

データ活用社会創成プラットフォームの推進に向けた 当面の整備方策

令和元年 7 月

データ活用社会創成プラットフォームの推進に関する有識者会合

— 目 次 —

1. データプラットフォームが求められる背景	1
(ア) 地方の産業、自治体を含む我が国全体でのデータ活用推進の重要性	1
(イ) データ活用の推進に向けた早急な活用環境の整備の必要性	2
2. データプラットフォーム推進における課題	2
(ア) 地方創生、地方活性化に向けたデータ活用のための人材に関する課題	2
(イ) データ活用のインフラ（基盤）の未整備と方法論の確立に関する課題	3
(ウ) データ流通における課題	3
(エ) エコシステムの機能に関する課題	4
3. 実現すべきデータ活用を促進する長期的な仕組み	4
(ア) 基盤整備と先導アプリケーション開発	4
(イ) データ活用を推進する体制の構築	5
(ウ) セキュリティの確保の仕組みの検討と、セキュリティルールの形成	8
(エ) エコシステムの形成（全国的なデータ活用のエコシステムを形成する）	9
4. 整備の現状	10
(ア) DIAS (Data Integration and Analysis System)	10
(イ) ABCI (AI Bridging Cloud Infrastructure)	10
(ウ) 広域データ活用のための IT 基盤整備	10
(エ) SINET (Science Information NETwork)	11
5. 当面の方策	11
(ア) 基盤整備と先導アプリケーション開発	11
(イ) データ活用を推進する体制の構築	13
(ウ) セキュリティに関する方策	13
(エ) エコシステムの形成	14
6. まとめ	14
7. 用語解説	16
8. 付属資料	
Society5.0 を実現するためのデータ活用による知識集約型社会の創成	
ーデータ活用社会創成プラットフォームの構築ー	20
9. 参考	
データ活用社会創成プラットフォームの推進に関する有識者会合	
開催要綱	21
データ活用社会創成プラットフォームの推進に向けた	
当面の整備方策（概要）	23

1. データプラットフォームが求められる背景

(ア) 地方の産業、自治体を含む我が国全体でのデータ活用推進の重要性

昨今、データ科学及び AI、IoT 技術の急激な進展やネットワーク環境の整備により、多種多様なビッグデータを集め、つなぎ、活用する、いわばデータ活用社会が到来しつつある。

このデータ活用社会は、Society 5.0 が目指す社会、すなわち、地域、年齢、性別、言語等による格差による課題を解消し、人々が効率的で質の高いライフスタイルの実現と、地域の特色を活かした多様な産業の活性化に貢献する社会の構築に繋るものである。

我が国では、「経済財政運営と改革の基本方針 2018」や「未来投資戦略 2018」において、今後のリアルデータの利活用の重要性が指摘され、その促進に向けた取組への期待が高まっている。また、地方創生の観点からは、地域での少子高齢化と人口減少が進んでいる中、情報通信技術によるデータ活用社会への移行の重要性がますます高まっている。「まち・ひと・しごと創生基本方針 2019」においても、新しい時代の流れを力にするために情報通信技術などの活用が重要であると指摘されており、「Society 5.0 の実現に向けた技術(未来技術)の活用等による新たなビジネスモデル等を構築し、地域企業の生産性向上と魅力的な雇用機会の創出を進める。特に、地域経済において中核的な役割を果たす企業の発掘・成長に対する重点的な支援、大学等とも連携した地域発のイノベーションの創出、地域のシーズ等を活用した創業の活性化、地域商社の取組拡大などを進める。」ことなどが当面の主要な取り組みに含まれている。

これらを受け、我が国を Society 5.0 と信頼性のあるデータ流通(DFFT:Data Free Flow with Trust)の先進モデル国としていくことが重要であり、世界に先駆けイノベーションに適した地としていくためには、知・人材が集積する大学等を中核として、あらゆる産業を知識集約型に変えていくことが必要である。また、これからの社会ではあらゆる分野におけるデータ活用が必須となることから、社会実装を指向した研究分野がより重要となる。

データを活用する上でその流通基盤となるネットワークは重要な要素である。我が国の大学など主要な学術機関に整備されている SINET(学術情報ネットワーク)は、100Gbps の帯域を持つ通信網であり、我が国全体に網目状に構築されているものであり、これは単位面積当たりの通信容量という観点で世界最先端のネットワーク網となっている。

また、わが国には世界的に見ても良質な様々なデータが存在しており、最先端のネットワーク上で、多種多様なビッグデータを集め、つなぎ、活用するデータプラットフォーム

ムの整備は、世界中から知と技術が集約されるポテンシャルの存在を積極的に活かすものであり、データ活用の国際的フロンティアとなり得るものである。

しかしながら、既存データ(教育、医療、農林水産、防災、エネルギー、スマートシティ等)の活用の現状は、産業構造や人口動態、人の流れなどの官民ビッグデータを集約し、可視化する地域経済分析システム(RESAS)などの整備が進められているものの、データを活用したい主体自らが、必要となる異分野のデータを把握し、それらのデータを活用する解析ノウハウを獲得し、さらに解析に用いる IT インフラを構築する必要があるなど、データ活用に至るまでの障害の大きさから、様々なデータが広く有効活用されているとは言い難い状況である。

(イ) データ活用の推進に向けた早急な活用環境の整備の必要性

データの活用を推進するためには、Proof-of-Concept (PoC) によりその効果を実証し、本格的な活用に迅速に移行することが重要であり、PoC をいかに迅速に行えるかは、データ活用の成否を左右することとなる。迅速な PoC を可能にするためには、分野にまたがる共用データ、データ解析ノウハウ、インフラなどをだれもが容易に利用できるようにすることが重要であり、様々な面での基盤と体制を整備することで、データ活用社会を作るための新しい取組を容易にする場を構築することが必要である。

また、データ活用を促進するためには、データを活用する人材の育成も重要である。研究者や実務者及び学生が、データ量の大小に関わらず、PoC のための環境を自身が有するデータと、利用可能な公開データを合わせて様々な解析を行い、より高い技術を習得するための「データ工房」を整備することが重要である。

2. データプラットフォーム推進における課題

(ア) 地方創生、地方活性化に向けたデータ活用のための人材に関する課題

データの活用を推進していくためには、データを実際に生成する場でもある地方・地域社会の現場の知識と、データ活用の知識・ノウハウの両者が必要となる。しかし、これらの現場には、アイデアがあっても実際にデータを収集・集積・解析する情報技術を保有する人材が少なく、また、データ活用の知識・ノウハウを持つ人材とのつながりも希薄であることも多い。さらには、医学医療領域のようにアカデミアでのデータ活用のための人材の常勤雇用枠がニーズに追いついていない領域もある。

また、個別の地方・地域社会でのデータ活用においては、データの適用範囲を当該地方・地域内に閉じて考えがちなため、波及効果も限定的でエコシステムも成立しない

ことが多いと考えられる。仮に、他地域でも利用できるノウハウやスキームが得られる場合でも、現状ではこれを広く展開する仕組みがないと考えられる。さらに、データ解析技術を持つ研究者等がコンサルティングを行う場合でも、活動がスポット的になりがちで、継続的な活動が難しい。データプラットフォームを推進するにあたっては、データ活用に係る国内外の動向や活用事例を適切に収集し、広く共有していくことが必要である。

(イ) データ活用のインフラ(基盤)の未整備と方法論の確立に関する課題

我が国では、現時点において、特に大量データの処理が必要な場合や、リアルタイムの処理が必要な場合に利用可能な、データ活用を試行錯誤する PoC を行うために容易に使えるインフラが存在しない。特に、リアルタイム処理を行うためのソフトウェアやアプリケーションが未開発である。

また、先に述べた地方・地域社会の現場では、データを活用することで何が実現できるかについて十分な知見がなく、実現したいサービスや解決したい問題はあっても、どのようなデータと解析技術があれば実現できるのかが分からないことも少なくない。

(ウ) データ流通における課題

データ活用により創出される価値の向上には、利用できるデータの量と質が十分確保されていることが不可欠であるが、個人情報保護とデータ活用を両立するためのポリシー等の方策が未整備であるため、適切な情報の保護・管理体制を整えたうえで活用を促進することが難しい。例えば、我が国においては、一定水準の医療が全国民に提供されており、そこでは個人の診療・診断データが蓄積されているなど、良質のデータが存在するが、これらのデータは個人情報を保護しなければならない観点から十分に活用することができておらず、イノベーションの創出を図る上では、個人情報の保護と活用のバランスがとれた方策等を検討していく必要がある。

一方、多種多様なデータ解析についての専門的な知識を有した人材も必要である。例えば、医療・健康等の分野でのデータ活用には、個々の適用分野に精通しつつデータ活用に取り組む人材が不可欠であるが、このような人材の重要性が十分に認識されておらず、ポストの確保が難しいという問題もある。

また、人材が不足していることもあり、どのようなデータが利用可能であるか、そのデータをどうすれば入手できるのかといった情報が広く共有されていない。

さらに、データを保有する研究者や企業等がデータを他に活用させることへのインセンティブがないために、有用なデータが流通しない。個人情報、プライバシーなどの問題によりデータの活用が進まないことや、データを収集・集積・解析するノウハウや技術課題解決のための適切な依頼先を選定する知見が共有されていないことも課題である。

なお、全国の大学や公的研究機関の「知」や「データ」を活用することが方策として考えられるが、従来の産学官連携の仕組みでは、どのような技術が必要か明確な場合にのみ機能するものが多く、潜在的な需要を拾い上げるのには適していない。

(エ) エコシステムの機能に関する課題

国際競争力に対する危機感やエコシステムへの期待が高まる中においても、データ活用の正の循環が機能しておらず、潜在的なデータ活用の需要を拾い上げるエコシステムが存在していない。データ活用にコストをかけることによって得られる利益は、データ活用が広がった後までを考えれば、その期待値はコストを大きく上回る。一方で、必ずしもデータ活用が成功するとは限らずリスクがあること、当初のデータ活用を担う主体が広範囲に広がるデータ活用を最後まですべて担うとは限らないこと、などの理由から、コストを負担して利益を生み出すエコシステムが確立されていない。

3. 実現すべきデータ活用を促進する長期的な仕組み

(ア) 基盤整備と先導アプリケーション開発

データ活用の方法論を確立するためには、データ提供者・利用者双方が利用可能な共通基盤を整備するとともに、先導的なデータ活用アプリケーション開発プロジェクトを推進し、成功事例を蓄積して、その資産の再利用することが必要である。具体的には、以下の、共通基盤の整備やアプリケーション開発の推進を通じて、全国に分散した設備や人材を適材適所で活用することにより、オールジャパンの体制で取り組むべきである。

① リアルタイムに大量のデータを収集・蓄積・解析可能な使いやすい共通基盤の整備

1) 基盤システム整備

仮想化技術の進展により、目的ごとに個別のハードウェア基盤を構築する必要性はなくなっている。また、広帯域で信頼性の高いネットワークを用いれば、利用の利便性は(非常に低遅延の応答が要求される場合などを除き)ハードウェア基盤の設置場所に依存なくなり、様々な用途で共用する効率的な基盤利用が可能となる。そこで、SINET を中心とした全国を結ぶ広帯域でセキュアなネットワークを拡張し、データ収集に活用するための整備を進めることが望まれる。また、SINET の民間利用拡大に向けた施策を検討すべきである。さらに、収集されたデータを蓄積・解析するための計算資源やストレージ環境を整備し、アプリケーション開発や PoC を速やかに実施できるようにすることが必要である。

2) データ解析システム(ソフトウェア)

様々なアプリケーションで共通して利用される解析ツール(ソフトウェア)を簡便に利用可能な環境を整備し、アプリケーションの速やかな開発を可能とする必要がある。

3) データ収集管理システム(ソフトウェア)

センサ等から連続的に生成される大量のデータをリアルタイムに解析するためには、ネットワーク経由で転送されるデータを確実に収集し、ストレージに蓄積するとともに、アプリケーションから利用可能とすることが必要である。これらの機能は異なるアプリケーション間で共通して必要とされるため、アプリケーション開発者が利用可能なデータ収集管理システムを整備することが必要である。

② アプリケーション開発支援

アプリケーションが対象とするデータ分析の特性に応じて、データ分析の専門家がアプリケーション開発者を支援する体制が必要である。また、データの収集・蓄積・解析の特性に応じて、計算機システムやソフトウェアの専門家がアプリケーション開発者を支援する体制が必要である。このような体制のモデルを構築し、多くの用途に広く利用可能とする必要がある。

(イ) データ活用を推進する体制の構築

全国的なデータ活用体制を構築するためには、データの整備・流通を図るとともに、データ活用のコンサルティング機能、データ提供者・データ解析技術者・データ利用者のマッチング機能を全国に構築することが必要である。全国展開に当たっては、先導的な機関による取り組みから始めてユースケースとしての事例を蓄積するとともに、具体的には、全国の大学がハブとなり、地方自治体や企業も含め展開するコミュニティを形成していくことが重要である。

① データの生成・収集・整備・流通

1) データに関する認識の共有及びインセンティブの付与

データの収集を促進する仕組みが必要である。これには、データを共有することの意義や利点を明確に示し、データ提供者にインセンティブを与えることなどが考えられる。データに DOI(Digital Object Identifier)を付加するなどし、さらにどのデータがどこでどのように使われ、そこからどのような二次データが生成された

のかなどのトレーサビリティを確保することが必要で、特に研究のコミュニティでは、データの提供による成果が明確化され、データ提供者の業績として評価されるようになることが重要である。また、提供されたデータの取り扱いルールを明確化し、データ提供者と利用者間で合意をとる仕組みが必要である。

さらに、これらの取り組みを通じて、信頼性のあるデータの流通(DFFT)をいかに実現するかが重要であり、そこにはデータが生成・収集される現場である大学・研究機関の役割が大きい。

2) 全国の大学が持つ共通するデータによる事例の蓄積

全国の大学がハブとなりデータ活用を推進していく上で、大学自身がデータの生成・収集・整備・流通の中心となりデータ活用を推進する事例をつみあげることが重要である。その促進が期待される共通のデータとして、大学のミッションである教育における学習データがある。これを一つのターゲットとして考えた場合、大学は毎年膨大な品質の良いデータを収集することが可能であり、学習データを用いて、学生の学び方や教員の教え方を改良することができる。大規模大学のみならず、地方の小規模な大学においてもデータ活用の促進が図られ、大学同士で授業や教育手法のシェアリングといったことも可能になると考えられる。このように、大学が持つ共通のデータにより事例の蓄積を積極的に行っていく必要がある。大学での学習データの活用技術は、初等中等教育から社会人教育へと幅広く全国展開ができる。

また、他にもより具体的な事例として、全国の大学病院が持つ Real World Data を有する国立大学医学部附属病院災害時 BCP システム「GEMINI」がある。現在は、大規模災害時の電子カルテの保全および災害直後の対策として、全国 2 か所のデータセンターに国立大学 42 大学 46 病院が SINET/L2VPN でセキュアに接続され、データセンターには各病院の電子カルテの完全コピーと、標準化された病名、検査結果、処方データのストレージである SS-MIX2 標準化ストレージが設置されている。これらを平時からデータ 2 次利用するフレームワークを構築し、データサイエンス教育や研究に活用することが望まれる。

3) データを共有するための体制の構築

民間の保有するデータを受け入れるための体制の構築、特にデータ利用の専門家を育成し、データ活用方策やオープン・クローズドデータの管理、データのアーカイブやキュレーションを実施する体制が必要である。

また、個別のデータの属性を示すメタデータや、データ取得に関する様々な

周辺情報を適切に共有する仕組みについても構築する必要がある。また、セキュリティやプライバシーに関する問題点を整理し、データ共有の基本的な指針を策定する必要がある。

4) 社会の変化に対応したデータ流通のあり方の検討

データの利用は、社会制度により定められるルールや、プライバシーに関する社会のコンセンサスを踏まえて行われなくてはならない。一方で、データ活用が広まることで社会制度自身やコンセンサスが変わっていくことも考えられる。PoCの実施に際しては、社会制度の変化と連携し、データの活用を推進していくことが必要である。また、今後データの活用範囲は大きく広がっていくものと考えられることから、収集したデータを収集時に提示した使用目的を超えて利用する場合も想定される。この際に、データ提供者に了解を取る仕組みなども検討していく必要がある。前述の教育機関における学習データの活用などの経験は、社会制度の改革も含めたデータ活用社会構築の優れた先進事例となり得る。

②他のデータ活用フレームワークとの連携

「データ統合・解析システム(DIAS:Data Integration and Analysis System)」における、大量の気象観測データや衛星画像データなどを解析し、ユーザーに活発に活用されている先行事例をはじめとして、公的データ、公共データ、研究データを中心としたデータの共有と活用フレームワークの連携等により、データの活用をより強力に推進していくためのあるべき姿について検討していく必要がある。

③ワンストップのマッチング・コンサルティング支援と人材育成

1) 大学・研究機関と民間の連携

大学や研究機関が進めるデータ収集・解析の取り組みを民間企業が社会実装することや、民間データを大学や研究機関の解析技術を用いて解析すること等、産学の技術やデータのマッチングを図り、産学共同研究をより一層推進する体制が必要である。

2) 実践的データ解析技術活用の場の形成

大学や研究機関が持つ知識や技術の社会実装を推進するためのコンサルティングサービスを提供する体制が必要である。また、その知識や技術の社会実装を通じて、博士課程を中心とする大学院学生等にデータ活用の現場で実践的な経験を積ませ、優れたデータ活用技術を持つ人材に育成するなど、データ活用

の実践の場に優秀な人材を提供していく体制が必要である。高度なデータ活用には様々な階層の技術が必要であり、応用分野ごとの研究者に活用を任せるのではなく、研究者以外の技術者を含めた社会全体のバランスのとれた体制を構築していくことが重要である。

一方、より高い価値を生み出すためにはよりよいデータ解析技術を用いることが重要である。このため、データ活用の実践や普及と、高度な解析技術の開発の研究を共に推進していく必要がある。

④地方の特性を生かしたハブの形成(マッチングハブの分散と連携)

1) 大学・研究機関・企業・自治体の地方のネットワークを生かしたデータ活用の推進

データの流通、データ提供者と利用者のマッチングにあたっては、大学・研究機関、民間企業、自治体等、地域に根ざした組織等が連携してマッチングハブを形成し、その地域の特性を踏まえたデータ活用を可能とすべきである。また、マッチングハブは、さまざまな法規制の問題や、セキュリティの取り扱い、プライバシーマスキングの技術なども含めて、知見を提供し、連携する組織等と共有していく必要がある。

2) ローカルな知見とグローバルな知見の融合。

データの流通やマッチングについて、地域毎に得られたノウハウ等を地域間で共有し、さらに全国的なノウハウを作り出す等、地域と国全体の知見を融合して、全国展開すべきである。そのためには、全国の大学、研究機関、民間企業、自治体等が利害を乗り越えて全体の利益につなげる方法論を確立すべきである。さらに国際的なグローバルなルールとの融合を図り、国際的に共有できる知見の構築に貢献するべきである。

(ウ) セキュリティの確保の仕組みの検討と、セキュリティルールの形成

①セキュリティ確保の仕組みの検討

個人の医療情報や学習情報等、社会的に機微なデータを扱う場合等も想定し、取り扱うデータに応じて、最適なレベルのセキュリティを確保する方策や、プラットフォーム運用上のセキュリティを高める方策を検討し、常にセキュリティレベルの向上に努めていく必要がある。

②セキュリティルールの形成

実際に機微な情報をデータ保有者が提供するためには、セキュリティ要件に関

して国際的にも広く認められた基準を設定する必要がある。このため、法制化等、社会的に認められるセキュリティレベルの基準を定める、個人情報保護のガイドラインを示すなどの方策が必要である。

一方で、社会にとっての有益度が高いデータが、セキュリティに関する制約により利用の優先順位が下がり、有効活用されないなどということが起きないような考慮も必要である。

(エ) エコシステムの形成(全国的なデータ活用のエコシステムを形成する)

①データ活用社会の継続的な維持/発展

共通基盤やデータ活用推進体制を維持するための予算施策と、全国的なデータ活用サービスを新たな産業として安定させるための施策を検討すべきである。

②共通基盤の利用拡大に向けた整理

SINET をはじめとする共通基盤の民間利用拡大に向けては、必要な基盤とその運用機関の組織体制等の在り方について検討すべきである。

③ステークホルダーとの連携

共通基盤の整備やデータ活用推進体制の構築は、日本学術会議を始めとする我が国の学術のステークホルダーと連携しながら、その方向性を検討すべきである。

④他制度・提言に係る他省庁との連携について

大学・研究機関を中心として、様々な用途に利用できる共通のデータ利用基盤を整備することで、省庁ごとや自治体単位で個別の用途ごとに構築されるデータ活用プラットフォームの受け皿となり、さらに先進的なデータ活用事例を創出してその成果を個別プラットフォームに展開する等、「まち・ひと・しごと創生基本方針」や「DFFT」など国のビジョンを踏まえた国全体での連携も視野に、データ活用を推進していく必要がある。

⑤安心して利用できる仕組み

データ利用の許諾に関わる手続きをデータ利用者・データ提供者双方が速やかに進められる仕組みが必要である。データ活用のための基準・規範を明確にするとともに、データ利用の手続きを円滑に行える仕組みを整備していく必要がある。

また、共通基盤を構成する計算機・ストレージ・ネットワーク等のシステムのセキュリティ対応状況や責任分界点をデータ提供者および利用者が統一したフォーマッ

トで入手し、自らのポリシーに合致するか否か判断できる仕組みが必要である。

さらに、共通基盤やデータ利用中の事故、不適切なデータ活用等が発生した場合の訴訟等のリスクや、その保障制度等の対策について検討し、利用者や提供者に提示する仕組みが必要である。

なお、これらの仕組みについては、技術の進歩や社会情勢の変化に応じて、継続的に見直し、修正すべきである。

4. 整備の現状

現時点で、データ活用を目指した分野にまたがる共通基盤として大学・公的研究機関が現に運用しているもの、あるいは整備を計画している事例には以下のようなものがある。

(ア) DIAS (Data Integration and Analysis System)

DIAS は多種多様かつ大容量な地球観測データ、気候変動予測データ等を統合的に組み合わせ、水循環や農業等の分野における気候変動の影響評価や適応策立案に資する科学的情報を提供するプラットフォームを実現している。DIAS 基盤システムは2015年10月時点で25PBのストレージを有し、莫大な量の地球環境データが蓄積されているだけでなく、さまざまなデータ処理アプリケーションや解析ツールも用意されている。

(イ) ABCI (AI Bridging Cloud Infrastructure)

ABCIは、国立研究開発法人産業技術総合研究所(以下、産総研)が構築し2018年8月に本格運用を開始した、AIとビッグデータのアルゴリズム、ソフトウェア、応用開発のためのオープンかつ専用の計算インフラストラクチャであり、世界トップクラスの計算能力とデータ処理能力を有し、産総研柏センター(東京大学柏第2キャンパス内)に設置されている。ABCIは産学官共同のAI研究開発を加速するオープンイノベーションプラットフォームとして位置づけられ、産総研地域センターのネットワークを活用した地域創生への貢献も目指している。

(ウ) 広域データ活用のための IT 基盤整備

2019年度予算として措置された「広域データ活用のための IT 基盤整備」により、全国利用を前提としたデータ活用社会創成プラットフォームの先導的システム整備が東京大学柏第2キャンパスになされる予定である。この予算は基盤整備のみを対象として

おり、前章で述べたアプリケーション開発あるいはデータ活用を推進する体制の構築について早急に検討を進める必要がある。

(エ) SINET (Science Information NETwork)

国立情報学研究所が構築・運用する広帯域なネットワークである SINET (学術情報ネットワーク) では、これら3つの共通基盤(プラットフォーム)が接続されるほか、全都道府県に配備された 100bps 以上のネットワークに全国の 900 以上の機関が接続され、さらに米国・欧州を經由して北半球を 100Gbps でつなぐリングネットワーク等の国際回線を運用するとともに、閉域性を確保したセキュアな通信環境を L2 レベルで実現する L2VPN を提供している。大量データの入出力があるフラグシップシステム等のスーパーコンピュータや、Spring-8 等の高度な共同利用施設を繋ぐとともに、地方にまで広帯域でセキュアなネットワークを提供する SINET は、世界に類を見ない強力なインフラであり、今後も継続的に増強を図っていくことでデータ活用を全国に広げるために重要な役割を果たすことが期待できる。さらに幅広い研究活動をサポートするため、SINET への新たなアクセス環境として、モバイル網を SINET の足回りとして活用する広域的なデータ収集基盤の構築が始まっており、2018 年度より、この広域データ収集基盤を活用した実証実験が開始されている。

5. 当面の方策

課題や現状を踏まえ、データ活用のためのアイデアを持っていたとしてもそれを実現するためのデータ収集・集積・解析する仕組みを実際に作ることが困難な地方大学や自治体等に、それを可能にする「場」を提供するとともに、解析されたデータから新たな「知」の創出を容易にするシステムとしての「データ活用社会創成プラットフォーム」の実現を目指す。リアルタイムデータの解析・活用を目的としたリアルタイム処理対応基盤ソフトウェアの研究開発や、技術の実証のための基盤システム整備、産学連携体制(コミュニティ)構築・強化、その中核としての大学の役割等に関し、当面の方策として、以下の事項について重点的、集中的に整備することが必要である。

(ア) 基盤整備と先導アプリケーション開発

- ①リアルタイムに大量のデータを収集・蓄積・解析可能な使いやすい共通基盤(様々なデータにアクセスできる環境整備を含む)の整備

1) システム整備

SINET の帯域増強ならびにモバイル網の拡充を推進し、全国から大量のデー

タを高速に収集可能な基盤を整備する。

また、東京大学が中心となって柏拠点に整備中のストレージ、解析用計算機の運用体制の構築と拡充を推進し、全国から収集される大量のデータの蓄積ならびにリアルタイム解析が可能な基盤を整備する。

さらに、安全かつリアルタイムなデータ収集・蓄積・解析を行うために、SINETならびにストレージ、解析用計算機を最大限に活用するための基盤ソフトウェアを整備する。

なお、システム整備においては、DIASをはじめとする他のデータ活用フレームワークとの連携も考慮する。

2) 解析ツール(ソフトウェア)

異なるアプリケーションにおいて共通して利用可能な解析ツール、データアナレーションツールのニーズを調査し、共通基盤上で利用可能とする。

3) データ収集管理システム(ソフトウェア)

アプリケーション開発者が、センサ等から連続的に生成される大量のデータをSINET(モバイル網も含む)を経由して確実に収集、蓄積し、解析するアプリケーションを開発するための機能を提供するデータ収集管理システムを開発し、共通基盤上に整備する。システムの開発・整備は、アプリケーションに必要とされる機能を調査しながら機能拡張を実施する等、アプリケーション開発者のニーズを取り入れながら進める。

②アプリケーション開発及び開発支援

アプリケーション開発支援の先導的なモデルを構築するため、データ分析、計算機システム、ソフトウェア等の専門的な知識・技術を持つ人材を地域に根ざした大学等において確保するための施策を検討する。また、アプリケーション開発者からのニーズや期待を集約・共有するためのコミュニティの形成を進める。

本基盤は、モバイル網を含む広域ネットワークと計算機、ストレージ資源を一体として扱い、個々のアプリケーションに仮想的・専用に使える広域処理環境を提供する点にこれまでにない特色がある。特にリアルタイム処理、大量のデータを収集しつつ行う処理、要求計算量が時間とともに大きく変動するような処理を効率よくサポートするためには、アプリケーションと連動した技術開発が必要であるが、アプリケーションで共通して利用できる部分が大きいため、先導的なアプリケーションを開発し、知見を利用者に提供していく。

(イ) データ活用を推進する体制の構築

①データ整備・流通

1) 大学が中心となった先導的な実施例

地域に根ざした大学が中心となってデータ整備・流通の先導的なプロジェクトを実施し、そこから得られる知見を収集し、問題点を把握する。

2) 国内外のデータ活用の動向把握

データ活用に係る国内外の動向を把握するとともに、上記1)のユースケースの掘り起こしを含めた事例の収集等を継続的に実施する。

②マッチング・コンサルティング機構

1) マッチング・コンサルティング機構の設計

大学・研究機関と民間企業の持つ技術やデータのマッチングを図るための具体的な実施方法を設計する。

また、大学・研究機関が持つ知識や技術の社会実装を推進するためのコンサルティングの具体的な実施方法を設計する。さらに、アプリケーション作成者やデータ提供者を広範囲に広げるためには、作成者・提供者が困りがちなセキュリティの取り扱いや、個人情報保護のガイドラインなどを相談できる、リスクコンサルティング機能もあわせて検討する。さらに、リスクを低減するための予算措置の可能性についても検討する。

③コミュニティの形成

1) 全国レベルの連携を実現するコミュニティ(コンソーシアム)の形成

地域に根ざした大学がハブとなり、先導的な取り組みによって得られたデータ整備・流通、マッチング・コンサルティングに関する知見を全国で共有する体制を構築する。

また、全国のデータ提供者、データ利用者、基盤提供者が参加可能なコミュニティ(コンソーシアム)を形成し、データ提供者およびデータ利用者からのニーズを基盤提供者が共有できる体制を構築する。

なお、これらの体制を実現するためには地域に根ざしたデータに精通する人材を大学等において確保する仕組みが必要である。

(ウ) セキュリティに関する方策

① 共通基盤のセキュリティ

共通基盤のセキュリティ対応状況や責任分界点をまとめたチェックリストの作成等、データ提供者および利用者が共通基盤のセキュリティレベルを理解し、安心して利用できるための仕組みについて検討する。

(エ) エコシステムの形成

① 体制

データ活用の共通基盤や推進体制における大学・研究機関と民間企業の関係(大学・研究機関主導でよいのか、設計当初より民の知見を入れるべきか)を整理し、検討する。また、産業界が所有するものを含むデータを、どのように集め世の中に提供していくのか、エコシステムを含めた議論をコミュニティで行っていく。

② 運営費用、事業の継続性の確立

先導的な取り組み(実証段階)から、共通基盤整備やデータ流通・整備体制の整備における民間資金の活用方法について検討する。

また、先導的な取り組み(実証段階)から、需要に応じた継続的な共通基盤整備の拡充方法についても併せて検討する。

③ 予算要求(受益者負担の在り方)

整備・運用費用については、将来を見据えた受益者負担のあり方についても検討を進める。

6. まとめ

Soceity 5.0 が目指す超スマート社会の実現に向けては、大学や研究機関等に蓄積される膨大なデータの中から必要となるデータを取得し、あらゆる分野と地域でデータを最大限に活用するデータ活用社会の創成が不可欠である。そのためには、日本全体を網羅する超高速ネットワークのバックボーンである学術情報ネットワーク(SINET)を最大限活用し、共用データ、データ解析ノウハウ、インフラなどをだれもが容易に利用できるようにすることが重要であり、47 都道府県に設置された国立大学を中心としたアカデミアが地方自治体や産業界との連携も視野に、これらを提供する開発共通基盤の整備を進めるとともに、共通基盤上でのアプリケーション開発を通した先導的な事例を積み上げ、その知見をコンサルティング機能やマッチング機能を介してあらゆる分野・地域に展開することが必

要である。これは、DFFT を体現する Society 5.0 の実現というフロンティアを開拓する最適な場の形成となり、我が国の国際的な優位性の確保につながる取組である。

本提言はそのための方策をまとめたものであり、政府、地方自治体、関係企業、ならびに大学・公的研究機関が協力して実行に移すことが求められる。

なお、本提言の内容は、今後の関連技術の進展、社会の変化等に応じて常に見直しを行い、進化・発展させていかなくてはならない。

用語解説

アノテーション・データアノテーション

あるデータに対してタグやメタデータと呼ばれる元のデータの注釈となる情報を付加すること。

アプリケーション

ある特定の機能や目的のために開発・使用されるソフトウェア。

エコシステム

企業等が互いの技術や資本を生かしながら、業界の枠や国境を超えて広く共存共栄していく仕組み。

オープンデータ

権利等に基づく制限を課されることなく、誰でも自由に入手、加工、利用、再配布などすることができるよう公開されたデータ。

キュレーション

収集した情報を選択、修復・分類、結合するなどし、高い価値を持たせること。

クローズドデータ

権利等に基づき制限され関係者のみが利用するデータ。

ストレージ

デジタルデータを保存する装置。磁気ディスク装置や半導体ディスク装置（SSD : Solid State Drive）、光学ディスク装置、磁気テープ装置などのことを指す。

トレーサビリティ

プログラムなどについて、対応関係や変更履歴などを記録し、追跡や検証ができる状態。

ハブ

中心地、結節点、集線装置などの意味。

ビッグデータ

従来のデータベース管理システムなどでは記録や保管、解析が難しいような巨大かつ多様なデータ群。

プライバシーマスキング

プライバシーの対象となる情報を保護するもの。

プラットフォーム

基盤となる装置やソフトウェア、サービス、あるいはそれらの組み合わせのこと。

マッチング

異なったものを組み合わせること。

メタデータ

データについてのデータ。あるデータそのものではなく、そのデータを表す属性や関連する情報を記述したデータ。

リアルデータ

健康情報、走行データ、製品の稼働状況等、個人・企業の実世界での活動についてセンサー等により取得されるデータ。

AI (Artificial Intelligence)

人工知能。高度に知的な作業や判断を実行するコンピュータを中心とした人工的なシステムのこと。

BCP (Business Continuity Plan)

企業や官公庁などで、通常業務の遂行が困難になる事態が発生した際に事業の継続や復旧を速やかに遂行するために策定される計画。

DFFT (Data Free Flow with Trust)

2019年1月にスイス・ジュネーブで開催された世界経済フォーラム年次総会（ダボス会議）にて、安倍内閣総理大臣がスピーチで提唱した言葉。「信頼性のある自由なデータ流通」と訳される。

DOI (Digital Object Identifier)

インターネット上の様々なドキュメント（電子データ）に恒久的に与えられる識別子。

Gbps (Giga bit per second)

ビットパーセカンド (bps) はデータ伝送速度の単位の一つで 1 秒間に何ビットのデータを伝送できるかを表す。毎秒10億ビット (1ギガビット) のデータを伝送できるのが1Gbpsである。

IoT (Internet of Things)

従来インターネットに接続されていなかった様々なモノが、ネットワークを通じてサーバーやクラウドサービスに接続され、相互に情報交換をする仕組み

L2・L2VPN (Layer2・Layer2 Virtual Private Network)

L2 (レイヤー 2) は通信プロトコル (通信手順/通信規約) の機能や役割を階層構造で整理したモデルを構成する層の一つで、回線やネットワークで物理的に繋がれた二台の機器の間でデータの受け渡しを行うもの。VPNは仮想専用線という意味でネットワーク上に仮想的な専用線を設けて、セキュアな経路を使ってデータをやり取りする通信環境。L2VPNはL2レベルで実装されたVPN。

PB (Peta Byte)

バイト (Byte) はデータ容量あるいは情報量を表す単位で、1バイトは8ビット、ペタバイト (Peta Byte) は 2^{50} バイトを表す単位。

Proof-of-Concept (PoC)

「概念実証」の意味で、コンセプト (概念) の実現可能性を検証することを指す。

Real World Data

医療分野において、診療報酬請求 (レセプト) データや、診療録 (電子カルテデータ由来)、健診データなどの実診療行為に基づくデータ、またはそのデータベースを指す。

SINET (学術情報ネットワーク)

日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所 (NII) が構築、運用している情報通信ネットワーク。

Society 5.0

サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会（Society）。狩猟社会（Society 1.0）、農耕社会（Society 2.0）、工業社会（Society 3.0）、情報社会（Society 4.0）に続く、新たな社会を指すもので、第5期科学技術基本計画において我が国が目指すべき未来社会の姿として提唱された。

SS-MIX2標準化ストレージ

「厚生労働省電子的診療情報交換推進事業」（SS-MIX：Standardized Structured Medical Information eXchange）において策定された、医療機関を対象とした医療情報の交換・共有のための規約を満たしたストレージ規格。

Society5.0を実現するためのデータ活用による知識集約型社会の創成 ーデータ活用社会創成プラットフォームの構築ー

データ活用社会における現状認識

- ICT機器の爆発的な普及や、AI、ビッグデータ、IoT等の社会実装が進むなど競争が激化、一部の企業や国のデータの囲い込みにより経済社会システムの健全な発展が阻害される懸念。
- 我が国が成長していくためには、デジタル新時代において、データを我が国全体の共同資産として、スピード感をもったデータ活用環境の整備が急務。
- Society5.0が目指すインクルーシブな社会を実現するためには、地域における知識集約の中核を担う大学を起点としてイノベーションの創出を図り、知識集約型社会を構築することが重要。
- サイバー空間とフィジカル空間が融合するデジタル新時代において、我が国に蓄積された農業、医療・健康の分野、教育データを含む多くの有用なビッグデータを共同で活用する上で、人材と技術を有する全国の大学を超高速・高信頼で網目状につなぐ国際的優位性をもつSINETを最大限活用することが重要。
- 異種データや異種知識の融合・活用を促進するための「場」として、様々な分野のデータ保持者、解析者、利用者が参画するコミュニティを形成するとともに、データ活用を旨とする利用者へのコンサルティングやアプリケーション開発支援が不可欠。

文部科学省における取組

- 経済財政運営と改革の基本方針2018や未来投資戦略2018において重要性が指摘されているリアルデータの利活用を念頭に、データ活用社会創成プラットフォームを推進するため、SINETを通じて収集されるリアルデータの集積や、解析結果を速やかにフィードバックする機能を備えたシステムを整備(2019年度予算)
- 文部科学省と大学コミュニティ、地域社会等が一体的に連携し、全国の国立大学等をハブとしたデータ活用社会創成プラットフォームの実現促進に向けた検討を行うための「データ活用社会創成プラットフォームの推進に関する有識者会合」を設置。
- 地域・産業・社会基盤を支える拠点となる大学を中心として、民間への利用拡大も視野に我が国全体の知識集約型社会の実現に向けた環境「データ活用社会創成プラットフォーム」を構築

データの高度利用環境(NII・東大に先行して整備)

【設備整備】

IoT接続 (モバイル)
AI特化スパコン
リアルタイム処理対応スパコン
高速/セキュアストレージ 等



SINETを通じて、全国のデータ収集・通信・解析環境をオンデマンドで活用。
高度・多様なデータ活用により新たな価値を創出。

利活用ニーズを踏まえたシステム整備・ソフトウェア開発

【大学等におけるデータ活用の潜在的なニーズ】

- ・地域農業・漁業・観光業のスマート化
- ・認知症・生活習慣病などの早期発見、予防方法の提案
- ・スポーツ科学への応用
- ・初中段階から高等教育、社会人教育に至る一貫した教育データの利用 等

データ活用社会創成プラットフォームの推進に関する有識者会合

リアルタイム処理対応基盤社会創成プラットフォームの実現に向けた実務的な検討を行う場

【主な検討課題】

- ・リアルタイムデータの解析・活用を目的とした基盤ソフトウェアの研究開発や技術の実証のための基盤システムの整備のあり方
- ・産学連携体制(コミュニティ)の構築・強化、その中核としての大学の役割等一体的な連携を確保する仕組み 等



大学等連携コンソーシアム

大学を中核としたデータ活用実務機関が連合したコンソーシアム

【主な取組】

- ・データプラットフォームの活用促進、データ活用ニーズ調査
- ・コミュニティ間連携の強化・促進等



データ活用社会創成プラットフォームの推進に関する有識者会合開催要綱

1 趣旨

Society5.0 が目指すインクルーシブな社会を実現するためには、我が国全体のネットワークインフラである SINET と地域における知識集約の中核を担う大学を起点としてイノベーションの創出を図り、知識集約型社会を構築することが重要である。

このため、「データ活用社会創成プラットフォーム」の実現促進に向けたデータの解析・活用を必要とする研究開発や技術の実証のための基盤システム整備、産学連携体制（コミュニティ）構築・強化、その中核としての大学の役割について重点的、集中的に討議することとし、当該取組の推進方策に関する助言を得るため、「データ活用社会創成プラットフォームの推進に関する有識者会合」を開催する。

2 検討事項

- (1) 我が国におけるデータ活用社会の推進に係るプラットフォーム（SINET の強みを活かせる多様なデータ利活用システム）の整備方策
- (2) 大学を中核としたデータ利活用環境の構築と発展の方策
- (3) 我が国の強み（ネットワークインフラ、知の集積拠点としての大学等）をいかした、データ活用による新たなイノベーションを創出するための具体的方策に係る論点整理
- (4) その他

3 実施方法

- (1) 別紙の構成員の協力を得て、上記 2 に掲げる事項について検討を行う。
- (2) 有識者会合に主査を置く。また、必要に応じて主査代理を置くことができる。
- (3) その他、有識者会合の運営に関し、必要な事項は主査が定める。

データ活用社会創成プラットフォームの推進に関する有識者会合構成員名簿

(50 音順)

相澤 清晴	東京大学大学院情報理工学系研究科教授
合田 憲人	情報・システム研究機構国立情報学研究所アーキテクチャ科学研究系教授
小池 俊雄	国立研究開発法人土木研究所水災害・リスクマネジメント国際センター長
後藤 厚宏	情報セキュリティ大学院大学学長
須藤 修	東京大学大学院情報学環教授
関口 智嗣	国立研究開発法人産業技術総合研究所理事
中島 直樹	九州大学病院メディカル・インフォメーションセンター長
中村 宏	東京大学大学院情報理工学系研究科教授
東野 輝夫	大阪大学大学院情報科学研究科教授
平藤 雅之	東京大学大学院農学生命科学研究科附属生態調和農学機構特任教授
○ 安浦 寛人	九州大学理事・副学長

○：主査

データ活用社会創成プラットフォームの推進に向けた当面の整備方策（概要）

（令和元年7月）データ活用社会創成プラットフォームの推進に関する有識者会合

1. データプラットフォームが求められる背景

（ア）地方の産業、自治体を含む我が国全体でのデータ活用推進の重要性

- ・データ活用社会は、**地域、年齢、性別、言語**による格差等の課題を解消し、**地域の特色を活（い）**かした多様な産業の活性化に貢献する **Society 5.0が目指す社会**であり、知・人材が集積する大学等を中核とした知識集約型拠点を形成・実現していく必要がある。
- ・これからの社会ではあらゆる分野でデータ活用が必須であり、学術コミュニティにおいても、**社会実装を指向した研究分野が重要**である。
- ・しかし、データ活用においては、必要となる異分野のデータの把握、それらのデータを活用する解析ノウハウの獲得、更に解析に用いるITインフラの構築を行う必要があり、**データ活用に至る障害の大きさから、様々なデータが広く有効活用されていない**。

（イ）データ活用の推進に向けた早急な利活用環境の整備の必要性

- ・データの活用を推進するためには、**Proof-of-Concept（PoC）**によりその効果を実証することが重要であり、PoCをいかに迅速に行えるかが活用の成否を左右する。
- ・データ活用を促進するためには、データ量の大小に関わらず、自身が有するデータと利用可能な公共データを合わせて様々な解析を行い、**より高い技術を習得するための「データ工房」を整備することが重要**である。

2. データプラットフォーム推進における課題

（ア）地方創生、地方活性化に向けたデータ活用のための人材に関する課題

- ・データの活用を推進していくためには、データを実際に生成する地方・地域社会の現場の知識と、データ活用の知識の両者が必要であるが、**アイディアがあっても実際にデータを収集・集積・集積・解析する情報技術を保有する人材が少なく、データ活用の知識・ノウハウを持つ人材とのつながりも希薄であることが多い**。
- ・地方・地域社会でのデータ活用においては、**活用の範囲を当該地方・地域内に閉じて考えがち**なため、波及効果が限定的でエコシステムも成立しないことが多く、**広く展開する仕組みがないと**考えられる。

（イ）データ活用のインフラ（基盤）の未整備と方法論の確立に関する課題

- ・現時点において、**大量データの処理、リアルタイム処理に利用可能なインフラが存在しない**。
- ・地方・地域社会の現場では、データを活用することで何が実現できるかについて十分な知見がなく、解決したい課題に対し、どのようなデータと解析技術があれば実現できるのかが分からないことも少なくない。

（ウ）データ流通における課題

- ・データ活用により創出される価値の向上には、利用できるデータの量と質が十分確保されていることが不可欠であるが、**個人情報保護とデータ活用を両立するためのポリシー等の方策が未整備であるため適切な活用を促進することが困難な現状**がある。
- ・データは、量・種類ともに多く、大量のデータを分析するためには専門的な知識を有した人材が必要であるが、**人材が不足していることもあり、どのようなデータが利用可能であるか、そのデータをどうすれば入手できるのかといった情報が広く共有されていない**。
- ・データを有する研究者や企業等がデータを他に活用させることへのインセンティブがないため、有用なデータが流通しない。
- ・従来の産学官連携の仕組みは、どのような技術が必要か明確な場合にのみ機能するものが多く、潜在的な需要を拾い上げるのには適していない。

（エ）エコシステムの機能に関する課題

- ・エコシステムへの期待が高まる中、**潜在的なデータ活用の需要を拾い上げるエコシステムが存在していない**。また、データ活用にコストをかけることによって得られる利益について、必ずしもデータ活用が成功するとは限らずリスクがあること、当初のデータ活用を担う主体が、広範囲に広がるデータ活用を最後まで担うとは限らないことなど利益を生み出すエコシステムが確立されていない。

3. 実現すべきデータ活用を促進する長期的な仕組み

(ア) 基盤整備と先導アプリケーション開発

- データ活用の方法論を確立するためには、データ提供者・利用者が利用可能な共通基盤を整備するとともに、**先導的なデータ活用アプリケーション開発プロジェクトを推進**し、成功事例を蓄積し、その資産を再利用することが必要である。具体的には、共通基盤の整備やアプリケーション開発の推進を通じて、全国に分散した設備や人材を適材適所で活用することにより、オールジャパンの体制で取り組むべきである。

(イ) データ活用を推進する体制の構築

- 全国的なデータ活用体制を構築するためには、**データの整備・流通を図るとともに、データ活用のコンサルティング機能、データ提供者/データ解析技術者/データ利用者のマッチング機能を全国に構築することが必要**である。全国展開に当たっては、先導的機関による取り組みからユースケースとしての事例を蓄積するとともに、**全国の大学がハブとなり、地方自治体や企業も含め展開するコミュニティを形成していくことが重要**である。

(ウ) セキュリティの確保の仕組みの検討と、セキュリティルールの形成

- 医療情報や学習情報等、社会的に機微なデータを扱う場合等も想定し、取り扱うデータに応じて、最適なレベルのセキュリティを確保する**方策や、プラットフォーム上のセキュリティを高める方策を検討し、**常にセキュリティレベルの向上に努めていく必要がある**。
- 実際に機微な情報をデータ保有者が提供するためには、セキュリティ要件に関して国際的にも広く認められた基準を設定する必要がある。このため、法制化等、社会的に認められるセキュリティレベルの基準を定める、個人情報保護のガイドラインを示すなど方策が必要である。
- 一方、**社会にとつての有益度の高いデータが、セキュリティに関する制約により有効活用されないなどということが起きないよう考慮する必要がある**。

(エ) エコシステムの形成（全国的なデータ活用のエコシステムを形成する）

- 共通基盤やデータ活用推進体制を維持するための予算施策と、全国的なデータ活用サービスを新たな産業として安定させるための施策を検討すべきである。
- データ利用の許諾に関わる手続をデータ利用者・データ提供者双方が速やかに進められる仕組みが必要であり、データ活用のための基準等を明確にするとともに、**データ利用中の事故や不適切なデータ活用等が発生した場合の訴訟等のリスクや、その保障制度等の対策について検討し、利用者や提供者に提示する仕組みが必要**である。

4. 整備の現状

- データ活用を推進していく上での現状の共通基盤として、①DIAS (Data Integration and Analysis System) や、②ABCI (AI Bridging Cloud Infrastructure) が運用されているとともに、全国利用を前提としたデータ活用社会創成プラットフォームの先導的システムの整備が予定されており、今後は、本設備を運用するためのアプリケーション開発やデータ活用を推進する体制の構築について検討を進めていく必要がある。
- これらの共通基盤（プラットフォーム）は、**全国の900以上の機関を100Gで接続し、広域かつ、セキュアな通信環境を実現するSINET（学術情報ネットワーク）に接続されるものであり、SINETを今後も継続的に増強を図っていくことでデータ活用を全国に広げるために重要な役割を果たすことが期待**される。

5. 当面の方策

- 課題や現状を踏まえ、データ活用のためのアイデアを持っていてもそれを実現するための**データ収集・集積・解析する仕組みを実際に作る**ことが困難な**地方大学や自治体等に、それを可能にする「場」を提供し、解析されたデータから新たな「知」の創出を容易にするシステム**としての「データ活用社会創成プラットフォーム」の実現を目指し、**リアルタイムデータの解析・活用を目的としたリアルタイム処理対応基盤ソフトウェアの研究開発や、技術の実証のための基盤システム整備、産学連携体制（コミュニティ）構築・強化、その中核としての大**学の役割等について、**重点的、集中的に検討・整備していく必要がある**。