

# ビッグデータの活用手法(IMDJ)の紹介

主任研究官 上田 章紘  
研究官 磯山 啓明

## 調査研究の目的

「ビッグデータの活用」は社会的な潮流となっており、様々な組織で取組が進められているが、活用が進んでいない組織では、効果的なデータ活用方法が分からないという悩みが少なくない。ビッグデータは大小様々なデータの集合であり、意図を持ってそれらにアクセスし、組合せなければ新たな価値を見出すことは困難である。そこで、ビッグデータの効果的な活用手法について、調査研究を行った。

## 調査研究の概要

Innovators Marketplace on Data Jackets (IMDJ) とは、東京大学大学院工学系研究科（システム創成学専攻）の大澤幸生教授によって考案されたデータ利活用方策検討手法（ワークショップ等における検討プロセス）である。本調査研究においては、IMDJ の検討プロセスを政策検討に活用する試行を行った。

有識者 10 名による IMDJ ワークショップ（イノベーションゲーム<sup>®</sup>及びアクション・プランニング）を開催した結果、イノベーションゲーム<sup>®</sup>からは、「空き家・空き地の発生箇所を予測するサービス」と「観光客毎の価値観に合った多様な観光ルート的设计、提供サービス」の 2 つのアイデアが得られ、これらをアクション・プランニングにおいてシナリオ検討を行った。

これらの試行結果を通じて考察した IMDJ の有効な活用場面及び効果的に行うための留意点を整理する。

## 1. 調査研究の目的

ビッグデータのデータ流通量が加速度的に増大する中、先進的な IT 企業は、ビッグデータをうまく使ってビジネスを成功させているが、ビッグデータを活用できていない企業も存在する。そこで、H26 年度に国土交通省関連企業にビッグデータに対する意識調査のアンケートを行った。

ビッグデータをうまく活用できていない理由を自由に記述してもらい、類似した回答を分類したところ、上位 3 つは以下の通りであった。

最も多かった理由は「データの統合分析環境が整っていない」ことであり、会社の体

制や設備が整っていないことが想像される。ただし、これは会社がビッグデータに対する意識を転換し、投資意欲を持って、人とお金を配分すれば、ある程度解消する問題である。

続いて2番目の理由は、「個人情報の取扱いが難しい」という問題であった。これも大きな問題であるが、平成27年9月に改正個人情報保護法が公布され、現在、対応が進められている。3番目の理由は「効果的なデータ活用方法が分かっていない」という回答であった。そこで、今年度の調査研究は、民間企業及び行政職員の今後のビッグデータの検討に資することを目的として、ビッグデータの活用手法を調査・試行することとした。

また、ビッグデータという言葉の流行や人工知能の発展などを背景として、ビッグデータはそれ自体で非常に価値があり、打ち出の小槌のように、自然とそこから新しい価値が生まれてくるかのような幻想が一部で広まっているようであるが、ビッグデータは、上に述べたような特徴はあれども、基本的にはただのデータであり、人間がそこに意図を持ってアクセスしない限りは、新たな価値は見出されない。すなわち、人間がデータをどのように活用するかが重要であるため、ビッグデータの活用手法に焦点を絞り、本調査研究を進めることとした。

## 2. IMDJについて

IMDJとは、「Innovators Marketplace on Data Jackets」の略称であり、東京大学大学院工学系研究科（システム創成学専攻）の大澤幸生教授によって考案されたデータ利活用方法検討手法（ワークショップ等における検討プロセス）である。大澤教授によると、各データに含まれる変数名等の概要情報を可視化したマップによって、データ間の潜在的な結合可能性が提示され、それを読み解くことにより、どのような知識が得られる可能性があるかを導出することができるようになる。それによって、要求（社会からの要請等）を満たすために必要なデータの組合せからステークホルダー、リソースにいたるあらゆる事物を考慮したデータ分析シナリオを構築支援するための一連のプロセス（ワークショップ）がIMDJであると解説されている。さらに、思い付きのアイデアを出すだけにとどまらず、様々な背景知識を有する参加者の間でゴールを共有し、目的達成のための論理的な解の創出を支援するプロセスであるとも指摘している。

### (1)IMDJ考案の経緯

もともと、大澤教授が、2000年頃より「チャンス発見学」という新分野を創始し、研究されてきた内容が起源となっており、「稀だけど重要な事象にいかにか気付くか」ということが、チャンス発見の本質であるとして、データの可視化に注目して研究をされてきた。この「チャンス発見学」の理論を議論の検討プロセスに落とし込んだものが、イ

イノベーションゲーム<sup>®</sup>と呼ばれる手法であり、民間企業の新商品開発や新サービスのアイデア導出に活用されてきたという経緯がある。

さらに、近年話題になっているデータの利活用方法の検討に拡張するため、後述する「データジャケット」という要素を手段として取り込み、現在のIMDJの形に改良された。IMDJとは、この改良されたイノベーションゲーム<sup>®</sup>とアクション・プランニングによって得られたアイデアの実現をワークショップ等で議論する一連の検討プロセスのことを言う。

## (2)IMDJのプロセス

IMDJは、複数のプロセスを経て新しいアイデアの創出から、その実現までの議論を後押しする検討プロセスである。本節では、各段階におけるプロセスの内容を説明する。

### ①事前準備

まず、IMDJの中核を成すイノベーションゲーム<sup>®</sup>を開始する前には、いくつかの事前準備が必要となる。

議論したいテーマと参加者が決定すると、テーマに即したデータジャケット（データの概要情報）を収集し、可視化する。ビッグデータに係るデータジャケットの一例を表1に示すが、ビッグデータの要素、特徴等をまとめたものになっている。データジャケットをベースとして議論を進めることが、イノベーションゲーム<sup>®</sup>の特徴となっている。

ビッグデータは、データ量が多いだけでなく、データの種類や含まれる変数名が多岐に渡ることも多く、データそのものをワークショップの際に参加者に見せながら議論するためには、パソコンを持ち込んでその都度データを画面上に映し出す必要も生じる。しかし、データによっては、全体像を簡単に映し出すことや、指し示すことが難しいもの、情報漏洩防止やプライバシーの観点からデータそのものの形では開示することができないものも存在する。このような際に、表1のようなデータの概要情報（データジャケット）で議論すれば、データそのものを見せるわけではないため、問題はクリアされることになる。

イノベーションゲーム<sup>®</sup>では、通常1つのテーマ当たり20～30のデータを収集し、データジャケットの形に整理する。

表1 データジャケットの例

タイトル	インフラ画像のモニタリング
概要	首都圏の主要道路の劣化診断のための撮影画像
データに含まれる変数	画像、撮影日時、位置情報
データの種類	画像データ、数値データ、自然言語データ
データの収集方法	専用車に内蔵したGPS及び車下に装着したカメラで収集している

②イノベーションゲーム<sup>®</sup>

ビッグデータは、データの要素等に多様性があるため、データ同士の掛け合わせにより、新たな発見が生まれやすいとされている。これに注目したのがイノベーションゲーム<sup>®</sup>の手法であり、データ間の関係性やつなかりに焦点を当てて議論が進められるように、①で収集したデータジャケット同士の関連性を可視化したマップ（キーグラフ<sup>®</sup>）を作成する。

キーグラフ<sup>®</sup>は、データジャケット（図中ではDJ）の概要及び変数等の情報を単語単位で切り分ける形態素解析を行ったものを、他のデータジャケットと共通する単語を抽出して、赤線で結合させたものである。図中の黒点はデータジャケットを表し、赤点は「頻度は低いがデータジャケットの関係性を表す重要語」を表す。ワークショップ中には、データジャケットの意味ある組合せを意識しながら、議論を進めることとなる。

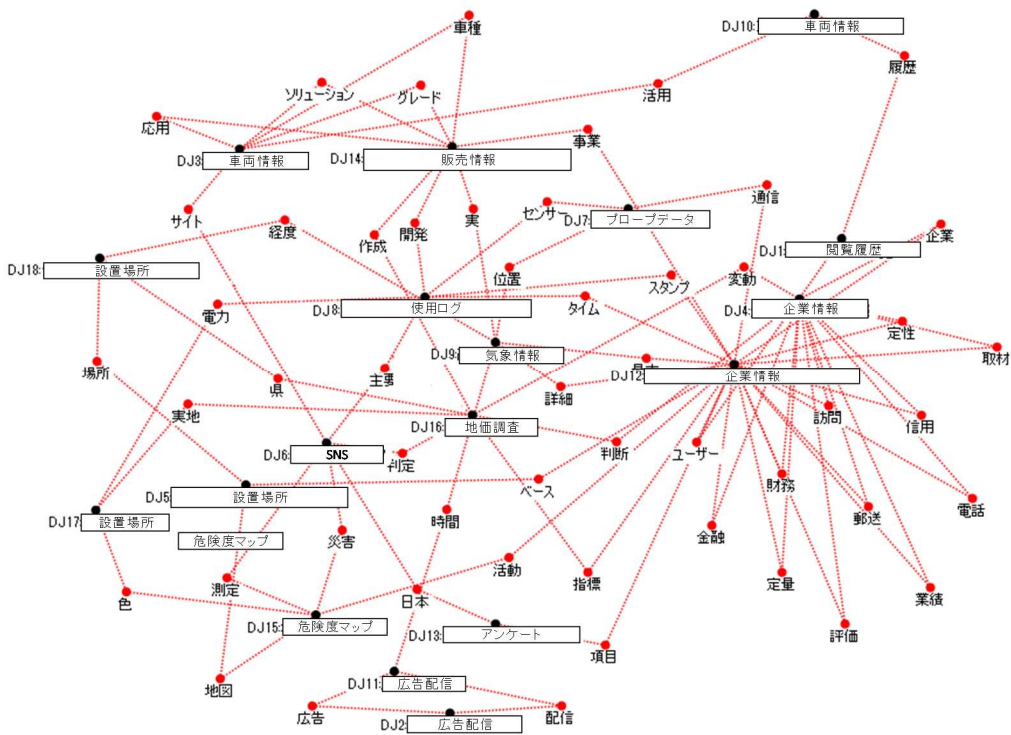


図1 キーグラフ<sup>®</sup>の例

イノベーションゲーム<sup>®</sup>の議論の進め方については、一例を図2に示す。例えば、「気象がどのように全国のビール消費者の嗜好に影響するかが知りたい」という要求に対しては、「全国の過去50年間の気象変化」と「全国の店舗における各商品の実績」のデータを使用することにより、「気象条件とビール消費動向の相関ルールが分かる」という解決方法が提案され得る。このように、イノベーションゲーム<sup>®</sup>の議論は、参加者の誰かが実現したい「要求」を提示し、これに対し、別の参加者がデータジャケットの組合せにより「データを用いた解決案」となるアイデアを提案する形で進められる。このプロセスは、ゲーム形式となっており、参加者は、それぞれが持つ社会的な役割に応じて様々な角度からアイデアを提案し、そのアイデアの実現可能性や有用性を評価し合う。

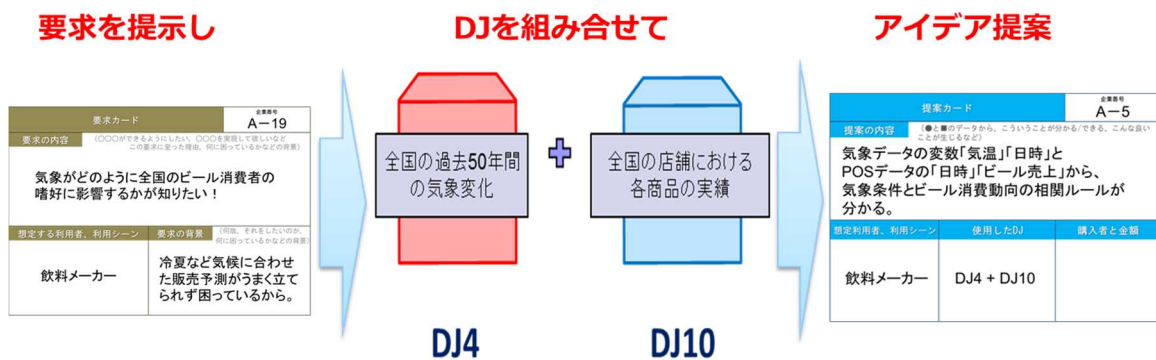


図2 イノベーションゲーム<sup>®</sup>の議論の流れ

### ③アクション・プランニング(アイデアのシナリオ化)

イノベーションゲーム<sup>®</sup>により、「要求」と「データを用いた解決案」のセットが多数提案された後に、アイデアをシナリオ化して検討を深めるものを参加者間で選定する。

選定されたものは、アクション・プランニングのプロセスによって、その解決のアイデアの実現に向けて、データの変数や要素の関連性、リスクを洗い出し、意志決定を行う際に生じる盲点をなくし、実行動を促すシナリオが策定される。

ここでは、データサイエンティストと呼ばれるようなデータ分析の専門家がデータを分析し、分析結果を利用する実施者の要求を考慮しながら具体化していくような進め方が望ましい。

## 3. 試行結果

本章では、今年度調査として行ったIMDJの試行結果について、簡単に報告する。IMDJは、本来は新商品開発や新サービスのアイデア導出のために用いられることが多いが、今回は行政機関による取組であるため、政策検討に適用して試行することとした。

## (1)IMDJ ワークショップ実施概要

### ①実施スケジュール

ワークショップの開催前に約2ヶ月強掛けて、メンバー選定、メンバーへの事前説明、テーマの絞り込み、データジャケットの収集・整理等を行い、平成28年1月28日(木)にIMDJワークショップを開催した。

当日は、以下のようなスケジュールで行った。

・導入説明	13:00～13:10
・自己紹介(専門、関心、本日の期待等)	13:10～13:30
・ウォーミングアップ(ルール確認)	13:30～14:00
・イノベーションゲーム <sup>®</sup>	14:00～16:00
・休憩	16:00～16:10
・アクション・プランニング	16:10～17:30
・まとめ	17:30～18:00

### ②参加メンバー

今回のIMDJでは、G空間情報(地理空間情報)に関連するビッグデータを取り扱うことを計画したため、G空間情報に関する知見の深い学識者4名と、関連の深い業界から民間有識者6名を選定し、ワークショップを開催した。メンバーは以下の通りである。

・東京大学空間情報科学研究センター	西澤 明	特任教授
・東京都市大学工学部都市工学科	今井 龍一	准教授
・東京大学空間情報科学研究センター	瀬戸 寿一	特任助教
・東京大学空間情報科学研究センター	秋山 祐樹	特任助教
・日本 IBM スマートシティ事業担当部長	森島 秀明氏	
・ゼンリンデータコムネットサービス本部 Web-GIS 事業部副部長	足立 龍太郎氏	
・日建設計総合研究所研究員	伊藤 慎兵氏	
・KDDI 研究所データマイニング応用グループ研究員	石塚 宏紀氏	
・アジア航測執行役員 社会基盤システム開発センター長	政木 英一氏	
・株式会社ケー・シー・エス代表取締役社長	天野 洋文氏	

また、事務局の技術協力として、東京大学大澤教授と研究室メンバーにも協力を仰いだ。

### ③IMDJ ワークショップのテーマ選定

今回のテーマ選定に当たっては、G空間情報関連のビッグデータを使用するという分野の大枠を決めた後に、参加者を選定し、参加者への事前ヒアリングを経て、ワー



ワークショップで議題とするテーマを選定した。

テーマは、「G 空間情報を活用した、次世代のまちづくり」として、これに関連する主に国土交通省が公開しているデータを議論の材料として政策検討を行うワークショップとした。

また、議論への参加者が組織外部の方々ばかりだったため、議論の方向性としては、社会厚生に役立つアイデアであって、自組織だけでは実現が難しく、かつ、国又は自治体の公的な機関が支援することが必要である事業の案を検討・提案するよう、参加者への事前説明時にあらかじめ依頼した上で、ワークショップに臨んでもらった。

#### ④データジャケットの収集・整理

今回の IMDJ ワークショップでは「G 空間情報を活用した、次世代のまちづくり」に関係するデータとして、39 個のデータジャケットを収集・整理した。表 2 は代表例を 4 つ抽出したものである。データジャケットは、ワークショップについての事前説明、専門や意向をヒアリングする際に、参加者から提案されたデータも多く含まれている。

表 2 使用したデータジャケット(代表例)

データ名	データの概要	変数名	データ特性	大項目	加工	利用可能なデータ粒度
パーソントリップ調査データ	パーソントリップ調査とは、一定の地域における人の動きを調べ、交通機関の実態を把握する調査。交通実態調査とも言う。個人の1日における移動状況を把握することにより、「どの交通機関が」「どのような人によって」「いつ」「どのような目的で」「使われているか」を調べることができ、バスや電車、地下鉄、乗用車などのいくつもの交通機関を総合的に把握することを目的としており、都市圏(大都市圏)で行われる。	発生・集中交通量、地域間交通量、住所、性別、年齢、職業、勤務先、通学先、自動車の保有台数、二輪車の保有台数、運転免許の有無、出発地、到着地、移動目的、移動手段、移動時間、車の運転有無、駐車場、駐輪場所、有料道路利用有無	更新間隔:10年 静的 蓄積期間: 1968年~ データ形状:表 ※東京都市圏の場合	静的 データ	不要	トリップ起終点の住所・施設名称
固定資産課税台帳データ	土地・家屋・有形償却資産といった固定資産を評価したデータ 市町村が、固定資産の状況及び固定資産税の課税標準である固定資産の評価を明らかにするために備えなければならない重要な台帳	所有者の住所及び氏名又は名称並びにその所在、地番、地目、地積及び基準年度の価格又は比準価格、家屋番号、種類、構造	更新間隔:1年 静的 蓄積期間:不明 データ形状:表	静的 データ	不明	住所
都市計画基礎調査データ	都市における人口、産業、土地利用、交通などの現況及び将来の見通しを各自治体が定期的に収集したデータ。都道府県が都市計画区域に関して5年ごとに実施する調査で、都市計画区域における人口、産業別就業人口、市街地面積、土地利用、交通量、地価など多種多様な項目が調査対象となっている	人口規模、将来人口、人口増減、通勤・通学移動、昼間人口、産業・職業分類別就業数、事業所数・従業者数・売上金額、区域区分の状況、土地利用現況 国有地の状況、宅地開発(電子化は状況農地転用状況、林地転用状況、新築動向、条例・協定、農林漁業関係 施設適用状況、建物利用現況、大規模小売店舗等の立地状況、住宅の所有関係別・建て方別世帯数等	更新間隔:おおむね5年 静的 蓄積期間: 1987年~ 2008年から データ形状:点、線、面	静的 データ	不要	旧市町村界、大字、町丁目・字及び都市計画区域界、線引き界など
日本の地域別将来推計人口	将来の人口を都道府県別・市区町村別に求めたもの 2013年での推計では、平成22(2010)年の国勢調査を基に、平成22(2010)年10月1日から平成52(2040)年10月1日までの30年間(5年ごと)について、男女年齢(5歳)階級の将来人口を推計している。毎月の出生・死亡・転入・転出を加減して算出された推計値をもととした人口数である。この数字には外国人も含まれている。	人口、都道府県、市区町村、年代、性別、将来の生存率、純移動率、子供	更新間隔:5年 静的 蓄積期間: 1997年~ データ形状:点	静的 データ	不要	都道府県、市町村

#### ⑤キーグラフ®の作成

④で収集した 39 個のデータジャケットをキーグラフ®に整理し、各データの関連性を可視化した。(図 3)





## (2)シナリオ検討結果

今回は、前半のイノベーションゲーム<sup>®</sup>で多数アイデアが出た中から、2つの政策アイデアに絞り、参加者を5名ずつの2グループに分けたアクション・プランニングの中で、シナリオ検討を行った。(当日の風景は図4の通り)

以下、2つのサービスに関する要求の背景、使用するビッグデータ、活用のシナリオという観点から、議論の結果を取りまとめる。



図4 IMDJ ワークショップの開催状況

### ①「空き家・空き地の発生箇所を予測する」サービス

#### (ア)要求の背景

今後の人口減少により、徐々に空き家が増え、地域として立ちゆかなくなる自治体が発生すると想定される中、人口減少地域を細かく推測できれば、効率的な都市計画が可能となる。また、地域のインフラに関わる事業者や、地域ビジネスに関わる事業者にとっても、将来の経営計画が立てられるため、空き家・空き地の発生箇所を予測するサービスは、官民ともに需要のあるものである。

#### (イ)使用するビッグデータ

平成27年5月に空家等対策の推進に関する特別措置法が施行されて以降、市町

村は空き家の実態把握を行う努力義務が課されたため、現在、市町村内における空き家調査を行っている自治体が多い。空き家調査結果が無い場合には、水道・電気・ガスの使用データにより、ある程度空き家の推定が可能であると想定されるため、代用することとする。これに、固定資産税、住民基本台帳のデータを使うことで、建物の築年数や構造形式、住民構成、年齢等を判別し、将来の空き家となるか否かの推定に用いる。さらに、地元の信用金庫には、各家庭の相続、跡継ぎに関するデータが蓄積されているため、これを考慮すると将来の空き家判定の精度が増すこととなる。

#### (ウ)活用のシナリオ

データを所持している自治体と、地域ビジネスを行っている不動産業者や、信用金庫、電気・ガス等のインフラ事業者の両者にとって、地域の今後のポテンシャルをミクロに分析することができるメリットがあり、協力関係を結べる可能性が十分にあると評価できる。

ただし、分析途中及び分析結果のデータについては、慎重に取り扱う必要があり、過疎が見込まれる地域の地元住民からは、結果の公表に対して反対を受ける可能性が高い。

### ②「観光客毎の価値観に合った多様な観光ルートの実設計、提供」サービス

#### (ア)要求の背景

遠方から来た観光客が有名な観光地だけ見て、あまり地域を周遊せずに帰ってしまう現状に対して、地域の観光関連業者にとっては、大きな観光資源が地域にあることは大きなメリットである一方、観光客を思うように取り込めていないジレンマともなっている。観光客に対して、地元民の知る観光の魅力や、タイムリーな観光サービスを届けたいと地域の観光関連業者は考えている。

#### (イ)使用するビッグデータ

携帯電話会社等に GPS 情報として蓄積されている過去の観光客の人流データの分析により、観光客がどのような足取りで帰ってしまっているかの分析が可能となる。また、現在、地域に訪れている観光客に対しては、各観光施設のリアルタイムの混雑データやお勧めの情報を各施設管理者等から吸い上げることで、適切なタイミングでの情報発信が可能となる。

#### (ウ)活用のシナリオ

旅行会社のパッケージツアーと単独で訪れている観光客との人流データを比較

分析することにより、旅行者の求める細かなニーズを地域に還元できる可能性があり、観光資源の開発にもつながるため、地域としてのニーズは高い。また、各観光施設のリアルタイムの混雑データやお勧めの情報に関して、自治体などが情報のプラットフォームを作り運用することができれば、観光客が地元を周遊している最中に、混雑情報や細かな観光情報を入手することができる。このサービスの実現には、多くの人を巻き込んで一体となった取組を行う必要があるため、行政が調整役を果たす必要がある。

## 4. まとめ

### (1)IMDJの有効な活用場面

IMDJの中核をなすイノベーションゲーム<sup>®</sup>は、データの組合せにより、新たな知見を得たい時に有効な手段である。キーグラフ<sup>®</sup>の作成により、データの繋がりを可視化することで、他のデータとの新たなつながりが模索できる。また、組織内に蓄積されているデータや行政等で公開されているオープンデータについても、他データとの組合せにより、既存の使い方とは違う、新たな利用価値が見出せる手法である。

IMDJは、イノベーションゲーム<sup>®</sup>で得られたデータを用いた解決案を、アイデアの創出にとどまらず、参加者間でゴールを共有し、目的達成のための論理的な解の創出を支援するプロセスとしてシステム化されている点も、有用性が高い。

### (2)IMDJを効果的に行うための留意点

イノベーションゲーム<sup>®</sup>もアクション・プランニングも、ワークショップの参加者は取組の実施者、協力関係にある者を中心としたメンバーで行うことが望ましい。取組の実施者がいることで、アイデアのシナリオ化の際には、より細かな課題まで網羅することが可能となる。また、利害関係又はライバル関係にある者が参加する場合、得られたアイデアの権利で調整が取れなくなる可能性があるため、同組織又は事前に協力関係が約束されている者を中心としたメンバーで行うことが望ましい。ただし、同組織、同部署内のみで開催すると、議論の幅が狭まるおそれがあるため、有識者やデータサイエンティストを適宜投入することも効果的である。

次に、IMDJの一連のプロセスは、イノベーションゲーム<sup>®</sup>の理念や議論展開のコツを掴むことが容易ではないため、未経験者が1回のワークショップで成果を残すことは容易ではない。丁寧な事前説明を行い、複数回行う中で議論を深めていくことで、より効果的なワークショップとなると考えられる。今回の政策検討の試行においては、ワークショップの開催が1日だけであったため、イノベーションゲーム<sup>®</sup>中に議論がうまく進まない場面が見られたことに加え、アクション・プランニングも深い検討を行うためには時間が足りない状況であった。

最後に、IMDJ のワークショップにおいては、テーマ設定の範囲が重要な鍵を握っている。テーマの範囲が狭すぎると、事前に想定されたシナリオから脱却する新たなアイデアが生まれにくく、テーマの範囲が広すぎると、当初の狙いと外れた議論が進んでしまう場合や、議論が集約されない場合が出てくるため、テーマ設定時には、特に注意を払う必要がある。