総務省 学術雑誌『情報通信政策研究』 第４巻第１号
Journal of Information and Communications Policy Vol.4 No.1

論文（査読付）

データ流通エコシステムにおけるプラットフォームを媒介とした価値創造

清水たくみ[1]、早矢仕晃章[2]、深見嘉明[3]、松島裕康[4]、坂地泰紀[5]、大澤幸生[6]

要　旨

様々な産業・経済システムのデジタル化が進展する中、デジタルデータを活用した価値創造への期待が高まっている。その流れを受けて、異なる事業者間でデータを交換・取引するデータ流通エコシステムが萌芽し、新たなイノベーションの源泉として注目され始めている。このようにデータ流通を通じた価値創造への関心は高まっているものの、法制度設計と技術仕様が複雑に絡み合うことから、データ流通エコシステム全体を包括的に捉えた議論が進んでいるとは言い難い。加えて、データ流通プラットフォームの仕組みやサービスについてのシステム面・ビジネス面に絞られた形での議論は存在する一方で、実際に取り扱われるデータ自体の特徴やデータ間の関係性などのデータ流通メカニズムについて十分には明らかになっていない。そこで本論文では、データ流通プラットフォームにおけるデータ連携・相互作用メカニズムを、実証分析に基づいて検討した。プラットフォーム上のデータネットワーク分析から、1）連携可能性を高めるデータの特徴、2）連携を促進するデータ共有条件、及び 3）データネットワーク構造の特徴に関する知見を得た。具体的には、「時間」や「場所」などに関わる変数が多様な異種データの連携可能性を高めることが示された。また、秘匿データと共有可能データの混在が密なネットワーク構造の構成を促し、異分野間データ連携の成立可能性を高めることが明らかになった。加えて、データネットワークは人間関係と似た「局所的に密、大域的に疎」な構造をしており、疎の部分を埋めるデータを提供することで、異分野をつなげるデータ活用を活性化できる可能性が示唆された。これら分析結果をもとに、データ流通プラットフォームの設計・運営及びデータエコノミー推進への示唆と課題を議論する。

キーワード：データ流通、データ連携、プラットフォーム、価値創造、ネットワーク

１．はじめに

デジタルトランスフォーメーション（DX）推進の必要性が広く謳われる中、様々な産業分野で情報技術を活用した業務再構築が積極的に取り組まれるようになりつつある。組織内ビッグデータであるデータレイクの構築や、増加した Internet of Things (IoT)デバイス

[1] 早稲田大学大学院経営管理研究科 准教授
[2] 東京大学大学院工学系研究科 システム創成学専攻 助教
[3] 慶應義塾大学大学院政策・メディア研究科 特任准教授、学習院大学経済学部 特別客員教授
[4] 滋賀大学 データサイエンス教育研究センター 准教授
[5] 東京大学大学院工学系研究科 システム創成学専攻 特任講師
[6] 東京大学大学院工学系研究科 システム創成学専攻 教授

から取得した多種多様なデータを活用したイノベーション創出への期待が高まっている。DX を通じた価値創造がもたらした新たな経済構造はデジタルエコノミーと呼ばれる。価値創造の中心を担うデータそのものの価値が強く認識されるようになり、データエコノミーへの注目・期待も高まっている。特定企業が集権的にコントロールするのではなく、行政を含む多様なステークホルダーが個々に消費者に対して便益を提供しつつも、必要に応じて有機的に連携するという分散型アーキテクチャを基盤とした社会への移行が期待される。

データは「21 世紀の石油」と形容されるなど[7]、2013 年前後のビッグデータの世界的なブームからデータの市場価値並びにデータ流通を核としたビジネスが活発に議論されるようになった。これまで組織内に死蔵されていた情報がデータ化されて流通することにより、新たなイノベーションの源泉となりうる可能性が注目されるようになった。実際に分散型アーキテクチャを前提とした企業間連携手法の開発や、手順や関連技術使用に対する標準化への取り組みも進められてきた。日本では、データ取引市場の存在を規定した 3 極モデルが世界に先駆けて提案されている[8]。このような中、多様な主体間でデータを交換・取引するデータ取引市場ならびにデータ流通プラットフォームを提供する企業が世界的に登場しつつあり、日本では EverySense (Mano, 2016)、Virtuora DX (Ejiri, Ikeda, & Sasaki, 2018)、D-Ocean (Yagihashi, 2019)などがデータ流通サービスとして展開されている。また、学術的側面からの関心も集まっている(Liang et al., 2018; Spiekermann, 2019; Stahl, Schomm, & Vossen, 2014)。

以上のようにデータ流通を通じた価値創造への関心は高まっているものの、それを支える環境の構築には法制度設計と技術仕様が複雑に絡み合う。セキュリティや個人情報保護、センサーデバイスの設計、データモデルや標準化など個別の観点における議論は活発ではあるものの、それぞれの議論から導き出された知見を総合的・有機的に組み合わせることは非常に困難な状況である。加えて、議論の前提となる技術面・法制度面での変化が速く、これも議論が複雑化する要因である。そして、データ流通エコシステムにおいて交換されるデータの特徴についての理解も限定的であり、プラットフォーム上の異なるデータ間の関係性についてはほとんど明らかになっていない。データエコノミーの実現と促進のためにはエコシステム全体の包括的な設計について議論が必要であるが、そのための検討材料はほとんど提供されていない。そこで本論文では、これまで議論されることが少なかったデータ流通プラットフォームにおけるデータ間連携・相互作用を、ネットワーク分析を通じて明らかにする。

本論文は以下の構成で進行する。まず、2 章にてプラットフォームを介したデジタルデータ流通に関する現在及び課題を概観し、本論文の論点を整理する。また同章にて、本研究の主たるアプローチであるネットワーク解析（共有されるデータ間のつながりに着目した分析）の有用性についても議論を行う。そして、3 章では本研究で対象とするデータ及び分析手法について詳述する。プラットフォーム上におけるデータの関係性の定量化、ネットワー

---

[7] The Economist: The world's most valuable resource is no longer oil, but data, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[8] 総務省　情報通信審議会　標準化戦略 WG 第 7 回 資料 7-1　標準化戦略 WG とりまとめ（案）：https://www.soumu.go.jp/main_content/000690783.pdf

ク指標について説明する。4 章では、連携可能性を高めるデータの特徴、連携を促進するデータ共有条件、及びデータネットワークの構造的特徴に関して、分析結果を示す。それら分析結果を踏まえ、5 章ではデータプラットフォーム運営とデータエコノミーについての示唆を議論し、6 章にて結論を述べる。

２．データ流通エコシステムの現在

２．１．データエコノミーとプラットフォーム

流通対象として議論の対象となるデータは、EU 指令などをふまえて以下の 3 つに分類されることが多い。公共財として提供される公共オープンデータ(Public Sector Information: PSI)、主に民間企業が事業活動を通じて生成する産業データ、個人の属性や活動に伴い生成される履歴など個人(データ主体)を特定するもしくは特定可能[9]な情報である個人情報である。これまで個別企業がその活動の中で生成する産業データは知的財産として取り扱われ、企業ないしは系列などの企業連合を超えて流通することはあまりなかった。数少ない例外として、EDI[10]など企業間取引に伴う帳簿連携や、法令で認められている電子的な財務状況の報告といった事例があるものの、多くは企業間どころか組織内の部門間ですらデータは共有されておらず、部門ごとに運用されてきた複数のレガシーシステムがサイロ化している。製造現場へのセンサー敷設／ロボット導入などが進められつつあるものの、同一ライン内や同一設備内ですらネットワークが閉じている場合も多い。このようなサイロ化が AI や IoT 関連技術の導入を妨げ、DX 実現の阻害要因となっている(藤野, 2020)。

このような日本企業の状況を打破しようと、Society 5.0 モデルに立脚した Connected Industries や第四次産業革命が提唱されている。これらの方針をもとに、政府はデータ共有基盤の開発や契約ガイドラインの整備といった産業データの流通促進事業を進めてきた。また、不正競争防止法にデータ不正取得などに対する民事措置が創設されたり、産業データ共有事業の認定制度が創設されたりと、産業データ流通ならびに利活用を促進する法制度の整備も進みつつある。

一方、顧客理解のために実施される市場調査や、顧客管理や選別を通じた生涯価値最大化を目指す Customer Relationship Management (CRM)など、マーケティング施策の一環として個人情報の収集と分析は長らく行われてきた。このような従来型の個人情報利活用は、分析からアウトプットまで全てを企業が一方的に設計、運用するという形態がとられており、消費者は受動的に情報を提供するのみといった立場に置かれてきた。

この状況は E コマースやクラウド技術を活用した Software as a Service (SaaS)の普及によって更に加速している。ウェブを介してサービスを提供する事業者の多くは、利用者のオンライン上の行動履歴を蓄積し、その分析結果をもとにしてサービスを提供している。一部のサービスにおいては、蓄積された履歴データをもとにした広告表示によって無料でのサ

---

[9] OECD (2013) Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data. https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

[10] Electronic Data Interchange(電子データ交換)の略称。専用回線／ネットワークなどを介して、標準化されたデジタルフォーマットの受発注書や請求書などの取引文書を交換することを指す。

ービス提供が実現されている。つまり、個人情報提供の対価としてサービスを提供するビジネスモデルを採用する企業も多い。この構造は、IoT 技術を活用して生活空間上に敷設された多数のセンサーデバイスからのデータ収集によって強化されることが予想される。また AI 技術の発展と普及は、オンライン／オフラインを問わず継続的に収集された個人情報を学習データとすることで、既存サービスの改善や新サービスの創出が期待されると同時に、自覚しないまま収集された個人情報が事業者にとって一方的に有利な形で活用されるなど、消費者の権利が毀損されるリスクが高まっている。実際にこれまでも、企業がマーケティング施策のために収集したプライバシー情報の漏洩といった人権侵害事案が生じている(Kemp, 2019)。このようなリスクは、情報の非対称性(Akerlof, 1970)、ネットワーク外部性がもたらす競争の阻害(Acquisti, Taylor & Wagman, 2016)や不完全な競争(Robertson, 2020)などによりもたらされると、OECD によってまとめられている[11]。そのため多くのデータ流通の振興策は、「安全」、「安心」、「品質の確保」、「信頼性向上」を前提とし、データ流通の増加に伴うリスクをいかにコントロールするかが重要な課題とされている。「世界最先端デジタル国家創造宣言・官民データ活用推進基本計画[12]」（以下「デジタル戦略」という）では、官民におけるデータの徹底活用を標榜し、その前提として「データの安全・安心・品質の確保を前提とした国際データ流通網等の実現」や「信頼性向上のためのデータ流通ルール整備」を推進するという構成となっている。このデジタル戦略を始め、政府・行政機関が示す戦略や方針には、個人の権利を保護しながらデータ利活用を進めるためのルールの提示に多くの分量が割かれている(板倉, 2020; 黒政, 2020)。

また、社会における AI の有効活用が叫ばれる中で、AI の開発と運用における倫理性の担保が強く求められるようになっている(中川, 2020)。その後 OECD の Recommendation of the Council on Artificial Intelligence[13]や G20 で採択された AI 原則[14]、総務省の国際的な議論のための AI 開発ガイドライン案[15]は、AI の積極的な開発・利活用を目指しつつ、同時に「不透明化や制御喪失など AI システムに関するリスクの抑制を図る観点から、関連する社会的・経済的・倫理的・法的な課題に対応することが必要」と指摘している。また、統合イノベーション戦略推進会議が決定した「人間中心の AI 社会原則[16]」は 7 つの原則で構成されるが、その内 6 つ（1.人間中心の原則、2. 教育・リテラシーの原則、3. プライバシー確保の原則、4. セキュリティ確保の原則、5. 公正競争確保の原則、6. 公平性、説明責任及び透明性の原則）までもが倫理的・法的・社会的な課題（ethical, legal and social implications: ELSI)に関連する項目である。国際的にも、AI によるイノベーションを通じた社会課題の解決が目指されると同時に、社会実装を進めるためにも、個人の権利を保護し、

---

[11] OECD. (2020). Consumer Data Rights and Competition - Background note. http://www.oecd.org/daf/competition/consumer-data-rights-and-competition.htm
[12] 高度情報通信ネットワーク社会推進戦略本部（IT 総合戦略本部）令和元年 6 月 14 日閣議決定 https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20190614/siryou1.pdf
[13] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
[14] G20 AI Principles, G20 Ministerial Statement on Trade and Digital Economy. https://www.soumu.go.jp/main_content/000625715.pdf (p.11-14)
[15] AI ネットワーク社会推進会議. 平成 29 年 7 月. 国際的な議論のための AI 開発ガイドライン案. https://www.soumu.go.jp/main_content/000499625.pdf
[16] https://www8.cao.go.jp/cstp/aigensoku.pdf

公正な社会環境を維持促進するための制度設計が大きな課題とされている。

データなどの情報財は複製コストが極めて小さく（限界費用がゼロ）、技術的には容易に複製・配布が可能な財の一つである。以前は情報へのアクセスに制約をかけて利用者を限定するケースが多かったが、デジタルデータの複製が容易であること、および個人情報の第三者提供など利活用に重点が置かれるようになった近時の社会的状況から、データが非排除性を有するものとなってきた。そのため、個人情報のネットワーク上への公開や、第三者への提供が一度なされると、そのデータがどのように流通し、利用されるかを把握・制御することは難しいとされてきた。それゆえ個人の権利保護のためには情報をデータ化しない、もしくはデータ化しても秘匿し続けることが合理的であると考える層が一定数存在する。その一方で、計算機やセンサー、データ通信技術の発展と費用の低廉化が進んだ結果、ますます多くの個人情報を活用したサービスアイデアが出現するようになった。それに伴い、蓄積されたデータが安全に活用されるための技術・制度設計が目指されるようになってきた。具体的には、コンピュータによる監視社会の到来に対抗するという文脈で提唱された自己情報コントロール権は、コンピュータ／ウェブサービスの普及とともに個人のプライバシーを保護しながら情報システムの恩恵を享受できるようなシステム設計へと、定義が拡張されてきた(山本, 2010)。

1980 年に勧告された OECD のプライバシー保護と個人データの国際流通についてのガイドライン[17]、いわゆる OECD8 原則には「個人参加の原則」が含まれ、2013 年の改正において国境を超えたデータ流通に対する対処を盛り込んでいる。この原則における「個人参加」とは、「自己に関するデータの所在及び内容を確認させ、又は異議申立てを保障するべき」というものであり、提供先組織による不当な取り扱いを抑止することを目的とする意味合いが強い。また、日本の個人情報保護法制においても、同様の取り扱いになっており(曽我部, 2018)、公権力による人権侵害の抑制という視点から議論がなされてきた(斉藤, 2019)。

しかし人間が予期できない結果を算出する AI の普及によって提供データがどのような目的に対して有効であるのか、あらかじめ予期することはますます困難になりつつある。また IoT 化の進展によるセンサーデバイスの増加によって、自身に帰属するデータがどのようなもので構成されているかを把握することも容易ではない。つまり、予め結んだ契約が前提となる自己情報コントロール権では、データの利活用拡大を促進するどころか抑制する効果が強くなる懸念がある。特に厳格な管理が求められる要配慮個人情報についても、与信・採用といった種類の重要場面におけるプロファイリングの実施に対して、本人関与の仕組みなどの整備が必要であるという主張もなされている(山本, 2019)。医療など本人が便益を享受できるユースケースを見据え、本人のプライバシーなどの権利を保護するための防御的な関与のみならず、個人情報を活用した便益創出の機会／権利を担保するという観点を、自己情報コントロール権の定義や運用に取り入れることが求められる。この構造は産業データにも同様に当てはまる。個人や組織が自身に帰属するデータから新たな便益を享受するためには、データの分析者と知識を交換し、共に用途を検討するというプロセスを、データ帰属者の権利保護を担保した上で実現できる法制度整備と技術仕様策定が必要となる。

---

[17] OECD (1980)「プライバシー保護と個人データの国際流通についてのガイドラインに関する理事会勧告」

2017 年に閣議決定された「未来投資戦略 2017[18]」では、産業データとパーソナルデータそれぞれの流通／利活用を促す方針が示されている。特にパーソナルデータをより広く流通させるために、Personal Data Store (PDS)ならびに情報銀行という概念が導入されている。PDS は「他者保有データの集約を含め、個人が自らの意思で自らのデータを蓄積・管理するための仕組み（システム）であって、第三者への提供に係る制御機能（移管を含む）を有するもの[19]」と定義され、様々なサービスの利用を通じて生成され、各サービス提供事業者が管理するデータの移転・活用について利用者自らがコントロールできる仕組みを提供する。情報銀行は「個人とのデータ活用に関する契約等に基づき、PDS 等のシステムを活用して個人のデータを管理するとともに、個人の指示又は予め指定した条件に基づき個人に代わり妥当性を判断の上、データを第三者（他の事業者）に提供する事業[20]」として規定され、個人が自らに帰属するデータの運用を信託するビジネスとして位置づけられている。日本では、2010 年代後半以降、個人が自らに帰属するデータについて、自らが新たな便益を享受するという目的の元、主体的に流通や処理をコントロールするための仕組みが整いつつある状況である。

一方、取引を通じたデータ利活用の拡大を目指した技術的な議論も活発に行われてきた。データ仲介者の役割や個人データの扱い方、マネタイズの仕組みを中心に、いくつかの特許が取得されるようになった。そして、ビッグデータブームと言われる 2013 年前後から「データ市場（data market、data marketplace、market of/for data など）」、「データ流通・データ取引（data exchange、data trading）」は急速に議論されるようになった。また、世界的な法制度の整備を受けて、今まで個々の組織や企業の中で収集・利用されていた多様なデータを、多対多で交換・取引するデータ市場ビジネスの社会実装が進んでいる。

同じ現象を測定したデータであっても、多様な表現形態が存在しうる。例えば、体温であれば測定単位(摂氏であるか、華氏であるか)、測定日時(年月日秒の表記や順番)、測定精度(用いられた機器に依存)などによって、同一事象が異なった形で表現されうる。複数の提供元から取得したデータを統合して解析しようとする場合、単位や表記方法の違いであれば変換すれば済むが、分類項目の違いや同一表記が示す意味が異なれば変換すら不可能である。民間企業間における単発の相対取引であれば、目的に応じて手変換するという対応が取れるが、公的部門のオープンデータや市場において不特定多数間で取引されるデータの場合、当初からデータ間の相互可用性が担保されていることが望ましい。

このような事態を防ぐため、データの共通仕様を策定する試みが長きに渡って推進されてきた。World Wide Web Consortium を始めとする様々な技術標準化団体が、データ記述手法／語彙／名前空間などの標準仕様の策定ならびに普及活動を展開している。また、公共データについては、内閣官房情報通信技術総合戦略室による地方公共団体向けの推奨デー

---

18 https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2017_t.pdf
19 内閣官房 IT 総合戦略室, (2017).「AI、IoT 時代におけるデータ活用ワーキンググループ中間とりまとめの概要」
https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/data_ryutsuseibi/dai2/siryou1.pdf
20 同上

タセット[21]や共通語彙基盤[22]などの仕様やガイドラインが策定されている。しかし、公的セクターにおいても、全ての主体が共通仕様に則ってデータを公開する、もしくはデータ連携を前提とした形態でデータを生成・管理することは難しい。業界ごと、もしくは系列や取引関係にある企業ごとに異なる慣習がある民間セクターにおいて、業界横断の共通仕様を採用するには高い障壁が存在する。この障壁は AI 解析用データ加工の経済合理性が高まり、DX の進展を通じた組織内データ仕様ならびに管理体制の集約化を通じて中長期的には解決される可能性がある。ただし、そうしたデータ仕様の標準化や、生成・公開・共有・利活用に至るオペレーション手順の確立支援など、公的セクターが果たすべき役割は少なくない。また、並行して技術的な解決策もプラットフォーム設計には導入される必要がある。

データは情報財であり、消費するまで中身を確認することができない。そのため、データ(の中身=content)に対して、それを注釈するデータ、つまりメタデータが付与される場合が多い。不特定多数間でデータが共有・取引されるデータプラットフォームにおいては、データの出し手、受け手が共通して理解可能な形式、かつ利活用者側がアプリケーション開発できるかどうかを判断できる内容を含んだメタデータが提供されなければならない。メタデータの生成、開示、評価といった取引を支援する機能をデータプラットフォームに実装する必要がある。

２．２．データプラットフォームにおける価値創造

現実の企業活動において、データの活用やそれに対する期待はどのように変遷し、経営に反映されてきたのであろうか。このことを直接観測することは困難であるため、本研究では日本の上場企業の決算短信をオルタナティブデータ[23]として分析した。決算短信は上場企業が四半期に一度発表する企業業績や経営方針に関する開示資料であり、文書中にはその企業が現在取り組んでいる経営課題、今後取り組む事業計画などが記載されている。ここでは、2013 年度から 2016 年度までの決算短信[24]を収集し、文書内に「データ」という語が出現する頻度を集計した（図 1)。ビッグデータブームが収束した 2013 年度以降も「データ」に言及する決算短信は継続的に存在し、データ戦略に対する関心が依然強いことが示唆される。また、集計結果を業種別に見てみると、情報・通信業のみならず電気機器、サービス業、小売業、卸売などでも頻繁に言及されている。

データエコノミーの発展を考える上で、データからどのように価値を創出するかは大きな課題である。2013 年頃のビッグデータ利活用への世界的な期待の高まり以前から、データの価値を検討する試みはいくつか行われていた。2010 年代に入ると、経済学の観点からデータの市場性や財としての性質が議論され始め(Aperjis & Huberman, 2012; Balazinska, Howe, & Suciu, 2011; Schomm, Stahl, & Vossen, 2013)、ゲーム理論を応用したデータの

---

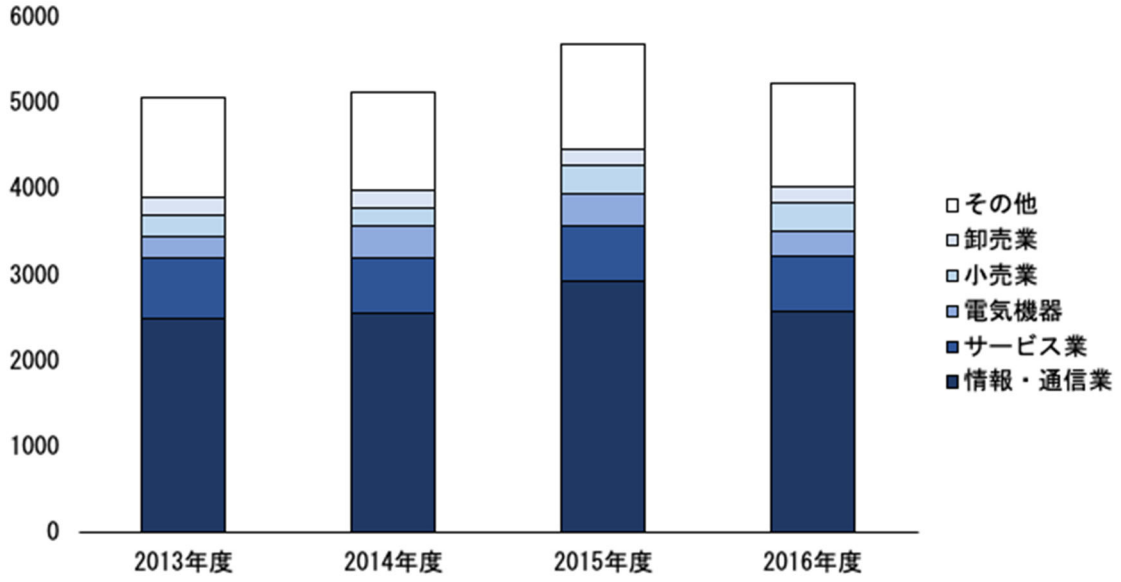
21 内閣官房 IT 総合戦略室による推奨データセット
https://www.data.go.jp/data/dataset/cas_20180312_0002

22 共通語彙基盤 https://imi.go.jp/goi/

23 オルタナティブデータとは、その名が示す通り「代替」データであり、直接観測が難しい事象を理解するための代わりのデータである。特に投資判断の材料となるとして、近年注目が集まっている。

24 2017 年度からは決算短信の書式が簡素化され、経年変化比較には適さないため除外。

図 1. 決算短信における単語「データ」の出現回数（業種別）



価格決定メカニズムや、プライバシーを考慮したデータ交換と分析手法なども提案されている。

本論文では、データや情報はつながることで価値を生むという観点から、データによる価値創造の可能性を検討する。新たな経済価値を生み出す有力な手段であるイノベーションは、既存の生産手段や資源などの新たな組み合わせ／新たなつながり（＝新結合の遂行）であることが指摘されてきた(Schumpeter, 1934)。個別には必ずしも意味・価値をもたないデータも、他のデータとの結合によって意味や価値の創発をもたらす可能性がある。國領(2013)は、そのようなデータ間のつながりを促進する基盤をプラットフォームと呼び、プラットフォーム上でのデータ・情報のつながりが人間の協働や新しい価値の創造につながりうることを示した。個人による価値創造や創造性を促す要因分析においても、個人の特性を分析するだけでなく、それら個人がどのような社会ネットワーク（＝人間同士のつながり）に埋め込まれているかを分析することの重要性が指摘されてきた(Burt, 1992; Granovetter, 1985)。

一方で、機密扱いの知的財産に関する情報や個人情報の流出、情報の独占性を失うことによるビジネス機会の損失などといった想定されるリスクは、データ連携の阻害要因である。そもそもデータ活用・連携の形態や、それによってもたらされる便益を具体的に想定することは難しいのが現状である。データ流通プラットフォームの構築・運営は萌芽的な取り組みであり、どのようなデータがどのような関係性によってエコシステムを形成しているのか、どのようなデータに潜在的に価値があるのか、どのようなデータがプラットフォームの中で分野横断的にデータ連携を促しているのか、などについて十分な知見が得られていない。そのようなデータ間のネットワーク構造や、価値の形成状況、ならびに価値創出の可能性を示すことは、プラットフォームへのデータ提供を活性化し、データエコノミーを促進することにつながる。

３．研究手法

以上の議論をふまえ、本論文ではプラットフォーム上の多様なデータの関係性をネットワークと捉えて分析することで、データ利活用による価値創造メカニズム解明を試みる。プラットフォーム上における多分野にまたがるデータ間の関係性や構造的特徴を解明するために、Hayashi & Ohsawa (2020)で用いられたデータ流通プラットフォームのデータを分析対象とし、プラットフォーム上のデータ群の特性及び関係性に着目したネットワーク分析を行うことで、潜在的に価値を持つデータの特徴について検討する。

３．１．研究対象とデータセット

本研究では、データジャケット（Data Jacket: DJ）によって収集されたデータ群を研究対象として用いた。DJ はデータ自体を共有するのではなく、データの使途や概要をメタデータの形式で記述する手法である(Ohsawa, Kido, Hayashi, & Liu, 2013)。これにより、誰がどのようなデータをどのような形で有しているのかということを理解可能となり、さらにはデータに対する評価や期待を共有し、異なる事業者間でデータを用いた新たな分析手法の提案や共創ビジネスを促進することが可能となる。

数あるデータ概要情報記述手法の中で DJ を用いることの大きなメリットは 2 つある。1 つは、オープンデータに代表される共有可能データのみならず、個人や企業が持つ秘匿データも対象としているところである。データ流通プラットフォームは多様な領域のデータを交換・取引・売買し、価値を生み出す場であることを考慮すると、DJ 記法によって収集されたデータ群は公開が困難なデータの概要を把握する貴重なサンプルとなる。DJ 記法には複数の共有条件を記述可能であるが、本研究では実験参加者の理解を優先させ、個人情報保護法などに定められた区別は用いておらず、共有可能データと秘匿データという 2 種類に分けて被験者に提示している。なお本研究における共有可能データは、第三者が自由に利用できるオープンデータならびに契約さえ結べば不特定多数に対して共有可能な情報を指す。秘匿データは主に個人や企業が有するデータであり、不特定多数には共有できないが、特定の対象に対してであれば共有について検討可能なデータを意味する。

もう一つのメリットは、変数に関する記述である。本論文における変数は、データセットが有するパラメータを意味する(Babbie, 2016)。気象データであれば、「日時」、「地域名」、「天気」、「最高気温」、「最低気温」などが変数であり、健康診断データであれば、「身長」、「体重」、「年齢」、「性別」及びその「検査項目」などが該当する。変数はデータ自体の構造を規定するものであり、他データとの結合性などの特徴を議論する上で重要な情報となる(Ridder & Moffitt, 2007; Weiner, Quwatli, Perkins, Lewis, & Callahan, 2006)。すでに提案されているメタデータにもタグのような形で変数に関する記述項目が存在するものもあるが、その利用方法は検索の手がかりとするなどに留まっている。

本研究では、DJ サイト[25]及び datajacket.org[26]にて 2013 年 9 月から 2018 年 8 月までにデータ市場及びデータ流通市場に参画している企業、研究者、データサイエンティストが登録した、変数を有する 1316 件の DJ を解析対象とした。表 1 は全データ及び共有条件別の

[25] https://sites.google.com/site/datajackets/
[26] https://datajacket.org/

サンプル数である。なお、共有条件はすべてのデータにて記入されているわけではないため、共有条件が記されていないデータは対象外とする。

表１．本研究で用いるデータ

	全データ	共有可能データ	秘匿データ
データの総数	1316	654	571
変数の総数	9022	4414	4214
変数の種類数	6594	3198	3206

３．２．分析手法

異分野データ連携においては、単一のデータから有益な知見を得るだけでなく、異なるデータを連携させることで新たな価値を創出するという考え方をもとに、共通する変数を持つデータは連携可能性が高いという仮定に基づいてデータのネットワーク構築をした[27]。すなわち、DJ 上の各データをノードとして取り扱い、データ同士が共通する変数を有している場合にリンクを有するネットワークを構築した。図 2 は本研究のネットワーク分析のプロトコルである。

どのような特性を持ったデータがデータ間のつながりを誘発し、新しいサービスやプロダクトにつながりうるイノベーションを促進するかについて考察するため、ネットワーク分析の指標として次数中心性（degree centrality）、媒介中心性（betweenness centrality）、拘束度（constraint）の 3 種を用いた。次数中心性は、ネットワークにおける次数（リンクを持つ数）の大きさを指標化したものである。ノードiの次数中心性$C_D(i)$は、次数をk、全ノード数を Nとした時に、以下の式で計算される(増田&今野, 2010)。

$$C_D(i) = \frac{k_i}{N-1}$$

媒介中心性は、あるノードが他ノードの最短経路上に位置する程度から、ネットワーク上の異なる集団をつなぐ度合いを指標化したものである。ノードiの媒介中心性$C_B(i)$は、$\sigma_{st}(i)$をノード sと tの最短経路とした時に、以下の式で計算される(増田&今野, 2010)。

$$C_B(i) = \sum_{s,t\ \in N} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

拘束度はノードが特定のネットワーク集団に拘束されている度合いを指標化したものであり、拘束度が低いノードほど多様な集団の仲介者としての役割を果たす。ノードiの拘束度C_iは以下の式で表される(Burt, 1992)。なお、p_{ij}はノードiとjの直接的な関係の強さである。

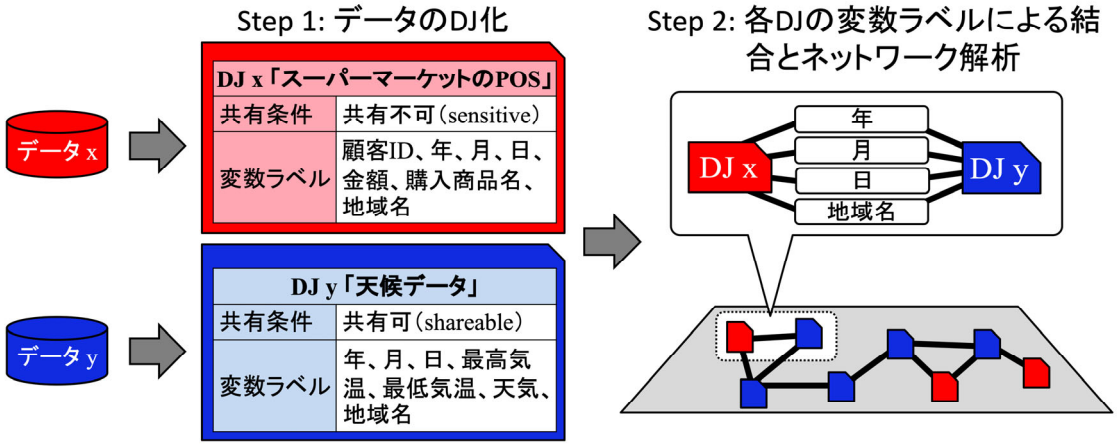
$$C_i = \sum_j c_{ij}$$

[27] データ流通エコシステム並びにプラットフォームにおいて、分野を横断して用いられる統一的なスキーマや変数の標準化は発展途上である。本研究ではバイアスを避けるため辞書の適用など、変数に対して特別な加工を行わずに分析を行った。

$$c_{ij} = \left(p_{ij} + \sum_{q} p_{iq} p_{qj} \right)^2$$

さらに、ネットワークのマクロ構造を定量化して捉えるため、平均次数、密度、クラスタ係数(Barabási, 2016)、及び次数親和性(Newman, 2002)を用いて大域的理解を試みる。

図２．ネットワーク解析のプロトコル



４．分析結果

まず 4.1 ではデータの内容に着目し、連携可能性の高いデータを明らかにする。4.2 ではデータの共有条件に着目し、秘匿データと共有可能データそれぞれの利用状況及びその関係性について分析する。4.3 では個別データの分析を超えて、データプラットフォームの全体構造を捉えることで、プラットフォーム上のデータ活用形態について考察する。

４．１．連携可能性の高いデータ

データの活用可能性を判断する一つの指標として、データがどれだけ多くの変数を有するかという点が考えられてきた。つまり、多くの変数を持つデータほど、データ活用・分析の幅が広がるというものである。しかし、データが持つ変数の数とデータネットワークにおける次数の相関係数はわずか 0.137 であり、多数の変数を有するデータほど他データとの連携可能性が高いという単純な関係性とはならないことが分かった。そこで、保有変数の数という単純な指標ではなく、次数中心性、媒介中心性、拘束度の 3 つのネットワーク指標を導入し、データ流通プラットフォームにおいて異なるデータとの連携可能性が高い、すなわちデータ連携によるイノベーションに貢献しうるデータについて検討する。

表 2 は各ネットワーク指標が上位 10 件のデータである（データ名が太字になっているものは、秘匿データであることを示す）。なお、拘束度は値が小さいほど多様なデータ群をつなぐ度合いが高いことを示していることに注意したい。本論文で構築したネットワークにおける次数中心性は、当該データがどの程度他データと同じ変数を共有しているかの度合いを表す。つまり、次数中心性で上位のデータほど数多くのデータと変数を共有しているため、当該データ単独では実現できないプロダクトやサービスを他のデータを組み合わせる

ことで展開できる可能性が高まると言える。

表２．ネットワーク指標の高いデータ上位 10 件

順位	データ名	次数中心性	データ名	媒介中心性	データ名	拘束度
1	駐車場利用数	0.216	交通機関	0.214	Twitter／Blog／掲示板	0.024
2	Facebook	0.178	世界幸福度	0.050	コワーキングスペース	0.025
3	公衆トイレ情報	0.174	Twitter ／ Blog ／掲示板	0.049	駐車場利用数	0.027
4	Twitter	0.173	Wi-Fi 利用状況	0.041	Wi-Fi 利用状況	0.028
5	健康食品／美容系通販	0.173	日本食売り上げ	0.034	Facebook	0.028
6	震源リスト	0.171	避難所情報	0.032	健康食品／美容系通販	0.029
7	オゾン層関連	0.168	体験共有アプリ	0.031	Twitter	0.029
8	紫外線関連	0.168	駐車場利用数	0.028	震源リスト	0.029
9	GPS による位置情報	0.168	産業別就業者数	0.028	携帯電話契約者	0.029
10	人事関連	0.167	地震関連	0.026	公衆トイレ情報	0.029

上位のデータを見ると、駐車場利用数データ、Facebook や Twitter などのソーシャル・メディアデータ、公衆トイレデータなどがある。例えば駐車場利用数データには、「駐車場利用量」、「緯度」、「経度」、「住所」、「日時」などの変数が含まれ、「緯度」、「経度」などの位置を表す変数は Facebook データや公衆トイレデータにも含まれている。次数中心性の高いデータ群は、「場所」や「時間」に関わる変数を通して様々なデータを結合しており、これらの変数は異種のデータの連携性において欠かせない変数であるということができる。

続いて媒介中心性について議論する。媒介中心性は、当該データが他のデータ間をつなぐ最短経路上に位置する程度を示している。媒介中心性が高いデータほど効率的に複数データの結合が実現できることを意味するため、データ結合を媒介してプロダクトやサービスを創出する可能性が高まると言える。媒介中心性上位のデータとしては、交通機関データ、世界幸福度データ、SNS 上のテキストデータ、Wi-Fi 利用状況データ、日本食売り上げに関するデータなどがある。例えば交通機関データには「交通量」、「時間」、「場所」、世界幸福度データには「幸福度」の他に「緯度」、「経度」、「国」、「パーソナル ID」などの変数が含まれる。これらデータからも、「場所」や「時間」に関わる変数を通して、異なるデータ群がつながる可能性が高まることが示唆される。

ネットワーク拘束度は仲介者度合いとしても知られており、当該データがどの程度一つのネットワーク集団に閉じているかの程度を示す指標となる。ネットワーク拘束度の低さ

が上位のデータほど多様なグループのデータ群をつなぐ役割（仲介者）を果たしていることを意味するため、通常は結びつかないデータ群をつなぎあわせ、新規性の高いプロダクトやサービスの創出につながる可能性が高まる。拘束度が低い、すなわち仲介者度合いが高いデータとしては、Twitter／Blog／掲示板データ、コワーキングスペースデータ、駐車場利用データ、Wi-Fi 利用状況データ、Facebook データなどが存在する。これらデータは「場所」に関わる変数を通してつながりを生み出すと共に、サービスの「利用」に関わる変数を媒介とすることで、新規性の高いデータ結合を生み出すことが期待できる。

４．２．共有条件とデータの連携可能性

前節では個別データの連携可能性に注目したが、本節ではデータの共有条件に着目してデータ連携の可能性を分析する。具体的には、共有可能データと秘匿データがそれぞれどのように／どの程度データ連携されうるかという点について検討する。

まず同条件のデータ間連携可能性について見ると、共有可能データ間のリンク数は 6663 であった一方、秘匿データ間のリンク数は 2460 であった。この結果は、共有可能データ間においてデータ連携可能となる組み合わせが多いことを示している。さらに、共有可能データと秘匿データの間をつなぐリンク数は 5146 であった。秘匿データは、秘匿データ同士での連携よりも、共有可能データとの連携可能性が高いことを示している。

続いて次数中心性の上位 10 データを見ると、6 件の秘匿データと 4 件の共有可能データが存在する。一方で、全秘匿データと全共有可能データの平均次数中心性を比べると、秘匿データ平均は 0.045、共有可能データ平均は 0.056 となっている。したがって、上位のデータに限った場合には連携可能性の高い秘匿データが多く存在する一方、全データ平均の観点からみると共有可能データの方が一般的により多くのデータと変数を共有していることを示している。

さらに媒介中心性の上位に位置するデータの大部分（上位 5 データ中 4 データ）は秘匿データであり、秘匿データを介して様々なデータが結合されうることが示唆される。同じくネットワーク拘束度に関しても上位 5 データ中 4 データは秘匿データである。一部の有益な秘匿データを介して異なる種類のデータ群が結びつくことで、新たなプロダクトやサービスの創出につながりうることが示唆される。

以上の分析内容から、秘匿データと共有可能データを組み合わせることの重要性が示唆される。リンク数の分析からは、全体として共有可能データの連携可能性の高さ、及び秘匿データ単独ではなく秘匿データと共有可能データ間の連携可能性の高さが示された。一方で中心性や拘束度といったネットワーク指標の分析からは、指標上位に来る一部の秘匿データ群の連携可能性の際立った高さが示された。これら結果から、一部の価値の高い秘匿データを核としながら、共有可能データを組み合わせてデータ連携を促進していくことの重要性が示唆された。

４．３．データネットワークの構造

本節では共有条件や変数といった個別データの特徴を超えて、データネットワーク全体のマクロ構造について考察する。まず本データネットワークの大部分は、多数の変数を有するデータ群によって構成されているのではなく、少数かつ多様な変数を持つデータ群によ

って構成されている点に注目したい。全データで 6594 の変数が観察された一方、その中で複数のデータに共通の変数はわずか 207 に留まっていた。つまり、標準的に採用されている変数はごくわずかであり、その他大多数の変数は出現頻度が少ない／ユニークな変数群である。

図３．データネットワーク構造図

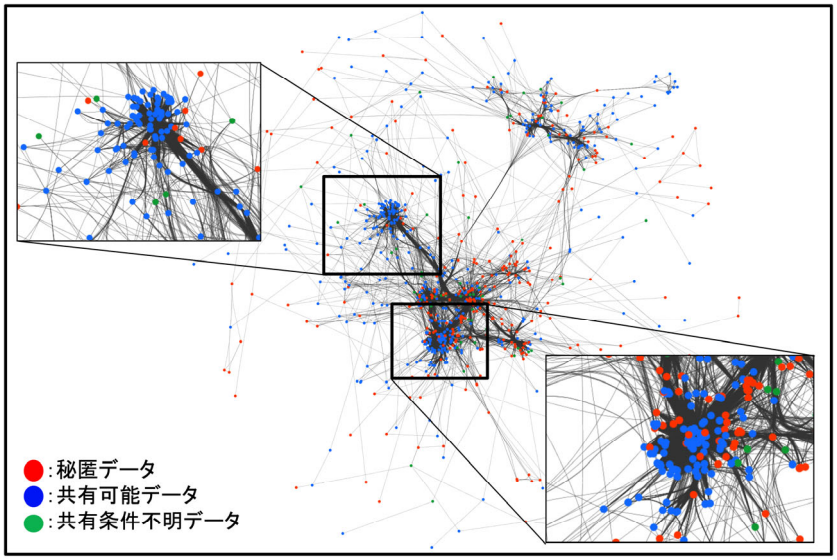


図 3 はデータ流通プラットフォームにおけるデータネットワークの構造図である。図中の青ノードは共有可能データ、赤ノードは秘匿データを表す。共有可能データ同士がつながってクラスタを形成している例と共有可能データと秘匿データが相互につながっている例を示している図中で一部拡大表示している。

表 3 はデータネットワークの特徴から、密度、クラスタ係数、次数親和性を算出したものである。最大連結成分を構成する 798 のデータによるネットワークでは、大域的に疎であり、局所的に密なつながりが観察された。さらに、次数親和性は極めて高い正の値であり、データネットワークは、自然界や工学的なネットワークと比較して、人間関係のネットワークに似ていることを表している。これは「類は友を呼ぶ」傾向がネットワークの随所で見られることを意味し、同じようなデータ同士が密なかたまりを構成してネットワーク上に偏在しているということになる。

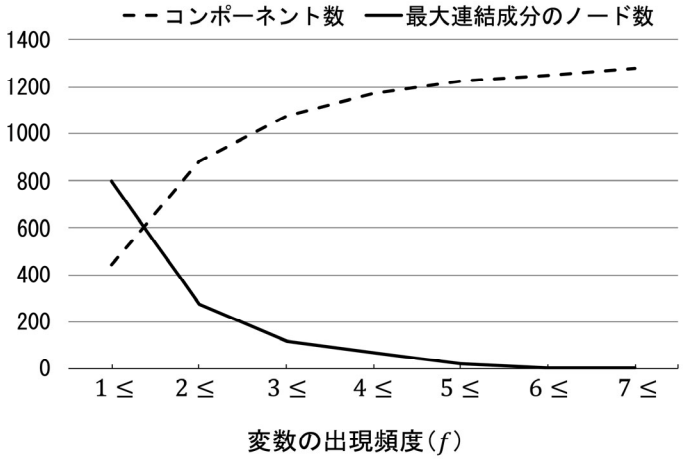
表３．データネットワーク構造の特徴量

	最大連結成分
密度	0.051
クラスタ係数	0.702
次数親和性	0.489

また、図 4 は変数の出現頻度(f)を変化させた際のネットワーク全体のコンポーネント数と最大連結成分のノード数の推移である。変数の最低出現頻度を上げていくことは、最大連

結成分に残るデータが標準化された変数で構成されることに対応している。$f \geqq 1$のときは、最大連結成分のノード数 798、ネットワーク全体のコンポーネント数は 443 である。最低出現頻度を上げていくと、コンポーネント数は増大し、最大連結成分のノード数は急激に減少していく。そして、$f \geqq 7$では最大連結成分のデータ数は 5 以下となり、ほぼすべてのデータが結合を失う。過度な標準化がデータの連携可能性を阻害する可能性が示唆された結果となった。

図４．変数の増減によるネットワーク構造の変化



５．議論

５．１．データプラットフォーム運営への示唆

前章の分析結果を踏まえ、本章では現在のデータ流通プラットフォームの課題とこれからの発展への展望を議論したい。はじめに、3 つのネットワーク指標を用いて中心的な役割や分野を横断して結合するデータについて議論した。今回対象としたデータは、基本属性などこれまで企業がマーケティング施策のために活用してきた情報に限らない種別を含んでいる。このような種類のデータは、データ帰属者にとってすら、その価値を正確に認識するのは難しい。特に近年増加する AI によるビッグデータ解析では、処理を明示的に設計したシステムとは異なり、メカニズムの理解が難しい機械学習アルゴリズムによって処理されることも多いため、帰属者にとっても活用の可能性や範囲を推定するのは難しい。加えて、データは組織内に閉じて活用されることが一般的であり、現状は市場で十分に流通しているわけでもないため、的確に価値を判断する材料が不足している。そのような制約を解消する有力な手段として、本論文ではプラットフォーム上のデータ間の関係性、及び各データがネットワーク全体の中でどのような位置に存在するかを可視化した。これにより、各データがもつ価値の可能性に対する理解を広げるとともに、利用者側のニーズとのマッチングを促進することが可能となる。このような可視化機能をデータ流通プラットフォームに実装することにより、データ供給ならびに利活用に至るマッチングの促進につながると考えられる。

本論文の分析では、秘匿データのみ／共有可能データのみを対象とするのではなく、秘匿

データと共有可能データを混在させることでより密なネットワークが構成されることが示された。つまり、これまで多く行われてきた特定主体間の限定されたデータの相対取引に、不特定多数で共有可能なデータを混在させることによって、多様な共有条件を有するデータの流通が促される。それによって、これまで取引関係が成立しなかった分野間でのデータ連携が実現し、イノベーション促進の効果が期待できる。一方で AI に関わる国際的な議論の中では、個人情報などの特定主体間で限定的に使用されてきたデータを暗号化等によってより柔軟に活用する方向も模索されてきている。そのような議論とも合わせて、データの共有・活用可能性を高めてデータ利用を促進していく枠組みが重要になってくると考えられる。

さらに、データを構成する変数に着目したとき、プラットフォームに含まれる変数は膨大かつ多様な低出現頻度の変数で構成されていることも明らかとなった。この特徴によって、データのネットワークは局所的に密、大域的に疎な構造となる。この結果はデータの世界において未だ取得されていない潜在的なデータの存在を示唆している。つまり、大域的に疎な部分を埋めるデータを新たに創出することで、異なる領域を結ぶ価値あるデータを市場に提供することが可能となる。さらに、プラットフォーム上のデータが、人間関係と同様の「局所的に密、大域的に疎」の構造を示すということは、データもタコツボ化している可能性を示唆している。つまり、そのままでは新たな用途の発見にはつながらず、公開のコストのみが発生する一方で、公開のメリットを生み出しづらい状況にあると言える。すなわち、価値を創出する可能性を有するデータが、狭いコミュニティの中で死蔵されているという可能性を示すといえよう。このタコツボ化の解消には、多種多様な領域のデータの共有を促進するようにプラットフォームを設計することが重要と考えられる。

ビッグデータ解析の対象として用いるのであれば、データ間の相互可用性の担保も重要である。本研究では、高頻度に出現する変数のみを使って分析を行うことで、標準に準拠したデータのみで連携するという状況におけるネットワーク形成状況をシミュレートしたが、このような条件では連携の発生が減少するという結果が示された。つまり、相互可用性の実現を重視するあまり、フォーマットや語彙といったデータモデルを無理に統一しようとすると、統一仕様への変換が困難なデータの排除につながることが示されたのである。過度に硬直的な標準仕様への準拠を求めることは、データ流通の抑制につながるため、できるだけデータの出し手に対する制約が少なくて済むようなプラットフォームの設計が必要となる。例えば、多様なデータを計算機処理できる形で流通させるための機能プラットフォームを、データカタログなどの設計において実装するなどの方法が考えられる。AI 技術の発展は、自然言語処理などの手法によって表記揺れを修正したり、表記は異なるが実質的に同一内容のデータのリンクを推薦したりするといった機能を可能にするだろう。このような機能を積極的にプラットフォームに実装すること、またメタデータも計算機処理に対応した設計とすることが重要である。

日本では情報銀行など関連法制の整備によってデータ流通プラットフォームの設置が進

んでいるが、これは国内に限った事象ではない。EU のデータ戦略[28]と標準化[29]を含む相互可用性実現のための試み[30]、米欧を中心にデータ流通プラットフォームの創業が相次いでいる中、国境を超えてプラットフォーム間でのデータ交換が活性化することで、より多様性のあるデータ流通が実現されようとしている。データならびにメタデータ自体の相互可用性のみならず、プラットフォーム間の相互可用性をどのようにして設計するかも重要となる。一方で、相互可用性を過度に追求することはデータの多様性を抑制する可能性もあるため、相互可用性をある程度担保しながらも表記などに柔軟性を許容できるような仕様であることが重要である。相互可用性実現のための標準化活動は、民間の企業コンソーシアムや標準化団体を舞台として推進されることも多い。そのため、仕様策定は各ステークホルダーのデータ利用実態に即して民間主導で行われる必要がある。公的部門は民間の活動支援、策定仕様の認証やデジュール化、運用・普及面でサポートする役割を担うことが重要であると考える。

データという財は実際に利用されるまでどのような、そしてどの程度の効用を実現できるか定まらないという特徴がある。この特徴が、データ保有者に提供を思いとどまらせる原因の一つとなっている。行政オープンデータ推進時に現場部署から挙がった声が、「手間をかけて公開用データを作成して、どのようなメリットがあるのか」であったことは、この特徴がデータ流通の大きな阻害要因となってきたことを物語る。データ流通を促進するには、プラットフォームへのデータ提供を促す要因の抽出・設計を進めていくことが必要である。

５．２．データエコノミー政策への示唆

パーソナルデータ分野においては GDPR 策定作業において、事業者間でデータを引き渡して利用することができるデータポータビリティの法制化が 2010 年代半ばには進展する(小向, 2018)など、法人間のデータ流通に対する法的・技術的な条件整備が進められてきた。GDPR に先立つこと 1 年、2017 年にオーストラリアは Consumer data right (CDR)を明文化し、米国でも 2020 年に The California Consumer Privacy Act (CCPA)が発行している。このような状況を踏まえ、データ流通市場は既に成長段階に達しているという前提での経済学的分析もなされている一方で(Spiekermann, 2019)、市場の成長を促すという観点での制度設計に関する議論・研究の蓄積は未だ十分ではない。数少ないデータ流通市場の構造ならびに成長動向を分析した研究として Stahl らの分析があるものの(Stahl et al., 2014)、そこで得られた知見をもとにした制度設計の議論は進んでいない状況である。またデータ流通市場自体を取り上げた研究は存在するものの、データ流通エコシステム全体を包括的にとりあげた研究や議論が進んでいるとは言い難い状況にある。

---

[28] A European strategy for data. COM(2020) 66 final.
https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[29] European Commission (2018). The Rolling Plan on ICT Standardisation
https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation

[30] European Commission (2017). European Interoperability Framework – Implementation Strategy. COM(2017) 134 final.
https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_1&format=PDF

GDPR はデータポータビリティを制度化したのと同時に、従来のデータ流通市場から個人情報を含むデータの流通を制限している。プライバシーデータに特化した市場構造を分析した研究もあるが(Niu, Zheng, Wu, Gao, & Chen, 2017; Shen et al., 2016)、プライバシーデータ流通に対する制限が市場に与える影響は未解明のままであり、GDPR が及ぼす経済的影響の分析はまだ十分に進んでいるとは言い難い。データという財の特性ゆえに、関連法制度の在り方や流通プラットフォームの設計は、市場の活性化や成長に多大な影響を及ぼす。それゆえに健全な市場を育成するためには、個人情報や知的財産の保護と取引の活性化のバランスを取ることを前提としつつ、データ帰属者が共有による便益をより多く享受できる形での施策立案、運営を戦略的に進めていく必要があるだろう。GDPR を策定した EU も、もともとはデータドリブン・イノベーションの実現を標榜し、データ戦略をまとめている。データ保護はデータ戦略の前提条件であり、それそのものだけが目的とされるものではないと考えられる[31]。

日本の個人情報保護法が GDPR を制定している EU との十分性認定を前提として制定・運用されており、第四次産業革命のグローバルな推進、Data Free Flow with Trust の推進といった活動が政府レベルで進められている一方で、実態としては日本においてパーソナルデータや産業データ利活用が進展するペースはあまり早くない。一方、EU においては非個人データの域内自由流通枠組みに関する規則[32]が策定され、米国では消費者プライバシー権利法の継続的な改定など、グローバルなデータ流通、つまり越境データを前提とした規範ならびに制度設計が進められている。これらの流れを総合的に捉えると、個人の権利保護を前提とした上でデータ流通の活性化を通じ、企業のみならず、データを提供する消費者自らがより多くの便益を享受できるための環境整備が重要であり、それはグローバルな視座に基づくプラットフォームの構築ならびに技術仕様の検討が不可欠である。

本研究では、データ流通プラットフォームで扱われるデータの分析を通して、データの利活用を促進する条件を検討してきた。データネットワークの分析結果からは、秘匿データのみでデータ流通プラットフォームを構築するのではなく、共有可能データと秘匿データが混在する環境を整えることでデータ間連携が促進されるという知見を得た。データ連携の促進には、流通するデータの仕様や内容とともに、プラットフォームや関連法制度をどのように設計するかが大きく影響する。その設計を支援する上では、既存データの実証分析だけではなく、実証分析を踏まえて構築したモデルに基づくシミュレーションが、法制度ならびに技術仕様を横断した形でのデータ利活用の進展を実現するための有望なアプローチである。市場の制度設計に対してシミュレーション・アプローチを採用した研究事例には、証券取引所におけるティックサイズ(取引価格の最小単位)変更の影響分析(水田,早川,和泉&吉村, 2013)や、自己資本比率規制の変更がもたらす影響分析(Hirano, Izumi, Shimada,

[31] A European strategy for data. COM(2020) 66 final. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[32] REGULATION (EU) 2018/ 1807 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 14 November 2018 - on a framework for the free flow of non-personal data in the European Union, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN

Matsushima, & Sakaji, 2020; 米納&和泉, 2016)などがある。これら金融市場に対して適用されてきたアプローチは、データ流通市場に対しても同様に適用可能であろうと考える。

６．おわりに

本論文では、プラットフォームにおけるデータ間の関係性についてのネットワーク分析から、データ流通プラットフォーム運営及びデータエコノミー推進への示唆と課題を議論した。これまで主流であった特定主体間の限定されたデータの相対取引に不特定多数で共有可能なデータを混在させることによって、多様な共有条件を有するデータの流通が促される。それにより、これまで取引関係が成立しなかった分野間でのデータ連携が実現し、イノベーション促進の効果が期待できるという示唆を導出した。この分析から、データ提供を促進するための環境整備にオープンデータ普及が有効であると同時に、提供者が自身で価値を見つけられないデータが共有されて価値を生み出すためのプラットフォーム設計が重要であることが示された。用途や処理手法が事前に予期できないというデータ特有の条件のもとでプライバシー保護とデータ流通を通じたイノベーション促進を両立するためには、法制度と技術仕様設計を統合的に検討し、構築していくことが必要である。このような包括的な議論を実現・活性化していくために、データ分析やシミュレーションなどを通じたデータ流通エコシステムの研究を引き続き進めていきたい。本論文のアプローチと考察が、我が国と世界的なデータ流通エコシステムの発展の一助になれば幸いである。

謝辞

本研究は JSPS 科研費 JP20H02384, 19K23235 の助成を受けたものです。

参考文献

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442-92.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500.

Aperjis, C., & Huberman, B. A. (2012). A Market for Unbiased Private Data: Paying Individuals According to Their Privacy Attitudes. *Working Paper. ArXiv:1205.0030*, 1–17.

Babbie, E. R. (2016). *The basics of social research* (7th Editio). Boston, MA: Cengage Learning.

Balazinska, M., Howe, B., & Suciu, D. (2011). Data markets in the cloud: An opportunity for the database community. *Proceedings of the VLDB Endowment*, *4*(12), 1482–1485.

Barabási, A.-L. (2016). *Network science*. Cambridge, UK: Cambridge University Press.

Burt, R. S. (1992). *Structural Holes: the Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Ejiri, Y., Ikeda, E., & Sasaki, H. (2018). Realization of data exchange and utilization society by blockchain and data jacket: Merit of consortium to accelerate co-creation.

In *IEEE International Conference on Data Mining Workshops* (pp. 180–182).

Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Hayashi, T., & Ohsawa, Y. (2020). Understanding the Structural Characteristics of Data Platforms Using Metadata and a Network Approach. *IEEE Access*, *8*, 35469–35481.

Hirano, M., Izumi, K., Shimada, T., Matsushima, H., & Sakaji, H. (2020). Impact Analysis of Financial Regulation on Multi-Asset Markets Using Artificial Market Simulations. *Journal of Risk and Financial Management*, *13*(4), 75.

Kemp, K. (2019). Concealed Data Practices and Competition Law: Why Privacy Matters. *University of New South Wales Law Research Paper*, 19-53.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A Survey on Big Data Market: Pricing, Trading and Protection. *IEEE Access*, *6*, 15132–15154.

Mano, H. (2016). EverySense: An end-to-end IoT market platform. In *Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services* (pp. 1–5).

Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, *89*(20), 208701.

Niu, C., Zheng, Z., Wu, F., Gao, X., & Chen, G. (2017). Trading data in good faith: Integrating truthfulness and privacy preservation in data markets. In *IEEE 33rd International Conference on Data Engineering* (pp. 223–226).

Ohsawa, Y., Kido, H., Hayashi, T., & Liu, C. (2013). Data jackets for synthesizing values in the market of data. *Procedia Computer Science*, *22*, 709–716.

Ridder, G., & Moffitt, R. (2007). The Econometrics of Data Combination. *Handbook of Econometrics*, *6*(B), 5469–5547.

Robertson, V. H. (2020). Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data. *Common Market Law Review*, 57(1), 161-190.

Schomm, F., Stahl, F., & Vossen, G. (2013). Marketplaces for data: An initial survey. *SIGMOD Record*, *42*(1), 15–26.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.

Shen, Y., Guo, B., Shen, Y., Duan, X., Dong, X., & Zhang, H. (2016). A pricing model for Big Personal Data. *Tsinghua Science and Technology*, *21*(5), 482–490.

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*(4), 208–216.

Stahl, F., Schomm, F., & Vossen, G. (2014). The data marketplace survey revisited. *ERCIS Working Paper*, *18*.

Weiner, M., Quwatli, Z., Perkins, A. J., Lewis, J. N., & Callahan, C. M. (2006). Limitation of a single clinical data source for measuring physicians' performance on quality indicators. *Journal of the American Geriatrics Society*, *54*(8), 1256–1260.

Yagihashi, T. (2019). Social Data Platform, D-Ocean. In *EEE International Conference on Data Mining Workshops* (pp. 45–48).

板倉陽一郎. (2020). プライバシーに関する契約についての考察（問答編). 情報通信政策研究, *3*(2), 95–102.

黒政敦史. (2020). 匿名加工されたデータの利活用に向けた課題. 情報通信政策研究, *3*(2), 171–194.

國領二郎. (2013). ソーシャルな資本主義. 日経BP.

小向太郎. (2018). データポータビリティ（特集 GDPRの適用開始をめぐって). *ジュリスト = Monthly Jurist*, (1521), 26–31.

斉藤邦史. (2019). プライバシーにおける 「自律」 と 「信頼」. 情報通信政策研究, *3*(1), 73–90.

曽我部真裕. (2018). 自己情報コントロール権は基本権か？. 憲法研究, (3).

中川裕志. (2020). AI 倫理指針の動向とパーソナル AI エージェント. 情報通信政策研究, *3*(2), 1–24.

藤野直明. (2020). 第4次産業革命とシステムの経済: パラダイムシフトとシステム変革への要請. 横幹, *14*(1), 50–63.

増田直紀, & 今野紀雄. (2010). 複雑ネットワーク—基礎から応用まで. 近代科学社.

水田孝信, 早川聡, 和泉潔, & 吉村忍. (2013). 人工市場シミュレーションを用いた取引市場間におけるティックサイズと取引量の関係性分析. *JPXワーキング・ペーパー*, *2*, 1–14.

山本龍彦. (2010). プライバシーの権利. ジュリスト, (1412), 80–90.

山本龍彦. (2019). 「完全自動意思決定」のガバナンス—行為統制型規律からガバナンス統制型規律へ？. 情報通信政策研究, *3*(1), 25–45.

米納弘渡, & 和泉潔. (2016). 人工市場を用いた自己資本比率規制に基づく市場リスク管理が複数資産市場に与える影響の分析. *JPXワーキング・ペーパー*, *18*, 1–12.

（掲載決定日：令和２年 10 月 12 日／オンライン掲載日：令和２年 11 月 20 日）