



# データプラットフォーム構想について

東京大学

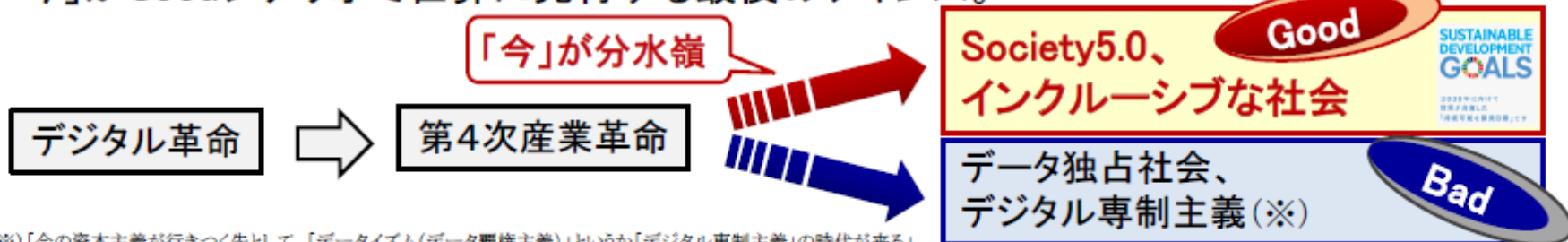
未来社会協創推進本部

データプラットフォーム推進タスクフォース副座長

中村宏

# 「Goodシナリオ」実現への未来投資戦略

- 「今」がGoodシナリオで世界に先行する最後のチャンス。



(※)「今の資本主義が行きつく先として、「データイズム(データ覇権主義)」という「デジタル専制主義」の時代が来る」  
DIAMOND online, 2018.2.27『ダボス会議の今年の主題はAIによる「デジタル専制政治」だった』(経済同友会小林喜光代表幹事へのインタビュー記事)

- ①高度なサイバーインフラ、②公的ビッグデータ、③大学の知識・ノウハウにより、誰もが高度なデータ利活用を素早く、ストレスなく実現できる環境を整備。

## ①オンデマンドの高度サイバーインフラ

IoT接続(モバイル)  
AI特化パソコン  
リアルタイム処理対応パソコン  
高速/セキュアストレージ 等

学術情報ネットワーク(SINET)を活用し、全国各地から利用可能で、高度なデータ利活用すぐに着手できるオンデマンドのデータ収集・通信・解析インフラを整備。

## ②公的ビッグデータとの接続

衛星画像 気象観測 医療介護

公的なビッグデータやリアルタイムデータとSINETでセキュアに接続。分野横断で、戦略的にデータを活用。

## ③知識集約産業のハブとなる大学

アイデア 社会課題

課題発見や高度なデータ利活用を担うAI人材の育成

心 社会システム ソフトウェア システム 超ビッグデータ解析 IoTチップ センサー 材料

一気通貫の知の集積

イノベーター

中小企業・ベンチャー

大学

誰もがアイデアや社会課題を持ち寄り、大学が持つ文理横断の広範な知の集積を活用。高度なサイバーインフラの取扱や超ビッグデータの処理・解析等のノウハウ提供、AI人材育成も担う。各地の大学が知識集約型産業のハブになる。

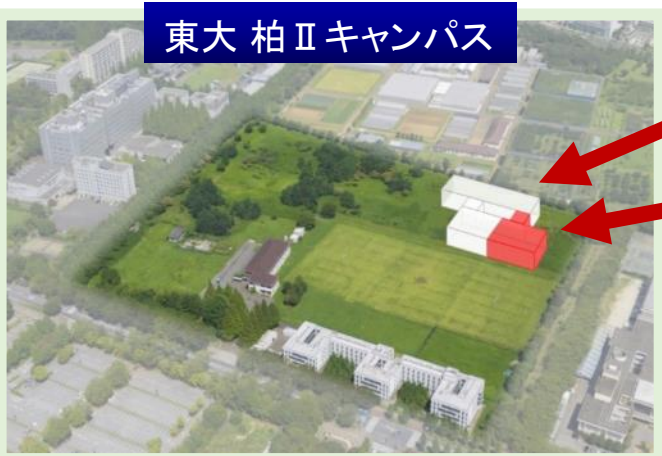
Society5.0を先導する「オンデマンド・データプラットフォーム」

# つくば-柏-本郷イノベーションコリドー

～ 柏キャンパスにおけるトップ研究機関連携と情報ネットワーク整備構想 ～

柏II：知識集約型社会へのパラダイムシフトを意識した先行的な取組を行う拠点として整備  
指定国立構想調書：「知識集約型産業集積形成のための拠点」

東大 柏II キャンパス



【産業技術総合研究所】  
グローバルAI研究拠点(仮称)  
データセンター  
【東京大学】  
産学協創プラットフォーム拠点  
以上 整備中

東大 柏I キャンパス



【東京大学】×【物質・材料研究機構】  
マテリアルイノベーション研究センター

【東京大学】×【国立情報学研究所】  
学術高速大容量ネットワーク拠点整備 構想

東大 柏の葉駅前キャンパス



【東京大学】×【産業技術総合研究所】  
先端計測技術オープンイノベーションラボラトリ

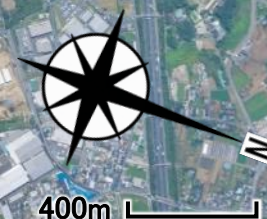
科学警察研究所  
千葉県警察第三機動隊

国立がん研究センター

柏インターチェンジ

常磐自動車道

つくばエクスプレス  
柏の葉キャンパス駅



財政制度等審  
議会・財政制  
度分科会  
(2017/10/4)  
五神総長資料  
を一部改変

# データプラットフォームとは

- 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」
  - 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
  - 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大学・研究機関や産学連携体に「場」を提供、「知」の抽出を容易に行えるようにする

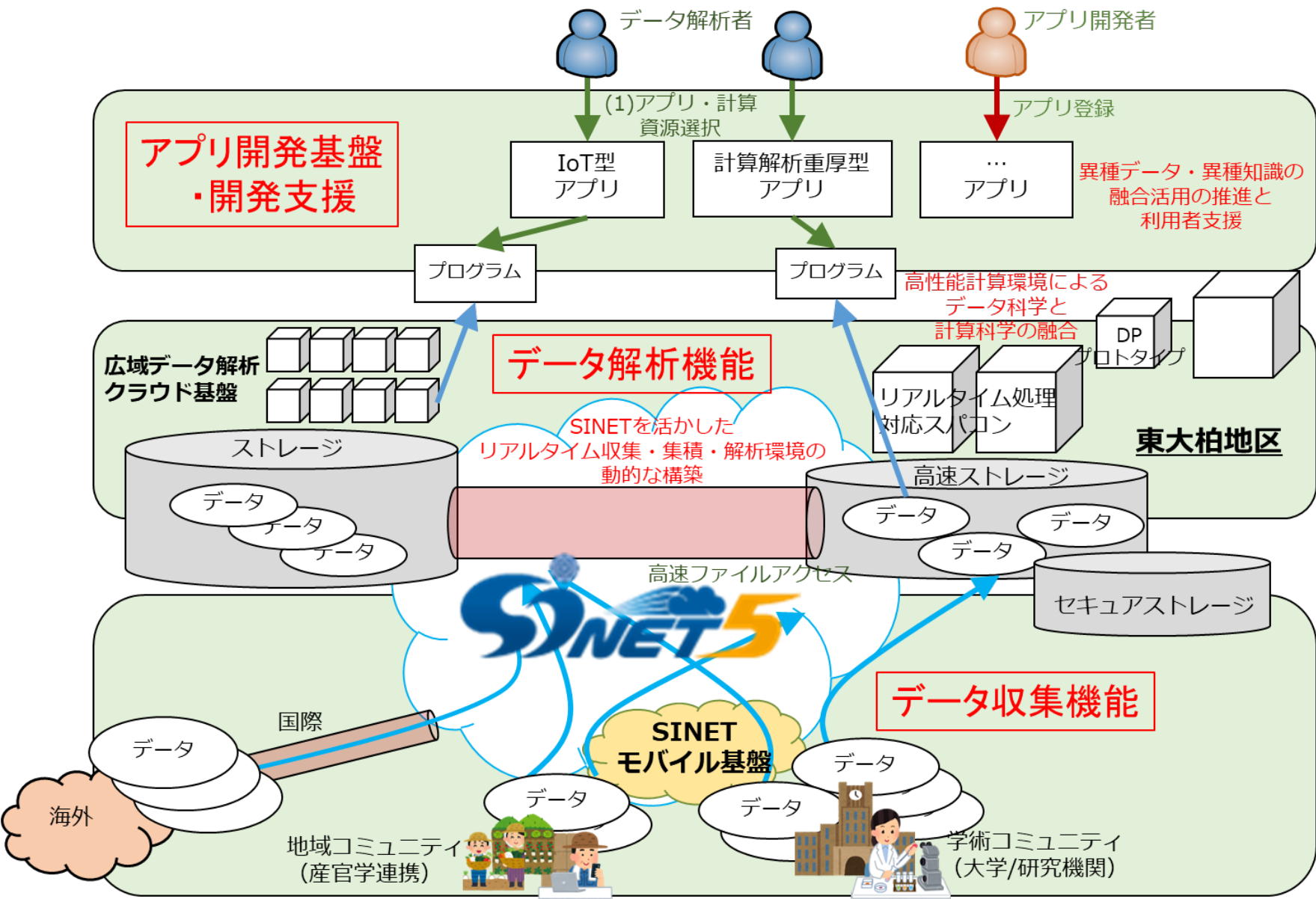
## データプラットフォームの3本柱

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
2. 高性能計算環境によるデータ科学と計算科学の融合
3. 異種データ・異種知識の融合活用の推進と利用者支援

# 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をセキュアにつなぐ
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
3. **異種データ・異種知識の融合活用の推進と利用者支援**
  - 様々な分野のデータ保持者、解析者、利用者が産学にまたがるコミュニティを形成し、新たな価値創造につなげる
  - データ活用を目指す利用者へのコンサルティングや開発支援を実施する

# データプラットフォームの概要



# SINETの優位性

- 任意の拠点間で高性能：全都道府県を100Gbpsでカバーし、大容量データ転送に適した帯域を絞らないアーキテクチャであるため、任意の拠点間で高性能に通信可能
- 通信サービスが豊富：インターネットサービスに加え、セキュアな通信環境を実現する各種VPNサービス、機動的に通信環境を設定するオンデマンドサービス等を利用可能
- 大学等にとって便利：アクセス回線（調達期間は6か月程度）を用意するだけで、多様な通信環境を高性能、迅速、低コストに整備可能

## 商用ネットワーク

インターネットサービス(低性能)

インターネットサービス

VPNサービス(企業内)

一般家庭用

一括で帯域を絞り込み

混んでくると  
超遅い

通信機器

↑

利用帯域

時間

企業用

各企業の最低保証帯域を確保

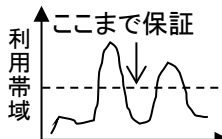
帯域が小さいと  
遅い

通信機器

↑

利用帯域

時間



## SINET

インターネットサービス(高性能)

各種VPNサービス(大学間等)

各種オンデマンドサービス

先端研究のために十分な帯域を確保

通信機器

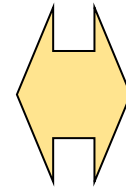
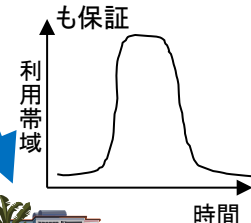
通信機器

大きなピーク  
も保証

↑

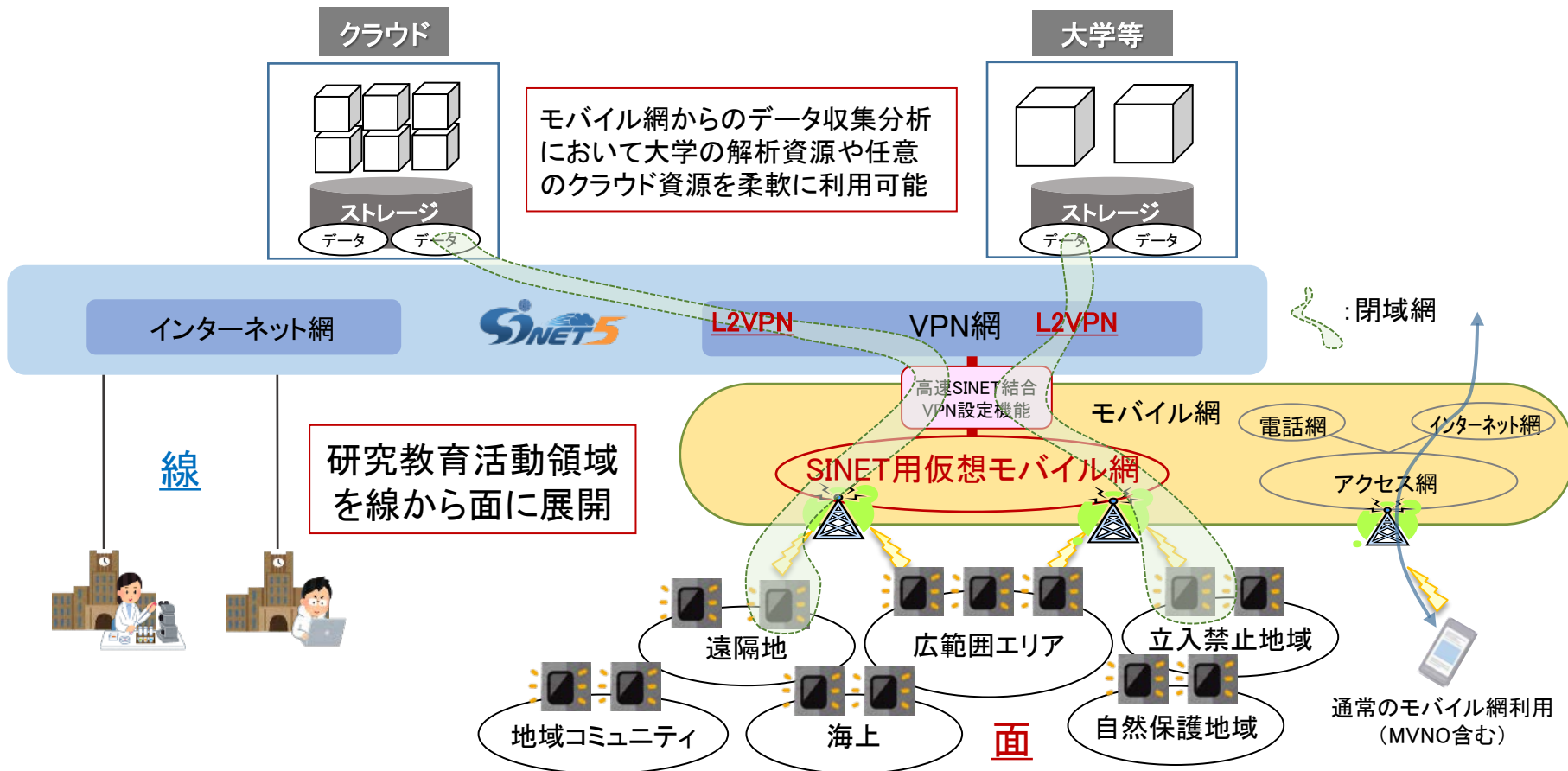
利用帯域

時間



# SINETモバイル基盤の優位性

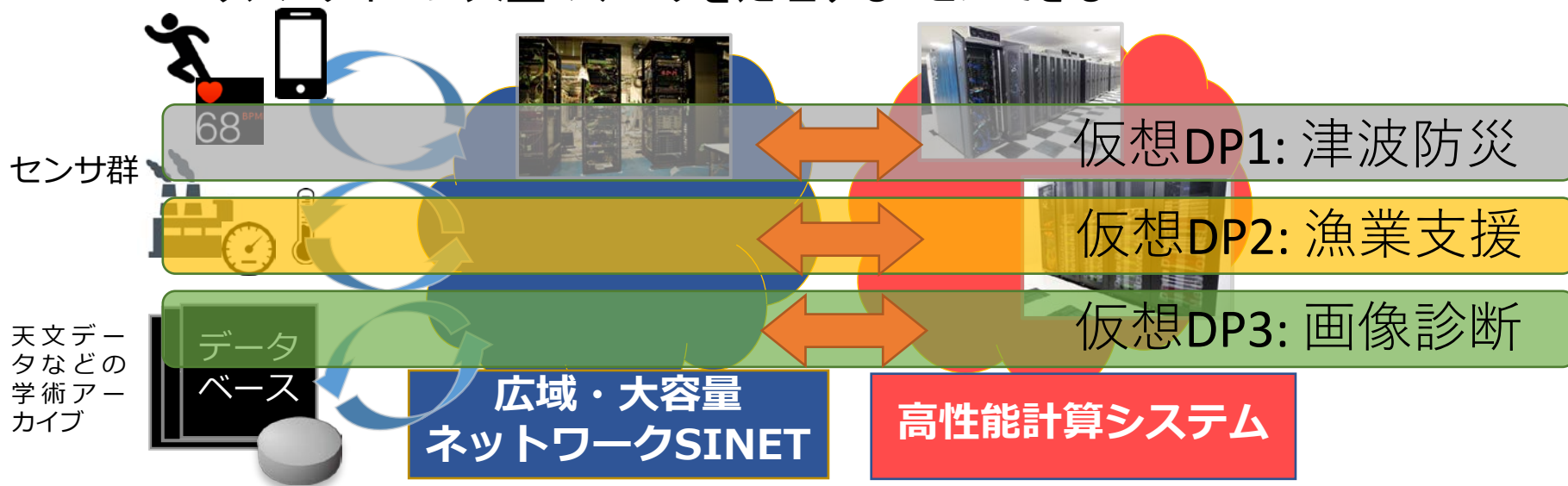
- セキュア：モバイル網の中にインターネットとは切り離されたSINET用の仮想ネットワークを形成し、この中に、研究プロジェクト毎の閉域ネットワークを形成
- 高性能：モバイル網で収集したデータを大学やクラウドの解析基盤に転送するために、SINETと東京と大阪において10Gbpsで接続してL2VPN（高性能かつセキュア）を設定（将来的には接続点の拡大を模索）
- 利用促進：新基盤の検証のために、一定の利用可能データ通信量を準備（公募で選定）





# 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



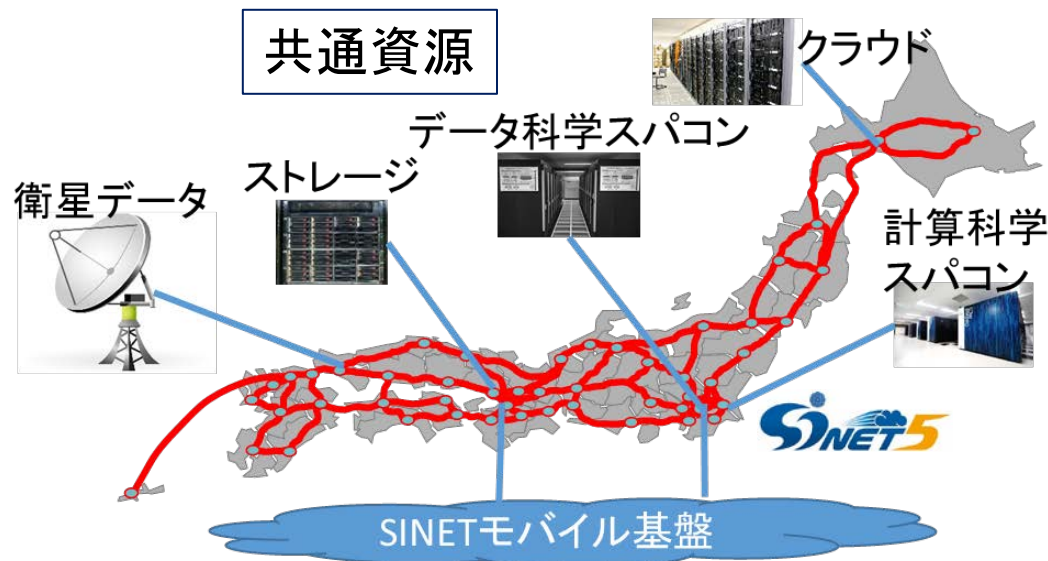
- 仮想DPはアプリが増えるのに応じて次々に**増やしたり**、需要に応じて**大きさ・構成を変えたり**、複数の仮想DPを**融合したり**することができる。
- 作りたいときにすぐに仮想DPを構築することで、**高速なプロトタイピング**ができる

時間勝負の競争に勝つために不可欠

# 「オンデマンド・データプラットフォーム」の構成・機能

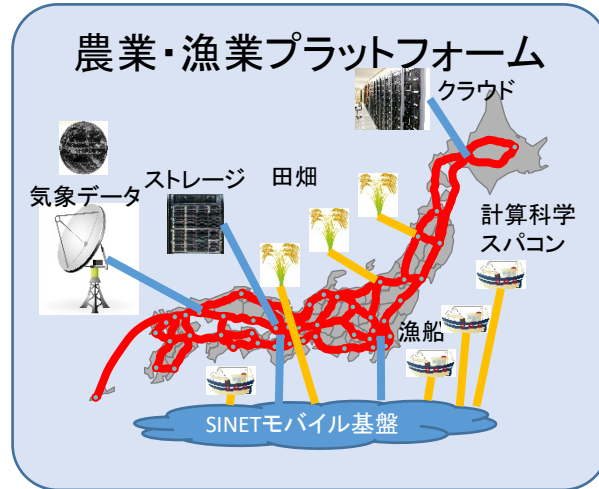
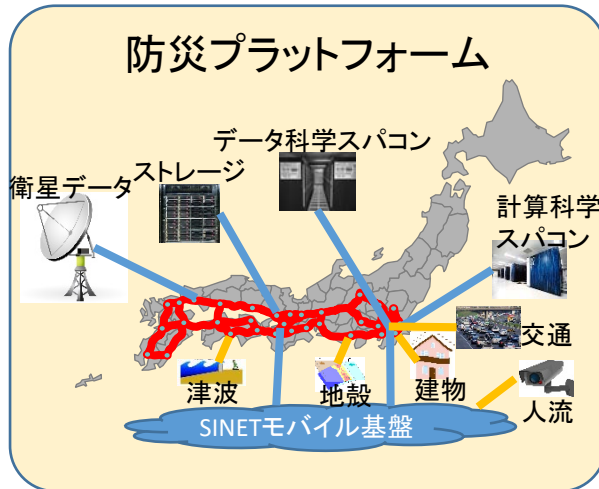
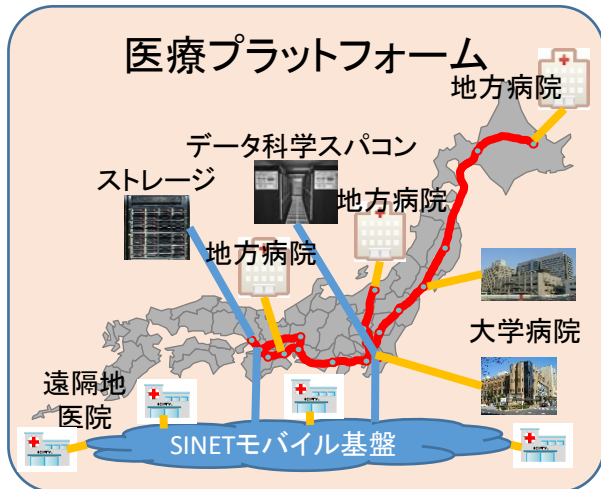
- ① IoTデータや大規模リアルタイムデータを円滑に扱えるセキュアな大容量通信回線【データ収集機能】
- ② 高度な解析を実現する高性能計算環境、ストレージ【データ解析機能】
- ③ 多様な応用を実現する基盤ソフトウェアと共用データ【応用開発基盤】

を一体的に提供。①～③を自在に組み合わせ、様々な応用を実装可能



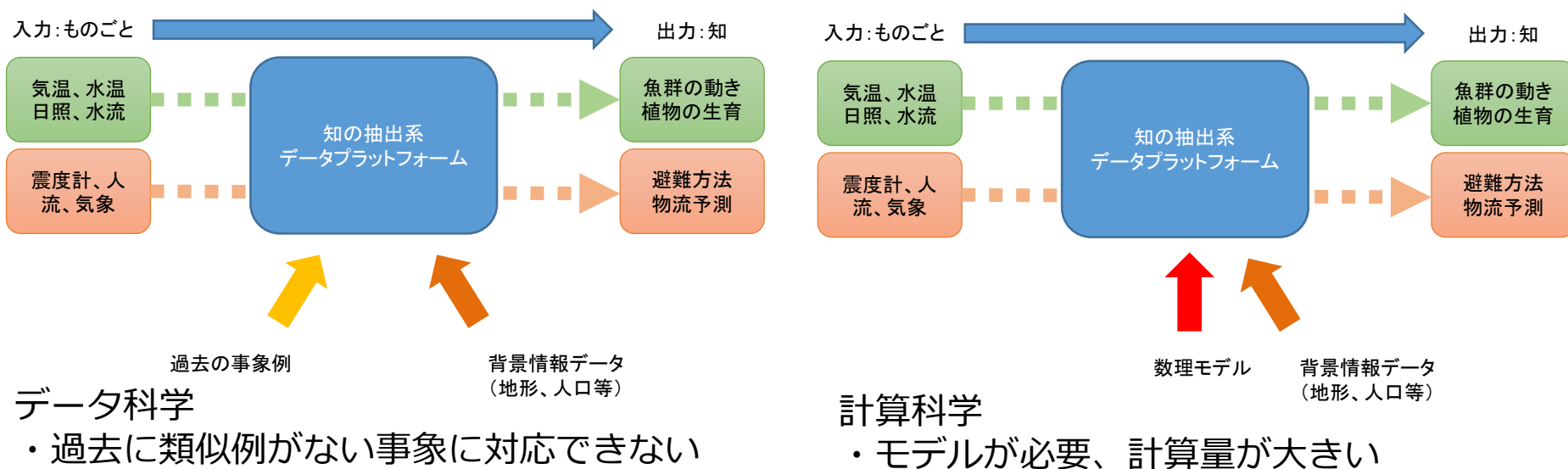
## 応用プラットフォーム

- 応用ごとに必要な機能を用途に応じて利用。オンデマンドで短時間に構築・拡張・融合可能  
自ら設備を持たなくとも誰もがデータ利活用



# 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（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



高い処理能力があっても出力を得られる応用多数

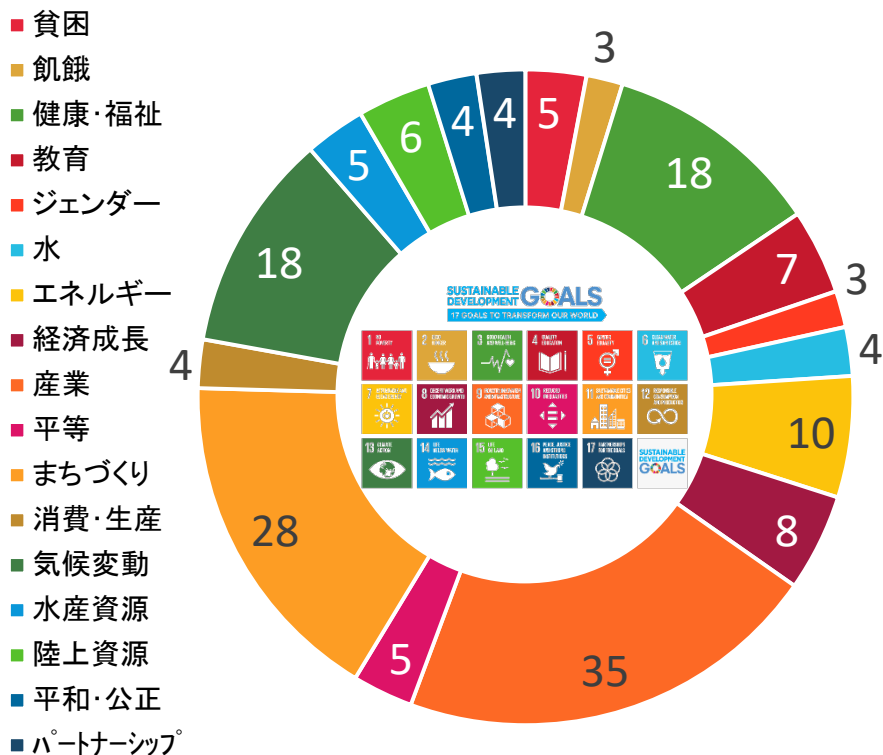
# 異種データ・異種知識の融合活用の推進と利用者支援

- さまざまな分野のデータ保持者、解析者、利用者が産学にまたがるコミュニティーを形成し、新たな価値創造につなげる
  - 多様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
  - コミュニティーを形成することで、データ活用を広げる
- このワークショップの目的
- 利用者支援
  - データ活用を目指す利用者へのコンサルティングや開発支援を実施する
  - 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

# 東京大学におけるデータ利活用型研究への取組

データ利活用により社会課題解決を図るプロジェクトを学内で募集(2018/4/24)  
1ヶ月程で文系含む22の部局から90件登録。SDGsの17目標をすべてカバー

SDGs目標をカバーするプロジェクト数



応用分野の例(件数)

- スマートシティ(14)  
—人流マネジメント、交通・物流最適化 等
- 防災(13)  
—リアルタイム地震波解析・予測、洪水予測 等
- 健康、医療(12)  
—VR・3Dシミュレーション応用 等
- 農林漁業(6)  
—養殖、農業におけるゲノム・環境等データ活用 等
- 地球環境(6)  
—気候予測、風力発電運用高度化、水資源管理 等
- ものづくり(4)  
—構造解析プラットフォーム、暗黙知の解析 等
- 社会科学(11)  
—ライフログ活用、デジタルアーカイブ、政策評価 等

※該当目標の申告があったプロジェクトのみ集計。複数回答。

特徴：応用の多様性、データの多様性  
蓄積データ、観測データ、シミュレーションデータ、それらの組み合わせ  
→ 広域性 + リアルタイム性 + 大量性