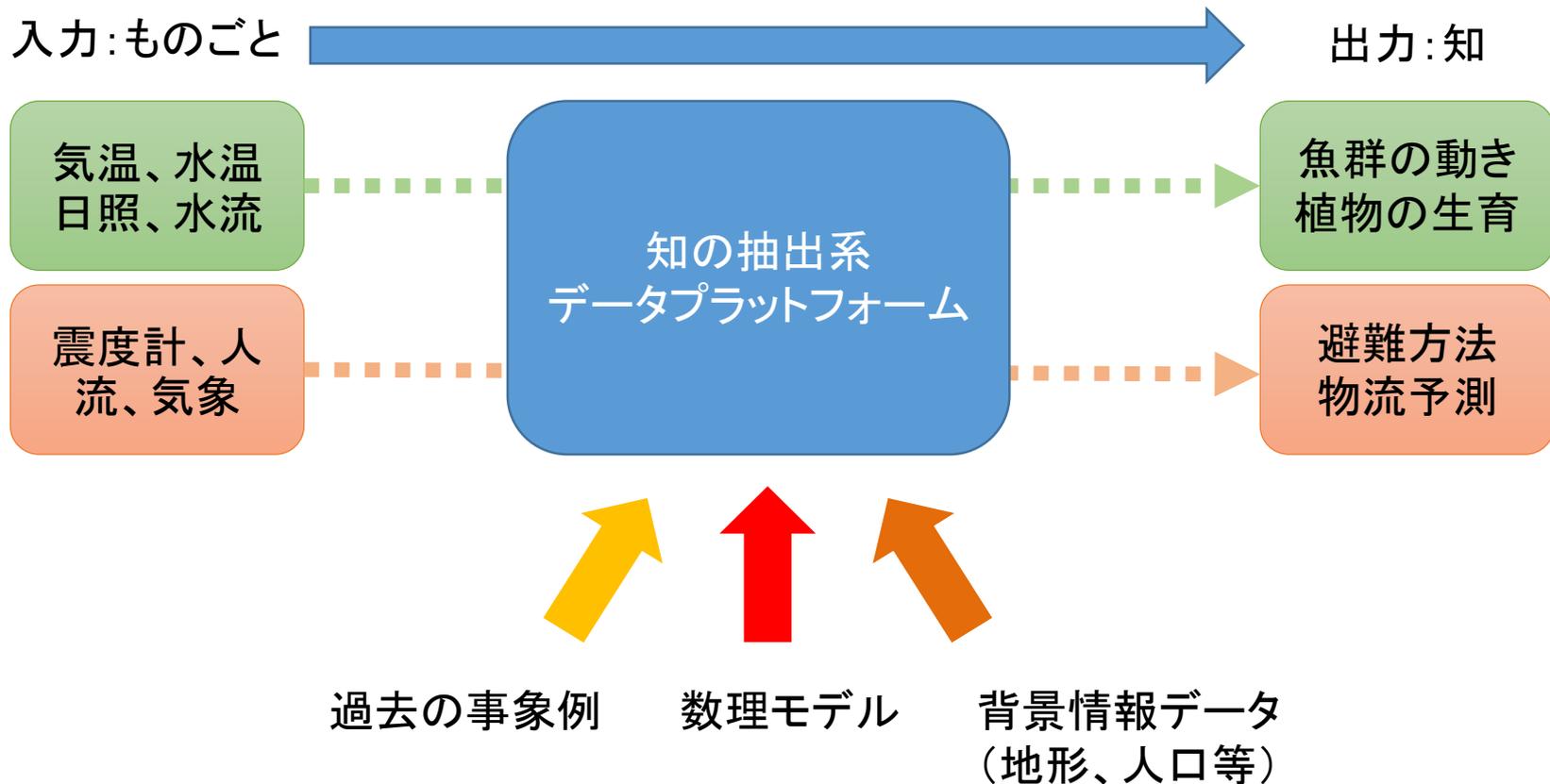


東京大学情報基盤センター 教員公募 別紙

学術高速大容量ネットワーク拠点 ～データプラットフォーム(仮称)～の構想

知識集約社会を支える「知」の抽出

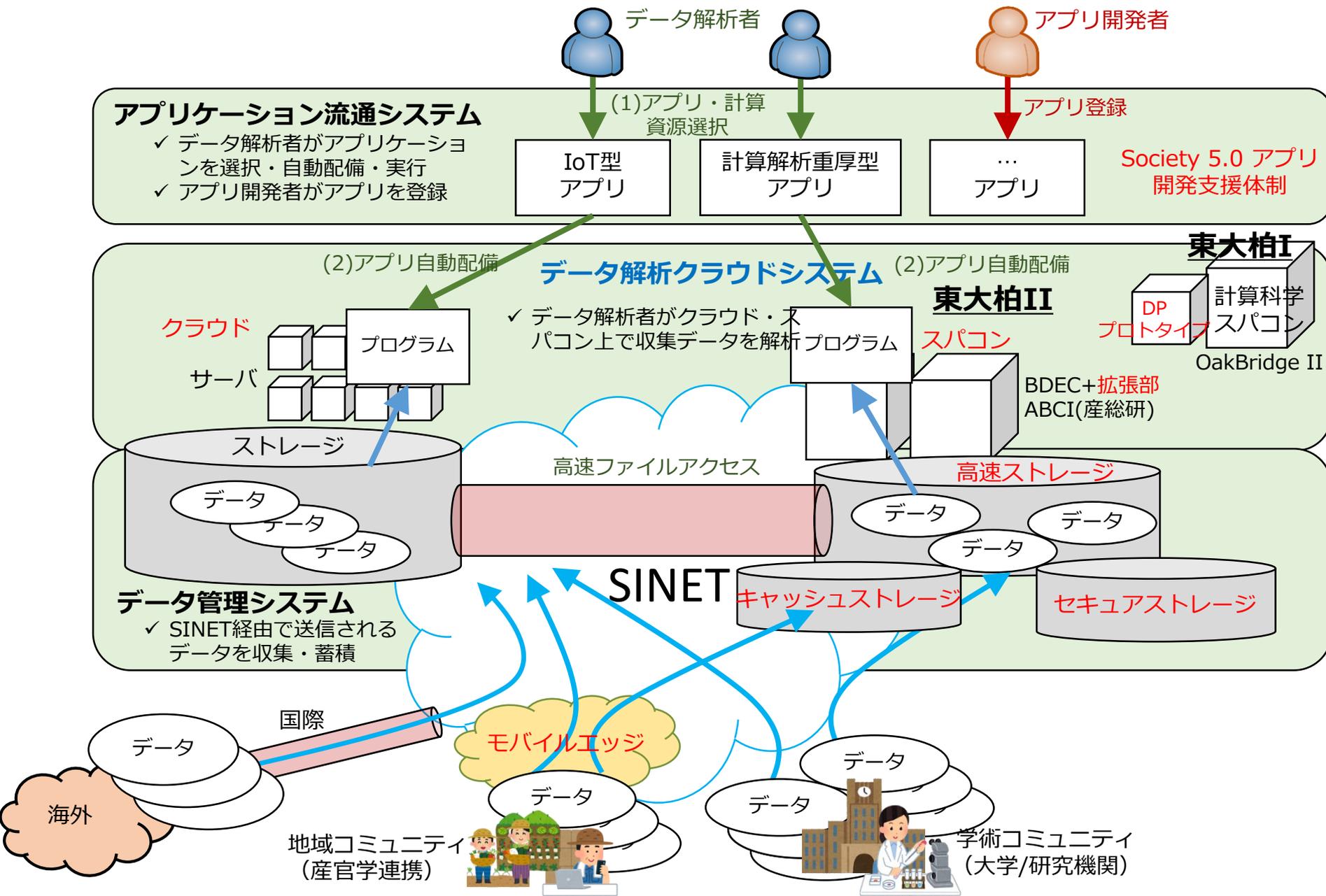
- データと知識に基づき、ものごとの結果起きることの予測や最適な判断（知）を提供し、産業活性化や防災などにつなげる



データプラットフォームとは

- 「知」の抽出により、Society 5.0における「ビッグデータ等の新たな技術をあらゆる産業や社会生活に取り入れてイノベーションを創出」するための「場」
 - アイデアはあっても実際にデータを収集・集積・解析する仕組みを作ることが困難な（地方）大学や産学連携体に「場」を提供、「知」の抽出を容易に行えるようにする
 - 短時間でのプロトタイピングを可能にし、機を逃さないシステム構築を支援する
- データプラットフォームの3本柱
 1. **SINETを活かしたリアルタイム収集・集積・解析環境の動的な構築**
 - 遠隔地のセンサーやストレージ、データプラットフォームの計算資源、ストレージをつないで、リアルタイムに入力から出力を得られるアプリケーションごとの収集・集積・解析環境（仮想データプラットフォーム：仮想DP）を、使いたいときに即時に構築する
 - SINETモバイル基盤によりセンサー等のデータを安定してセキュアにつなぐ
 2. **高性能計算環境によるデータ科学と計算科学の融合**
 - データ科学、計算科学の手法を融合し、さらに国内最高の計算環境を用いて他に無い高精度の予測を行えるようにする
 3. **開発支援とコミュニティの形成**
 - データ科学、計算科学、活用可能データなどの知見を基にコンサルティング・開発支援を行う
 - データ保持者、データ利用者等のコミュニティを形成し、新たなデータ活用につなげる。

データプラットフォームの概要

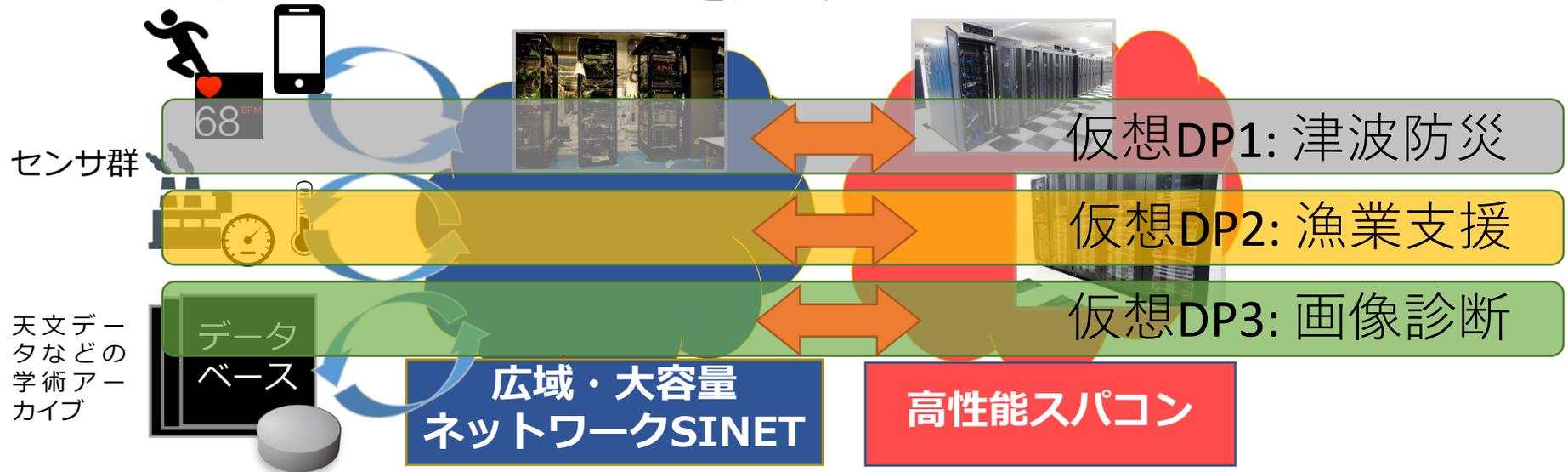


データプラットフォームの機能

- データ収集
 - モバイルエッジなどからSINETを介してデータを集める
 - 解析に必要な大量データを収集するネットワーク環境を提供
- データ集積
 - さまざまなデータ: フィールドから収集するデータ、シミュレーションで生成するデータを集積
 - 高速ストレージ: 大量のデータを格納
 - セキュアストレージ: セキュリティが必要なデータを格納
 - キャッシュストレージ: リアルタイムに集まるデータを欠落無く格納
- データ解析
 - データ科学、計算科学を融合してデータを解析
 - 国内最高レベルの計算機環境を提供

SINETを活かしたリアルタイム収集・集積・解析環境の動的な構築

- 共通（汎用）の基盤（スパコン、ネットワーク、ストレージ等）の上に、アプリケーションごとのリアルタイム収集・集積・解析環境である**仮想DP(仮想データプラットフォーム)**を、**オンデマンド**に構築
- 仮想DPは、**性能が確保**されている。**他と隔離**されているため安全性も担保される。
 - リアルタイムに大量のデータを処理することができる

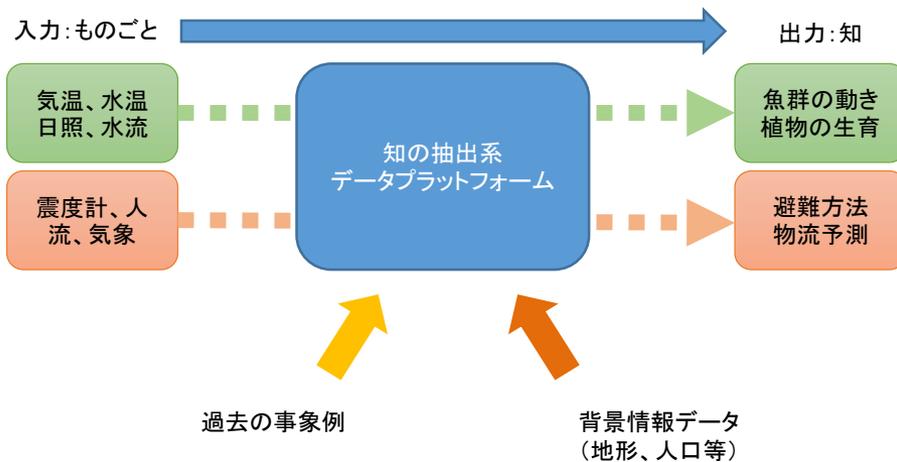


- 仮想DPはアプリが増えるのに応じて次々に**増やしたり**、需要に応じて**大きさ・構成を変えたり**、複数の仮想DPを**融合したり**することができる。
- 作りたいときにすぐに仮想DPを構築することで、**高速なプロトタイピング**ができる

時間勝負の競争に勝つために不可欠

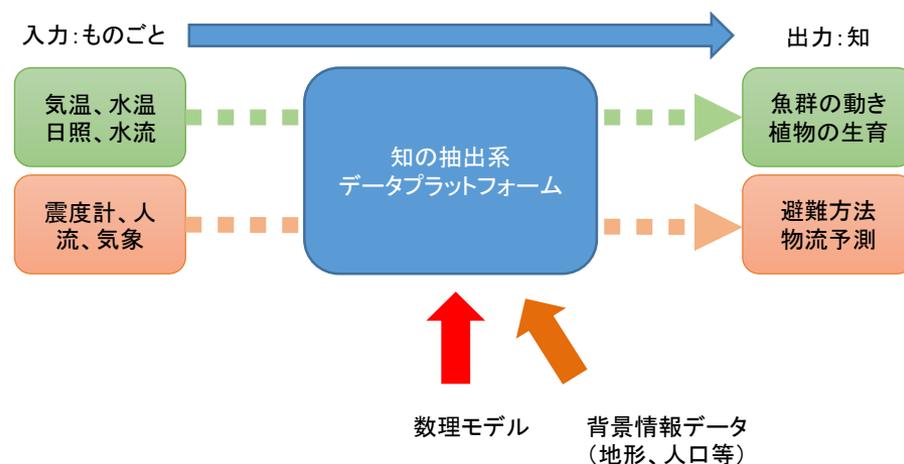
データ科学と計算科学

- データ科学（機械学習など）
 - 大量の事象の情報（教師データ）を基に、帰納的に結果を予測
 - 教師データの量、正確さ、分散が重要である。
 - 教師データから大きく外れた入力に対応できない。
- 計算科学（シミュレーションなど）
 - 数学的モデルに基づいて、物事の過程や結果を予測
 - 仮定した条件について、シミュレーションにより結果をある程度正確に予測可能
 - モデルが無いと計算できない。個々のケースの予測に大量の計算が必要な場合も



データ科学

- 過去に類似例がない事象に対応できない



計算科学

- モデルが必要、計算量が大きい

高性能計算環境による計算科学とデータ科学の融合

- 計算科学での計算量をデータ科学により削減する
 - 例：災害発生時に、リアルタイムでシミュレーションしても計算が間に合わないとき、一部をデータ科学で予測して結果を早く得る
- データ科学に必要な教師データを、計算科学によって生成する
 - 例：過去に観測データが得られていない大きな災害時に何が起きるのかをシミュレーションにより計算し、機械学習の教師データとする
- 計算科学による予測とデータ科学による予測を組み合わせる
 - 単純なモデルに従う物理現象では計算科学は正確
 - 人の行動予測などでは、データ科学が有利
 - 両者を組み合わせることにより、より複雑な事象の予測ができるようになる。

開発支援とコミュニティの形成

- Society5.0 アプリ開発支援体制
 - データ解析のためのアプリケーションの開発支援
 - データ解析手法に詳しくない利用者に、何をどのように使えばよいかをコンサルティング
 - データ科学と計算科学の融合を実現
 - 既存・新規開発ソフトウェアにより利用者のニーズを満たす
- データ保持者、解析者のコミュニティを形成
 - 様々なデータを複合的に利用するには、ウェットな人のつながりが重要
 - -どの部分で競争し、どの部分で協調するかが要になる
 - コミュニティを形成することで、データ活用を広げる。