

教育関連データの二次利用の実現に向けた データベースの構築に関する調査研究

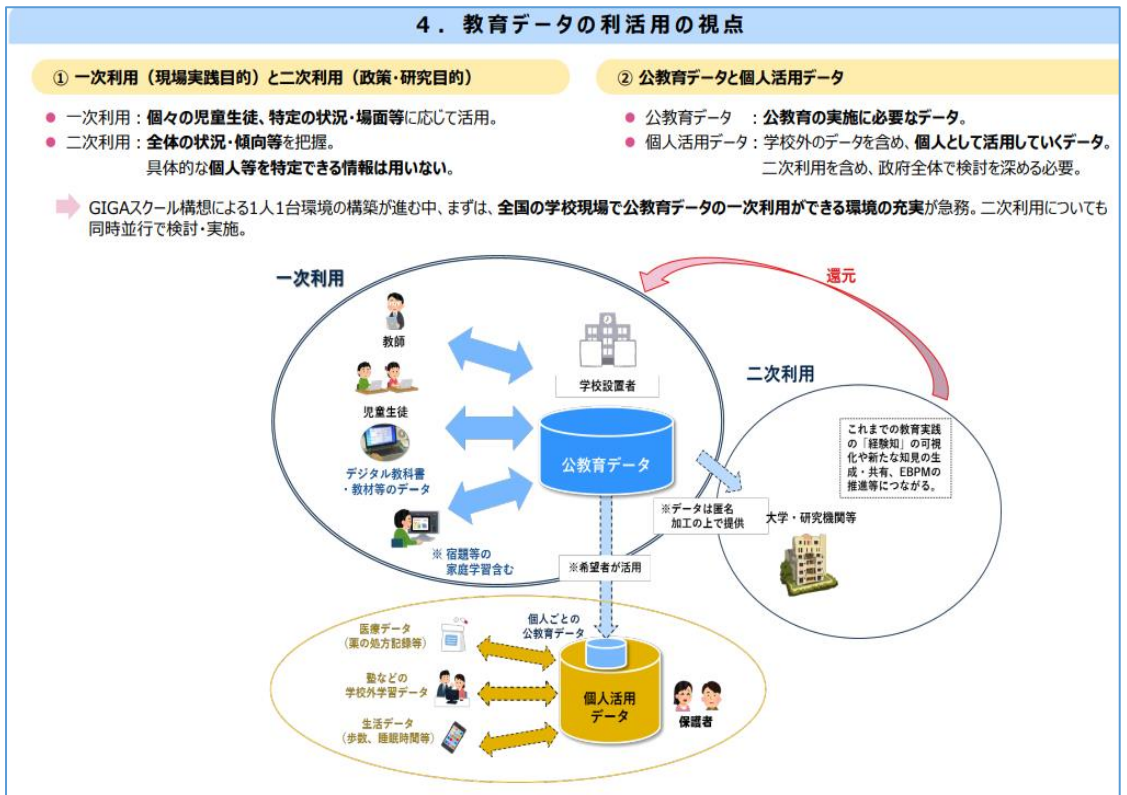
2024/3/29

はじめに

教育関連データの有効活用と二次利用を通じて、教育政策の策定と研究を強化し、教育の質を向上させるためのデータベース構築を目指している。

背景	<ul style="list-style-type: none">EdTechの急速な発展が教育政策・研究のためのデータ活用を促進データ保護、プライバシーの懸念などの問題が二次利用の拡大を阻害
目的と範囲	<ul style="list-style-type: none">教育データ活用の課題解決と効果的なデータベース構築への基本情報提供利害関係者のニーズ分析から法的・倫理的課題抽出まで幅広く調査
経緯と現状	<ul style="list-style-type: none">教育データの生成と収集の急増初期段階にあるデータ活用の課題と可能性の探索
教育関連データの二次利用基本定義	<ul style="list-style-type: none">新たな目的での既存データ活用プライバシー保護、データセキュリティと法律・ガイドラインの遵守が重要
調査目標	<ul style="list-style-type: none">政策立案と研究に貢献するデータベースの構築教育の質向上と学習者一人ひとりに合わせた教育提供への貢献

■主な調査対象 ・ 政策立案者および研究者 / 教育関連事業者は支援的役割



※文部科学省・教育データ利活用に関する有識者会議
教育データの利活用に係る論点整理（中間まとめ）概要資料（2021年3月）

教育データの有効活用が教育政策と研究の推進に必須であり、データ保護と技術進展を活かした活用が新たな機会を生む。

教育データの現状とその重要性

教育データは、教育の質を向上させ、学習経験を個別化するための貴重な資源。

・課題

データ集約性の欠如

アクセス制限

分析能力の不足

潜在的価値の実現を阻害

関連技術と動向

・活用技術

ビッグデータ分析

人工知能 (AI)

クラウドコンピューティング

等

これらの技術は、データ処理能力の向上と高度なデータ分析を可能に。
ブロックチェーン技術の導入によりデータの安全性と透明性が確保される可能性。

二次利用の社会的・教育的意義

目的

学習者のサポート向上、教材開発、政策立案等、多岐にわたる用途に役立つ。

社会的意義

データドリブンなアプローチにより、学習者一人ひとりのニーズに合わせた教育が実現可能。教育政策の立案に実証的データを基にした政策が可能。

教育的意義

教育プロセスの革新を促進。学習者の成績データ、フィードバック等の分析を通じて学習効率の最適化や個別の学習プラン作成が可能。

法的・倫理的考慮事項

課題

関係者の同意の取得、プライバシー保護、データセキュリティ

必要

明確な法的枠組みの下での倫理的ガイドラインの策定と遵守。

目的

教育データの二次利用に関する現状分析と戦略的枠組みの提供。
教育データの有効活用による教育政策の最適化と教育研究の推進。

教育データの有効活用を目指し、政策立案・研究支援のデータベース構築方向性を探求。事業者も支援的に参画。

目的

教育データの政策立案と研究への有効活用を目指し、関連データベース構築の方向性を探求。

対象と範囲

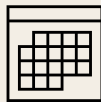
- ・国内の教育機関、研究機関、教育関連企業。
- ・先進事例と具体的施策に焦点。

期待成果

- ・現状分析とデータベース構築提言。
- ・プライバシー保護と二次利用ガイドラインへの提案。

調査期間

- ・2カ月。計画立案から最終報告書作成まで。



課題と制約

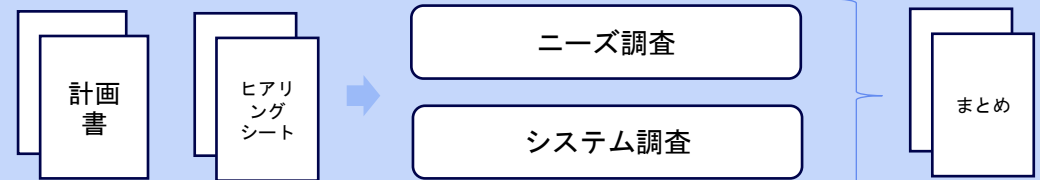
- ・時間制約による深い分析は困難。広範囲なアクセスと詳細ケーススタディの限界。

主要調査項目

- ・ **利害関係者のニーズと期待**
地方公共団体、研究機関、事業者などの要望と期待調査。
- ・ **データの種類と利用可能性**
二次利用に適した教育データの特性と活用法探索。
- ・ **教育データの機会と課題**
教育データの再活用の可能性と直面する課題の検討。
- ・ **データ保護と法的枠組み**
現行法規とプライバシー保護の調査。
- ・ **技術的ツールとデータベース**
二次利用を支援する技術やデータベースの研究。



調査イメージ



海外の教育データの二次利用事例では、政策策定、研究進展、実践改善への貢献が示されている。

・アメリカとイギリスの事例から学ぶ、データを活用した教育改善のアプローチ

国名	データ収集システム	利用事例	プライバシー保護対策
アメリカ	EDFacts、CCD、CRDC、NTPS 及び、SEDA（収集システム上のデータを研究用に成形したデータセット）	<ul style="list-style-type: none"> ・ 成績データ分析 ・ 社会経済的状況との相関分析 ・ 特別支援教育の評価 ・ 公立学校の人種的分離解消提案 	<ul style="list-style-type: none"> ・ データ匿名化 ・ アクセス権限の厳格な管理
イギリス	ONS SRSを通じて教育省データへアクセス可能	<ul style="list-style-type: none"> ・ 高等教育データ ・ 個別学習者記録 ・ 全国生徒データベース ・ 学校職員調査 	<ul style="list-style-type: none"> ・ UK GDPR及び2018年データ保護法に基づくガイダンス ・ ONS承認研究者スキームを通じたアクセス制限



アメリカ

教育省が直接関与して広範囲にわたるデータ収集システムを構築し、これらのデータを教育の質向上、政策策定、研究の促進に活用している。



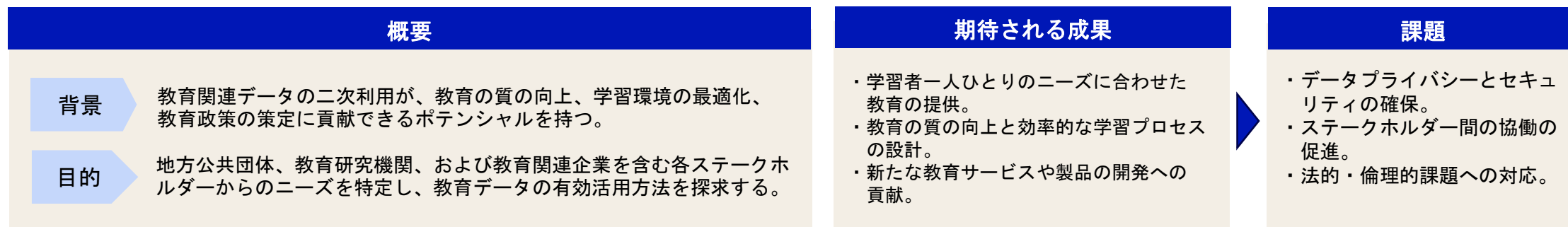
イギリス

Office for National Statistics Secure Research Service（ONS SRS）を通じて、より広範囲のステークホルダーが教育データにアクセスできるようになっており、プライバシー保護の観点からも、ICOによる指導とガイダンスが提供されている。




展望

教育データの二次利用は、教育改善への貢献、政策策定への情報提供といった重要な役割を果たす。未来への展望は、データ駆動型教育への移行と、個人情報保護の強化がある。

教育データの二次利用は、教育の質を向上させ、ニーズに合わせた教育の提供を可能にするが、プライバシー保護と各関係者の連携が重要。



■主なヒアリング調査結果

教育関連組織	ヒアリング先	ニーズ
地方公共団体	尼崎市 つくば市 戸田市 	地方公共団体は、児童生徒の出欠情報、成績、進学率、学校施設の利用状況などを含むデータの二次利用を通じて、地域における教育の質の向上、学習環境の最適化、教育政策の策定に役立てることに関心が高い。
教育研究機関	国立教育政策研究所 京都大学学術情報メディアセンター 	教育研究機関は、効果的な教育方法の開発、児童生徒の行動や成績の分析、教育プログラムの評価などに教育データの二次利用を求めている。
教育関連企業	スタディプラス スズキ教育ソフト システムディ 	教育技術（EdTech）企業を含む教育関連企業は、新しい教育ツールやプラットフォームの開発、パーソナライズされた学習経験の提供、効果的な学習管理システムの構築に教育データを活用することを目指している。

※本調査では、「政策立案・研究」を核とした教育データの有効活用に関心があるが、教育関連企業、特に教育技術（EdTech）企業は、教育プロセスのデジタル化において中心的な役割を担っていることから、これらの企業のニーズと期待を探ることは、教育政策の策定支援や研究推進に資するデータベース構築の方向性の探求に良い洞察を提供すると考えられる。

教育関連データの活用で個別化された学習支援を実現し、教育内容のカスタマイズと学習成果の向上を通じて教育サービスの質を高める。

概要

教育関連データの二次利用が、教育の質の向上、学習支援の最適化、新しい教育サービスの開発にどのように貢献できるか、具体的なデータ項目と活用事例を通じて紹介する。

効果的活用のための主要なデータ項目イメージ

データ項目	内容
学習者の進捗データ	成績、試験結果、課題の進捗など
学習行動データ	オンライン学習プラットフォームでの活動 (例: 閲覧履歴、クイズ回答履歴)
教育プログラムの評価データ	プログラムの効果や評価に関する情報
出欠や参加データ	学習者の出席状況や教育イベントへの参加状況
学習者の属性データ	年齢、性別、学年、居住地などの基本属性

活用事例

- ・ **カスタマイズされた学習経験の提供**
学習者の成績とオンライン活動のログを分析し、個別の学習ニーズに合わせた教育コンテンツやサポートを提供。
- ・ **教育の質の向上**
学習者のフィードバックと教育プログラムの評価データを用いて、教育方法の改善や効果的な学習プログラムの開発。
- ・ **新たな教育サービスの開発**
教育データを市場のトレンド分析に利用し、ニーズに応える教育製品やサービスの開発。

課題と対応策

- ・ **データプライバシーとセキュリティ**
データの匿名化、安全なデータストレージとアクセス制御の確保。
- ・ **データの統合と分析**
異なるソースからのデータ統合、高度な分析技術の活用。
- ・ **ステークホルダー間の協力**
教育関係者、開発者、政策立案者間でのデータ共有と協力体制の構築。

教育データの二次利用には、プライバシー保護と有効活用を両立させる明確な法的枠組みと実践的課題の解決が不可欠。

現行法制度の概要

個人情報保護法に基づく法的枠組みの説明
教育データの利用に関する基本的要件
(利用目的の明確化、保有制限、適切な管理など)

二次利用の条件とプロセス

- ・ 利用目的の明確化、データの匿名化、個人情報保護
- ・ データ提供時の同意取得

教育データの適切な二次利用

・ 学術研究や統計作成など、特定の利用目的でのデータ利用が許可されていること
法令に基づく業務範囲とデータ保護措置の必要性。

教育分野特有の法的要件

- ・ 文部科学省による指針の紹介
- ・ 個人情報保護の適用と児童生徒のプライバシー保護

直面している課題

- ・ データ利用目的の具体性欠如
- ・ 法的制約への理解不足
- ・ ステークホルダー間の倫理観の相違

実践的課題への対応

- ・ 利用目的の明確化
- ・ 法的枠組みに対する理解の深化
- ・ ステークホルダー間のコミュニケーションの改善

- ・ 匿名加工情報・仮名加工情報とは？

匿名加工情報とは、特定の個人を識別することができないように個人情報を加工し、当該個人情報を復元できないようにした情報のことを指します。
仮名加工情報とは、一定の措置を講じて個人情報を加工し、他の情報と照合しない限り特定の個人を識別できないようにした情報を指します

■個人情報保護委員会 <https://www.ppc.go.jp/>

教育データの二次利用には、セキュアで柔軟なデータベースや適切なガバナンス構造・機能が必須。プロセスの最適化により、持続的価値を実現する。

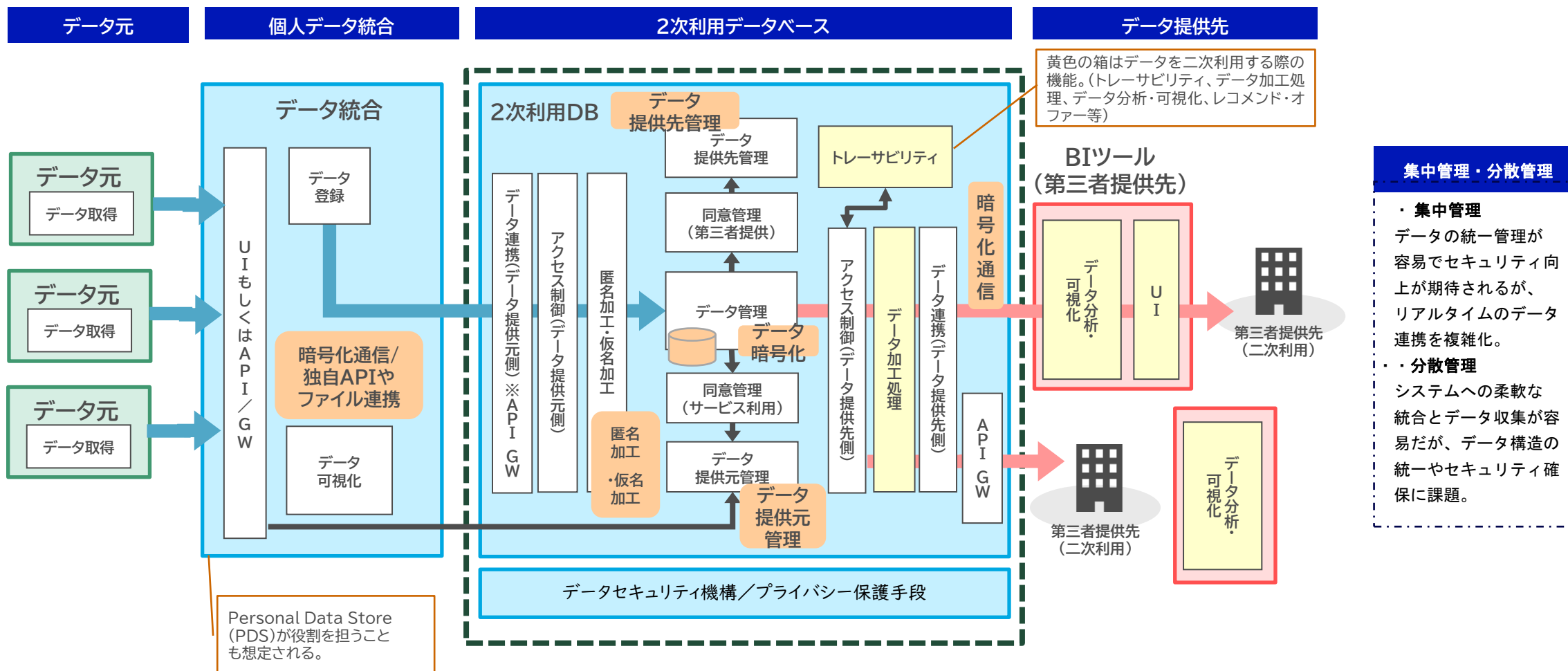
運用フロー

- 1. データ収集** 学校や教育施設からのデータを収集する。
- 2. データの匿名化または仮名加工** 個人を特定できる情報を除去し、データのプライバシーを保護する。
- 3. データ統合と前処理** 収集したデータを統合し、分析に適した形式に加工する。
- 4. データベースへの格納** 加工されたデータをデータベースに保存する。
- 5. セキュリティとアクセス管理** データベースのセキュリティを確保し、適切なアクセス権限を設定する。
- 6. データ分析と活用** 分析ツールを用いてデータを分析し、教育政策やプログラム開発に活用する。
- 7. レビューと更新** データベースと業務プロセスを定期的にレビューし、必要に応じて更新する。

二次利用のためのアーキテクチャ

- データ収集と統合** 複数ソースからのデータ収集と一元管理への統合。
- データクリーニングと前処理** 品質向上のためのノイズ排除、欠損データ処理、外れ値検出。
- データストレージ** クラウドベースやデータウェアハウスを用いた安全な保存。
- セキュリティとアクセス制御** 個人情報の保護とアクセス管理の強化。
- 分析とモデリング** 機械学習や統計モデルを活用した有意義な情報の抽出。
- パーソナライズド学習支援** 学習者一人ひとりに合わせた学習経験の提供。
- 透明性と説明可能性** 分析やモデリングの結果を理解しやすい形で提供。

教育データの二次利用には、効率的な運用フロー、堅牢なアーキテクチャ、柔軟なシステム構成が必要。



教育データの二次利用を目的としたデータベースの構築と運用は、長期的な視野と効果的な予算計画に基づき、その投資価値と正当性を明確にすることが重要

概要

教育データの二次利用を支えるデータベースの構築と運用に関わるコストを、プロジェクトの規模やデータ複雑性、システム要件に基づいて試算。

コスト要因

主要なコスト要因には、システム開発、データ収集・統合、セキュリティ対策、プライバシー保護、分析・モデリング、コミュニケーション、トレーニング、クラウドサービスが含まれる。

試算結果

※10万アカウントを想定

追加コストの発生も予想されるため、柔軟な予算計画が必要。

項目	内容	想定コスト
開発費用	システム開発/導入費用	1.5億円~2億円
	データ収集と統合費用	
	データセキュリティ対策費用	
	プライバシー保護対策費用	
分析・モデリング費用	分析およびモデリング費用	5000万円~1億円
運用コスト(4~5年)	コミュニケーションおよび報告費用	1.5億円~2億円
	トレーニングとサポート費用	
	クラウドサービス費用	
	その他の経費:	

受益者負担のコスト概要

- ・二次利用事業者は、データ利用ライセンス料、API利用料、データ提供者への手数料、カスタム機能追加の費用、データセキュリティ対策費用、トレーニング費用、サポート契約費用など、特定のコストを負担する可能性がある。
- ・負担額は事業利用者の利用量やニーズに応じて変動し、具体的な分担額は交渉によって決定される。

提案と展望

- ・投資の正当化には、データベースが教育分野にもたらす改善点や効率化、長期的な利益の提示が必要。
- ・初期投資の回収期間、教育の質の向上、アクセスの容易さ、データ利用範囲の拡大など、長期的な効果の見込みを提示。
- ・技術進化や政策変更への適応、継続的な改善とアップデートの重要性を強調し、教育データの二次利用促進と教育質の向上に寄与するデータベースの構築と運用の価値を強調。

本システム構成調査にご協力頂いた企業
 ・大日本印刷株式会社 / TIS株式会社 / 株式会社ミルウス

教育データの二次利用が教育改善に大きな潜在力を持つが、安全性・倫理性の確保とステークホルダー間の連携が成功の鍵である。

調査結果の要約

教育データの潜在性

教育データの二次利用は、**学習成果と教育方法の革新的な改善**を可能にする。この潜在性を解き放つことは、パーソナライズされた**学習体験と教育政策の質の向上**に繋がる。

必要なシステム基盤

効果的かつ倫理的なデータ利用のためには、安全かつ堅牢な**データベースが絶対に必要**。信頼性の高い基盤こそが、教育データの価値を最大限に引き出す。

ステークホルダーの視点

教育者、技術者、政策立案者はデータの価値を高く評価しているが、**プライバシー保護とセキュリティ確保には特に注意が必要**。

提案

セキュアなデータベース開発

本調査では、教育データの集約、匿名化、分析をサポートする、**セキュアでスケーラブルなデータベースの設計と実装を提案**する。

倫理ガイドラインの確立

教育データの倫理的利用を確保するため、**プライバシー保護、同意プロトコル、透明性のあるデータ利用ポリシー**を含む、**明確なガイドライン**を作成する。

参加や関わりを深めること

教育コミュニティ全体との継続的な対話を通じて、データベースの開発を進め、**児童生徒、保護者、教育者、政策立案者のニーズ**に応える。

今後の展望

データベースの進化

教育データの真の潜在能力を解放する鍵は、**データベースの継続的な実装と改善**にあります。技術の進歩と共に、本調査は教育の質を根本から変えることができる。

技術革新

AIやブロックチェーンなどの新技術の探求は、データベースの機能、セキュリティ、アクセシビリティを向上させ、**教育データ利用の新たな可能性を開く**。

法令やガイドライン

教育データを倫理的かつ効果的に利用するための法令やガイドラインの立案には、**立法・行政機関との協力が不可欠**。これにより、革新とプライバシー保護のバランスを見つけ出し、教育データの未来を安全かつ効果的に形成する。データ駆動型の教育によって、個別化された学習体験と、教育の全レベルでの情報に基づいた意思決定が可能になり、教育の未来が大きく拓ける。

デジタル庁
Digital Agency