

# (コロナ対応、ポストコロナで判明したことを中心とした) 学術情報基盤のあり方について

田浦 健次郎

東京大学情報基盤センター長

学際大規模情報基盤共同利用共同研究拠点長

今後のHPCI第2階層システム検討委員会（取りまとめ係）

# Executive Summary (I)

- 3月～現在にかけて「コロナ対応」のための授業オンライン化が多く大学で行われた
- ZoomなどWeb会議システムの役割が大きいが、多くの大学で導入済みの、学習管理システム（LMS）やGoogle/Microsoftのクラウドも大きな役割を果たした
- 新しいシステムよりも、導入済み教育ICTシステムがこれを機に学内に普及したという側面が大きい
- Web会議、既存教育ICTどちらも「米国パブリッククラウド依存」は鮮明

# Executive Summary (II)

- 今後必要な学術情報基盤
  - データ連携、分散したデータと高性能計算機の連携
  - 組織をまたがった統一的な認証、柔軟なデータ共有
- コロナ感染拡大防止のための取り組み、研究にもデータ共有・連携が欠かせない

# 以降の話

- 授業オンライン化
- そこで使われた情報基盤
- 今後必要な学術情報基盤

# 授業オンライン化について

- 教室授業よりも優れた点が数多く認識された
  - 教員よりも学生が肯定的 [https://www.nii.ac.jp/event/upload/20200605-5\\_Uehara.pdf](https://www.nii.ac.jp/event/upload/20200605-5_Uehara.pdf)
  - 意外なところでは質問がしやすい
- 実施が難しい活動も（当然）ある
  - 試験準備（機密性の高い会議）
  - 試験実施（監視・不正防止）
  - 機材や薬品など登校が必須の実験
- 問題点
  - 友達との交流や情報交換がないことによる不安・ストレス
  - 課題の過多によるストレス

満足している理由は？

- ・移動が不要で効率がいい！！！！（多数）
- ・質問しやすい
- ・自分にあった環境で受けられる
- ・身だしなみにかける時間0
- ・ペーパーレス



[https://utelecon.github.io/events/2020-luncheon/luncheon\\_10\\_slides.pdf](https://utelecon.github.io/events/2020-luncheon/luncheon_10_slides.pdf)

# 来学期（9, 10月～）と将来の授業について

- 来学期は教室、オンライン「混合」（すべてオンラインの選択肢はないため）
- 自宅から出られないストレスはない（と期待される）状況で、「混合」が「すべて対面」より良いかを問われる学期になる
- オンラインの利点を拡大させる準備に
  - 大人数演習での介入自動化・自動採点など教育支援、個別化支援
  - 幅広いコンテンツ共有（学内の学部間、大学間）

# オンライン授業で使われた情報基盤

- Web会議

- Zoom, Webex, Google Meet, Teams, etc.
- Web上で提供されたクラウドサービス

- 共有ストレージサービス

- Google DriveやMicrosoft OneDrive
- Web上で提供されたクラウドサービス

- 学習管理システム（LMS）

- 独自実装, オープンソース（Moodle）など様々
- 提供はオンプレ、IaaSクラウド、クラウドサービスなど

## 基本パターン

LMSで授業のZoom URLを通知  
教材をLMSやGoogle Driveなどで配布

# 結局何が必要だったか

- 「不特定多数」「組織内限定」の安全なデータ共有

- 必要場面例

- 授業のZoom URL通知
- 授業録画提供、試験問題配布、etc.

これらはいずれもオンライン授業に限らず、研究基盤でも必須の共通の機能である

- システム

- LMS：大学共通アカウントで認証
- Google, Microsoftクラウド：組織向け契約があり、組織内限定のデータ共有機能がある

- 高性能Webサービス・ストレージ

- 普段の負荷 << 負荷集中時の負荷
- 授業開始時の集中アクセス（～千人/分）
- 試験開始時の集中アップロード

# パブリッククラウド頼りとなった場面

- 試験問題配布、答案アップロードなど、**短い時間の大量のアクセス**を低遅延で実行
  - 数MBの画像 x 3000人の答案を試験終了間際にアップロード
- 講義録画など**大量データ**の蓄積、提供

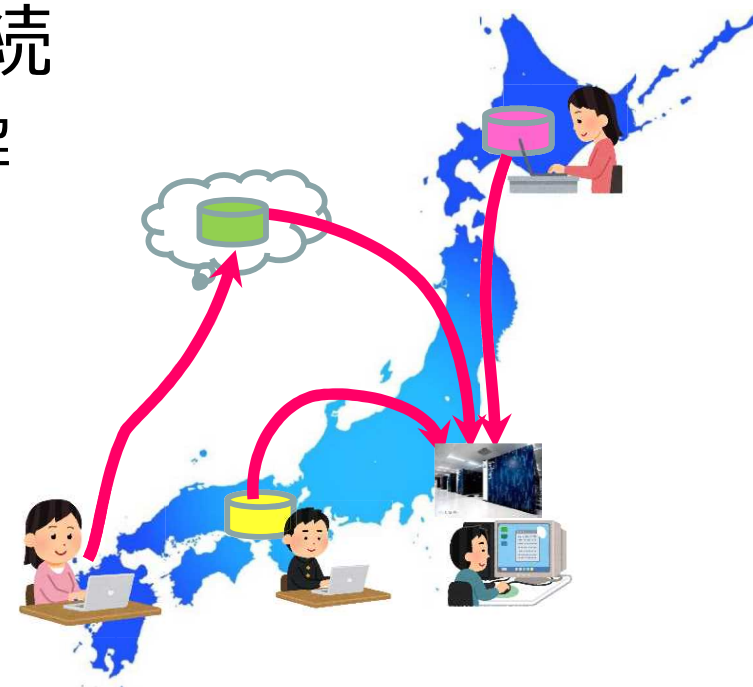
# 今後の情報基盤に必要なこと

- データ中心の基盤





- ビッグデータプロジェクトの半永続的なホスティング
- データ共有範囲を「組織間にまたがり」柔軟に設定可能
  - 大学間、大学-企業間
- 学内、IaaS, クラウドストレージの混合で実装

- 高性能計算基盤とのシームレスな接続

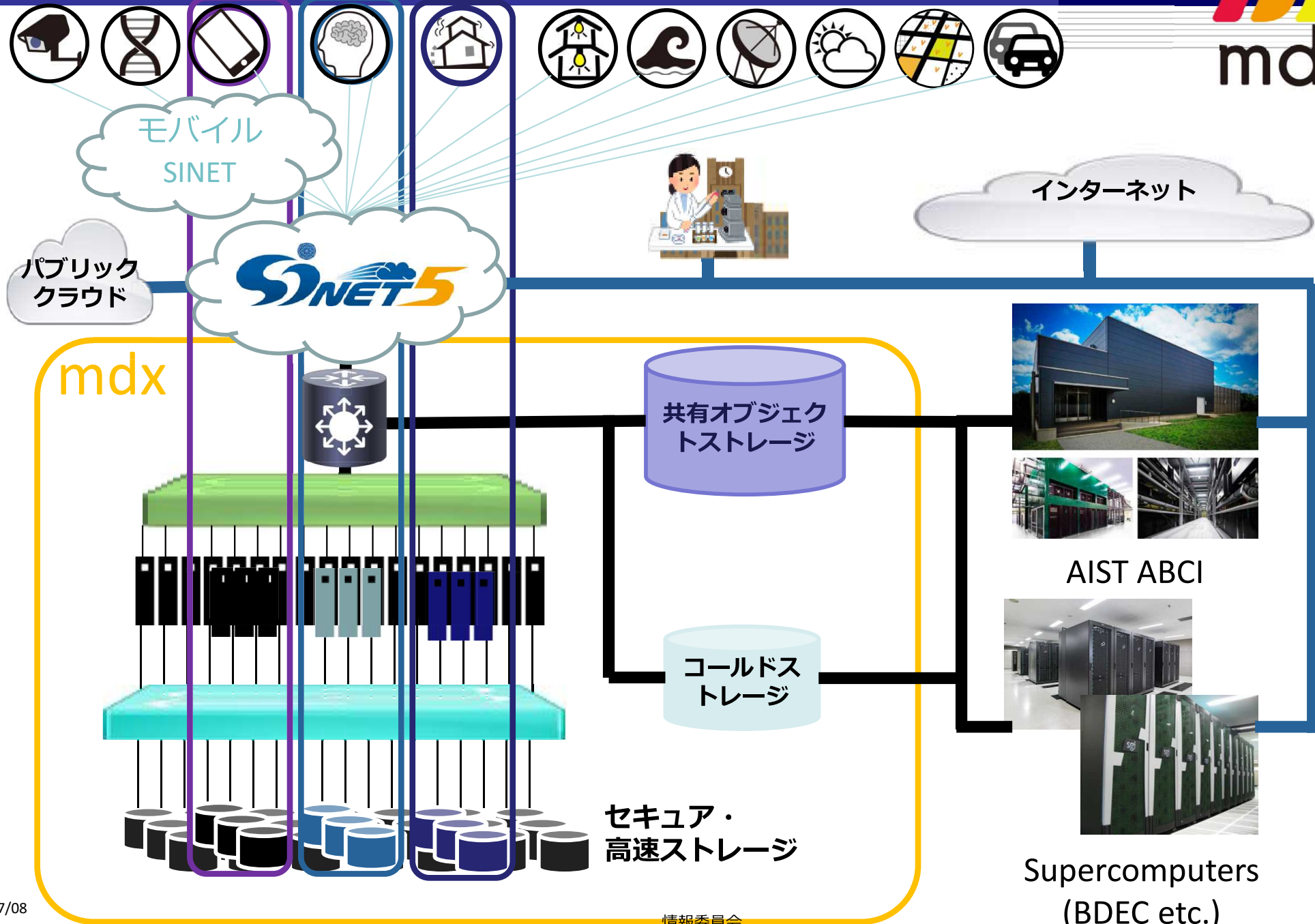
- データの閲覧からWebベースの簡便な解析へシームレスに移行できる
- さらに任意のプログラム（スパコン）による解析に移行できる
- 他大学含めどのストレージのデータもシームレスにアクセスできる



# 既存・進行中エフォート

-  **GakuNin** (NII 学認)
  - 認証を大学間で連携できる
-  **GakuNin RDM** (NII Gakunin RDM)
  - 研究データ管理のストレージ
  - 柔軟にデータ共有・公開の範囲を指定可能な基盤
-  **mdx** (NII, 9大学, AIST mdx)
  - 高性能IaaS環境
  - ストレージ・ネットワークの分離によってプロジェクト間をセキュアに隔離
-  **JHPCN** (8大学共同利用共同研究拠点)
  - スパコンを中心とした資源を提供、共同研究実施

# mdx 仮想プラットフォームの概要



仮想プラットフォーム

モバイル  
SINET

パブリック  
クラウド

SINET5

インターネット

mdx

共有オブジェク  
トストレージ

コールドス  
トレージ

セキュア・  
高速ストレージ

AIST ABCI

Supercomputers  
(BDEC etc.)

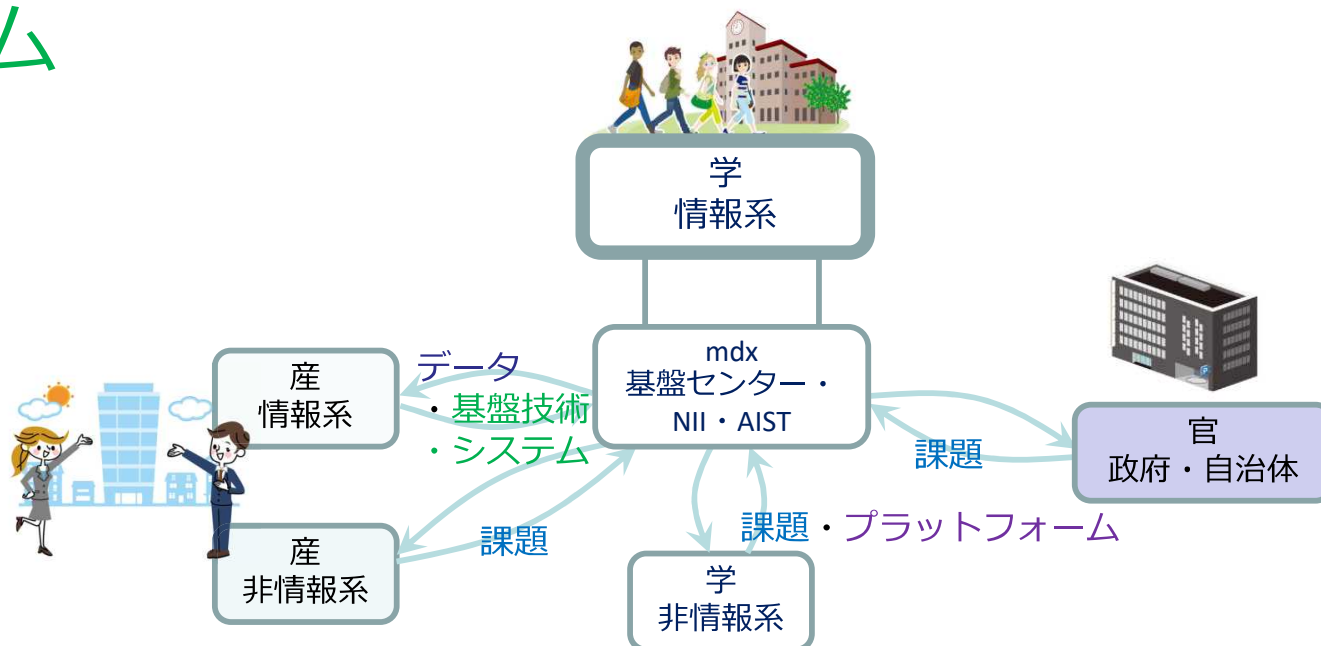
# 様々な連携を想定



- mdx 基盤システム稼働は2020年度末
- 課題持ち込み（分野の、地方の、企業の、…）
- データ提供・サービス（収益）化模索
- 分野データの整備・プラットフォーム構築
- データ科学・AI・機械学習
- 基盤技術・システム  
研究・開発

• ...

コロナ感染拡大防止関係のデータを  
集約して産学官連携を生み出す基盤  
とすることができないか



# mdx : 大学の学術情報基盤の新しい形を目指して

- データ科学・活用のための基盤
- 大学をまたがった連携を想定 (e.g. Mobile SINET, Gakunin RDM, 大学・研究所スパコン)
- これまでのHPC用途のクラスタとは異なる
- 「単一OS環境 + バッチスケジューラ」ではない
  - 分野データプラットフォームのホスティング (連続稼働)
  - 多様なソフトウェア構成の許容
  - 長年にわたるデータの蓄積・利用
  - 高いデータセキュリティ・隔離への要求
  - IoT・センサなど外部データのストリーム (実時間) 処理
  - JupyterLabなど対話的・探索的利用 (AI, 機械学習)
  - データ検索・発見のための利用
  - 対話的利用から高性能環境へシームレスな移行・連携

# 今後の学術情報基盤

- これら既存・進行中のエフォートに、パブリッククラウドも組み合わせ
  - セキュアで柔軟なデータ共有
  - 高性能計算機との連携
- を持続的に提供することで大学の基盤力強化、産業界の強化（特にクラウド、共有ストレージ、LMSなどの基盤ソフトウェア）をはかる

# このような環境を持続的に提供するには

- 強力なパブリッククラウドベンダー上の資源を用い、Gakunin RDMやown Cloudなどのオープンソースをベースにスケーラブルなデータ共有環境を構築、維持できるソフトウェアベンダーと、それら基盤の開発区に加われる大学のシニア技術スタッフが協力・人材交流をしながらオンプレ資源との統合、ソフトウェアの進化、新しいサービスの創出を担うことが必要
- 赤字が現在日本に欠けているもの

現実には、研究を犠牲にして日々の問題に教員が対応しており、国内にはクラウドストレージのようなサービスを担えるベンダーはGoogle/Microsoft以外にいない状態。国内ではLMSのようなサービスをスケールさせることもままならない業者がLMSを納入・運営している。