Innovators Marketplace® on Data Jackets

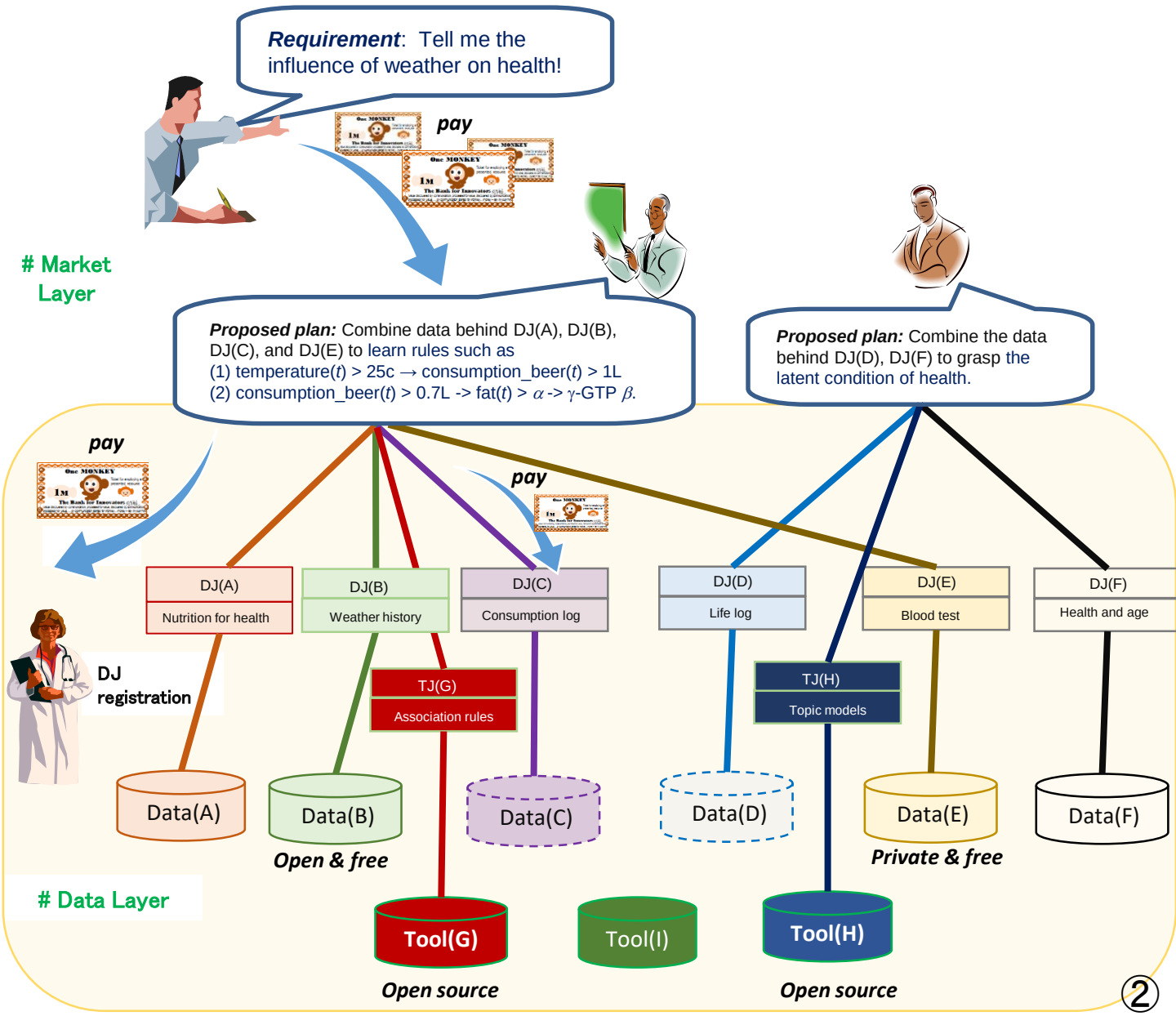
(Ohsawa, Kido, Hayashi, Liu, Data Jackets for Synthesizing Values in the Market of Data, *Procedia Computer Science* 22, 709-716 (2013)) : *Photo*: a snapshot from METI project

Contact: Ohsawa Lab, The University of Tokyo
info@panda.sys.t.u-Tokyo.ac.jp

Innovators Marketplace on Data Jackets as a place for creating scientific solutions

To make a social environment where analysts and actors in businesses and sciences can get data they need, we are on the way to (re)design the ***Market of Data***, where users and/or providers of data can externalize and share the value of data via buying/selling in reasonable conditions, e.g., for a reasonable price – or free it can be expected that the data give merits to everyone rather than to particular rich people.

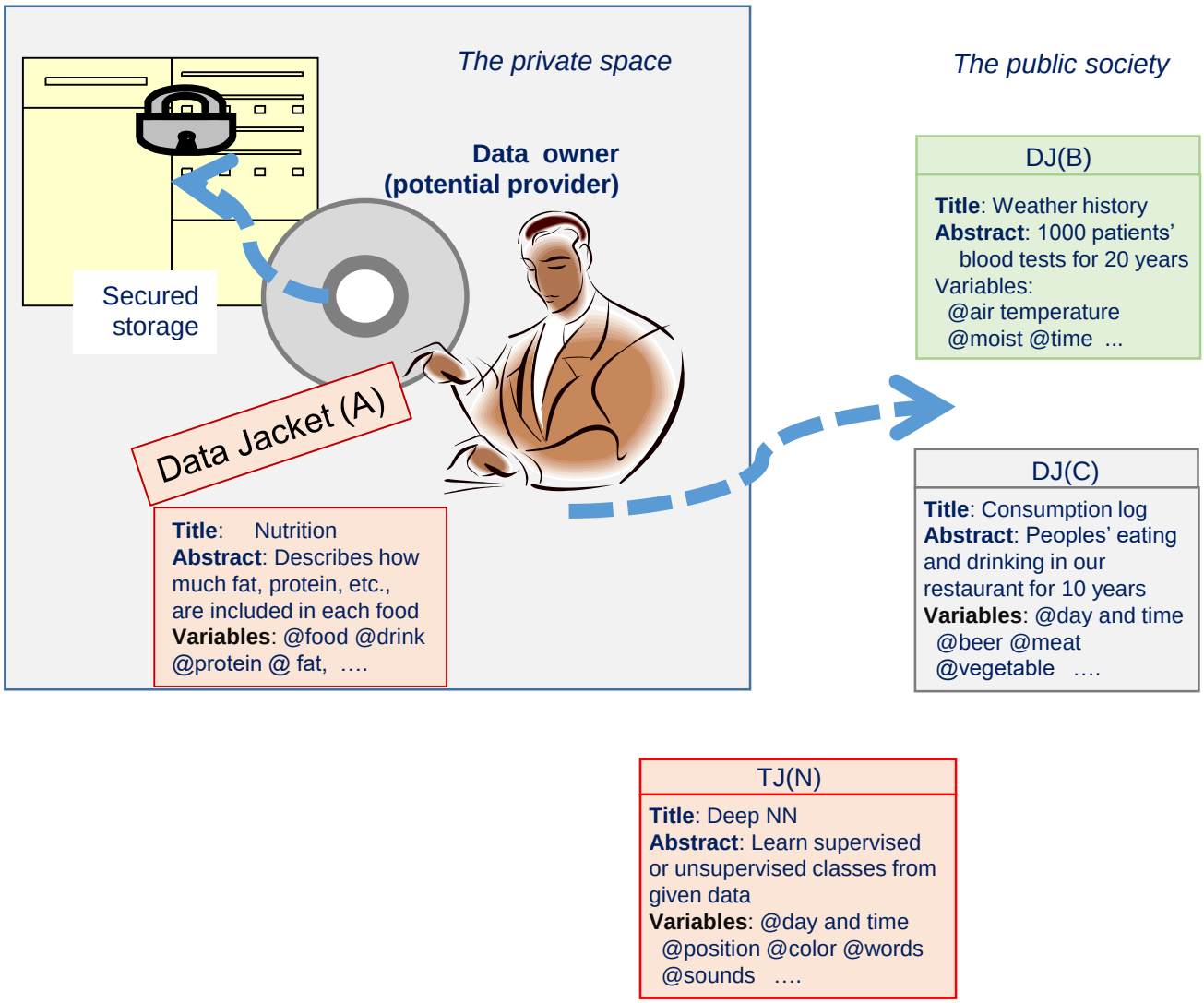
The ***Innovators' Marketplace on Data Jackets*** (**IMDJ**) is a systematic design of the market of data, where tools and new technologies are combined to achieve data driven innovations via communications about and for requirements in the market.



Procedure of Innovators Marketplace on Data Jackets (IMDJ)

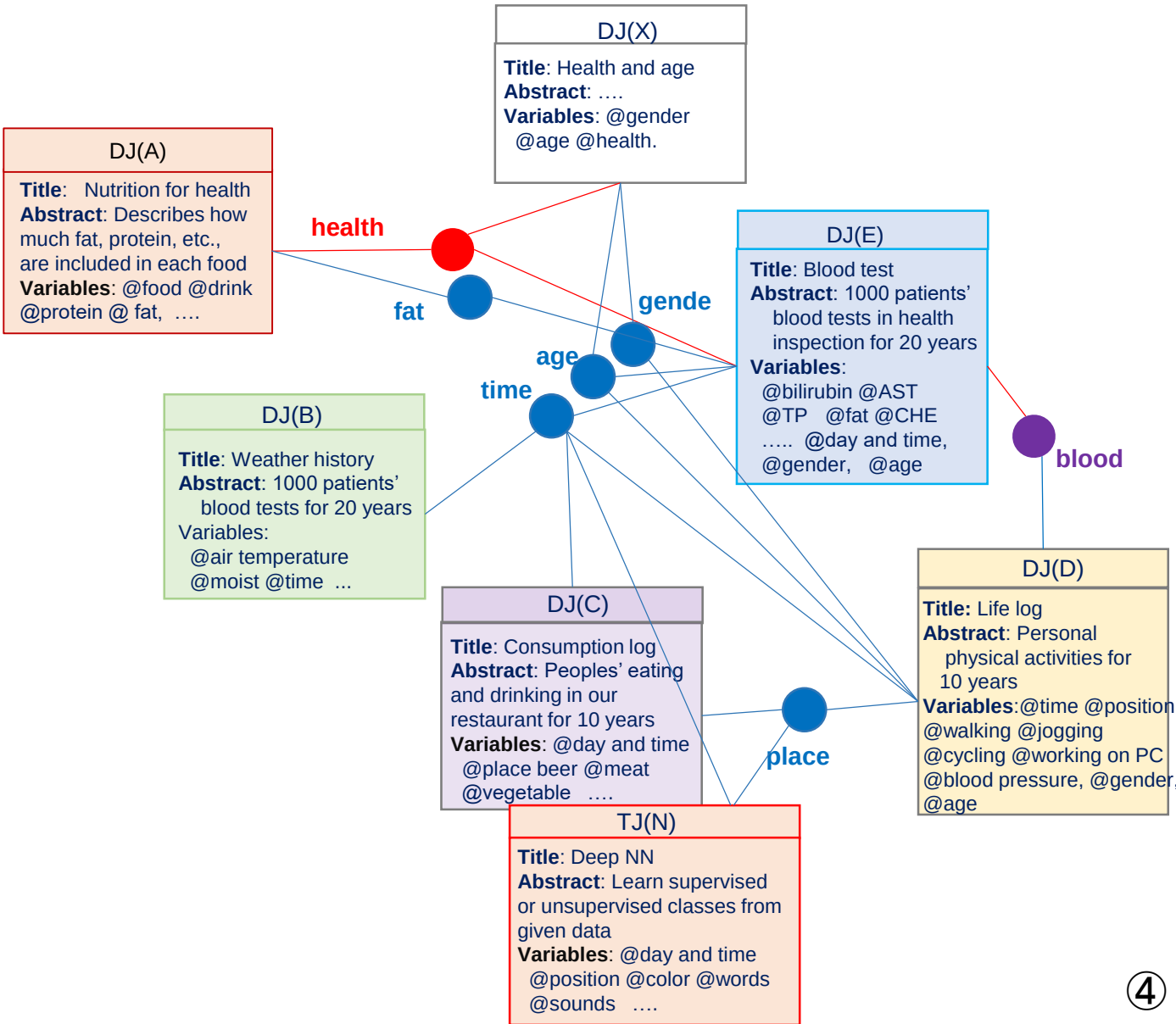
(*Step 1*) *Preparation*

The owner of data may hide and lock one' own data somewhere. Only digest, i.e., the abstract and the names (*not* the values) of variables in the data are put on DJs and shared (DJ: a Data Jacket)



(*Step 2*) *Visualization*

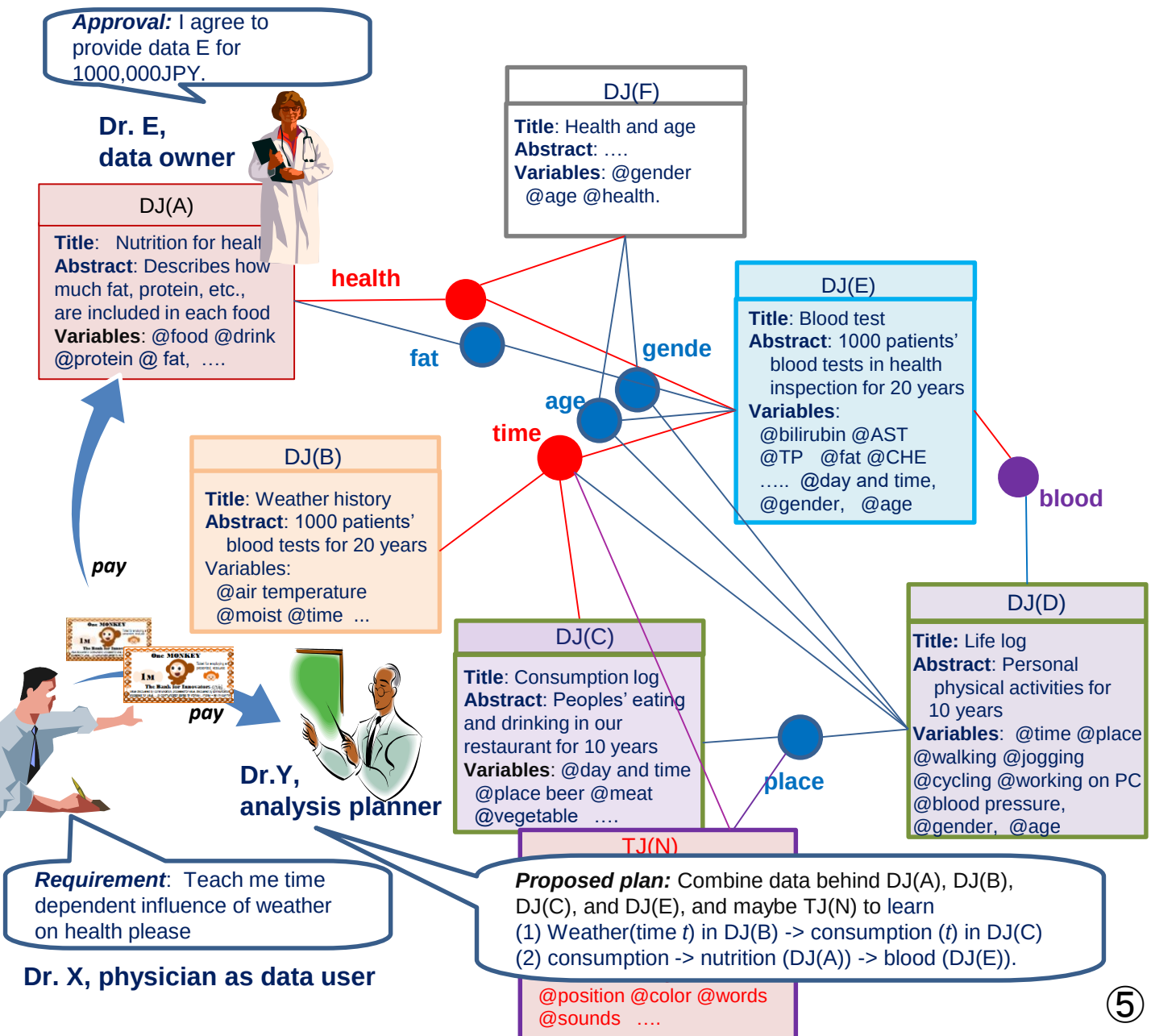
DJs are linked via features (variables, words, etc). Known relations among features of data, may be also described by the owner of the data. One may explicitly declare links to other data (e.g., by RDF) which can be reflected to visualization.



Procedure of Innovators Marketplace on Data Jackets (IMDJ)

(*Step 3*) *Communication, Evaluation, and Transaction*

Owners of data, users of data, and analysts (or planners of analysis) communicate to evaluate the value of data, considering utilities i.e., user values.



The gaming rule of IMDJ

Visit us at
https://sites.google.com/site/datajackets/

Provide your Data Jackets and/or Tool Jackets, to find you business opportunities.

	Each plays the role of a stakeholder of the goal we set for today's IMDJ		
Game start (5 min.)	***user***: business people, habitants, etc.	***creator*:** creative data broker, consultant, or data scientist	***domain expert*:** data owner
	・Declare your domain of business (as real as possible), including "students" or "just consumers". ・Receive the initial property (10M)		
On the way of the game DJ: Data Jacket	• Present **requirements**/ criticisms to others' ideas: In presenting a requirement, one should **speak and put** the filled ***yellow sticker*** at a position on the game board close to relevant words. • Pay for ideas with pricing by negotiation. Payment to inventors means consultation fee, a sign of your interest in the idea's creator and owners of data corresponding to the used DJs. • Put a ***small sticker*** with your own name, on the idea or DJ you buy	Wait 10 minutes • Propose an **idea, i.e., an analysis plan** to satisfy consumers' requirement, **orally putting** a ***blue square sticker*** written "know/do…. by combining ①, ③, ⑬, … (the ID numbers of DJs/TJs)" at a position of the game board close to corresponding DJs and/or requirements. • Add new DJs with ***red stickers*** if necessary • Put sequential numbers to the ideas	• Recommend the ***data*** **(*tool*)** represented by your DJ (TJ) for creators or consumers. • Revise/add DJs/TJs, reflecting the negotiation
Ending of the game	Make a presentation about DJs and TJs you bought, to show you expect some merit by the purchase.	Compete on the total property (the amount of money you get in the game) .	

Case 1: Exchange of data and thought for traffic safety

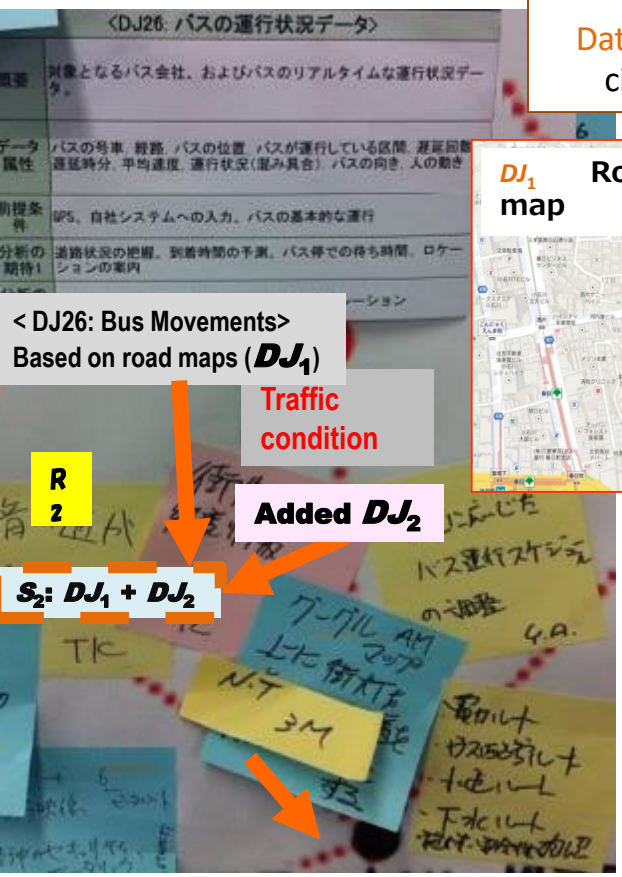
Realized by Teruaki Hayashi and Kenshin Ikegami

Requirements are presented first, to which solutions are presented by combining DJs. DJs may be added if necessary.

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

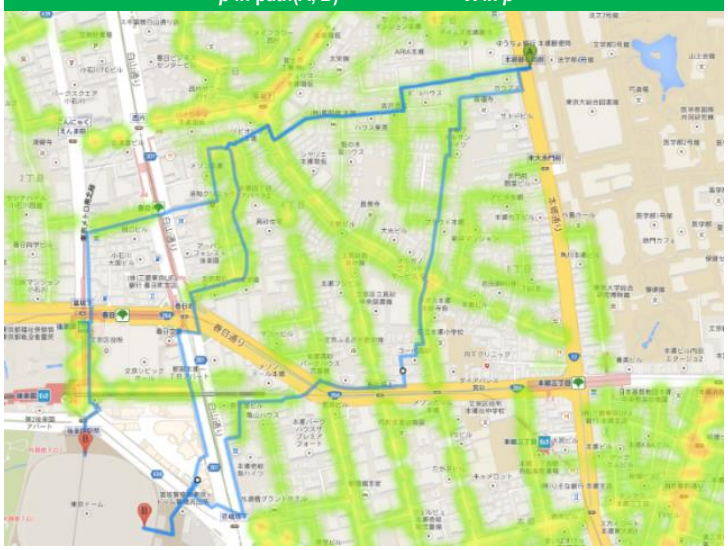
Data for realizing S_2: {DJ_1: map data, DJ_2: Location of city lights}



照明種別c	管理No	柱種c	適合ランプc	町名c	丁目	番地	号	緯度	経度	光束
街路灯	1	ポール大型	CM220	湯島	1	13	2	354209902	1394603203	19800
街路灯	2	ポール大型	CM220	湯島	1	12	5	354209792	1394601561	19800
街路灯	3	ポール大型	CM190	湯島	1	12	3	354209675	1394559999	20800
街路灯	4	ポール大型	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	ポール大型	CM220	湯島	1	11	10	354210382	1394557204	19800
街路灯	7	ポール大型	CM220	湯島	1	7	12	354210922	1394554851	19800
街路灯	8	ポール大型	CM220							19800
街路灯	9	ポール【139	HF125							6000
街路灯	10	ポール大型	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							6000
街路灯	16	東電共架	HF125							6000
街路灯	17	ポール【139	HF200							9900
街路灯	18	東電共架	HF125	湯島	1	7	9	354209141	1394554870	6000
街路灯	19	東電共架	LED33W	湯島	2	30	12	354223533	1394556237	2635

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.

Case 2: Tangled String, born in IMDJ

Presented by Yukio Ohsawa and Teruaki Hayashi http://www.panda.sys.t.u-tokyo.ac.jp

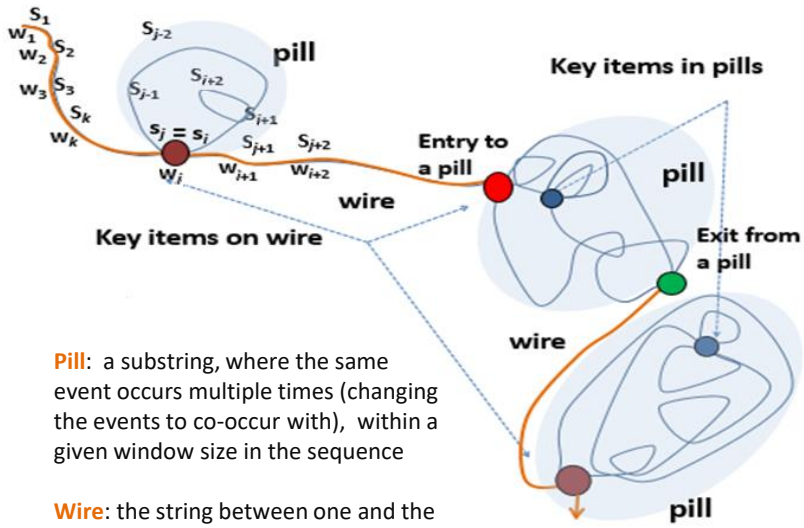
IMDJ is the engine to create new tools and analysis scenarios. Requirements are casted, and solutions are presented and evaluated by the extent to meet the requirements in IMDJ.

The origin of Tangled String, in IMDJ

Requirement R_1: Collect credible and persuasive information.
Solution S_1 : Extract high impact information
Data (Jackets) for realizing S_1: **{DJ_1: Text of communication, DJ_2: facts for supporting/negating messages}**

Tangled String, as a product of IMDJ: a tool for detecting switches of trends. This method fits, if the latent dynamics are hard to explain based on the assumption that each segment of time is ruled by a limited focus of topics.

Ohsawa and Hayashi, "Tangled string for sequence visualization as fruit of ideas in innovators marketplace on data jackets", Intelligent Decision Technologies, to appear in 2016



Pill: a substring, where the same event occurs multiple times (changing the events to co-occur with), within a given window size in the sequence

Wire: the string between one and the next pill

Detection of the end and the start of a pill:

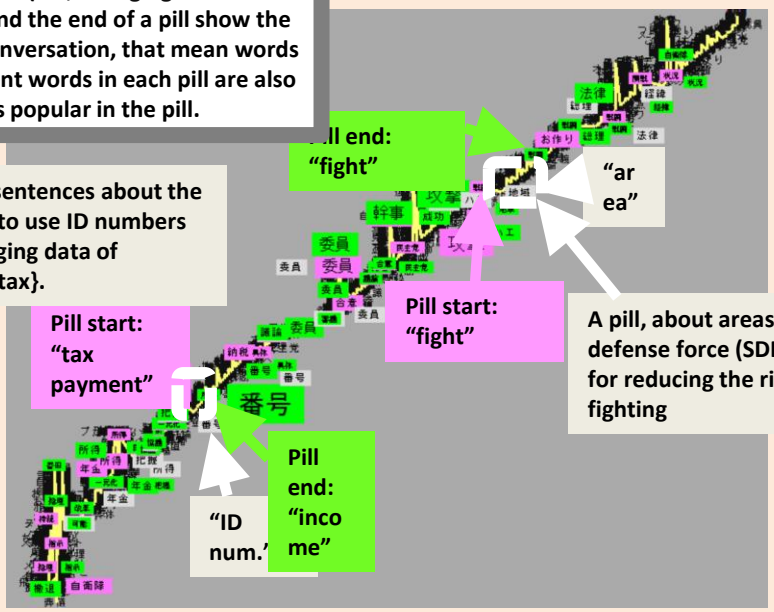
(1)The tangled string is created, by making each item (s_j) in the string take the same position as a previous item (s_i) within N past neighbors of the same token (name of the item e.g., s_i : "hello", s_j : "hello").

(2)Each multiply connected segment (i.e., where each pair is connected via multiple paths) is extracted as a pill

(3)For each pill, all included items (tokens) are assigned the same ID number (ID is w_i for item s_i , s_{i+1} , s_{j-1} , … , s_j in this figure).

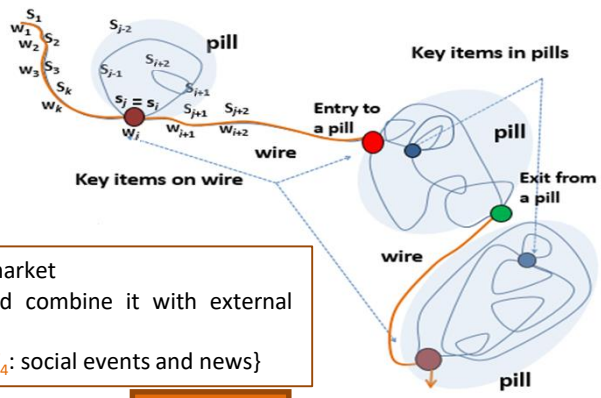
(4)An item s_i in a pill, of which the ID is different from the previous (next) one s_{i-1} (s_{i+1}) , is the start (end) of the pill.

A pill means a segment where same words are repeated with involving various interests (i.e., changing the events to co-occur with). The start and the end of a pill show the opening and the closing of conversation, that mean words that switch the trend. Frequent words in each pill are also shown, to represent concepts popular in the pill.



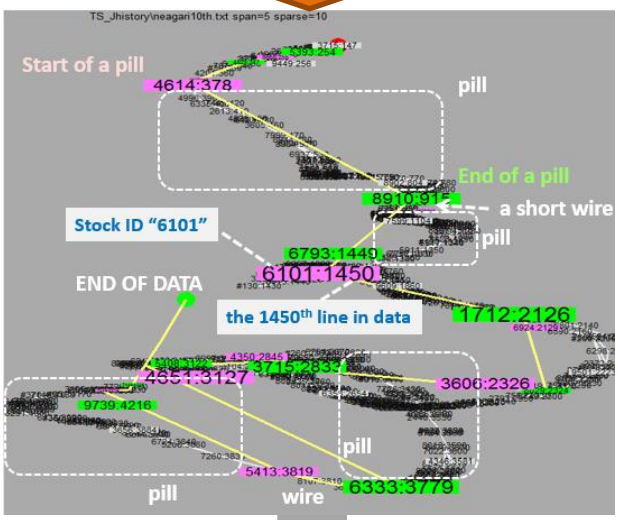
Tangled String, Applied for the dietary debate by Prime Minister Koizumi and Mr. Okada debate (2004)

Case 3: Tangled String, reused in finance and retail via IMDJs



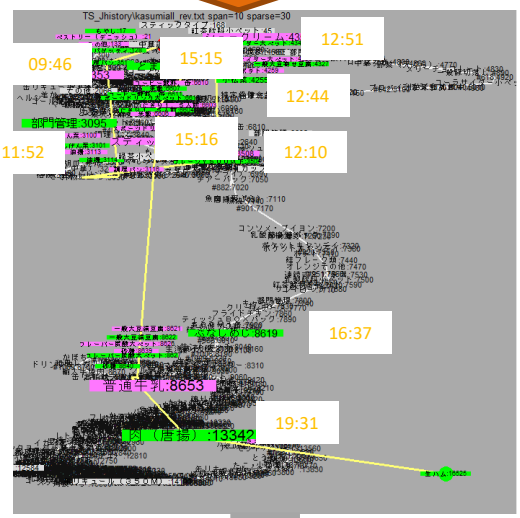
Requirement R_2: detect tipping points of consumer behaviors in the market
Solution S_2: obtain a high-impact event in a sequence by TS and combine it with external information
Data for realizing S_2: {DJ_3: log of consumptions or purchase history, DJ_4: social events and news}

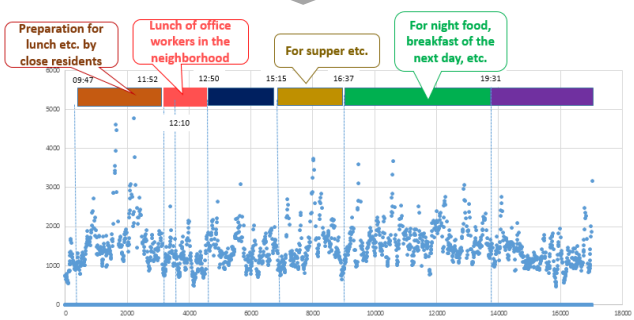
Stock market



The strength of TS: supported by the quantitative financial analyst ranked No.1 (T. Yoshino) in Japan: "TS will be quite a useful method for catching the timing for selling. The results in this figure are useful from the viewpoint of financial analysis."

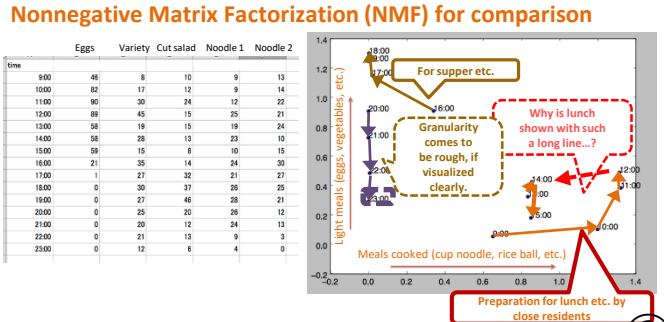
Super market





The strength of TS: supported by managing staff members of the supermarket, who tell "the switching points of segments correspond to our daily experiences," and "the information shown by TS is as fine as 10 minutes of granularity, which is useful for controlling the shifts of selling workers in the retail store."

Nonnegative Matrix Factorization (NMF) for comparison



* NMF failed to show the up and down of prices in the case of stock market

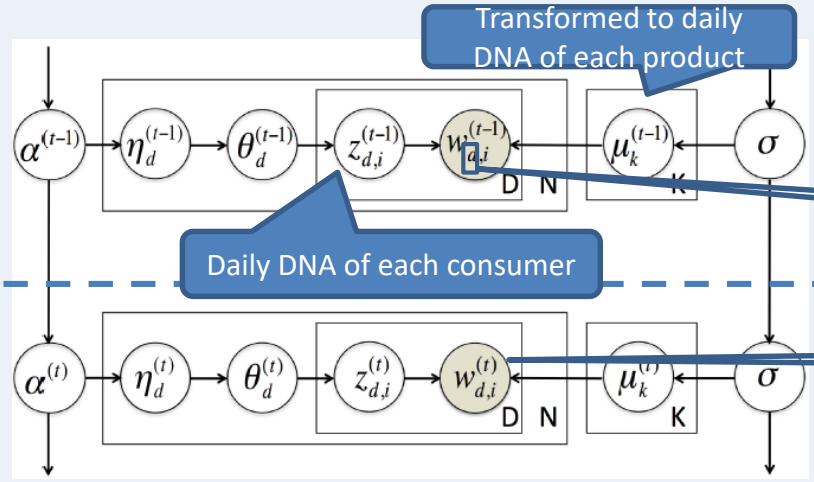
Ohsawa, Y.,、"Tangled String Diverted for Evaluating Stock Risks - A by Product of Innovators Marketplace on Data Jackets" MoDAT2015 (Workshop on Designing Safe and Secure Life on the Market of Data), in IEEE International Conference on Data Mining, Atlantic City (2015)

Case 4: Dynamic Product DNA and Customers DNA

Emoto, M., and Ohsawa, Y., Proposal of extracting purchase behavior and product DNA using Topic Model, IEICE-AI2016-33, IEICE-116, no.460, pp.51-55 (2017)

Requirement R_2: detect tipping points of consumer behaviors in the market
Solution S_2: Extract changes in topics of customers behavior, using Dynamic Topic Modelling (DTM) to extract document-topic distribution. At the same time, extract changes in product features, by extracting term-topic distribution.
Data for realizing S_2: {DJ_3: log of consumptions or purchase history (POS), TJ_1: DTM (Blei, 2006)

Blei's Dynamic Topic Modeling: an existing tool for extracting dynamic changes in topics from a sequence of documents. This is based on the dynamic Bayesian model, where distribution of topics are assigned to each document (corresponding to each purchase basket including items) and each term (corr. purchased item), with modeled temporal continuity.

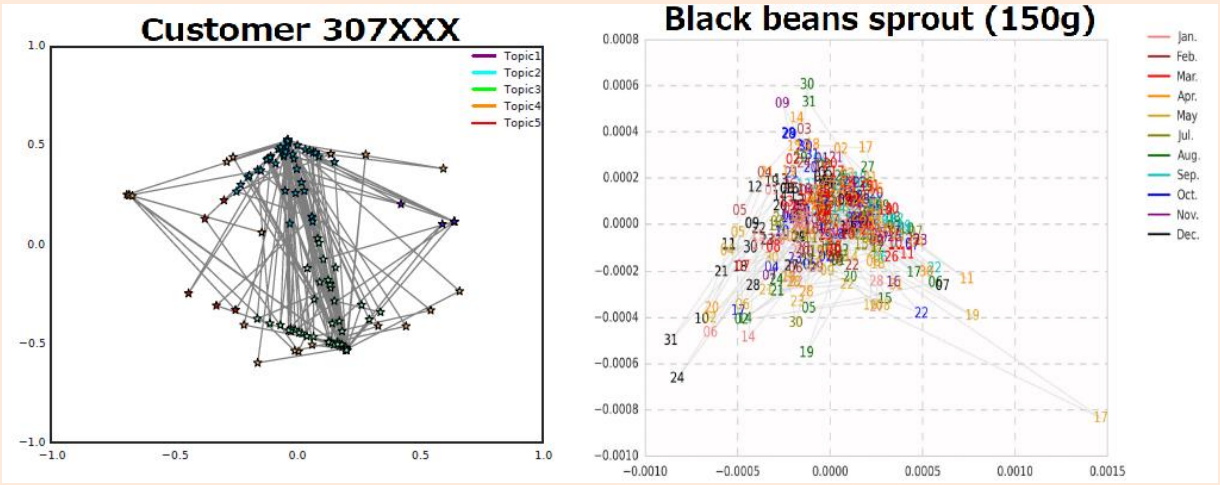


α : parameter for setting topic distribution, with temporal continuity
θ_d : Topic distribution (probability of each topic) of document d
$Z_{d,i}$: a topic of the i th word of document d
$w_{d,i}$: the i th word of document d
μ_k : the word vector of the k-th topic
σ : std of the change of μ_k from time t-1 to $t$ga

Day 1:
d_1: beer, wine, cheese
d_2: water, chips
d_3: eye drop, milk
d_4: socks, t shirt

Day 2:
d_1: milk, water, bread
d_2: magazine, chips
d_3: coke, chips,
d_4: Sandwich, juice

Visualization on POS data



Shifts of consumers' interests (i.e., DNAs) are now explained visually from product DNA

Your Solutions

Solution	Data Jackets (DJs)					Tool Jackets (TJs)		

Your Purchase

Your merit	**Solutions and/or data you bought**		