T., Ohsawa, Y.: Knowledge Structuring and Reuse System Design Using RDF for Creating a Market of Data, SPIN2015, pp.607-612, 2015.
- Hayashi, T., Ohsawa, Y.: Retrieval System for Data Utilization Knowledge Integrating Stakeholders' Interests, AAAI Spring Symp., 2018.

# Tool 2. Variable Quest

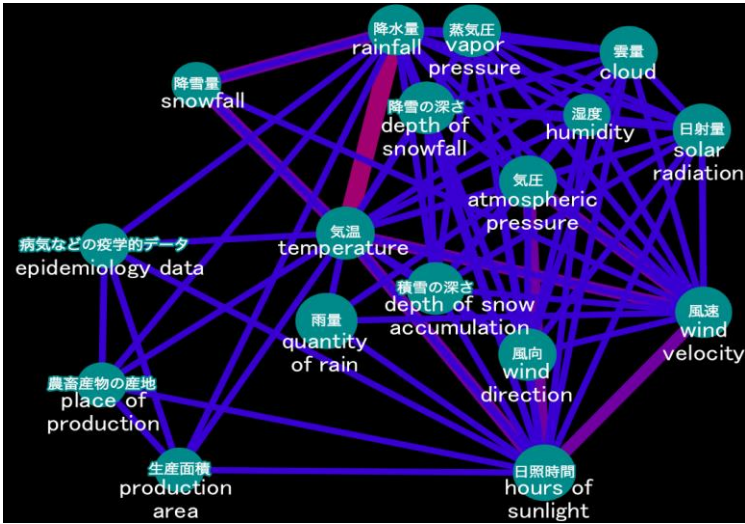
- ❑ **VARIABLE QUEST (VQ)** is a matrix-based inferring method of **variable labels (VLs)**, which are the names/ meanings of variables in Data Jackets.
- ❑ VQ infers related VLs from the free text queries whose VLs are missing or unknown, focusing on the similarity of outlines of data and the co-occurrence of VLs.

Japan weather data (temperature, weather etc.)

Considering Model 1  
(similarity of ODs)



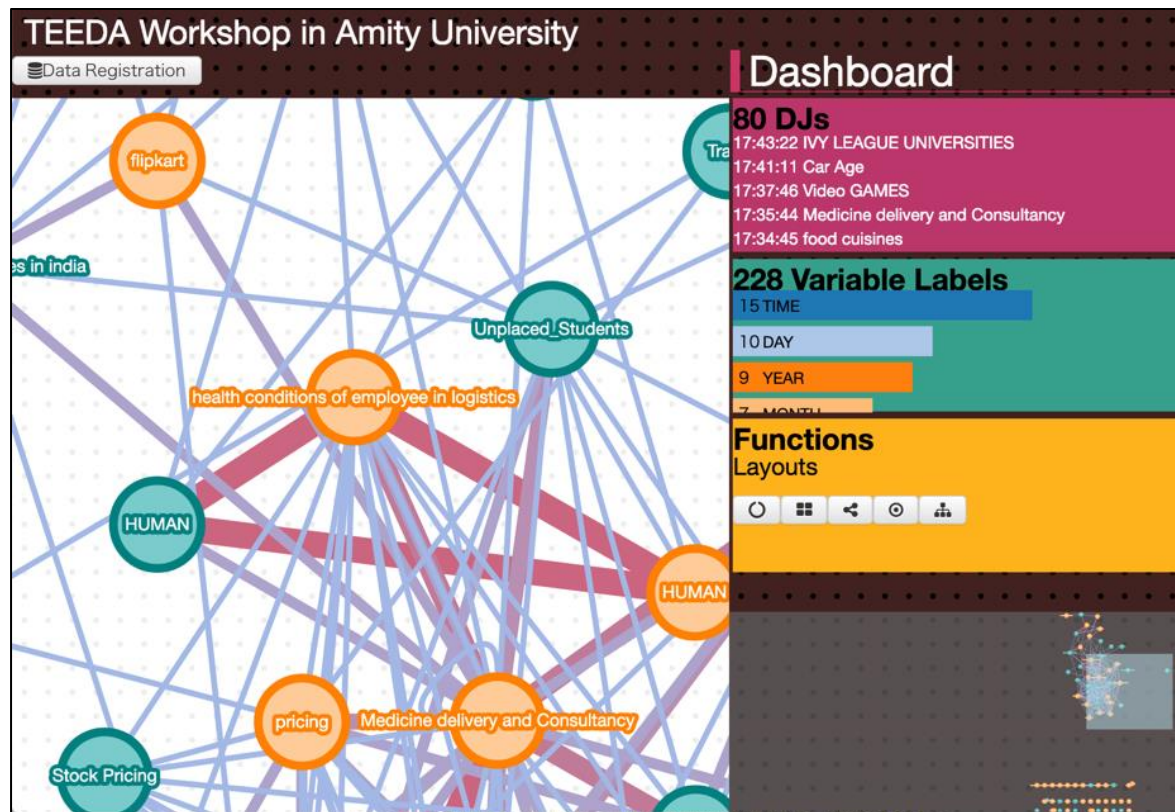
| Inferred VLs      | Similarity | Inferred VLs      | Similarity |
|-------------------|------------|-------------------|------------|
| weather           | 0.318      | solar radiation   | 0.293      |
| hours of sunlight | 0.313      | depth of snowfall | 0.293      |
| rainfall          | 0.307 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



1. Hayashi, T., Ohsawa, Y.: Matrix-based Method for Inferring Variable Labels Using Outlines of Data in Data Jackets, The Pacific-Asia Conference on Knowledge Discovery and Data Mining 2017 (PAKDD2017), 2017.
2. Hayashi, T., Ohsawa, Y.: VARIABLE QUEST: Network Visualization of Variable Labels Unifying Co-occurrence Graphs, IEEE-ICDM Workshops 2017, pp.577-583, 2017.
3. 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.

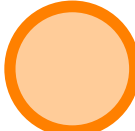
# Tool 3. Data Matching System: TEEDA

- ❑ Developing a matching system which externalizes needs of the data users and visualizes the potential linkages with the data of the data holders
- ❑ TEEDA: Treasuring Every Encounter of Data Affairs



## Node

 : Data wanted

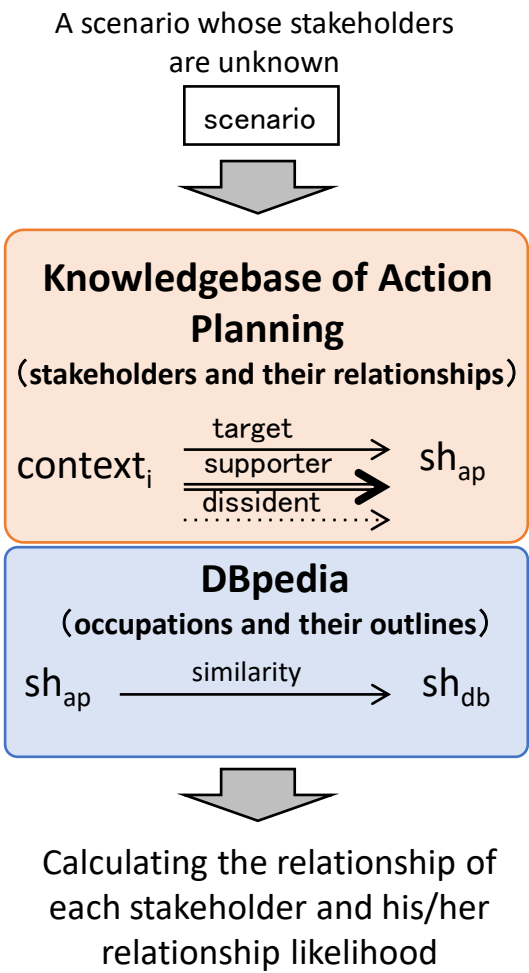
 : Data provided

## Link (Edge)

 : A link is assigned when data have a common variable with each other  
(Thickens according to the number of common variables)

# Tool 4. Human Resource Finder

- Even if the same stakeholders, the relationships of them with a business might change depending on a context the business.
- Human Resource Finder is the recommender system of stakeholders and their relationships with a data business scenario.





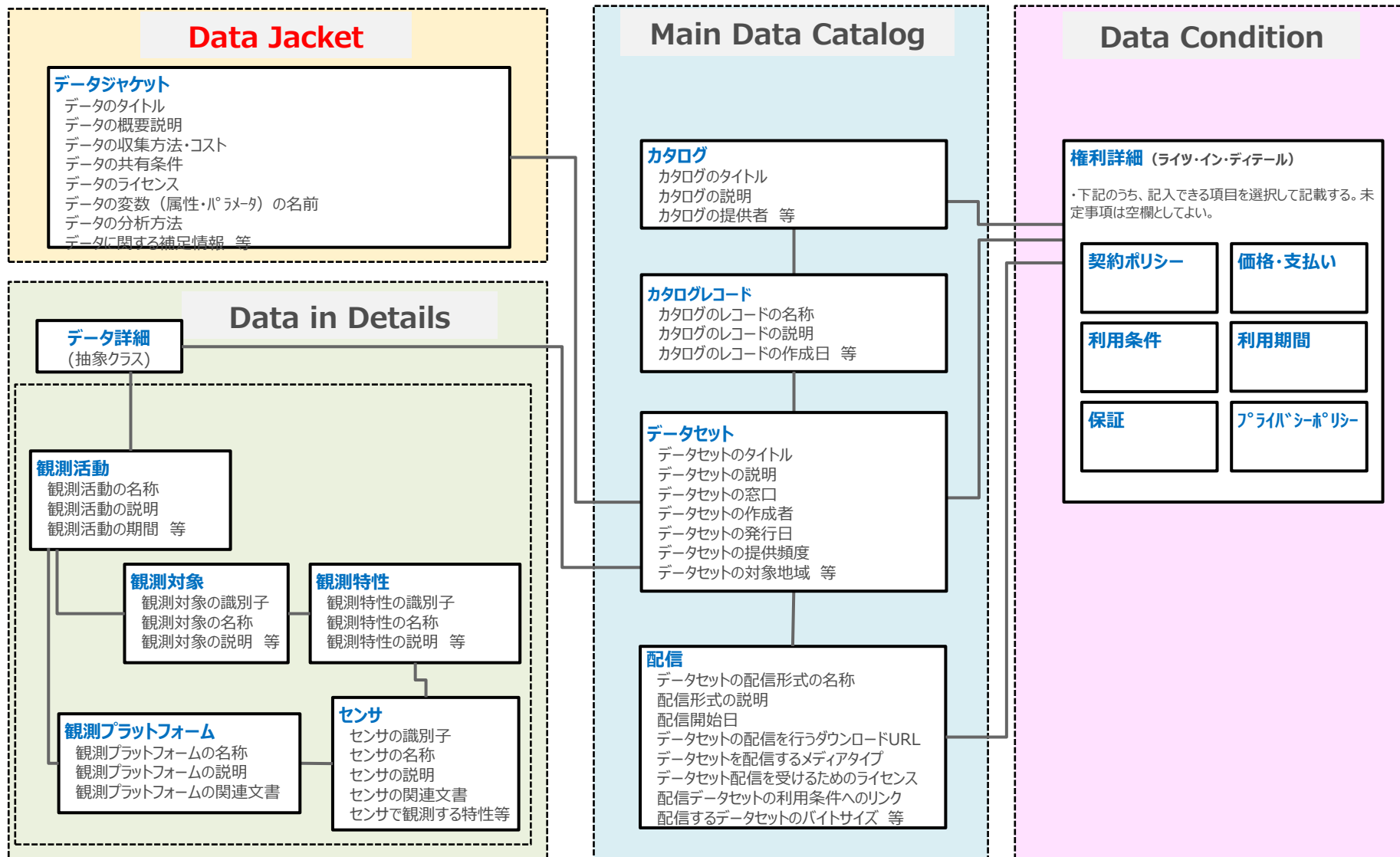
# Tool 6. Virtuora DX by Fujitsu

- ❑ Digital transformation system Virtuora DX introduces a part of DJs and DJ Store functions.
- ❑ Fujitsu deploys IMDJ/AP as a consulting service based on Virtuora DX.
  - ❑ Data Co-creation Project in Tokyo Marunouchi Area
  - ❑ Marunouchi Data Consortium
  - ❑ And so on...

The screenshot displays the Virtuora DX web application. The top navigation bar includes 'データジャケット | プライベート | パブリック | サービス事例 | その他' and a user profile 'normal1\_o1 / innovator'. A search bar is on the left. The main content area is titled 'データの繋がり' (Data Connection) and features a grid of data categories: 施設 (1), 避難所 (1), どこ (3), 良い (3), 災害時 (3), 単位 (6), 性別 (6), 年齢 (6), イベント (6), 店舗 (6), 利用者 (10), and 救急 (10). The 'イベント' category is selected. Below the grid, a network diagram shows connections between various data points, including 'イベント', 'JR東京駅利用客...', '監視カメラ映像', 'モバイル端末未利用者...', '災害時のニーズ', '災害情報', '入店カウンタ...', '個人行動データ', and 'イベントとモールの売...'. On the left, a detailed view of 'ABCビルテナントPOSデータ' is shown, including a description of the data, its format (xlsx), and its update time (2018/04/06 14:16:32).

Copyright 2018 FUJITSU LIMITED

# DJ as a Standard for Data Exchange with Trust

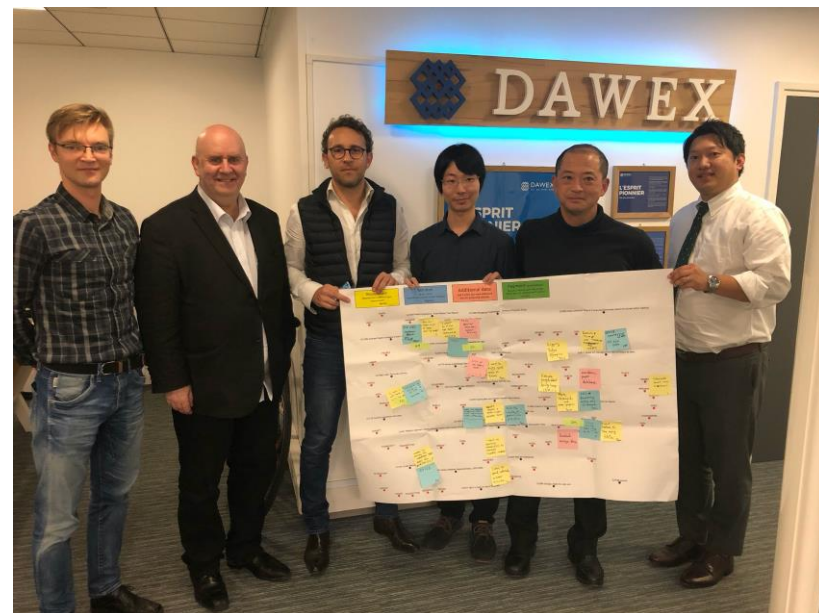


# 付録) データジャケットの国際的外部評価

## -Dawex社における事業者視点における実験による評価-

- Dawex社は、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。
- 2020年1月に同社を訪れ、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。

- データ市場ビジネス概説 (Dawex社)
- IMDJに関する説明 (以下、東大)
  - ◆ 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データ会社、丸の内データコンソーシアム等の紹介
- IMDJワークショップ
- Web-IMDJ、DJストア、Variable Quest、TEEDAのデモ



- IMDJにおけるデータの追加、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、人の思考を更新してゆくインタラクティブフェーズに興味が寄せられた。
- 支援技術、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技術面について多く議論した。
- 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はこの観点でDawexシステムと相性が良い。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。