

データの経済的価値と法的保護 実証分析とケースによる考察

渡部 俊也¹ (東京大学)、 平井 祐理² (文部科学省)

要 旨

データ政策を検討するうえで、基礎となるのが、データの価値をどうとらえるのか、どのようなデータに経済的な価値があるのかという考え方である。一般的にデジタルデータを多量に保有しているからといって、そのデータから利益を生み出せるとは限らず、多くの条件が整って初めて企業の利益や競争力につながり、産業競争力に結び付く。その要件を踏まえたデータの保護やデータ利活用のインセンティブシステムを設計することがデータ政策として重要となる。

本稿ではそのような観点から、データについての経済的価値に着目して、質問票調査に基づく実証分析による検討結果を示し、あわせて具体的な事例として、機械学習に提供されるアノテーションデータにおける価値について考察を加えた。

これらの結果をもとにデータの法的保護の在り方について検討した。

キーワード：データの価値、データの保護、機械学習、実証分析、データの来歴管理

1. はじめに

データについての政策は、産業振興のみならず安全保障や人権問題などにもかかわることから、体制の異なる国や地域によって大きく考え方が異なる。そのなかでも先駆けて総合的な政策を制定し、国際的に大きな影響を与えてきたのは EU である。特に 2018 年に施行された「一般データ保護規則 (General Data Protection Regulation (GDPR))³」は各国に大きなインパクトを与えた。GDPR は、個人データに関する EU の伝統的考え方として、基本的人権としての位置づけを与える体形を取っており、この考え方をもとに個人データやプライバシーの保護を厳格に規定している。これによって、個人データの流通や、個人データを扱うデジタル・プラットフォームの事業のガバナンスを実現しようとしたものである。

一方 EU は最近では、個人データに加えて産業データ等を対象とする政策にも踏み込んで取り組んでおり、2018 年には非個人データの EU 域内自由流通枠組み規則 the Regulation on the free flow of non-personal data (FFD)⁴を制定している。これは IoT やロボティクスの一層の進展により産業データの利用可能性が大幅に高まる中で、データローカライゼーションを禁止しようとしたものである。

¹ 東京大学 未来ビジョン研究センター 教授

² 文部科学省 科学技術・学術政策研究所 (元・東京大学 未来ビジョン研究センター)

³ Regulation (EU) 2016/679.

⁴ Regulation (EU) 2018/1807.

そして2020年の2月19日に、“A European Strategy for Data”を発表して注目された。この提言では、ヨーロッパの単一のデータベースを作成し、個人データと非個人データについて、企業が簡単にアクセスできる、かつ安全性の高いシステムの構築を最終的な目的としている。そのうえでの問題意識として、クロスセクターのデータ共有が進まないことやデータの公益性の観点から必要なデータにアクセスできない現状について課題とし、これらを解消するための「信頼に基づくデータガバナンス」の必要性を指摘している。この目的を果たすために、EUのデータガバナンス法を制定し、価値あるデータについては、データガバナンス法を適用するという計画を示している。

我が国は昨年G20大阪サミットの議長国として「信頼ある自由なデータ流通」(DFFT)のコンセプトを提案しているが、その後発表された前述の“A European Strategy for Data”を受けて、様々な議論がなされている。EUの動きを踏まえ「我が国としてもDFFTの考え方に沿った国際的なルール作りを加速することが重要である」とする方針が知的財産戦略計画⁵において示されているほか、A European Strategy for Dataについては、日本経団連⁶や自由民主党⁷、さらには米国の団体などからも反応が示されている。このように、EUその他の国際的なデータ政策の検討をうけて、日本における産業データ政策について何をどのように進めるかが重要となっている。

そのような政策を検討するうえで、基礎となるのが、データの価値をどうとらえるのか、どのようなデータに経済的な価値があるのかという考え方である。一般的にデジタルデータを多量に保有しているからといって、そのデータから利益を生み出せるとは限らず、多くの条件が整って初めて企業の利益や競争力につながり、産業における競争力に結び付く。その要件を踏まえた保護やインセンティブシステムを設計することがデータ政策として重要となる。

A European Strategy for Dataでも「デジタルエコノミーにおけるデータの重要性の分析、および既存の政策のフレームワークのレビューが必要である」と述べられているように、我が国においても同様の視点でデータの重要性、特に経済的価値についての分析や検討を行う必要がある。本稿ではそのような観点で、データについての経済的価値に着目して、実証的な分析による検討結果を示し、あわせて具体的な事例としてアノテーションデータにおける価値について考察を加えた。これらの結果をもとにデータの保護の在り方について検討した。

2. 既往の研究

デジタル革命によって重要性を増すデータの起源としては、3つのソースが考えられる。一つは天候や地形など自然を起源とするデータ、そして2番目は企業内の生産活動、運輸・物流、建築土木などの産業活動を起源とするデータである。そして最後が個人の活動を起源とするデータである。1番目のデータは、通常はそのデータを取得した組織がデータ利用す

⁵ <https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20200527.pdf>

⁶ 欧州委員会“A European strategy for data”への意見(経団連)

<https://www.keidanren.or.jp/policy/2020/050.html>

⁷ 知的財産戦略調査会提言2020年5月21日

<https://www.jimin.jp/news/policy/200198.html>

ることが想定され、2番目のデータの場合、データ発生源の企業がデータ利活用を行うことが想定される。3番目の個人の活動を起源とするデータについては、多くは個人情報保護法によって、データの利活用には個人の承諾が必要となっている。

これらのうちリアルデータといわれることもある産業データは、おもに2番目の企業等の組織（政府自治体も含む）による様々な活動、具体的には生産活動や運輸・物流などのデータを指す。一方これらのデータが単独で利用されるケースに加えて、自然を起源とするデータと組み合わせて用いられ、または個人情報情報を完全に匿名化することにより産業データと同様に取り扱いが可能なデータとも組み合わせて用いられることもある。産業界におけるこれらのデータの利活用は、生産性向上やマーケティング、研究開発など多様な目的のもとに実施される。それらのプロセスが適切に行われ、条件が整えば、企業の利益や競争力の向上に結び付くと考えられる。したがってこれらのデータについての経済的な価値が、データをマネジメントするどのような組織能力や、具体的な条件が整ったときに向上するのかを明らかにすることが重要となる。

このような観点での研究としては、最近の経営学分野の研究において”big data analytics capability” や”data analytics competency”などデジタルデータに関する組織能力の新たな概念が提案されてきた。Wamba et al. (2017) は、データマネジメント、インフラ（技術）、社員の技能（スキルや知識）がビッグデータ解析の主要な要素であるとし、これらを用いてパフォーマンスに結び付けることができる企業の能力を”big data analytics capability”と定義している。また、Ghasemaghaei et al. (2018) は、”data analytics competency”を”a firm’s ability to deploy and combine data analytics resources for rigorous and action-oriented analyses of data”と定義し、データの品質、データの大きさ、分析スキル、ドメイン知識、ツールの精巧さという5つの指標を用いている。

これらの実証分析研究では”big data analytics capability”や”data analytics competency”が企業パフォーマンスや、企業的意思決定パフォーマンスに正の影響を与えていることを示している。これらの研究結果からは、なんらかのビジネスパフォーマンスに結び付けるためにデータ利活用を試みる企業は、まずこれらの能力が備わっていることが必要であることを示している。逆に同じデータへアクセスできたとしても、これらの能力が備わっていない限り、ビジネスパフォーマンスに結び付かない可能性があることも示唆している。

最近では、特定の業界でのデータ利活用によるビジネスパフォーマンスについても研究がなされている。観光サービスの例では、セキュリティとプライバシーの懸念の払しょくなどの条件が整うことにより、データがビジネスパフォーマンスを向上させることを示した研究などが行われている（Yadegaridehkordi et al. (2020)）。

業界の相違や企業戦略の相違はデータ利活用に対する姿勢にも大きく影響を与えるとみられるため、その意味でデータの経済的な価値もそれを利活用する組織によって異なることになり、その意味では相対的なものとなる。データ政策の策定のためには、これらを踏まえた価値あるデータの要件を示すことが重要である。

3. 質問票調査によるデータの価値を示す要因の探索

日本企業に関するデータマネジメントの現状についてはいくつかの実証分析が行われている。本章では、著者らが中心となり独立行政法人経済産業研究所（RIETI）が実施した平

成 29 年度「データ利活用に関するアンケート調査」を用いた一連の分析（立本ら（2018）、渡部ら（2018）、Hirai et al.（2019））をもとにして、一部分析をし直した結果を示す。このアンケート調査は 6278 社を対象に実施され、562 社から有効回答を得た。

図表-1 には、データが他の経営資源と比べてビジネスパフォーマンスにどのような影響を与えているのかについて、産業分野ごとの比較を行った結果を示す。産業ごとの特徴値の比較を行った結果からは、データ利活用の成果についても特別に優越している産業が存在しているわけでないことが分かっているが、一方でデータの総容量とビジネスパフォーマンスとの関係については、反応係数がマイナスの産業群と、反応係数がプラスの産業群があることがわかった（図表-1）

ここで注目されるのはマイナスになるケースで、工業資材や知的財産権などの経営資源では、利活用されていなければ利益に直接結びつかない場合もあるが、マイナスになることは通常考えにくい。これらの経営資源とは異なり、データに関してはその利活用方法が不適切であれば、総容量が増すほど利益などのビジネスパフォーマンスを引き下げてしまうことが示唆された。

さらに詳しくこのデータの役割を、総容量だけでなく、データのマネジメント能力や体制、さらには契約の習熟度などを変数として、製造業該当企業（製造業を行っている企業）と製造業非該当企業（製造業を行っておらず、非製造業（サービス業）のみを行っている企業）について回帰推計を行った結果を図表-2 と図表-3 に示す。

被説明変数として、「具体的成果有無」「具体的間接的成果有無」の 2 つを用いた。具体的成果有無は、データ利活用のこれまでの成果として具体的成果（売上やコストダウンといった利益等）が得られていると回答した場合は 1、そうでない場合は 0 とした。具体的間接的成果有無は、データ利活用のこれまでの成果として具体的成果もしくは間接的成果（事業活動に役立つノウハウやアイデアの獲得等）が得られていると回答した場合は 1、そうでない場合（まだ成果は得られていないと回答した場合）は 0 とした。

説明変数としては、「ビッグデータの体制整備」「高度なデータ処理・解析の体制整備」を使用した。これらはそれぞれ「ビッグデータの利活用を行える体制が整備されている」「ディープラーニング等の高度なデータの処理・解析を行える体制が整備されている」という記述に対しての 5 段階評価（1「全くそう思わない」～5「強くそう思う」）であり、上述の big data analytics capability や data analytics competency に相当する変数である。

また、「データ総容量」は、当該企業に利用権限があるデータのうち、利活用を行っている、または、今後利活用することを期待しているデータの総容量についての 3 段階評価（1「1 台の PC で管理できる程度」、2「数台のサーバで管理できる程度」、3「専用のサーバ室、サーバセンターで管理する程度」）である。「契約書のひな型」は、データ利活用の利害関係者との契約書のひな型についての 4 段階評価（1「契約書のひな型はない」、2「契約書のひな型を作成している途中である」、3「すでに契約書のひな型はあるが、それを使いこなしていない」、4「すでに契約書のひな型があり、それを使いこなしている」）であり、この変数は当該企業の契約の習熟度を表していると思われる。

その他、「1 部上場ダミー」（当該企業が東証 1 部上場企業であれば 1、そうでなければ 0）、「従業員数」、「海外進出国数」（当該企業及び当該企業のグループ企業が保有する研究開発拠点、販売拠点、生産拠点等が海外に進出している国数）、「保有特許件数」（6 段階評価：1

「0 件」、2「1～9 件」、3「10～99 件」、4「100～999 件」、5「1000～9999 件」、6「10000 件以上」)、「実際のデータ利活用率」(当該企業に利用権限があるデータ総容量のうち、実際に利活用を行っているデータの容量についての 5 段階評価: 1「20%未満」、2「20%以上～40%未満」、3「40%以上～60%未満」、4「60%以上～80%未満」、5「80%以上」)、「担当者数」(全社的なデータ利活用を推進する担当者数)を制御変数とした。また、製造業該当企業についての分析においては、サンプルの中に製造業と非製造業の両方を行っている企業が存在するため、「非製造業ダミー」(非製造業も行っていれば場合は 1、そうでなければ 0)も制御変数として加えた。

図表-2、図表-3 において、モデル 1～3 は被説明変数を「具体的成果有無」とした結果、モデル 4～6 は被説明変数を「具体的間接的成果有無」とした結果である。また、モデル 1 及び 4 では「データ総容量」、「契約書のひな型」、その他の制御変数を投入、モデル 2 及び 5 ではこれらに加えて「ビッグデータの体制整備」を投入、モデル 3 及び 6 では「高度なデータ処理・解析の体制整備」を追加している。

製造業該当企業ではデータ総容量やデータの処理解析能力が 5%水準レベルではあるがモデルによっては有意となることがあるものの、全体としては、被説明変数を左右する明確な要因が見当たらない。一方、契約の習熟度を表す変数は安定して正に有意であった。

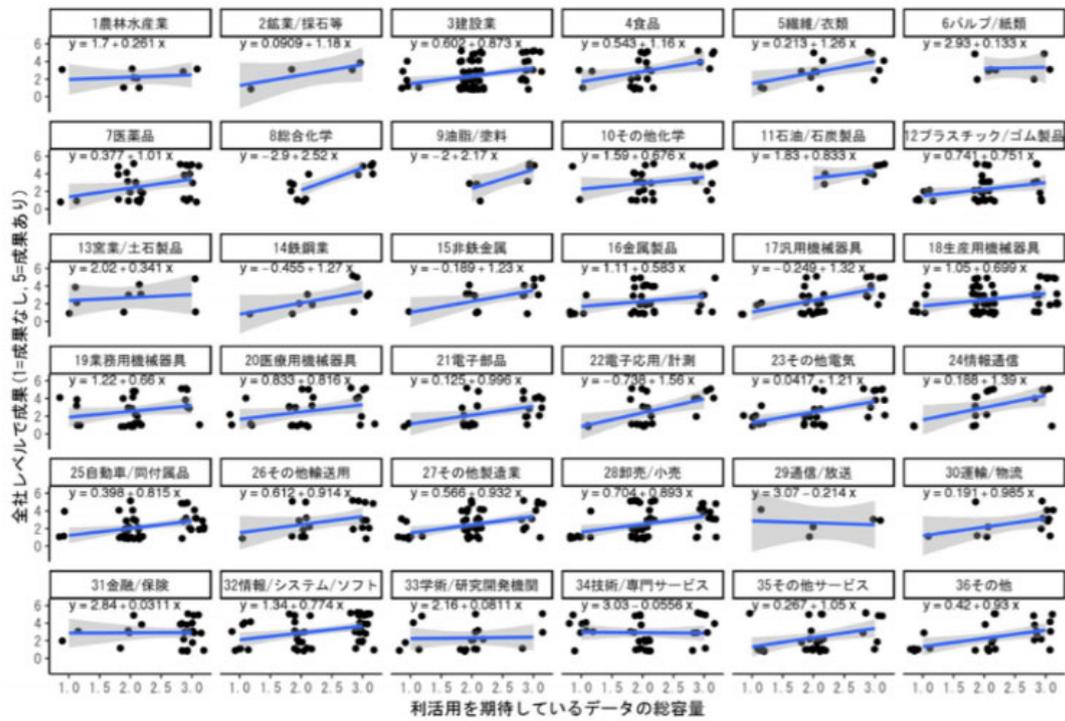
製造業非該当企業では具体的・間接的成果に対してはビッグデータの体制整備やデータの処理解析能力は安定して正に有意だが、これらの能力が実際のビジネスパフォーマンスに結び付くかどうかは明確でない。一方、データ総容量はビジネスパフォーマンスに対しては全く影響がみられなかった。しかし製造業該当企業と同様、契約習熟度に関する変数については安定して正に有意な結果であった。

これらの分析結果をまとめると、データ利活用による成果を得るためには、①契約の習熟度はまず極めて重要であること、②製造業非該当企業について、また特に間接的效果を得られるかどうかという側面では、ビッグデータの体制整備や高度なデータ処理や解析能力を備えた体制などが重要であることが明らかとなったといえる。一方、製造業該当企業に関してはデータの総容量や体制整備の影響ははっきりしないが、契約の習熟度のみは重要であることが示された。

製造業該当企業と製造業非該当企業の結果に差異がみられることに関していえば、パーソナルデータの利活用やマーケティングデータの利活用は一部の非製造業で進展している一方、製造業における産業データ(リアルデータ)の利活用は、この調査が行われた時点では、未だ着手したかしないかといった段階であるとみられることから、パフォーマンスについてもはっきりした結果が認められなかったとも考えられる。

一方、製造業該当企業・製造業非該当企業を問わず契約の習熟度を代理する変数については、直接的間接的ビジネスパフォーマンスに対して、常に安定して正の有意な効果を与えている。この結果からはデータ利活用の際には、企業等のデータ利活用の初期段階から、ある程度体制が整って本格的にビジネスパフォーマンスに寄与する段階に至るまで、企業は自社だけでデータ利活用が完結せず、複数他社との連携によってデータ利活用を進めることで初めてパフォーマンスに結び付けることができるということを示唆していると思われる。

図表-1 データの総容量とビジネスパフォーマンスの関係 (立本ら (2018) より)



図表－2 製造業該当企業のロジスティック回帰分析の結果

	被説明変数：具体的成果有無								
	モデル1a			モデル2a			モデル3a		
	B	標準誤差	有意確率	B	標準誤差	有意確率	B	標準誤差	有意確率
ビッグデータの体制整備				0.323	0.213				
高度なデータ処理・解析の体制整備							0.432	0.218	*
1部上場ダミー	-0.988	0.516		-1.052	0.526	*	-1.107	0.537	*
従業員数	-0.0003	0.0002		-0.0003	0.0002		-0.0003	0.0002	
海外進出国数	0.048	0.025	*	0.049	0.025		0.052	0.025	*
保有特許件数	0.451	0.231		0.440	0.234		0.415	0.233	
データ総容量	0.695	0.354	*	0.630	0.362		0.710	0.362	
実際のデータ利活用率	0.248	0.138		0.227	0.141		0.217	0.142	
担当者数	0.011	0.021		0.011	0.022		0.010	0.022	
契約書のひな型	0.484	0.143	**	0.423	0.151	**	0.437	0.148	**
非製造業ダミー	1.013	0.412	*	0.957	0.416	*	0.954	0.418	*
定数	-5.467	0.993	**	-5.802	1.035	**	-6.007	1.060	**
Cox-Snell R2 乗	0.196			0.204			0.210		
Nagelkerke R2 乗	0.286			0.298			0.306		
N	228			226			226		

	被説明変数：具体的間接的成果有無								
	モデル4a			モデル5a			モデル6a		
	B	標準誤差	有意確率	B	標準誤差	有意確率	B	標準誤差	有意確率
ビッグデータの体制整備				0.307	0.215				
高度なデータ処理・解析の体制整備							0.050	0.220	
1部上場ダミー	-1.132	0.516	*	-1.218	0.525	*	-1.175	0.520	*
従業員数	0.0002	0.0002		0.0001	0.0002		0.0002	0.0002	
海外進出国数	0.157	0.057	**	0.151	0.057	**	0.155	0.057	**
保有特許件数	-0.144	0.218		-0.120	0.222		-0.152	0.220	
データ総容量	0.254	0.323		0.249	0.328		0.282	0.324	
実際のデータ利活用率	0.515	0.142	**	0.497	0.142	**	0.512	0.143	**
担当者数	0.020	0.050		0.017	0.048		0.022	0.051	
契約書のひな型	0.499	0.191	**	0.435	0.196	*	0.496	0.194	*
非製造業ダミー	0.395	0.405		0.332	0.411		0.399	0.409	
定数	-1.925	0.784	*	-2.380	0.851	**	-2.027	0.855	*
Cox-Snell R2 乗	0.226			0.236			0.229		
Nagelkerke R2 乗	0.308			0.322			0.313		
N	228			226			226		

**p<0.01, *p<0.05

図表-3 製造業非該当企業のロジスティック回帰分析の結果

	被説明変数：具体的成果有無								
	モデル1b			モデル2b			モデル3b		
	B	標準誤差	有意確率	B	標準誤差	有意確率	B	標準誤差	有意確率
ビッグデータの体制整備				0.360	0.225				
高度なデータ処理・解析の体制整備							0.134	0.221	
1部上場ダミー	-0.454	0.687		-0.292	0.695		-0.361	0.699	
従業員数	0.0004	0.0002	*	0.0004	0.0002	*	0.0004	0.0002	*
海外進出国数	0.015	0.060		0.015	0.060		0.008	0.062	
保有特許件数	-0.235	0.261		-0.208	0.263		-0.232	0.260	
データ総容量	0.476	0.380		0.255	0.407		0.377	0.414	
実際のデータ利活用率	0.108	0.168		0.073	0.174		0.090	0.172	
担当者数	-0.015	0.024		-0.018	0.024		-0.015	0.024	
契約書のひな型	0.556	0.172	**	0.544	0.175	**	0.553	0.172	**
定数	-3.312	1.036	**	-3.659	1.076	**	-3.364	1.039	**
Cox-Snell R2 乗	0.220			0.235			0.222		
Nagelkerke R2 乗	0.317			0.339			0.320		
N	132			132			132		

	被説明変数：具体的間接的成果有無								
	モデル4b			モデル5b			モデル6b		
	B	標準誤差	有意確率	B	標準誤差	有意確率	B	標準誤差	有意確率
ビッグデータの体制整備				1.003	0.286	**			
高度なデータ処理・解析の体制整備							0.870	0.275	**
1部上場ダミー	0.204	0.570		0.217	0.596		0.436	0.613	
従業員数	0.0001	0.0001		0.00004	0.0001		0.0001	0.0002	
海外進出国数	-0.010	0.060		-0.024	0.061		-0.064	0.068	
保有特許件数	0.016	0.211		0.136	0.224		0.098	0.227	
データ総容量	0.590	0.346		0.134	0.385		0.124	0.383	
実際のデータ利活用率	0.120	0.161		0.028	0.175		0.008	0.178	
担当者数	0.065	0.065		0.097	0.069		0.085	0.066	
契約書のひな型	0.446	0.177	*	0.413	0.188	*	0.422	0.182	*
定数	-2.065	0.843	*	-3.152	0.971	**	-2.690	0.914	**
Cox-Snell R2 乗	0.185			0.274			0.260		
Nagelkerke R2 乗	0.252			0.375			0.355		
N	132			132			132		

**p<0.01, *p<0.05

4. 機械学習に用いるデータの価値

第3章においては一般的にデータ利活用について、経済的価値との関係でどのような要件が影響するのかについて実証分析を参考に検討した。その結果はデータを単に多量に保有しているということはほとんど意味がなく、なんらか「ビッグデータの体制整備」「高度なデータ処理・解析の体制整備」などの組織の能力と、データ利活用に関する「他組織との連携を円滑に進める契約に関する能力」があって、はじめて価値あるデータとなるということが示唆された。

ここではさらにデータの価値について具体的な事例をとりあげてその要件を検討する。もっともデータ利活用といっても様々な利活用方法があり、高度な機械学習などを用いて推論エンジンを開発し、サービスを実装する場合もあれば、そのデータを単純集計ただけで分析などに用いる場合もある。しかし第3章で示したように高度なデータ分析体制がビジネスパフォーマンスに結び付くということからも、今後ますます機械学習によるデータの利活用などが重要になっていくことが予想される。ここでは特に機械学習に用いるデータの価値、特に機械学習特有のアノテーションデータの価値について検討する。

機械学習においては、大量の学習データの統計的分布からパターンを作り出す学習プロセスと、その学習プロセスで生成した推論モデルに、分類や識別をしたいデータを入力して当てはめて、結果を導く推論プロセスの2つに分けられる。

このとき学習プロセスに用いられる学習用のデータがまず必要になるが、その際に、機械学習によってその統計的特徴を認識できるようにするために、テキストや音声、画像など様々なデータにタグを付けたデータで学習させることが行われる。このタグをつけたデータのことをアノテーションデータと称する。例えば複数のソースから特定の画像データを収集して学習用データとして用いようとする場合、それぞれの企業のデータの品質をどのように把握するのかという課題が生じる。例えば人間を正しく識別していないアノテーションデータを用いて機械学習を行った場合、得られた推論エンジンを用いて自動運転を制御すれば事故が起きる可能性がある。このように、学習用データに誤りがあったり、偏りがある場合、そのデータを機械学習に用いて得られた推論エンジンでサービスを行った場合に生じる事故については重大な責任問題につながり得る。

この点の法的責任については様々な議論がなされているが、経済産業省が取りまとめたAIとデータに関する契約ガイドライン⁸においては「提供データが期待されたものではなかった場合の責任」について留意点が示されている。具体的には「データ提供型契約が有償契約である場合、データの品質について問題があれば民法上の瑕疵担保責任（契約不適合責任）の適用があると考えられる。もっとも、提供データの品質についての問題といっても様々な内容があるため、提供データの正確性、完全性、有効性、安全性、第三者の知的財産権の非侵害等について、どの範囲でデータ提供者が責任を負うのか契約で明確にしておくことが望ましい（例えば、表明保証条項を用いることが考えられる）。」（提供データの品質 30P）との記載があるが、この際の「データの正確性、完全性、有効性、安全性」については詳しい定義や表記方法についての検討はこの時点ではなされていない。特にアノテーションが付与されたデータの場合、その来歴において正確性や完全性について、だれが責任を負ってどのような方法で行ったかについて明らかにする必要がある。さらに複数事業者がそれぞれ取得したデータを一つのプラットフォームに格納して提供する場合では、それぞれのデータの来歴が管理されていることと、その管理内容がそれぞれのデータにメタ情報として表記されていることが、重要な情報となる。この際来歴管理がされていないデータを使った場合のリスクや、来歴の中でだれがそのプロセスの責任を負っているのかなどによって、データを利用するかしないかの判断も変わってくる。極端なケースにおいてはデータが改ざんされているなどの場合は、それをを用いて生成した推論エンジンの利用に際して深刻な問題が生じ得ることになる。

このような来歴管理や法的責任を含むデータの品質については、アノテーションデータにおいては特に問題になりやすいという側面があるが、機械学習において入力されるデータ全般においても、程度の差こそあれ問題になりえる。この点これらのデータの価値につながるデータの特性であるといえるが、どのような特性なのかについては、まだ概念の整理が十分行われているとはいえない。

⁸ AI データ契約ガイドライン（経済産業省）

<https://www.meti.go.jp/press/2018/06/20180615001/20180615001.html>

しかし比較的近い概念としては、データの経路を追跡することを指す **data lineage** や、データの完全性（不正に変更されていないこと）や具体的なデータの来歴（起源、正確に計算されたこと）を証明することを指す **data provenance** などがある。しかしいずれの場合も、来歴管理を含む法的責任を明らかにした情報を、データに付すことの手続きを明らかにしているものではない。より経済的価値のある情報財とするためには、この意味での管理手法を整備し、その手続きを標準化することが必要であると思われる。

同時にこのようにして標準化されさらに経済的価値を生み出す必要条件を備えたデータについては、一般のデータに比べてより多くの投資が行われ、かつ経済的価値の発現の条件を整えている情報財として、広義の知的財産として経済的取引が円滑に行われるように、その流通や利活用を促進し、漏洩などを防止する法的保護の仕組みが整備されることが望まれる。

5. 分析検討結果を踏まえたデータの保護の在り方

以上述べてきた経済的価値につながるデータの性質を前提に、データの法的性格をどのようにとらえるべきかが論点となる。データは企業利益の源泉であると位置づけできるとして「データを物権的な構成を有する知的財産権の対象とするべき」、さらには「データを会計上の資産とみなすべき」との主張もみられる。

しかし先述の実証分析によって得られた結果を踏まえると、少なくとも、データを大量に保有することが、そのまま企業にとっての価値につながるということにはならないことは明らかである。データ自身がそのまま経済的価値につながると見做すことはできないのであれば、無条件にデータを保護してもその意義は少ない。一方第4章で検討したアノテーションデータなどについては、来歴が管理され改ざん防止が保証されているなどの要件を備えている場合は、そのための投資も行われていることになり、容易にコピーされて他者に無条件で利用をされてしまえば、データセットに経済的価値を付加するために負担した投資が回収できないといった問題が生じ得る。データは知的財産権として保護されるプログラムなどとは異なる無体物であり、特段の法的保護がない限り、仮にこのようなアノテーションデータが流出してしまった場合にも、差し止めなどの措置をとることができない。

このようなケースにおける保護方法としては、営業秘密として保護することが考えられる。もともと企業内で発生するデータに関しては技術ノウハウと関係しているケースも多く、営業秘密として管理されているケースも少なくない。より有効に多くの企業の有するデータを統合して利活用すべきという主張はEUなどでも広くみられるものの、この点、我々の実証分析でもデータを通じた技術ノウハウの流出に関する懸念が根強くあることが示されている（Hirai & Watanabe, 2016）。

しかし他方で、技術ノウハウとは異なり、前述したように、データは複数の組織が関与して利活用が行われるという特徴を有する。このことから、データ提供に際して営業秘密として管理していることを条件とした契約を行うことが重要となる。実証分析の結果に示した契約の習熟度がデータ利活用のパフォーマンスにも反映しているという結果も、このような背景を考えると妥当な結果であると解釈することができる。

これらの結果を踏まえると、データの性格は知的財産権というよりは、むしろ技術ノウハウなどの営業秘密などに近いものであると考えられる。しかし多数の不特定事業者にデー

タの利活用を促すなどのケースにおいては、不特定組織に提供されることで営業秘密としての要件を満たさなくなる。このようなケースにおいても、一定の要件で管理されている場合には、不正な取得に対しては差し止め請求を認めるなどの保護がなされるべきであると考えられる。

このような考え方に沿って、我が国では 2019 年に不正競争防止法が改正され限定提供データ制度が利用できるようになった。「限定提供データ」とは、業として特定の者に提供する情報として電磁的方法により相当量蓄積され、及び管理されている技術上又は営業上の情報（秘密として管理されているものを除く。）をさす⁹。その際の管理とは、平成 30 年の改正によって、価値あるデータのうち、一定の要件を満たしたデータを「限定提供データ」とし、悪質性の高いデータの不正取得・使用等を不正競争防止法に基づく「不正競争行為」と位置づけることにより、救済措置として差し止め請求権が可能となっている。

その際の管理方法としては、データ保有者とデータの提供を受けた者以外の者に対するアクセス制限が施されていることなどで要件を満たすとされる。この制度はまだ各国とも本格的に試みられていないデータ保護について、先駆けて不正競争防止法による行為規制による対応を試みたものであるということができ、一方現時点でこの制度は日本法のみであることや、効力についても通常の正当な事業活動を阻害しない範囲で、悪質性の高い、不正取得・不正使用等への救済措置として、必要最小限の民事措置（差し止め請求、損害賠償額の推定等）に限られていることなどは考慮する必要がある。

本稿で議論したようなアノテーションデータのような投資が行われた経済的価値の高いデータについて、営業秘密の管理を超えて広く提供して利活用を促そうとした場合、現在の限定提供データの保護で十分なのかという点では、いずれ議論が必要と思われる。

またデジタルデータの利活用ではクロスボーダー取引も頻繁に行われているが、限定提供データの漏洩が海外で生じた場合のエンフォースメントには困難が予想される。今後、限定提供データについては国際的にも適切な保護がなされることが望ましいが、この点日本法だけの対応である問題をどのように扱うかは論点となる。EU データ法など今後外国においても、同様の行為規制による保護が検討されることは望ましいことであると思われるが、限定提供データの制度の国際的認知は進んでおらず、行為規制による国際的なデータ保護の先行きの見通しは明らかでない。

限定提供データによるものか営業秘密としての保護を前提とするかを問わず、他者へのデータ提供を伴う場合は契約による債権債務関係の構築が必須となるが、無体物であるデータの契約上の取り扱いや、機械学習特有のデータに推論エンジンの性能が依存するという性質などから、契約においても配慮すべき事柄が少なくない。

この点、先述した経済産業省が 2018 年 6 月に策定した、「AI・データの利用に関する契約ガイドライン」は、データの利用等に関する契約、及び AI 技術を利用するソフトウェアの開発・利用に関する契約の主な課題や論点、契約条項例、条項作成時の考慮要素等を整理したものであり、データに関する契約慣行を定めていくうえで重要なガイドラインである。このガイドラインは英訳もなされており、クロスボーダー取引においても利活用が可能であると考えられるが、その国際的認知もまだそれほど進んでいない。

⁹ 【第 2 条第 7 項（定義）】

その点今後の法政策として考えられるのは、まず、日本の限定提供データ制度や「AI・データの利用に関する契約ガイドライン」について国際的認知をすすめ、考え方の普及を促すことが考えられる。また同時に価値あるデータの要件を、本稿で述べたような経済的価値の要件をもとに、さらに精緻化していくことと、そしてこれらの要件を備えた標準化などが進むことを前提に、限定提供データの効力を再検討して、要件を満たす客体が明確に定義されている場合は保護強化が図られることなども検討されるべきであろう。

6. まとめ

本稿では主に産業データに関して、どのような要件でデータが企業にとって経済的価値があるのかについて、質問票を用いた実証分析および機械学習に用いられるアノテーションデータにおける価値についての考察をもとに、価値あるデータとは何か、またその法的保護はどのように行われることが必要なのかについて、現状の制度を踏まえて検討した。

データにかかわる法政策の難しさとしては、そもそも客体の定義が困難であることがある。多様なデータを客体として想定しなくてはならないため保護水準を検討しにくいといった状況も生じる。この点価値あるデータの要件の精緻化は、今後のデータ政策策定の第一のポイントになるだろう。またデータは企業のデータマネジメント能力との組み合わせで価値を発現するという相対性もあることから、その相互作用の見極めを進めることも、データ政策の第二のポイントになるだろう。

謝辞

本稿第3章は、独立行政法人経済産業研究所（RIETI）におけるプロジェクト「企業において発生するデータの管理と活用に関する実証研究」

(https://www.rieti.go.jp/jp/projects/program_2016/pg-04/005.html)の成果の一部である。

参考文献

1. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356-365.
2. Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, 27(1), 101-113.
3. Yadegaridehkordi, E., Nilashi, M., Shuib, L., Nasir, M. H. N. B. M., Asadi, S., Samad, S., & Awang, N. F. (2020). The impact of big data on firm performance in hotel industry. *Electronic Commerce Research and Applications*, 40, 100921.
4. 立本博文、平井祐理、渡部俊也「データ利活用に関する質問票調査を用いた産業別比較」RIETI ディスカッション・ペーパー 18-J-032 (2018)
<https://www.rieti.go.jp/jp/publications/summary/18120001.html>
5. 渡部俊也、平井祐理、阿久津匡美、日置巴美、永井徳人「企業において発生するデータの管理と活用に関する研究」RIETI ディスカッション・ペーパー 18-J-028 (2018)

https://www.rieti.go.jp/jp/projects/program_2016/pg-04/005.html

6. Hirai, Y., Tatsumoto, H., & Watanabe, T. (2019). The effect of big data and advanced analytics in Japan. In ISPIIM Connects Ottawa, The International Society for Professional Innovation Management (ISPIIM).
7. Hirai, Y., & Watanabe, T. (2016). Empirical study regarding the leakage of technological know-how in Japanese firms. In PICMET '16 Conference, Portland International Conference on Management of Engineering and Technology (PICMET)