

AIとデータのエコシステム実現を目指して

AI データ活用コンソーシアム

AI Data Consortium

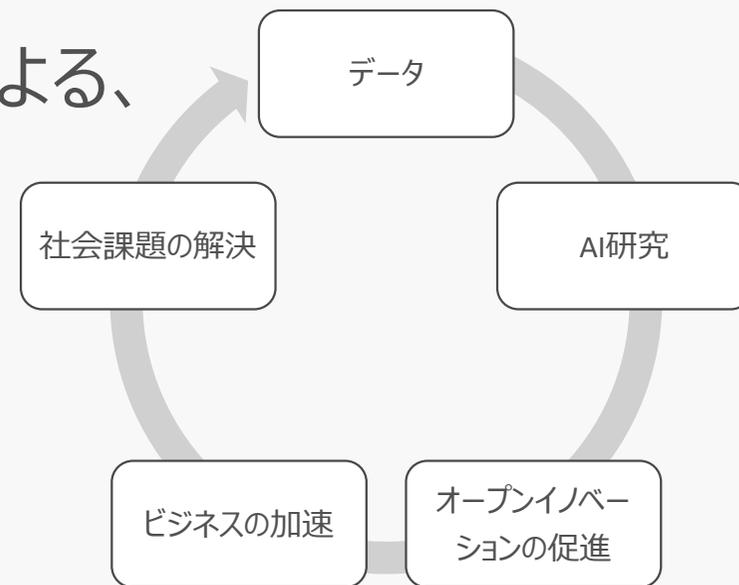
日本マイクロソフト株式会社

業務執行役員 NTO 田丸 健三郎

エコシステムの欠如

細分化されたデータ提供者、異なるライセンスの考え方、個人情報、計算リソースとの連携など、AI研究、オープンイノベーション、そしてソリューション化（商用化）には解決すべき多くの課題が存在

- データ提供、データ活用における契約プロセスの標準化とデータ共有基盤の構築
- 個人情報・プライバシーに配慮したデータ活用モデルの実現
- 多種多様なデータのストア、データの商取引の実現による、持続可能なデータ活用基盤の実現が重要



国内における課題（データ）

- AI研究およびビジネス活用を前提とした業種を超えた多様なデータの共有、取引基盤の欠如
 - 先行する米国、中国
- 圧倒的に不足する日本固有のデータ（海外データによる偏ったAIモデル）
 - 多様性に富む日本語のデータ（自然言語データ）
 - 地域差、年代差、性差、業界差
 - 日常会話、ビジネス会話
 - 日々増え続ける固有名詞、変化し続ける用法
 - 日本文化固有の画像データ
 - 海外で作成された画像データセットは日本固有の物を含まない（茶碗、鳥居、だるま、赤べこ、けん玉）
 - 日本の地域固有データ
 - 道路交通、気象、地理、電力・水道、人口動態

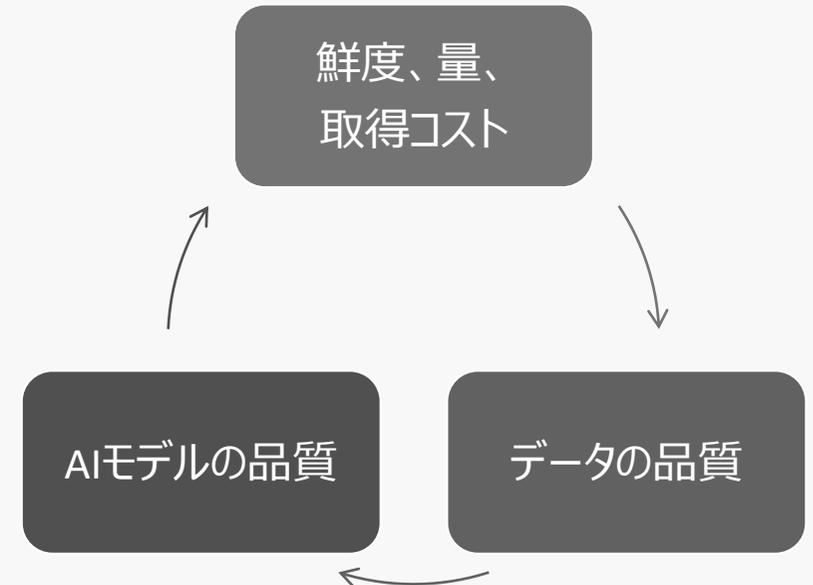


ASEAN特許審査協力
(ASPEC)



国内における課題（商流）

- 現在はデータホルダー個別に費用、契約などの交渉、手続きが必要。
多様なデータを複合的に必要とするAI研究を阻害
- データの商取引におけるモデルの欠如
 - AI活用を前提とした商取引のガイドラインが十分に整備されておらず、また事例も非常に限られることから、公正な取引環境の実現が求められる。
- 価値の確定までの時差
 - 「物」と異なり、保証が難しいデータの品質



国内における課題（連携）

データを持つ企業の課題解決と研究促進の相互補完を実現

企業

- データはあるが、AIで紐解くことができない



研究機関・AIスタートアップ

- AIの知見・頭脳はあるがデータがない

『コンプライアンス確認を目的とした膨大な顧客との会話（音声）データがある。このデータを活用して年齢、地域（方言）に依存しない高精度な音声認識を実現し、コンプライアンス確認の自動化を実現したい。』

AI研究に寄与できるデータ共有基板があれば、保有データを開放したい。

『レジで使用するポイントカードの膨大なデータがある。AIによる様々な可能性について知見を得たい。』

AIの研究にも役立てることができるのであれば、データを開放し、国内のAI研究に貢献したい。我々は、研究成果を有償で使用させてもらえれば問題ない。

『国語研究、辞書編纂資料など膨大なデータがデジタル化されることなく、破棄されつつある。膨大な資料を自然言語研究に活用してほしい。』

膨大な日本語資料を現代のAI技術を用いて分析できないか。
様々な自体に対応した文字認識技術の促進。

個人情報の扱い、プライバシー侵害リスク、ビジネス活用とライセンス問題（契約・課金）

増え続けるデータと求められるリアルタイム性、難しい利用条件とビジネス活用

分散した元データ。時間的猶予が限られる貴重なデータとアノテーション。

データとAIと社会課題

重要性を増す複数のモダリティによるモデル構築

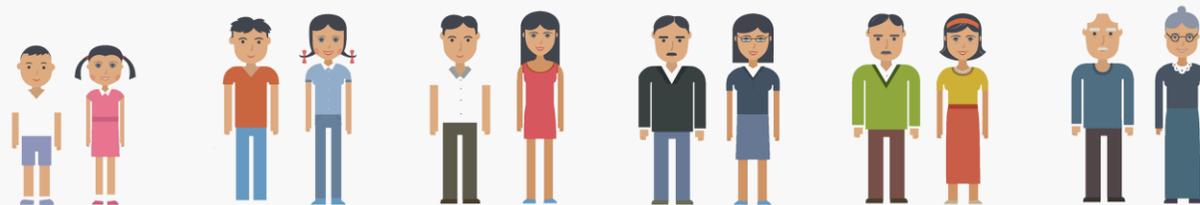
- マルチモーダル深層学習で重要となる複数の異なるデータ
- ヘテロジニアスなデータによる複合モデルの構築
- 継続的な更新が必要なデータ

テキストデータ、音声データだけでは「やばい！」の意味を評価できない。行動データとの組み合わせで初めて「やばい！」の評価ができる。



やばい！
このラーメンやばいよ！

美味しい！
このラーメン美味しいよ！



変化し続ける言葉。10年前の10代と現在の10代では使用する言葉が大きく異なる。AIモデル品質の維持、向上には継続的なデータ収集、更新が欠かせない。

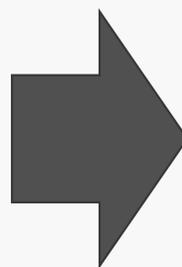
ヘテロジニアス データとAIによる都市課題の解決

気象、交通、モバイル空間統計、GISなどの様々なデータによるモデル構築とサービスの提供を実現



東京都下水道局『東京アメッシュ』
数分後の降雨状況の予測を可能に

<http://tokyo-ame.jwa.or.jp/>



都市の動態モデルによる人・物の予測

混雑前の移動（時間差移動）、混雑した経路の回避

AIによる社会課題解決への貢献

様々なデータの提供、共有基盤の構築を通して、AI研究と社会課題の解決のための取り組みの接続を実現。



脳性麻痺、聴覚障害者の話す言葉を認識できるASR（Automatic Speech Recognition）エンジンは存在しない



支援者の多くは、発話困難な障害者の話す言葉には共通点があると感じている。



会話データの収集と最新のAI研究による発話困難な障害者のコミュニケーション、社会参画を支援

重要な3つのテーマ

求められるAIにおけるデータ活用における課題の検討および解決

知的財産

学習で使用するデータ、AIにより生成されるデータの権利関係の検討。

学習結果であるモデルの知財をどのように考えるのか。また、生成系アルゴリズムによる学習データと生成されたデータにおける権利関係の検討

法令

海外で進むゲノム、個人情報を含むデータ活用における法令、ガイドラインの調査、および検討。

国際的に競争力のあるAI研究、およびデータ活用における課題の検討、提言、法令解釈について作業を行い、AIにおける国際競争力強化に貢献。

データ共有基盤

多種・多様なデータを共有する上で求められるサービス基盤の検討。

完全にオープンなデータから、機密性の高いデータ、NDAなどの契約が必要となるデータ、一部有償のデータをホストする上で求められる基盤の検討。また、データ活用における利便性についても検討。

契約の標準化、関係省庁との調整

想定される知財、個人情報、プライバシーなどを考慮した手続きの確立。円滑なデータ、AIモデル取引基盤の実現

データ収集・活用

基盤となるデータ収集、およびデータホルダーリレーションの構築。AI研究、イノベーションによるビジネス活用の推進。

AIモデル、データ共有・取引基盤の構築

多様なトランザクション要件を満たした、多様なデータのストア・共有、AIモデル実行基盤の構築

「AI・データ契約ガイドライン」に基づく手続き、サービスの提供

「AI・データ契約ガイドライン」で想定している4つの視点と3つの契約形態に対応したデータ共有基盤の構築

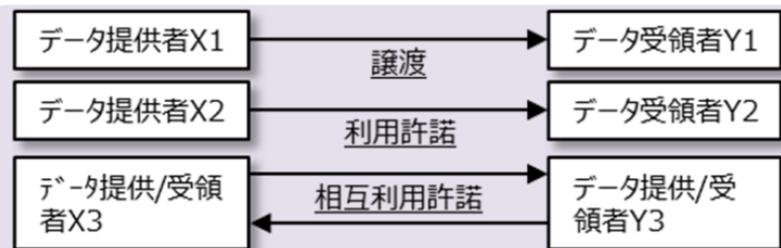
【4つの基本的視点】

- ① データ流通・利活用の重要性と課題：データは保有するだけでは大きな価値がなく、利用する方法を開発することで価値が創出される。契約に際しては、データの利用権限および発生した利益を、適切に分配することが重要。他方、データの流出や不正利用のリスクへの配慮も必要。
- ② 契約の高度化：本ガイドラインは、あくまで契約で定めておくべき事項を示すことにとどまる。したがって、契約当事者が協議し、本ガイドラインを参考としつつ、データの創出や利活用に対する寄与度等を考慮し、取引の実状に応じて契約を高度化させていくことが望ましい。
- ③ イノベーションの促進：多様な立場に配慮したデータ契約の考え方や契約条項例等を示すことにより、データ利用の促進を図り、オープン・イノベーションを促進する。
- ④ 国際協調：クロス・ボーダー取引が一般化する状況下、データの越境に関する問題も考慮する。

経済産業省『AI・データの利用に関する 契約ガイドライン』

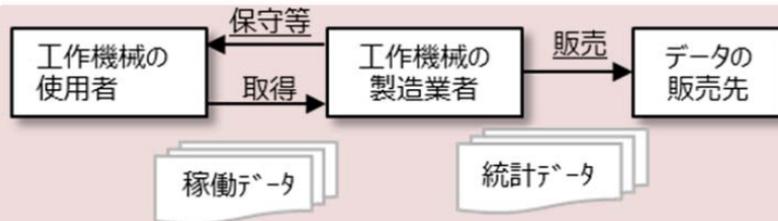
契約類型を3つに整理し、それぞれ、構造・法的性質、課題、法的論点、適切な取決め方法、モデル契約書案等を整理

「データ提供型」契約



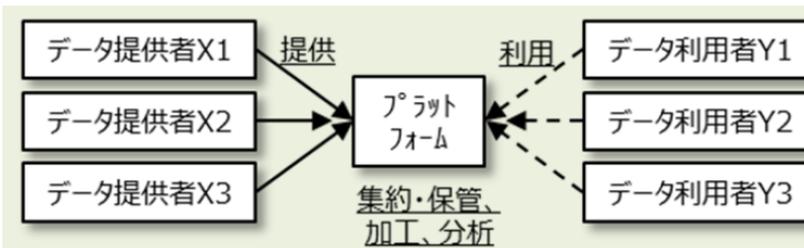
・データ提供者から他方当事者に対してデータを提供する際に、他方当事者のデータ利用権限等を取り決める契約(対象データをデータ提供者のみが保持しているという事実状態が明確である場合)

「データ創出型」契約



・複数当事者が関与することによりデータが新たに創出される場面において、データ創出に関与した当事者間で、当該データの利用権限を取り決める契約

「データ共用型」契約



・プラットフォームを利用したデータの共用を目的とするタイプの契約

AIとデータの統合基盤

Data



AI

Dynamic Web API / Notification (e.g. Webhook)



Data Store and Computation

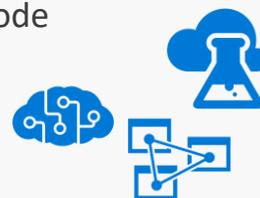
Dynamically Optimized Data Store

- Optimized for
 - ✓ Cost
 - ✓ Latency
 - ✓ Bandwidth
 - ✓ Utilization
 - ✓ Data Type
 - ✓ Size
- Versioning
- Change Tracking
- Event Based Data Factory



Computation Node

- **Sandbox**
- Dynamic Training
- Dynamic Model Update
- Dynamic Biz Logic Update
- Dynamic Visualization
- Flexible Experiment (e.g. Jupyter Notebook)
- ML/DL Flexibility
 - From traditional ML to latest DL algorithms



ID and Management

Security

- Authentication
- Access Control
- Logging
- Role Management
- Auditing



Licensing

- Subscription
- Contract Management
- Metering
- Cost Charge Management



Resource Management

- Provisioning
- Catalogue
 - Data
 - AI/ML Model
 - Tools
- Classification
- Ownership

Authority Separation (inc. Organization): Resource, Permission and ID Management

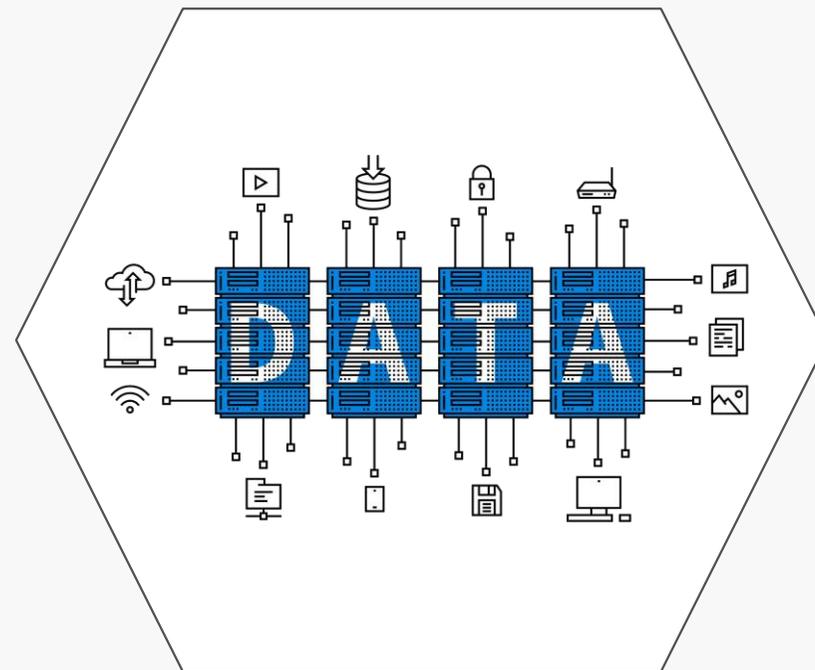
契約手続きを媒介：手続きの簡素化



データ流通基盤との連携と組織内活用



AIデータ活用コンソーシアム



コンソーシアムが実装するデータ流通基盤を
組織内データ共有基盤としても活用可能に
(会員に無償ライセンス)

研究者

データ
ホルダー

ITベンダー
企業ユーザー

会員間連携による課題解決（マッチング）



AI
スタートアップ

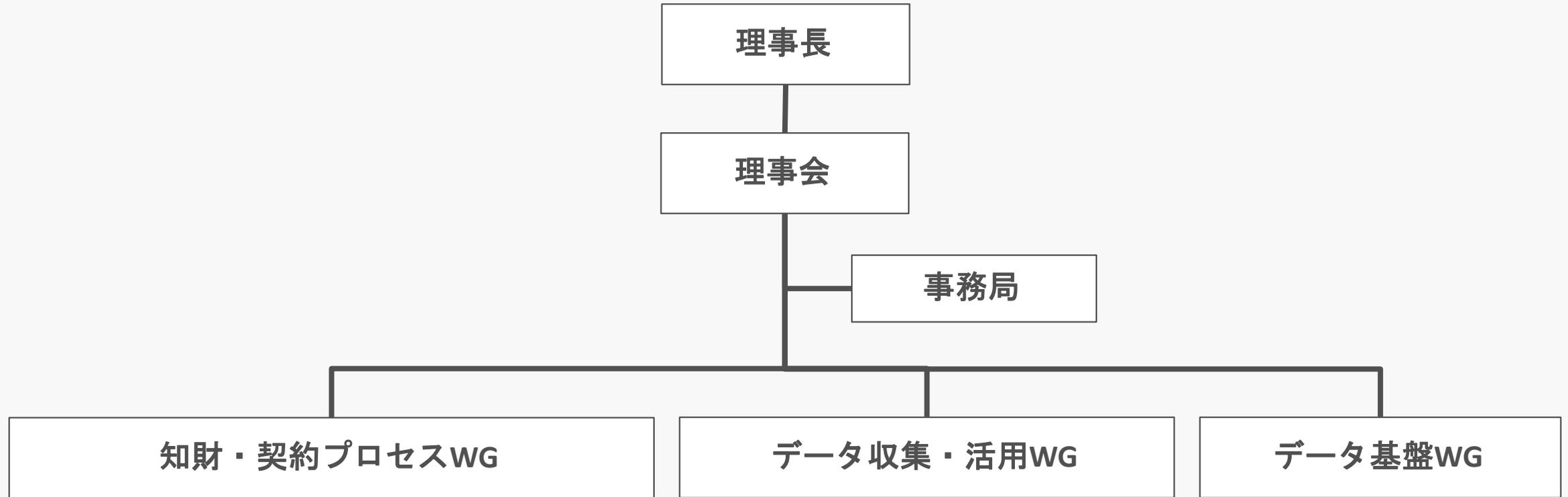
ユーザー
企業



研究者



組織体制



■ データ提供者、データ活用者、およびデータによる生成されたAIモデル利用者間における知財、ライセンスの考え方を整理し、AIオープンデータ協議会が提供するデータ共有基盤を通して信頼できる手続きを実現する。

■ データが含む個人情報の扱い、プライバシーへの配慮について検討を行い、契約およびデータ基盤で求められるサービスを明確にする。

■ データ提供者と連携し、当データ基盤がイノベーションおよびビジネスに寄与するための計画の策定、実施を行う。当活動を持続可能なものとする為の施策を実施する。

■ データが含む個人情報の扱い、プライバシーへの配慮について検討を行い、契約およびデータ基盤で求められるサービスを明確にする。

発起人案

- 長尾 真
元京都大学総長 / 元独立行政法人情報通信研究機構理事長 / 前国立国会図書館長 / 前国際高等研究所所長
- 坂村 健
東洋大学情報連携学部INIAD学部長 / cHUB（学術実業連携機構）機構長 / YRPユビキタス・ネットワークング研究所 所長 / 公共交通オープンデータ協議会 会長
- 渡部 俊也
東京大学 政策ビジョン研究センター 教授/副センター長
大学執行役・副学長
工学系研究科技術経営戦略学専攻 教授（兼務）
- 杉山 将
理化学研究所 革新知能統合研究センター センター長、
東京大学 大学院新領域創成科学研究科 教授
- 井佐原 均
豊橋技術科学大学情報メディア基盤センター センター長・教授
- 前川 喜久雄
国立国語研究所 音声言語研究領域 教授
- 安岡 孝一
京都大学 人文科学研究所 教授
- 藤原 洋
株式会社ブロードバンドタワー 代表取締役 会長兼社長 CEO
- 清水 亮
ギリア株式会社 代表取締役社長
- 中 俊文
野村證券 営業企画部 営業企画課 課長代理
- 石井 靖乃
公益財団法人 日本財団 ソーシャルイノベーション部 公益事業部 部長

- 小野 雄次郎
一般社団法人 日本支援技術協会
- 吉川 健一
株式会社ブリックス 代表取締役社長
- 田丸 健三郎
日本マイクロソフト株式会社 業務執行役員 NTO
- 柳原 尚史
株式会社 Ridge-i 代表取締役社長
- 下川 和男
イースト株式会社 代表取締役会長

協働

- 公共交通オープンデータ協議会

協賛：

- Deep Learning Lab

アドバイザー：

- 東京都
- 文部科学省
- 総務省
- 経済産業省
- 国土交通省

スケジュール

2019

1月

- 設立準備会合、組織概要の策定
 - 目標
 - 活動
 - 規約
 - 組織
- 発起人会議

2月

- 一般社団法人として登記
- 設立総会
- 記者説明会

3月

- 作業部会の設置

4月

- 実作業開始
 - データホスティングデザイン、実装開始
 - クラウド環境構築

『AIデータ活用コンソーシアム』

お問い合わせ先：
info@aidatacon.jp