

論文の内容の要旨

論文題目 データ市場における知識構造化に基づく
データ利活用シナリオ検討プロセスの研究

氏 名 早矢仕 晃章

本研究の長期的な目的は、データ市場におけるデータ駆動型イノベーション（Data Driven Innovation : DDI）に貢献することである。データ市場とは、データの公開・共有を強制するのではなく、自由市場の原理で利用者が必要なデータを選び、所有者と交渉の末に入手できるプラットフォームである。市場とは提供者と利用者が接し、その相互作用の中で商材の価値を評価し、その評価に合う条件を設定して取引を行うイノベーションの場である。つまり、データ市場はデータを商材として扱い、ステークホルダー間のコミュニケーションによって価値を決定し、取引を行う場である。

近年、異なる領域のデータを組み合わせて新たな知識を獲得し、意思決定に役立てることへの期待が高まっているものの、データの共有及び公開には高い社会的障壁がある。企業の機会損失やプライバシー侵害のリスク、さらにデータの取扱いについての各国の法制度の違いなどの問題がある。このような状況において、データの公開ではなく、市場における交換という戦略によってDDIを推進しようとする様々な形態のデータ市場がWebを中心に萌芽し始めた。しかし、Webのみをプラットフォームとするサービスは、データの表層的な情報を陳列するだけに留まっており、ステークホルダー間の十分なコミュニケーションが期待できず、イノベーションの場としての市場の機能が有効に働いているとは言い難い。また、データの蓄積方法は積極的に議論されてきたが、既存の知識やモデルでは扱えないデータを含むデータ利活用知識の蓄積方法は十分に議論され

てこなかった。つまり、データ市場におけるDDIを促進するためには、データ市場の仕組みを理解するとともに、利用価値のあるデータの利用方法を知り、どのような仮説が検証可能かということを議論するプラットフォームと、データ利活用知識を蓄積し再利用するための技術が必要である。

そこで本研究では、DDI創出環境としてのデータ市場の仕組みとデータ利活用方法の検討による価値化プロセスを理解し、提案手法の有用性を評価するため、実際のデータ市場に参画するステークホルダー（実業家、研究者、分析者など）を参加者とした実験的データ市場を設計した。そして、データ市場に関わる様々なステークホルダーの意図や目的が反映される要求とデータ利活用案をデータ利活用知識として構造化し、ステークホルダー間の創造的コミュニケーションによるデータ価値化とデータ市場創出を支援する技術を開発した。

本研究における実験的データ市場の設計には、基礎技術としてデータジャケット(DJ)、データ利活用方法検討ワークショップInnovators Marketplace on Data Jackets (IMDJ)、データ利活用シナリオ創出手法アクション・プランニング (AP) を用いた。本研究のコア技術であるDJとは、データの中身ではなく、データの概要情報(データ内の変数名、保存形式、収集方法など)を共有し、価値を検討可能にする方法である。個人を識別する情報を含む共有不可能なデータでも、DJにすることでリスクを低減させて情報が共有可能となる。例えば、商品の購買履歴データには氏名、性別、支払金額などの個人を識別する情報が含まれるため、一般公開することはできない。しかし、購買履歴データを「氏名」、「性別」、「支払金額」といった変数名としてメタデータ化すれば、個人を識別する情報を秘匿のまま、データに関する情報を共有できる。実験的データ市場におけるDJの利用とは、データ交換・売買以前のデータに対する期待の高さを表し、データの利用価値を保有者に提示することに相当する。つまり、データ利活用方法の提案によってデータの利用価値が定まり、需要が生じる。その需要によって供給の必要性が生じ、取引のための条件（価格）が調整されるという市場の原理が働く。すなわち、データの概要情報の共有によって今まで秘匿であったデータの価値の評価が可能となる。

まず、本研究では、データ利活用知識の蓄積方法として、DJだけでなく、過去のIMDJにおいて議論された要求・データ利活用案・データの間接関係をデータ利活用知識として構造化し、検索システムData Jacket Store (DJストア) を実装した。データ利活用知識構造化により、ユーザーが自分と異なる視点を持つ過去のユーザーが考案したデータの用途を発見したり、別の人が考案したデータ利活用案に注目することで意思決定に役立つDJを探し出すことが可能となることを評価実験により示した。以上により、過去に検討されたデータ利活用知識の構造化と再利用は新たな知識獲得に有用であることが分かった。特に、実験的データ市場において既存の知識や情報だけでは解決できない問題に直面した際に、データに関する情報の陳列だけではなく、データ利活用知識の構造化による検索システムが利用価値の高いデータの発見と問題解決を促し、データに対する新

たな需要を喚起する可能性が示唆された。さらに実験では、オープン化できないデータほど、提案者及び利用者にとって問題解決及び新ビジネス創出において有用性が認められる可能性が高くなることが示唆された。また、検索結果においても秘匿データの方がユーザーの興味・関心の度合いも高いことが分かった。以上の結果から、データ市場はオープンデータに代表される公開可能データのみに関じられた場ではなく、公開が難しい個人や企業のデータ及びその保有者を巻き込むイノベーションの場として機能し得ることが分かった。

しかし、データ市場はデータの組合せのみを議論する場ではない。続いて、データ利活用に関わる諸要素が意思決定者のデータ利活用方法検討のプロセスにどのように現れるのかという、シナリオ生成プロセスとシナリオの構造化について議論した。事業計画立案時の筆記行動に着目した実験により、実験的データ市場においてデータに文脈を付与するシナリオ生成プロセスには仮説推論における非単調性が現れることが分かった。また、データの組合せだけでなく、ステークホルダーやリソースといったデータ利活用に関わる諸要素の関連性を考慮した検討が重要であるという示唆が得られた。

以上に得られた知見を元に、データ利活用知識を拡張し、シナリオ創出手法APによって生成されたシナリオの構造化と再利用の仕組みを提案した。そして、ステークホルダー表出と関係推定システムResource Finderを実装し、過去に検討されたデータ利活用シナリオの構造化が、文脈によって異なるステークホルダーのシナリオへの関係を推定するのに有効に作用することを実験的に評価した。さらに、新たにデータを取得する意思決定者を支援する変数ラベル推定方法VARIABLE QUESTを提案した。実験では、データ概要の類似性と変数ラベルの共起性を考慮することで、変数ラベルが未知のデータ概要からそのデータに含まれる可能性の高い変数ラベルが推定可能であることを示した。膨大なデータから必要な知識を発見することが困難であるように、データ市場において複数の領域にまたがって存在するデータ、ステークホルダー、変数など、データ利活用に関わる全ての要素を考慮することは難しい。それ故、意思決定者の異なる価値観や多様な背景知識、意図に対応して構造化された知識ベースとそれを検索するシステムが必要となる。データ利活用知識だけでなく、ステークホルダー及び変数ラベルを含むシナリオの再利用により、潜在的なビジネスパートナーや取得すべき変数についての情報をユーザーに提示できる可能性が示唆された。

最後に、実験的データ市場の枠組みを拡張し、データの入手、分析、課題発見とフィードバックという実社会とのインタラクションを含んだ実装的データ市場を観察する二つの応用実験を行った。一つ目の実験では、データ利活用案の創出、分析シナリオの生成、そして実データ分析によって結果を得る一連の過程を観察した。実際のデータ分析は様々な試行錯誤によって結果を得るプロセスであり、実装的データ市場における分析計画と実データ分析の間にはギャップが存在することが明らかとなった。しかし、上流設計部にあたる分析シナリオを十分に検討していれば、比較的评价の高い分析結果が

得られる可能性があることが分かった。二つ目の実験では、分析結果を元にした行動によってステークホルダーから新たな情報と有益なフィードバックが得られることが分かった。また、一般に共有できないデータでも、当該データの活用方法を示したシナリオをデータ保有者に提示すれば、範囲を限定してデータの共有が可能となるという重要な示唆を得た。すなわち、行動しながら新しい知識を取り込むことでシナリオを修正し、分析結果を精緻化するプロセスを経ることが、データ市場における新たなステークホルダーの存在を掘り起こし、データの新しい利用方法の発見を促すものと考えられる。

本研究の提案手法により、データ市場において既存の知識やモデルでは扱えないデータの利活用法及び潜在的なステークホルダー、変数ラベルの発見が支援されるとともに、データ利活用知識ベースが更新されることが分かった。また、それらの手法を利用して新たに知識を獲得する意思決定者の行動が改善されることが示された。本研究の提案手法及び得られたデータ市場のモデルによって、データに関する情報及び知識の蓄積が可能となったことで、従来のIMDJ及びAPをデータ市場創出支援技術として大きく改良したといえることができる。また、本研究の提案手法によって、事業者の新しいデータの発見や異なる事業者とのインタラクションによって新規事業創出が促進されたことが報告されており、データ市場の創出と発展に貢献したものといえることができる。