

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、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、**社会にとっ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に考慮**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、全国的なデータ利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利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利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、広帯域かつ、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**。