

# データの価値と利用方法発見のための創造的コミュニケーションとメタデータ記述方法の提案

## Creative Communication and Description of Meta-data for Discovering Latent Values of Data

早矢仕晃章\*<sup>1</sup>  
Teruaki Hayashi

大澤幸生\*<sup>1</sup>  
Yukio Ohsawa

\*<sup>1</sup> 東京大学大学院 工学系研究科  
Graduation School of Engineering, the University of Tokyo

By the advances of information devices such as smart phones, the amount of data acquired daily has become enormous. It is important to obtain information for the decision making by discovering the value of data and utilizing them. In the current situation, although a movement to open data is coming out, a lot of data have not made public yet. In this paper, we propose a method to discover the latent value of data, by Innovators Marketplace on Data Jackets (IMDJ). IMDJ is the workshop-styled game for discovering the value of data and finding the solutions, by creating new ideas by combining existent data. The data referenced in IMDJ are not necessarily open. The data are described on Data Jackets (DJs) which are the digests of data. It is possible to visualize the relationship of the data, whose contents are not open, when they are described as DJs, and it supports players of IMDJ to discover the value of data. In this study, we analyze the communication of IMDJ, and clarify the characteristics of the data which are highly evaluated.

### 1. はじめに

2012 年から、ビッグデータという言葉が流行し、データという言葉が社会的に認知されるようになった。データは日常生活の様々な場面で用いられるようになり、スマートフォンを始めとするパーソナルな情報端末の普及により、より多くのデータが取得できるようになってきた。このように膨大かつ様々なデータが日々蓄積されるようになったことで、データの利用方法と新しい価値を発見し、意思決定に役立てることが重要となってきた。しかし、POS データなどのデータは必ずしも有効に分析かつ活用されているわけではなく、有用であると考えられるが活用方法が見出されていないデータなどが存在する。また、CKAN\*<sup>1</sup> や Linked Open Data[大向 13]などで世界的にデータを公開する動きが出てきているが、重要なデータは公開されていないのが現状である。

2013 年から大澤は、データの市場を創造するためのデータジャケット[Ohsawa 13]というコンセプトを提案している。データジャケットとは、データに含まれる変数やデータの形式などを説明・要約したデータ、すなわちメタデータである。中身が公開されていなくても、データの説明・要約であるデータジャケットを公開することにより、データの中身を理解し、価値を策定し、データを合理的な値段で売買することが可能になる。本論文は、データジャケットを用い、データの利用方法についてアイデアを創出する Innovators Marketplace on Data Jackets のワークショップを提案し、ワークショップにおけるコミュニケーション及び高評価ソリューションとデータの特徴を見ることで、データの活用方法及びデータの価値が策定されるプロセスについて分析を行う。

連絡先:

早矢仕晃章, 東京大学大学院 工学系研究科 システム創成学専攻, teru-h.884@nifty.com

大澤幸生, 東京大学大学院 工学系研究科 システム創成学専攻, ohsawa@sys.t.u-tokyo.ac.jp

This research was supported by JST, CREST

\*<sup>1</sup> <http://ckan.org/>

### 2. 先行研究

#### 2.1 Data Jackets

データジャケット(以下, DJ)とは, あるデータがどのようなデータであるのかを説明・要約したメタデータである。DJ に記述されているメタデータを読むことにより, データの中身が公開されていなくても, そのデータがどのような形式で保存されたもので, 誰がどのような意図で取得したのか, またデータ取得にかかったコストなど理解することが可能になる。CD や DVD のジャケットのように, 中身については購入しないと参照できないが, 中身に関する説明をジャケットとして記述することにより, 中に入っているコンテンツを理解できるというところから着想を得たコンセプトである。

また, DJ として記述されたメタデータをテキストマイニングのツールなどを用いることにより, 各 DJ 間の公開可能変数を元にデータの関係性を可視化することができる。DJ 同士のつながりを可視化することで, 価値を判断する主体である人間の発想とデータの価値策定を支援することが可能になる。データの価値が定まれば, データの市場が形成され, データ保有者の重要なデータの公開や販売の動機にもつながることが期待できる。

図 1 は例として, 可視化ツール KeyGraph[Ohsawa, et al. 98]を用いてデータの関係性をノードとリンクで表現したものである。KeyGraph は定められたテーマに基づいて集められた情報群(データ)から, 頻度は低い意思決定において重要と考えられるキーワードを抽出し, 関係性を可視化するアルゴリズムである。図では, 黒ノードに DJ が現れ, その横に表形式でデータの中身がメタデータとして記述されている。人間はデータ同士のつながり及び DJ の記述内容から, データの利用方法を考案する。次節では KeyGraph などのツールを用いて作成したシナリオマップを用い, データの利用方法及び価値を策定するための, 市場を模したワークショップ形式のゲームである Innovators Marketplace について説明する。

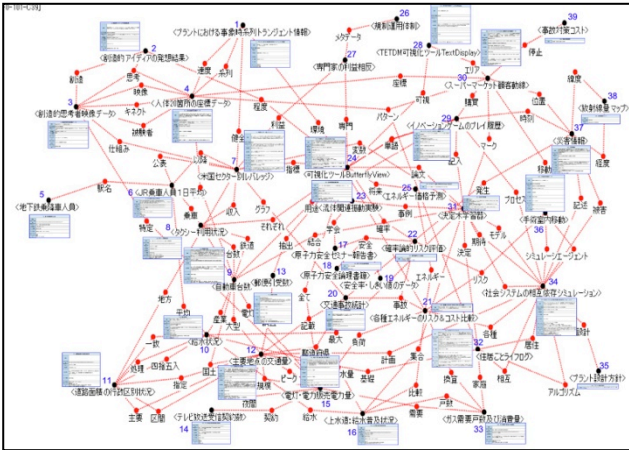


図1 DJ 同士の関係性をシナリオマップとして可視化した例 (可視化ツールに KeyGraph を用いた。黒ノードが各 DJ を表し、赤ノードに現れる単語を介し、DJ 同士がリンクでつながれる。)

## 2.2 Innovators Marketplace

集団によるコミュニケーションを元にアイデア創出及び創造的問題解決を行う手法に、発想法というものがある。Osborn によって考案されたブレインストーミング [Osborn 79] に始まり、KJ 法 [川喜田 67] や NM 法 [中山 80] などが開発され、様々なビジネスやプロジェクトでの創造的アイデア発想に利用され、学術分野での研究も盛んに行われてきた。その後、新しい視点を提供する仕掛けや思考の枠組みを制限することで創造力を発揮するような問題解決方法だけではなく、論理的な思考や客観的なデータに基づいて、人間の意思決定を促す手法が考案され始めた。

Innovators Marketplace (以下、IM) [Ohsawa & Nishihara 12] はアイデア発想に必要な基礎的な情報の収集から、ワークショップにおけるアイデア発想、そして精緻化というアイデア収束プロセス全体を含めた、イノベーションを促進させる方法論である。古典的なアイデア発想法とは異なり、計算機と人間の協創により、意思決定において重要な事象を発見するというモデルによる創造技法である。テーマに基づいて集められた情報群 (データ) から、可視化ツールによりシナリオマップを作成する。参加者はマップ中のキーワードやデータを組み合わせることで、意思決定において重要な判断材料となる事象、状況またはそれらについての情報をアイデアとして創造する。IM では参加者が担当するロールにより様々な角度からアイデアを吟味し、実現可能性や有用性をゲーム中の架空の通貨によって評価するというアイデアを淘汰する仕組みがある。また、専門家やビジネスパーソン、研究者との効果的なコミュニケーションの場としての役割も果たす。発散的にアイデアを創出するだけでなく、ゲーム中のコミュニケーションおよび収束的思考を用いて消費者の要求を考慮したアイデア創出を支援するのが IM の特徴である。

## 2.3 IMDJ (Innovators Marketplace on Data Jackets)

前節で説明した IM は、既存の技術に関する情報をシナリオマップにおいてノードとして表出させることにより、ノード同士の組み合わせから、技術の組み合わせを探索し、創造的な問題解決を行うというものである。2013 年から、IM のシナリオマップにデータの関係性を表出させることにより、データの組み合わせから有用なデータを発見するワークショップとして Innovators Marketplace on Data Jackets (以下、IMDJ) [Ohsawa, et al. 13] が実施されてきている。IMDJ では、メタデータとして記述された

DJ を用いることによって、データ同士の関係性を可視化させている。DJ に記述されたデータの公開可能変数同士の共起性などから、DJ 同士の関係性を導いている。シナリオマップ上にノードとして DJ を表示することで、関連するデータのつながりの理解を促進させ、データの利用方法や価値策定の支援を行う。

IMDJ は以下の手順により進行する。IMDJ のゲーム部分の実施時間は 2 時間程度である。

1. 参加者から DJ を収集する。
2. 可視化ツールにより、DJ の関係性を可視化したシナリオマップを作成する。
3. シナリオマップを A0 大の模造紙に印刷する。
4. 参加者に、データ及びソリューションを評価するための架空通貨を配布する。
5. IMDJ のゲームを開始する (約 2 時間)。
6. 参加者は、「利用者 (消費者または自身の社会的な立場)」から、設定されたテーマに基づいて要求を出す。要求は黄色の付箋に記入し、シナリオマップ上に貼り付ける。この作業を 20 分から 30 分程度行う。
7. 参加者は、「発明者 (問題解決を行うためのアイデアを創出する人)」として、「利用者」が創出した要求を満たすソリューションを提案する。ソリューションは青い付箋に記入し、シナリオマップ上に貼り付ける。ソリューションを提案する際には、問題解決を可能にするデータ、またはソリューションを支持するデータの DJ 番号を記入する。ソリューションを出した「発明者」は、データを保有している「データ保有者」にゲーム中の架空通貨を支払う。「データ保有者」は IMDJ において DJ を記述した参加者を指す。支払う架空通貨の値段については、「提案者」と「データ保有者」の交渉で決定する。自身が「データ保有者」である場合は、架空通貨を払う必要はない。組み合わせるデータがシナリオマップ上に存在しない場合、新たにどのようなデータが必要であるのかを「追加 DJ」として赤色の付箋に記入し、シナリオマップ上に貼り付ける。
8. 参加者は「発明者」として、自分自身の出した要求についてソリューションを創出することができる。その際にも、問題解決を可能にするデータ、またはソリューションを支持するデータの DJ 番号を記入する。ソリューションは白色の付箋に記入し、シナリオマップ上に貼り付ける。
9. 参加者は引き続き「利用者」としての立場から要求を出すこともできる。
10. 参加者は「利用者」の立場から、提案されたソリューションをゲーム中の架空通貨により評価する。支払う架空通貨の値段については、「利用者」と「発明者」の交渉で決定する。
11. ゲーム後、参加者は「利用者」の立場から、自身が購入したソリューションを元に、どのように課題が解決されたのかを発表する。「利用者」の勝者はプレゼン後の投票で決定する。「発明者」と「データ保有者」としての勝者は、手持ちの架空通貨の量で決定する。

## 3. DJ のメタデータ記述

本論文におけるメタデータとは、DJ に記述されるデータの中身に関する説明・要約を意味している。DJ は、データを保有するステークホルダーが自身の意思で記述することを原則としている。また、国や各自治体の統計データや、個人が公開している研究データなど、すでにウェブなどに一般公開されているデータについては、一般の参加者が入力してもよい。その際はデータの所在などを URL で明示する必要がある。本研究では、

実験的にウェブ上にデータジャケットの入力フォーム<sup>\*2</sup>を開設しており、専門家やデータサイエンティストなどから広く DJ を募集し、収集している。入力項目は以下の 13 項目である。

1. データジャケット名
2. データ/ツールの概要説明
3. データ/ツールの所有者とその所在
4. データ収集方法やコスト
5. データの共有について
6. データ/ツールに関する追加情報
7. データの種類(時系列, 画像など)
8. データの形式(txt, csv など)
9. 分析・シミュレーションプロセスへの入力(データが含まれる変数/ツールの入力変数)
10. 分析・シミュレーションプロセス
11. 分析・シミュレーションプロセスの結果
12. 分析・シミュレーションプロセス以外に期待する分析
13. 自由記述(データに関する補足事項)

以上の 13 項目の中で、公開可能な部分のみを「データ保有者」は記入し、DJ を作成する。図 2 は DJ 入力フォームを元に作成された DJ の一例である。

なお、DJ の記述にツールを含んでいるのは、ツールはデータを処理するためのデータ群であるためである。

9. 自動車台数	
概要	事業用とは、道路運送法第 11 条により国土交通大臣の免許を受けられた運送事業者がその事業のために使用する自動車をいいます。(1)「被けん引車」とは原動機のない車体のみの車で、トレーラーなど。(2)「特種用途車」とは救急車、消防車、タンクローリー、コンクリートミキサー車、冷凍車など特種の用途に使用される。(3)「大型特殊車」とはロードローラー、フォークリフトのように特殊の製造をしたもの。(4)「小型二輪車」とは総排気量が 250cc を超えるオートバイ。(5)「軽自動車」とは総排気量 660cc 以下(二輪車は 250cc 以下)のものとなります。
在り処・所有者	<a href="http://data.linkedopendata.jp/dataset/yokohama-statistics-book-11">http://data.linkedopendata.jp/dataset/yokohama-statistics-book-11</a>
収集方法やコスト	Web でフリーで提供 資料:国土交通省 関東運輸局 神奈川運輸支局 登録部門、財政局 主税部 税務課
共有姿勢	
追加情報	注 1) 区分した区の台数は、新区になって登録されたものは新区に、それ以前に登録されたものは区分前の区に加算されています。 注 2) 軽二輪の数値については、「市税課課税課」を資料とし、次年度当初の軽二輪と軽三輪の合計です。
データ種類	時系列, 数値
データ形式	表形式 (RDB)
分析・シミュレーション入力	地域名: 年度 H12-H23 総数: 貨物用, 乗合, 乗用ごとに: 普通台数, 小型台数, 被けん引車台数 総数: 特殊については: 特種用途台数, 大型特殊台数, 小型二輪台数, 軽自動車台数

図 2 データジャケットとして表現されるメタデータの記述例

## 4. 実験

### 4.1 実験方法

本論文では、インターネット上に公開している記入フォームから、実験 IMDJ への参加者及び一般のデータ提供者によって入力されたデータジャケットを用いた。利用した DJ は、「自動車の点検履歴データ」、「日本の火山データベース」、「Web サイトのアクセス解析データ」など無作為に 45 個を抽出した。これらの DJ を入力とし、KeyGraph により可視化したシナリオマップを用いて、被験者 12 名による IMDJ を行った。テーマは「2020 年

の東京オリンピックを成功させるためのアイデア創出」である。実験者は実験前に課題内容を被験者に教示し、作業時間は 1 時間程度とした。最初の約 15 分を「要求」創出段階、残りの約 45 分を「ソリューション」提案及び評価段階とした。また、被験者全員の同意のもと、作業中のコミュニケーション内容はビデオカメラによる録画と録音を行い、分析対象とした。

### 4.2 実験結果

各プレイヤーは約 15 分の「要求」創出段階において平均して 2 個の要求を出し、約 45 分の「ソリューション」創出段階においては平均 2 個のソリューションを提案した(表 1)。図 3 は IMDJ 後のシナリオマップである。また、マップ上の小さい付箋は購入されたデータまたはソリューションを指している。

表 1 要求及びソリューション数

項目	総数	付箋の色
要求	24	黄色(大)
他プレイヤーの要求に対するソリューション	22	青色(大)
自身の要求に対するソリューション	2	白色(大)
追加データジャケット	4	赤色(大)



図 3 実験後の IMDJ シナリオマップ

### 4.3 考察

まず、IMDJ 実験中に行われたコミュニケーションに着目し、対話データから、各要求が出されてからソリューションが創出するまでの時間を算出し、グラフ化した(図 4)。グラフの横軸はソリューション創出時間と要求創出時間の差、縦軸はソリューションの評価者数を表す。ここで、横軸の時間を、要求が創出されてからソリューションが創出されるまでの思考の時間を表していると考えると、およそ 700 秒から 1800 秒の間に評価者数が 1 人以上の高評価ソリューションが点在していることが分かる。また、この時間内に存在するソリューションとそれ以外の時間のソリューション創出時間には有意な差は見られなかった。また、要求創出時間においても有意な差は見られない。つまり、IMDJ のゲーム中に要求やソリューションが出された時点によらず、要求として問題提起がなされてから、評価に値する解決方法が創出されるまでには、ある程度の時間が必要である可能性が示唆される。

続いて、ソリューションに用いられた DJ から、元となるデータの価値付けについて分析を行った。本論文では、DJ の利用頻度から、データの価値評価を行うものとした。つまり、ソリューション創出において組み合わせられた頻度を DJ の評価値とし、元

\*2

[https://docs.google.com/forms/d/1GQ-YwAIYoeus7Q\\_woRCD3mKoMIwAJD2zaXAXYwU6Qag/vie wform](https://docs.google.com/forms/d/1GQ-YwAIYoeus7Q_woRCD3mKoMIwAJD2zaXAXYwU6Qag/vie wform)



となるデータの価値とした。そこで、ソリューション創出において1回以上利用されたDJを高評価群、一度も利用されなかったDJを低評価群として分析を行った。シナリオマップ作成に利用した45個のDJに記述されている13項目のメタデータの記述項目数を高評価群と低評価群に分けて比較を行ったところ、有意な差は認められなかった(高評価群平均:6.7, 低評価群平均:6.9)。つまり、メタデータの記述項目における情報量がデータの評価値に影響を与えている可能性は低いことが示唆される。

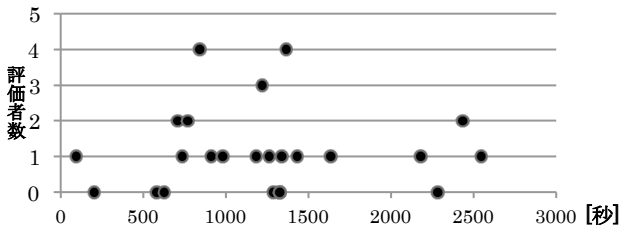


図4 要求とソリューションの創出時間の比較グラフ

続いて、高評価群と低評価群において、データ共有の可否及びデータの形式について比較を行った。図5を見ると、一度以上利用されているDJを含む高評価群では、90%以上が一般に共有できないデータであることが分かる。つまり、ソリューションを創出する上で、一般に共有できないデータに利用価値が見出されているということになる。データ形式については図6の通り、「CSV・XLSなど」の割合が高いことが確認された。

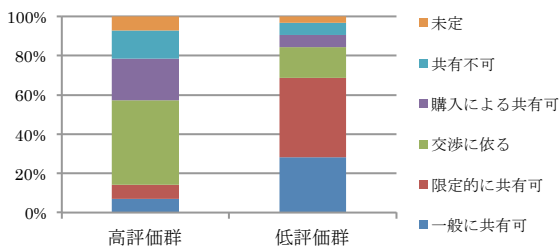


図5 データの共有の可否比較

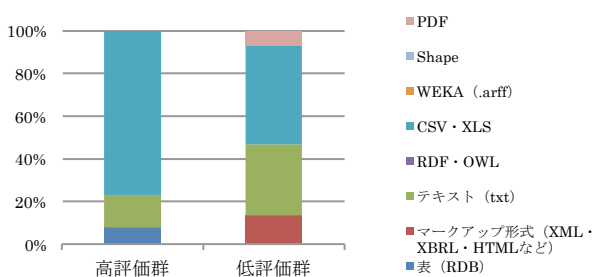


図6 DJのデータ形式比較

## 5. 結論

実験によって、メタデータの情報量が多ければデータの価値が高いという直観的な理解は成立しないことが示唆される。また、IMにおける複数の参加者による問題解決では、コミュニケーションが重要であることが分かっている[Hayashi & Ohsawa 13]。要求が出されてからある一定の時間を経て、ソリューションが提案されたという結果は、要求に対して即座に応えるのではなく、参加者同士の多様な視点により要求を分析し、解決すべき問題を明らかにするという創造的なコミュニケーション・プロセスが

実現していたと考えることができる。データを組み合わせることによりデータの利用方法を発想するというIMDJにおいても同様であると考えられる。

また、本実験により、共有や公開が不可能なデータでも、メタデータとして記入することでデータの価値を策定したり、交渉することが可能であるという示唆が得られた。また、一般的に共有できないデータほど、それらを組み合わせたソリューションが高評価を得る傾向にあることが確認された。データ保有者が、それらのデータを用いて問題解決を行うことが可能であると認識すれば、積極的なデータの共有または、売買によるデータ市場創造の可能性があると見えるだろう。なお、高経年化原子力システムのためのIMDJ実施結果によれば、IMDJにより高評価が与えられたDJについては、実施中に用いられた用途(ソリューション内容)を見た参加者は該当するデータを共有するための有益情報(所有者や入手方法など)を新たに提供している[大澤 14]。

一方で、人間が価値を見出すデータ形式には限界があることも示唆される。本実験にて集め、価値が見出されたDJは、人間にも計算機にも可読なCSV・XLS形式が主であった。計算機がデータ同士の複雑な関係性や構造を読み込み、人間が理解できる形式に可視化するための可視化方法及びメタデータによる記述と構造化が今後の重要な課題であるといえる。

## 6. 今後の展望

今回は一つのIMDJワークショップにおいて分析を行った。テーマによって利用されるDJが異なる可能性も無視できない。そのため、今後も様々なテーマに基づきIMDJを行うことにより、データのどの部分が注目され、ソリューションを導く上での利用価値が発見されたのか、実験を進め明らかにしていく予定である。

## 参考文献

- [大向 13] 大向一輝: オープンデータと Linked Open Data, 情報処理, 54(12), pp.1204-1210, 2013.
- [Ohsawa 13] Ohsawa, Y., Kido, H., Hayashi, T., Liu, C.: Data Jackets for Synthesizing Values in the Market of Data, 17<sup>th</sup> International Conference in Knowledge Based and Intelligent Information and Engineering Systems – KES 2013, *Procedia Computer Science* 22, pp.709-716, 2013.
- [Ohsawa 98] Ohsawa, Y., Benson, N. E., & Yachida, M.: KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In Proc. *Advanced Digital Library Conference (IEEE ADL'98)*, pp.12-18, 1998.
- [Osborn 79] Osborn, A. F.: *Applied imagination: Principles and procedures of creative problem solving*, Charles Scribner's Sons, 1979.
- [川喜田 67] 川喜田二郎: 発想法, 中公新書, 1967.
- [中山 80] 中山正和: NM法のすべて, 産能大学出版, 1980.
- [Ohsawa & Nishihara 12] Ohsawa, Y., & Nishihara, Y.: *Innovators' marketplace: Using games to activate and train innovators*, Springer-Verlag, 2012.
- [Hayashi & Ohsawa 13] Hayashi, T., Ohsawa, Y.: Processing Combinatorial Thinking: Innovators Marketplace as Role-based Game Plus Action Planning, *International Journal of Knowledge and Systems Science*, 4(3), pp.14-38, 2013.
- [大澤 14] 大澤幸生: 市場メカニズムを模倣したシステム安全評価に資するデータ共有・活用手法の研究, 平成 25 年度高経年化技術評価高度化事業報告書, 2014.