

# データベースシステムが これまで以上に 時代の先端を拓く時代に突入

喜連川 優

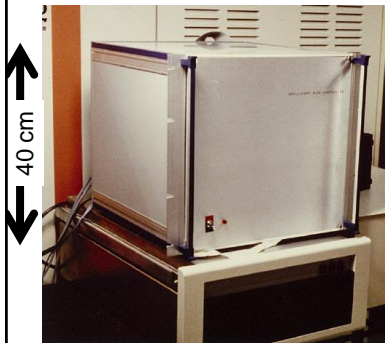
国立情報学研究所 所長  
東京大学 特別教授

## お話すること

- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に

# ポスグレとの色々な接点

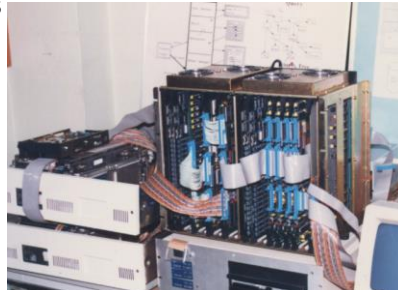
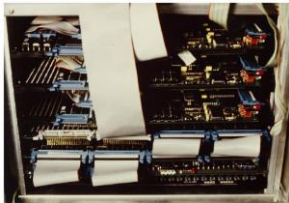
## Functional Disk System 機能ディスク

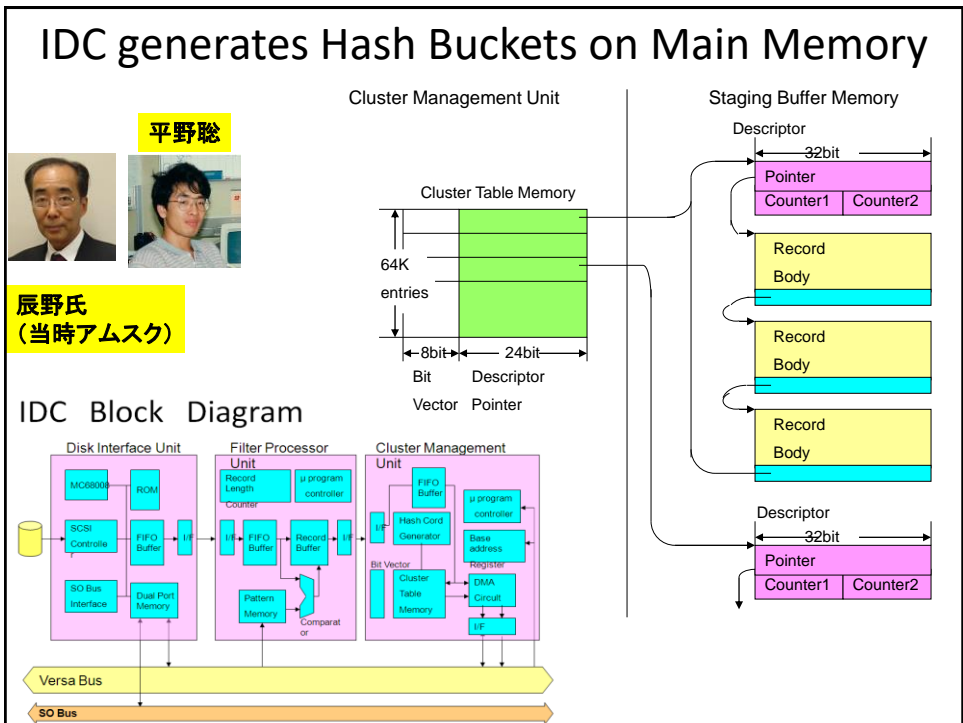
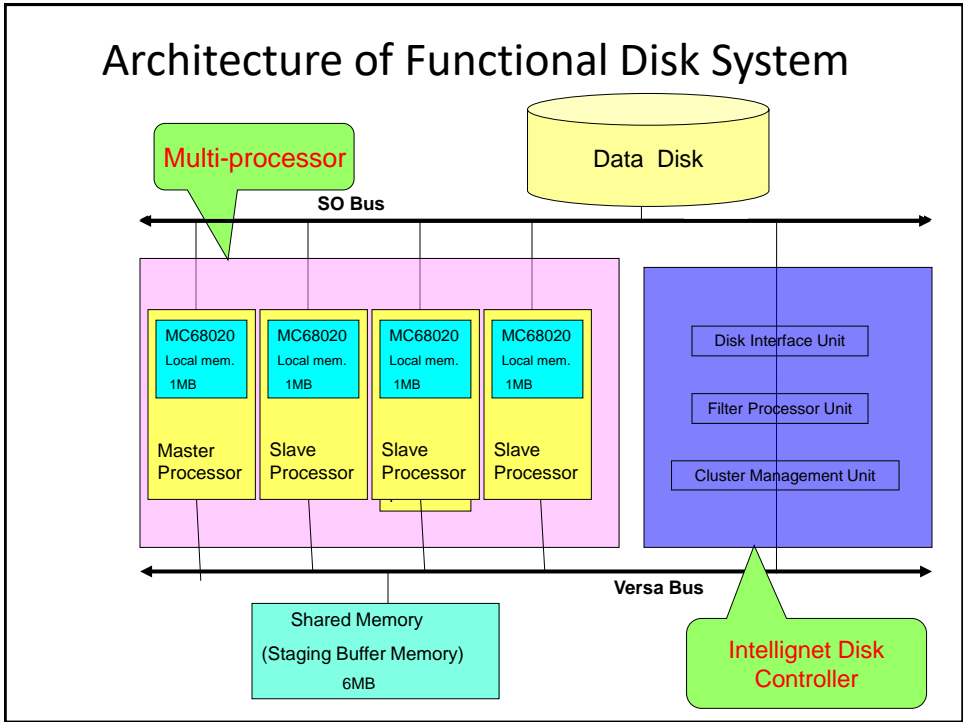


Masaru Kitsuregawa, Miyuki Nakano, Mikio Takagi:  
Performance Evaluation of Functional Disk System (FDS-R2).  
**ICDE 1991**: 416-425

Masaru Kitsuregawa, Miyuki Nakano, Mikio Takagi: Query  
Execution for Large Relations on Functional Disk Systems.  
**ICDE 1989**: 159-167

Masaru Kitsuregawa, Miyuki Nakano, Lillian Harada, Mikio  
Takagi: Functional Disk System for Relational Database.  
**ICDE 1987**: 88-95

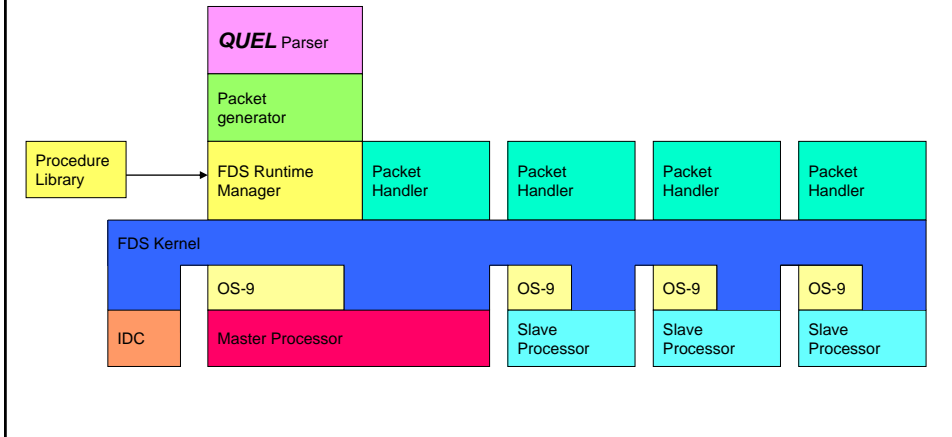




# 機能ディスクシステムのソフトウェアアーキテクチャ



中野先生



## Performance of Functional Disk System (圧倒的勝利)

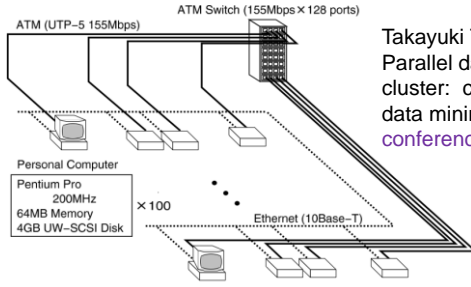
Masaru Kitsuregawa, Miyuki Nakano, Lilian Harada  
Mikio Takagi: Functional Disk System for Relational Database. ICDE 1987: 88-95

Query Type	System	Performance
(1) simple retrieval QUERY: range of t is tenKtuple retrieve into TEMP(t.all) where a0<1000;	No index (sec)	
	U-Ingres	64.4
	C-Ingres	53.9
	ORACLE	230.6
(2) duplicate elimination QUERY: range of t is oneKtuple retrieve unique into (t.all)	U-Ingres	236.8 (sec)
	C-Ingres	132.0
	ORACLE	199.8
	FUNCTIONAL DISK	0.975+a+b
(3) complex query with join QUERY: range of t is tenKtuple1 range of w is tenKtuple2 retrieve into TEMP(t.all,w.any) where (t.a0=w.a0) and w.a0 < 1000	U-Ingres	10.2 (minutes)
	C-Ingres	1.8
	ORACLE	>300
	FUNCTIONAL DISK	6.18+a+b (sec)
(4) aggregate operation QUERY: range of t is tenKtuple retrieve into sum(t.a1 by t.a2) 100 partitions	U-Ingres	174.2 (sec)
	C-Ingres	484.8
	ORACLE	1487.5
	FUNCTIONAL DISK	3.243+a+b (sec)

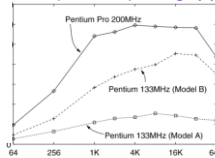
Table.1 Performance Evaluation of FDS with Wisconsin Benchmark

a: query translation time      b: time to put results back to disk

# NEDO100: 100 node PC cluster =TPC-D and Association rule mining=



Takayuki Tamura, Masato Oguchi, Masaru Kitsuregawa:  
Parallel database processing on a 100 Node PC cluster: cases for decision support query processing and data mining, *Proceedings of the 1997 ACM/IEEE conference on Supercomputing*, pp.1 - 16



# IBM DB2が 我々の方式 を。

Anastassia Ailamaki, David J. DeWitt, Mark D. Hill, Marios Skounakis: Weaving Relations for Cache Performance. *VLDB 2001*: 169-180

**JOIN OPERATOR.** The adaptive dynamic hash join algorithm [23], which is also used in DB2 [20], was implemented on top of Shore. The algorithm partitions the left

[23] M. Nakayama, M. Kitsuregawa, and M. Takagi: Hash-Partitioned Join Method Using Dynamic Destaging Strategy. In *Proceedings of the 14th VLDB International Conference*, September 1988.

Masaru Kitsuregawa, Masaya Nakayama, Mikio Takagi:  
The Effect of Bucket Size Tuning in the Dynamic Hybrid GRACE Hash Join Method. *VLDB 1989*: 257-266

Masaya Nakayama, Masaru Kitsuregawa, Mikio Takagi: Hash-Partitioned Join Method Using Dynamic Destaging Strategy. *VLDB 1988*: 468-478

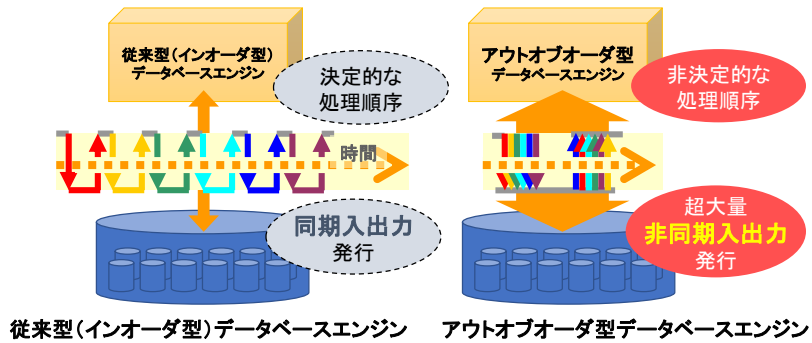


これがポスグレに入っているの  
か良く存じません。

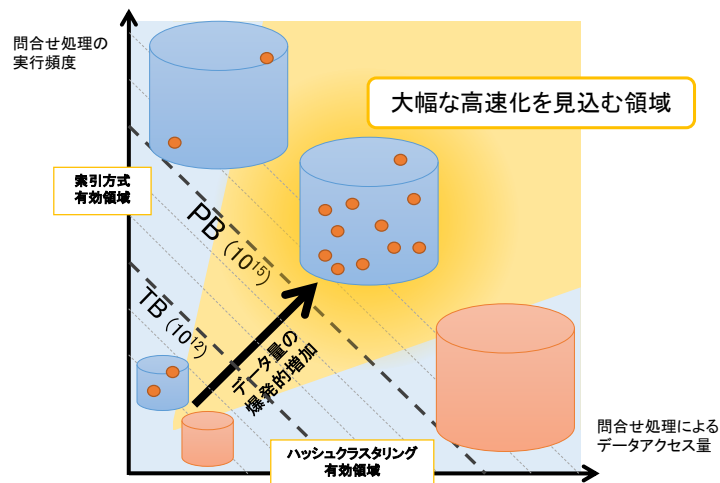
「ポスグレにhash joinが入って  
グッと  
パワフルになって嬉しかった」と  
石井様より伺いましたが、  
導入は いつ頃でしょうか？

## アウトオブオーダー型データベースエンジンOoODE

- アウトオブオーダー型クエリ実行
  - クエリ実行処理を動的タスク分解
  - 非同期入出力を高多重発効
  - 入出力完了を契機として並列演算実行

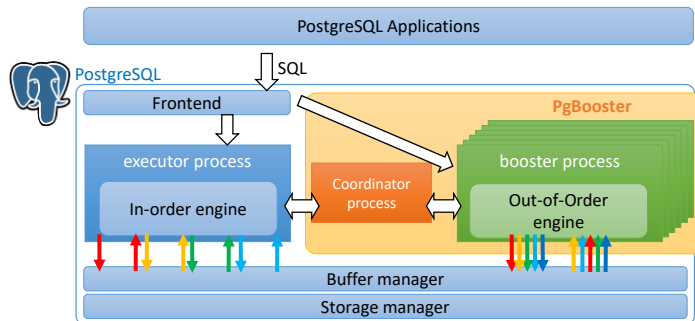


## アウトオブオーダー型データベースエンジンの適用領域

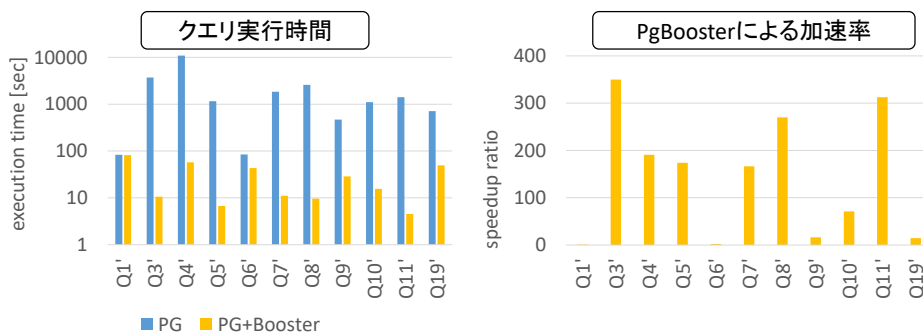


## PgBooster: PostgreSQLベースの加速機構試作実装

- 現有資産を活かしながら実行を高速化
- コンパクトな開発規模
  - 約7,000行 (PostgreSQL本体: 230万行)



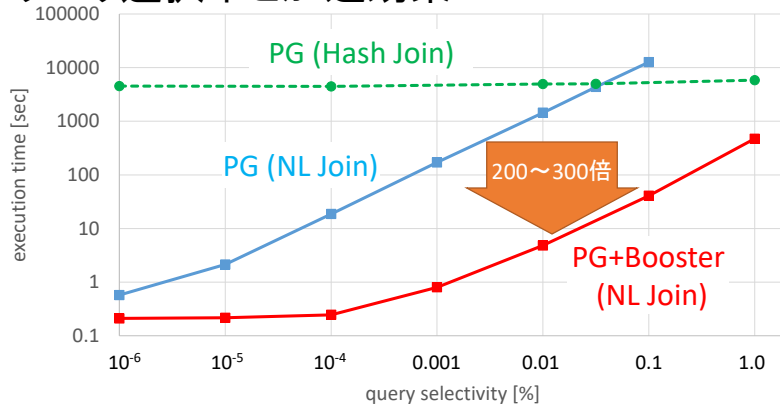
## TPC-Hデータセットによる性能評価



- TPC-H Benchmark SF=1000 (1TB in raw) を用いた性能測定
  - クエリ選択率を規定より小さく設定
  - 通常版PostgreSQL (PG)、PostgreSQL+PgBoosterを比較
- 選択性を有する複数のクエリで100倍以上の性能向上を確認
  - 最大で350倍の性能向上



## クエリ選択率と加速効果



- 選択率 10<sup>-6</sup>% から1% では常にPG+Boosterが最速
- 選択率10<sup>-4</sup>%から1%ではPG(NL Join)に対して200~300倍高速化
  - 入出力帯域を活用した分クエリ実行時間が短縮
- 選択的クエリにおける加速効果を確認

## Stonebraker氏と！



26

Stonebraker氏のTuring賞受賞の背景  
最初の大規模オープンソースソフトウェア！！

27

最近の喜連川のメッセージ  
「ソフトウェアは出来るだけ作るな！」

オープンソースの重要性  
メンテ

## お話すること

- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に

## データ駆動：ビッグデータとAI でSociety5.0を創る

喜連川 優

国立情報学研究所 所長  
東京大学 特別教授

NII

日本でコンピュータ（IT）だけを  
研究している唯一の国研

NII

National Institute of Informatics  
国立情報学研究所  
(文部科学省)



30

## 沿革・歴代所長

NII

年月	事項
昭和51 (1976) 年 5月	東京大学情報図書館学研究センター発足
昭和58 (1983) 年 4月	東京大学文献情報センター設置 (情報図書館学研究センターを改組)
昭和61 (1986) 年 4月	学術情報センター (NACSIS) 設置 (東京大学文献情報センターを改組)
平成 6 (1994) 年11月	千葉分館 (千葉県千葉市) 竣工
平成 9 (1997) 年 3月	国際高等セミナーハウス (長野県軽井沢町) 竣工
平成12 (2000) 年 4月	<b>国立情報学研究所 (NII) 設置</b> (学術情報センターの廃止・転換)
平成16 (2004) 年 4月	大学共同利用機関法人 情報・システム研究機構 国立情報学研究所設置



初代  
猪瀬 博

任期：  
2000年4月 - 2000年10月



第2代  
未松 安晴

任期：  
2001年4月 - 2005年3月



第3代  
坂内 正夫

任期：  
2005年4月 - 2013年3月



第4代 (現在)  
喜連川 優

任期：  
2013年4月 -



2020年  
NII20周年!



31

基盤>全体 NII

# 学術情報基盤

15の研究施設

**Research**  
研究

アーキテクチャ  
科学研究系

コンテンツ  
科学研究系

情報社会  
相関  
研究系

大学院  
教育

情報学  
プリンシプル  
研究系

学術コンテンツ事業

**Service**  
事業

学術情報ネットワーク事業

1. SINET
2. クラウド
3. 認証
4. セキュリティ
5. 学術コンテンツ
6. オープンサイエンス (研究データ基盤)

35

## 多様な研究分野を支えるNII

NII

NII

情報学      学術情報基盤

36

基盤&gt;全体

## NII事業の全体像

NII

- ・ 超高速ネットワーク、オープンサイエンス、サイバーセキュリティ、クラウド、認証に関するサービスを提供し、全国の大学や研究機関等の研究教育環境を高度化

## オープンサイエンス

- ◆ オープンサイエンスの推進
- ◆ 学術情報流通とオープンアクセスの推進

## サイバーセキュリティ

- ◆ 大学間連携に基づく情報セキュリティ体制の基盤構築
- ◆ 研究用データの公開

## クラウド活用支援

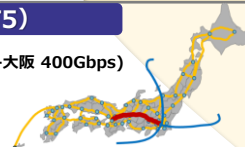
- ◆ クラウド利活用によるIT経費削減・研究教育環境の高度化
- ◆ 直結クラウドによる利用の促進

## 学術認証

- ◆ 電子証明書による安全な認証の推進
- ◆ 大学間認証連携による各種資源の相互利用の促進

## 学術情報ネットワーク (SINET5)

- ◆ 全ての都道府県を超高速の100Gbps回線で網羅 (東京-大阪 400Gbps)
- ◆ 米国、欧州、アジアと超高速の100Gbps回線で接続
- ◆ 多様な通信サービスを最新ネットワーク技術で提供



37

基盤&gt;1.SINET

## SINET5の概要

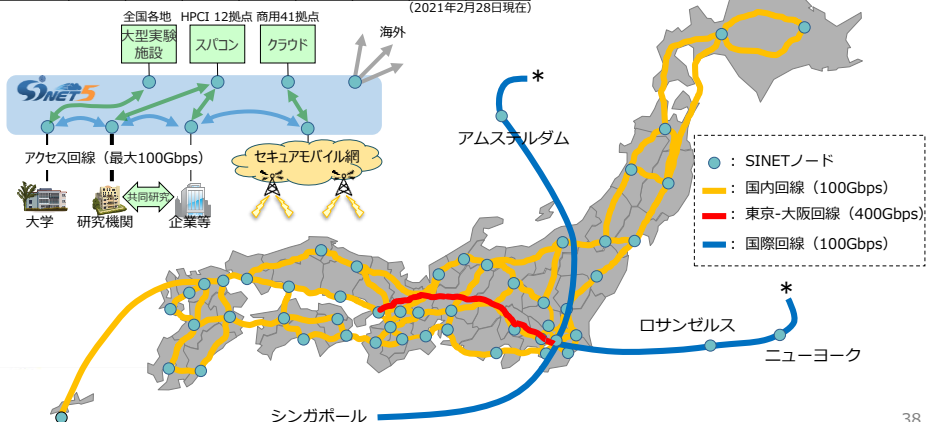
NII

- ◆ 日本全国の大学・研究機関等が利用する学術専用の情報通信ネットワーク
  - ・ 全都道府県にSINETノードを設置、任意の拠点間や海外への100Gbps通信環境を提供
  - ・ 民間企業も大学等の共同研究契約があれば利用可能

	国立大学	公立大学	私立大学	短期大学	高等専門学校	大学共同利用機関	その他	合計
加入機関数	86 (100%)	86 (91%)	421 (68%)	82 (25%)	56 (98%)	16 (100%)	222	969

(2021年2月28日現在)

SINET5



38

基盤> 1.SINET **SINET利用例 - 高エネルギー研究** **NII**

◆ 小林・益川理論の検証を目的としたBelle実験、ニュートリノ観測を目的としたスーパーカミオカンデ実験、ヒッグス粒子の発見を目的としたLHC (スイス) のATLAS実験等、が発生する大容量データの転送やセキュアな通信環境形成に利用されており、ノーベル物理学賞受賞に影ながら貢献

提供：高エネルギー加速器研究機構、東大宇宙線研究所、東大素粒子物理国際研究センター

CP対称性の破れ **Belle** 筑波

**Belle II (soon)** 筑波

ニュートリノ **Super Kamiokande** 神岡

ヒッグス粒子 **CERN LHC ATLAS** スイス

大容量データ

KEK

SINET VPN

東北大学

超高速国際回線

東京大学

その他の大学

各大学で並行解析

大阪大学

名古屋大学

超高速でセキュアな通信環境

LHC: 大型ハドロン衝突型加速器  
ATLAS: 高エネルギー陽子衝突反応測定器

39

**SINET5からSINET6へ** **NII**

- SINET6では、①400Gbpsの全国展開、②SINET接続点の拡大、③超高速モバイルと有線の融合、④エッジ機能配備とサービス拡大、⑤国際回線の増強等を目指します
- 同時に、オープンサイエンス実現のための研究データ基盤の本格展開を目指します

SINET5 (2016~2021年度)	SINET6 (2022~2027年度)
<ul style="list-style-type: none"> <li>• 全国100Gbps (東阪は400Gbps)</li> <li>• 4G モバイルSINET</li> <li>• ルータによるVPNサービス</li> <li>• 国際回線の全100Gbps化</li> </ul>	<ul style="list-style-type: none"> <li>• 全国400Gbps化 + SINET拡張DC</li> <li>• 5G モバイルSINET + ローカル5G</li> <li>• NFVとルータによる柔軟なサービス</li> <li>• 国際回線の帯域強化と対地拡大</li> </ul>

研究データ基盤

コンテンツ基盤

学術認証基盤

クラウド活用基盤

VPNサービス、オンデマンドサービス等

4G モバイルSINET

欧州

米国

アジア

研究データ基盤

シームレスセキュアクラウド基盤

高度学術認証基盤

有線・無線融合ネットワークサービス基盤

ローカル5G

5G モバイルSINET

欧州・アフリカ

米国・南米

アジア・オセアニア

図はイメージ

● : SINET DC  
○ : 拡張 DC

● : 400Gbps国内回線  
○ : 拡張 DC

● : 100Gbps国内回線  
○ : 拡張 DC

● : 100Gbps国際回線  
○ : 拡張 DC

41

# コロナ禍における 超短期間での大学講義の オンライン化に関する 情報共有シンポ (2020年3月から)

約5000講義/7大学

43

NHK おうちで学ぼう！ for School

解説委員室 NEWS COMMENTATORS BUREAU

新着

解説アーカイブス これまでの解説記事

「"やんちゃ"とデータが救う コロナ禍の大学教育」(視点・論点)

2022年01月12日(水)

国立情報学研究所 所長 喜連川 優

国立情報学研究所 所長 喜連川 優



## 「4月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム」NII

- **主催：国立情報学研究所（NII）大学の情報環境のあり方検討会**
- **遠隔授業等の準備状況に関する情報を出来る限り多くの大学間で共有することを目的に、大学等関係者が現状の取組や課題等を発表するシンポジウムを開催**
- **オンライン会議形式。後日、国立情報学研究所ホームページにて、映像及び資料を掲載。参加者は、2,000名を超える回も。**
- **今後も定期的に開催予定**

### ■ 開催経過

- 2020年3月26日 第1回 オンライン開催
- 2020年4月3日 第2回 オンライン開催
- 2020年4月10日 第3回 オンライン開催
- 2020年4月17日 第4回 オンライン開催
- 2020年4月24日 第5回 オンライン開催
- 2020年5月1日 第6回 オンライン開催
- 2020年5月8日 第7回 オンライン開催
- 2020年5月15日 第8回 オンライン開催
- 2020年5月29日 第9回 オンライン開催
- 2020年6月5日 第10回 オンライン開催
- 2020年6月26日 第11回 オンライン開催
- 2020年7月10日 第12回 オンライン開催
- 2020年7月31日 第13回 オンライン開催
- 2020年8月21日 第14回 オンライン開催
- 2020年9月4日 第15回 オンライン開催
- 2020年9月11日 第16回 オンライン開催

2021年9月17日 第40回 オンライン開催

### ■ これまでの発表大学等

- 北海道大学
- 室蘭工業大学
- 東北大学
- 群馬大学
- 千葉大学
- 東京大学
- 電気通信大学
- 新潟大学
- 名古屋大学
- 滋賀大学
- 京都大学
- 大阪大学
- 大阪教育大学
- 徳島大学
- 香川大学
- 愛媛大学
- 広島大学
- 九州大学
- 熊本大学
- 九州工業大学
- 北九州市立大学
- 神田外語大学
- 杏林大学
- 芝浦工業大学
- 帝京大学
- 慶應義塾大学
- 東京医療保健大学
- 東京通信大学
- 東邦大学
- 日本体育大学
- 法政大学
- 早稲田大学
- 星槎大学
- 藤田医科大学
- 立命館大学
- 人間環境大学
- 大阪工業大学
- 関西医科大学
- 近畿大学
- 関西学院大学
- 神戸女子大
- 北九州工業専門学校
- 熊本高等専門学校
- 前田高等学校
- 青山学院中等部
- 川崎北高等学校
- 日出学園高等学校
- 日中
- 早稲田大学高等学院
- 京都市立西京南高等学校
- 校附属中学校
- 埼玉県立川越南高等学校
- アランブロン国際中学校
- 校高等学校
- 関西創価高等学校
- 東京学芸大附属小金井小学校
- TWOLAPS TRACK CLUB
- 北京大学
- ハーバード大学医学部
- ノースカロライナ州立大学
- タイ国立電子コンピューター技術研究センター
- 韓国教育芸術情報院
- シドニー大学
- マサチューセッツ工科大学
- マサチューセッツ州立大学
- 高知県教育委員会
- 熊本市教育委員会
- 広島県安芸太田町教育委員会
- 奈良県立教育研究所
- 国立情報学研究所
- 国立教育政策研究所
- 一般社団法人授業目的公衆送信補償金等管理協会
- 総務省
- 文部科学省
- 文化庁
- 経済産業省

※予定



サイバーシンポジウムのURL：  
<https://www.nii.ac.jp/event/other/decs/>

## NII

## 多様なテーマの取り上げ

教授法  
海外  
メンタル  
運動

遠隔・対面ハイブリッド 高等教育 学生の声 学生の気持ち 教員の気持ち サークル活動 非常勤の先生方のご苦労 学生に係るデータの取り扱い 実験の講義 海外事例（北京、チェコ、ドイツ、ハーバード、ノースカロライナ、オーストラリア、韓国、タイ、メキシコ）試験・モニタリング・落語 ウィズコロナ初中教育 教育委員会 大学付属事例 中学事例 データダイエットの定量的効果測定 学生の声から 学習解析 見るニューノーマル 遠隔講義支援 図書館再開 教育データ 大学事例 教職員向け 高等学校 中学校 初等中等教 高等学校 超入門（LMSと学習解析） 国際特別レクチャー Fail Smart 成績評価 遠隔子守 大学間連携 実験 看護系・医療系・語学系 事例紹介 文科省 経産省 バーチャル運動 こそ大事 授業設計 超入門 学習管理システム（LMS） 著作権 新入生支援 成績・評価 教育データ VR サポート体制 ピーク負荷メンタル 授業設計 高専 高大接続 オンライン会議サービス セキュリティ ネットワーク 合理的配慮 初等中等教育 私立大学 情報保障



## 7大学とNIIの協力

NII

**高等教育を止めない！  
世界に一切遅れない！**

3月末、4月、5月

**IT支援者** は 膨大な数の学生と教員の  
遠隔講義対応に眠る時間無し(温かい目を)

**「壮烈な戦い」** (東大5000,東北大4400講義)

48

NII

## Fail Fast, Cheap, Smart

目的：

誰もやったことの無いことなので、大きな大学が率先して**失敗し**、その経験知を皆と共有しよう！

「ここは何かを教えてくれると思わないで下さい。勘違いしないでください。皆で情報共有する場です。」

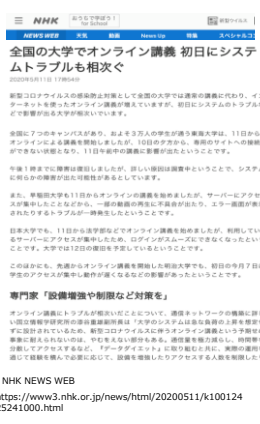
49

# オンライン授業 障害記事の度、何故を講演してもらおうNII

## どんまい ファイト！！



- ・ 大規模な遠隔講義を開始するなかで、システムの不具合が多く発生。報道相次ぐ
- ・ 大きなシステムの起動ではよくあること。とりわけ、多くの大学で財政的制約や人材が払底する中で、避けられない。トラブルとは言え、**「FAIL CHEAP」**で乗り切った！
- ・ 東北大はそもそも本番ではなく練習の際の不具合



東北大ネット講義で一時、アクセス障害(4/20)

50  
全国の大学でオンライン講義 初日にシステムトラブル相次ぐ(5/11)

# 著作権問題 5月に間に合う！ NII

- ・ 北大、東北大、東大、名大、京大、阪大、九大の総長、並びに、国立情報学研究所所長の名を連ねた要請文をSARTRASに提出。
- ・ 遠隔講義シナジウムでも再三取り上げられる中で、4月28日より令和2年度は遠隔講義においても著作物の無償利用可能。



51

## 通信インフラをリセットして見つめる

通信量の増大
データダイエットご協力  
のお願い(5/8)

### 1600万人が入るか？

夜間ピークの通信量 +0% (5/11週との比較)  
昼間帯の通信量 +8% (5/11週との比較)  
夜間帯の通信量 +15% (2/25週との比較)  
昼間帯の通信量 +48% (2/25週との比較)

NTT東日本  
<https://www.ntt-east.co.jp/aboutus/traffic/20200428/>

- オンライン授業は通信量（データ量）が軽くなるように。
- 通信量の削減は必要となる教育や学習への合理的配慮等必須の分野に。
- 不要なカメラはオフにしましょう
- 授業の録画はクラウドに保存する必要はありません
- 学生の主体的な学習にサポートは必ずしも必要ではありません

国立情報学研究所  
National Institute of Informatics

## 当時のツールmax2000人超え NII

### 【第5回】2020.4.24 参加状況

登録者数内訳：2790名登録

Webex + Zoomオンライン参加者数：2332名視聴

### 【第6回】2020.5.1 参加状況

登録者数内訳：2306名登録

Webex + Zoomオンライン参加者数：1930名視聴

国立情報学研究所  
National Institute of Informatics

# 内閣官房教育再生 デジタルタスクフォース


**文部科学省**
MINISTRY OF EDUCATION,  
CULTURE, SPORTS,  
SCIENCE AND TECHNOLOGY-JAPAN

[> サイトマップ](#)
[> English](#)
[文字サイズの変更](#)
小 中 大
サイト内検索

[会見・報道・お知らせ](#)
[政策・審議会](#)
[白書・統計・出版物](#)
[申請・手続き](#)

[トップ](#) > [今日の出来事](#) > [令和2年9月](#) > [教育再生実行会議 高等教育ワーキング・グループ第1回を開催し、萩生田大臣が出席「ニューノーマル\(新たな日常\)における大学の姿」を中心に議論](#)

**● 教育再生実行会議 高等教育ワーキング・グループ第1回を開催し、萩生田大臣が「ニューノーマル(新たな日常)における大学の姿」を中心に議論**

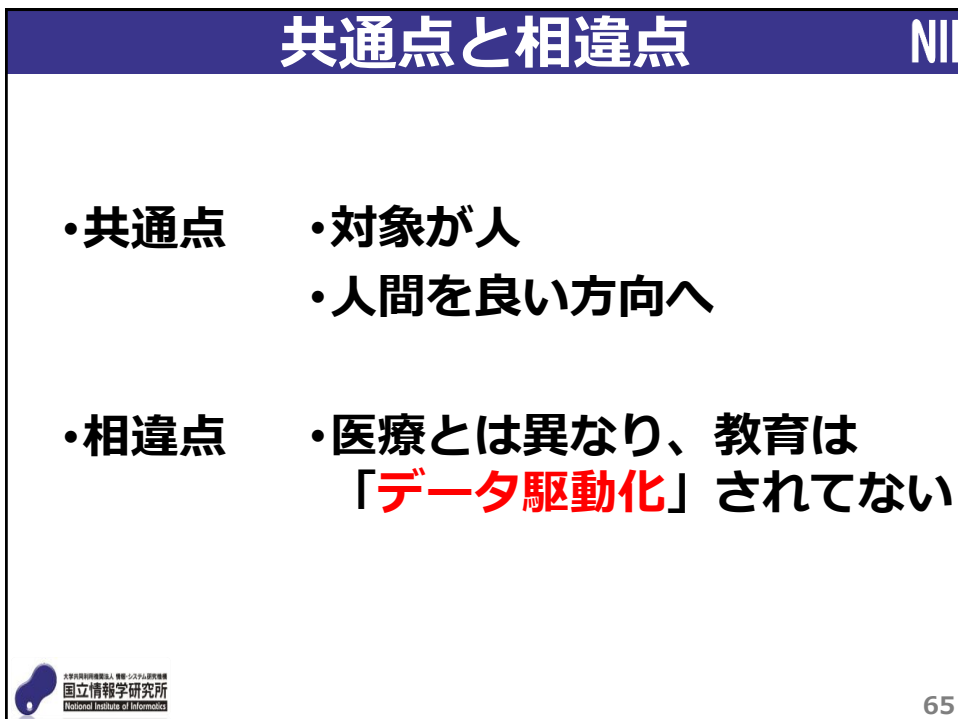
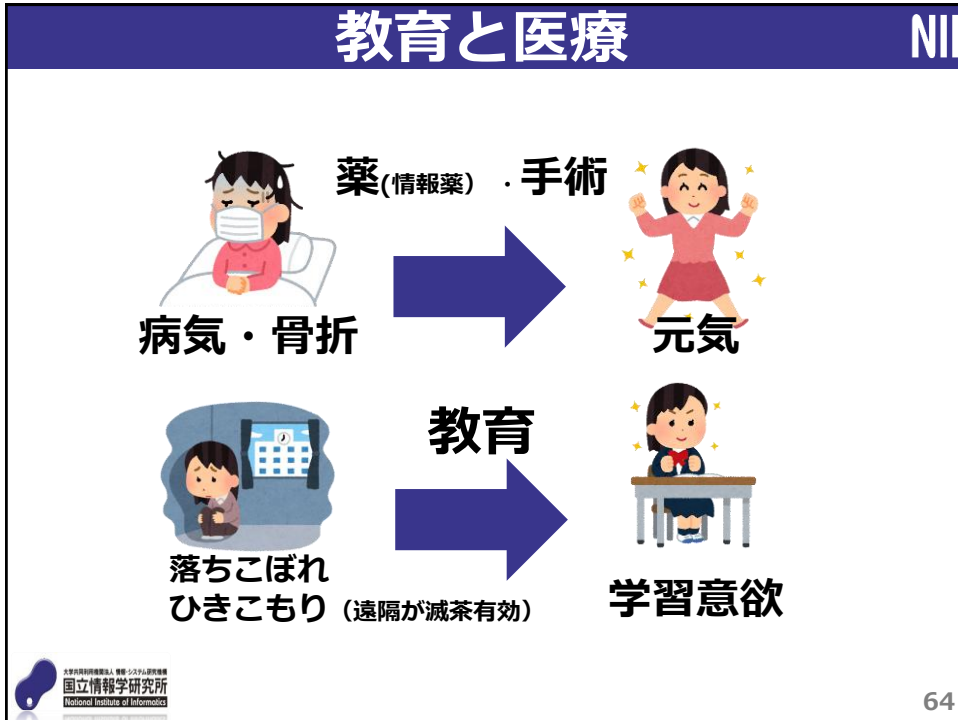
9月14日(月曜日)  
 教育

9月14日、教育再生実行会議 高等教育ワーキング・グループ(WG)の第1回が開催され、鎌田座長、オブザーバー、WG委員、本体会議の有識者の皆様等、リモートも含めて24人に出席いただき、ポストコロナ/ウィズコロナを念頭に、対面と遠隔の双方の良さを組み合わせたハイブリッド型の教育など、ニューノーマルにおける大学の姿やその実現方策について議論いただきました。

まず、情報・システム研究機構国立情報学研究所長の喜連川委員から、これまで16回にわたり開催された「大学等遠隔授業に関する取組状況共有サイバーシンポジウム」の成果を踏まえ、オンライン授業のメリット・デメリット、ハイブリッド教育への期待や課題について発表いただきました。※

今週のトピックス





66

文部科学省  
MINISTRY OF EDUCATION,  
CULTURE, SPORTS,  
SCIENCE AND TECHNOLOGY (JPKE)

> サイトマップ > English 文字サイズの変更 中 大 サイト内検索

会見・報道・お知らせ | 政策・審議会 | 白書・統計・出版物 | 申請・手続き | 文部科学省の

トップ > 今日の出来事 > 令和3年6月 > 教育再生実行会議において第十二次提言「ポストコロナ期における新たな学びの在り方について」が鎌田座長から菅総理に手交されました

● 教育再生実行会議において第十二次提言「ポストコロナ期における新たな学びの在り方について」が鎌田座長から菅総理に手交されました

6月3日(木曜日)  
教育

6月3日、第48回教育再生実行会議が菅総理出席の下、総理官邸で開かれ、萩生田大臣、鎌田教育再生実行会議座長などが出席しました。

会議では第十二次提言として「ポストコロナ期における新たな学びの在り方について」が取りまとめられ、鎌田座長から菅総理に提言が手交されました。

今回の提言は、一人一人の多様な幸せと社会全体の幸せ「ウェルビーイング」の実現を目指した学習者主体の教育への転換を軸としており、その実現の方策として、

○初等中等教育に関しては、データ駆動型の教育への転換による学びの変革の推進や、少人数によるきめ細かな指導体制と教師の質の向上等

○高等教育に関しては、遠隔・オンライン教育の推進、グローバルな視点での新たな国際戦略等

今週のトピックス

文部科学省

「深海」まだ知らない世界がそこへ  
—2021年から「国産海洋科学の10年」が始まる—

映画「深海のサバイバル」タイアップ企画 対談

「小学校社会科教育支援動画 国会へ提出」を公開しました

外国人児童生徒教育に関する動画

## お話すること

- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に

# データ駆動：ビッグデータとAI でSociety5.0を創る

喜連川 優

国立情報学研究所 所長  
東京大学 特別教授

New IT Infrastructure for the Information Explosion Era  
Research Project funded by MEXT Grant-in-Aid for  
Scientific Research in Priority Areas (FY 2005-2010)

*inf@-plosion*

情報爆発時代に向けた新しいIT基盤技術の研究

平成17年発足 文部科学省特定領域研究

領域代表

喜連川優

東京大学 生産技術研究所

70

© 特定領域研究「情報爆発IT基盤」



# 科学の新潮流：データ(駆動)サイエンス NII



## The FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY JONAH BECK, STEPHEN HANDEL, AND KRISTIN THOMAS

Tony Hey



2009年10月1日

計算からデータへのシフト

支配方程式の無い世界への挑戦

## The Evolution of Science

- **Observational Science**
  - Scientist gathers data by direct observation
  - Scientist analyzes data
- **Analytical Science**
  - Scientist builds analytical model
  - Makes predictions.
- **Computational Science**
  - Simulate analytical model
  - Validate model and makes predictions
- **Data Exploration Science**
  - Data captured by instruments
  - Or data generated by simulator
  - Processed by software
  - Placed in a database / files
  - Scientist analyzes database / files



71

# データ駆動 と AI

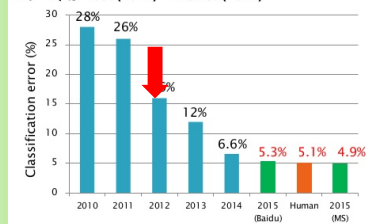
2012年

ビッグデータイニシアティブ



深層学習のパワー

エラー率が16% (2012) → 4.94% (2014)



Wu et al., "Deep Image: Scaling up Image Recognition", 2015.  
He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Classification", 2015.

ビッグデータとAIが並走

**Data Fuel AI**

72

# *Data Fuels AI*

73

# *Data Fuels AI*

*データはAIの燃料*

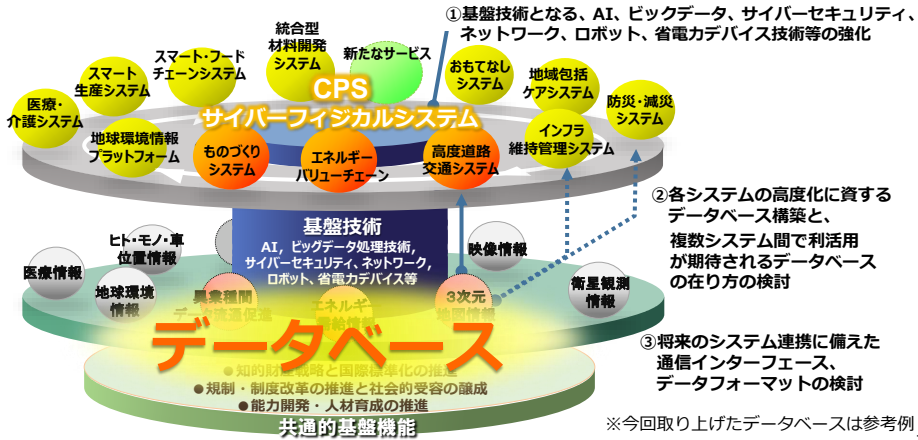
74

# Society 5.0 プラットフォーム

情報、通信、デバイス技術と制御技術の飛躍的進歩により、サイバー空間とフィジカル空間を融合させ、新たな価値を創出 ⇒産業価値はコンポーネントからシステムへ

## ●「Society 5.0」プラットフォーム構築

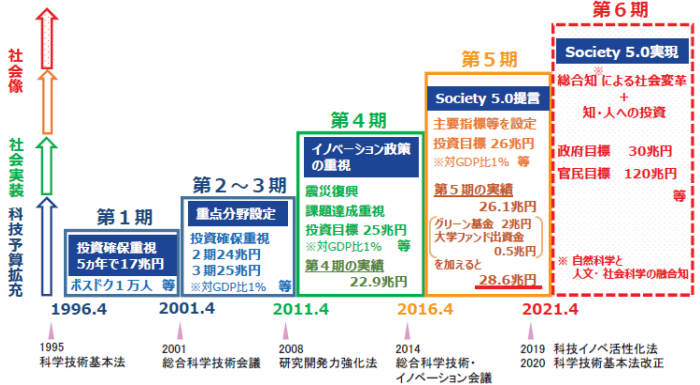
総合戦略2015で定めたシステムのうち「高度道路交通システム」「エネルギーバリューチェーンの最適化」「新たなものづくりシステム」をコアシステムとして開発。他システムと連携協調を図り、新たな価値を創出。



# 6期科学技術基本計画

## 科学技術・イノベーション基本計画策定の経緯

- 科学技術基本法制定(1995年)に基づき、基本計画を5年毎に策定
- 第1～3期では科学技術予算拡充、第4期では社会実装を重視、第5期では「Society 5.0」を提言
- 第6期は基本法を改正(2020年)、基本計画の対象に「人文・社会科学の振興」と「イノベーションの創出」を追加。本格的な社会変革に着手



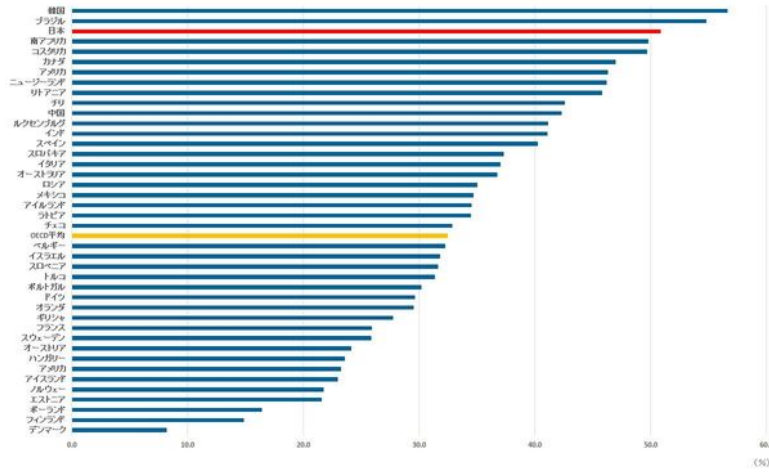
**大きくサイエンスのゲームが変わった  
スパコンからデータへ！**

**コロナ社会の支配方程式は無い！**

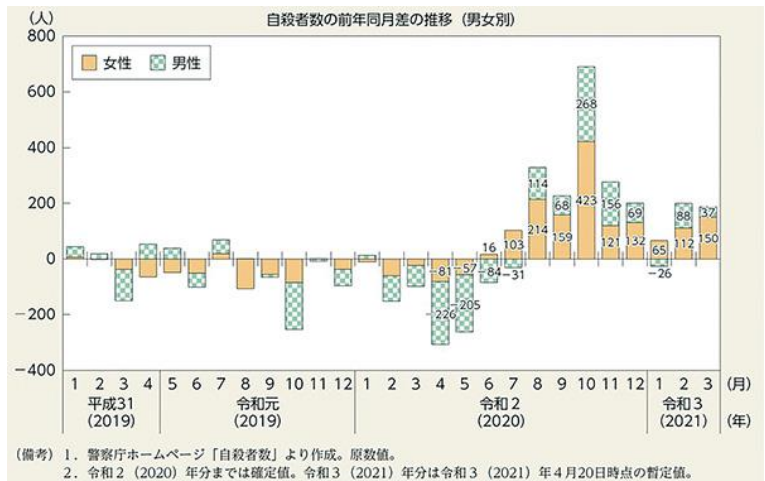
**データ駆動化が必須**

# ひとり親世帯の貧困率 (OECD加盟国最悪)

【表3】 ひとり親世帯の貧困率の国際比較 (OECD)



# 自殺率



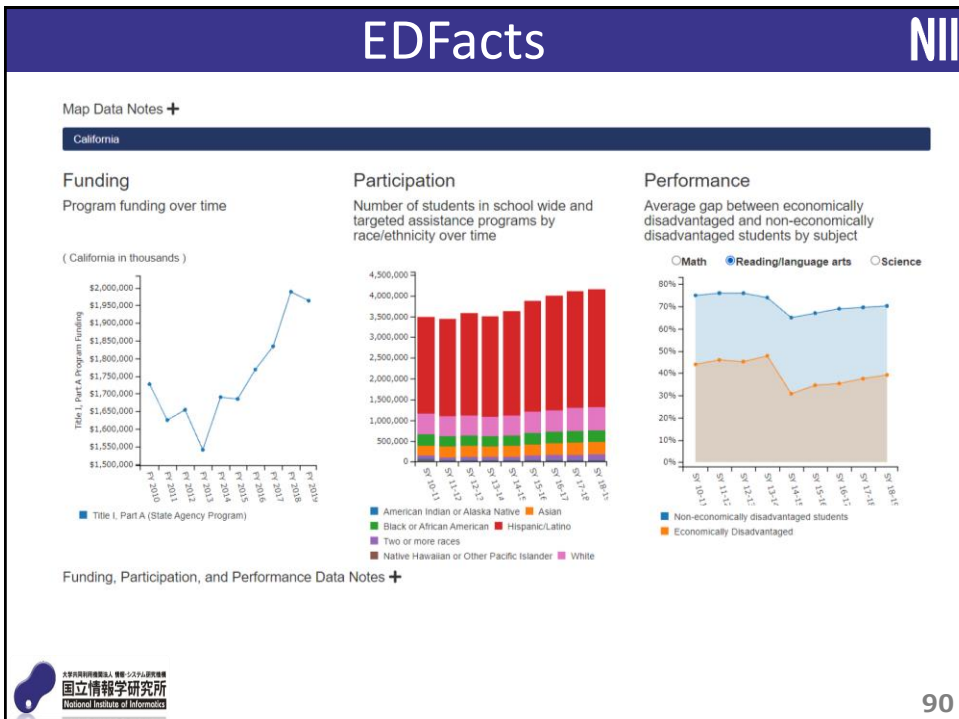
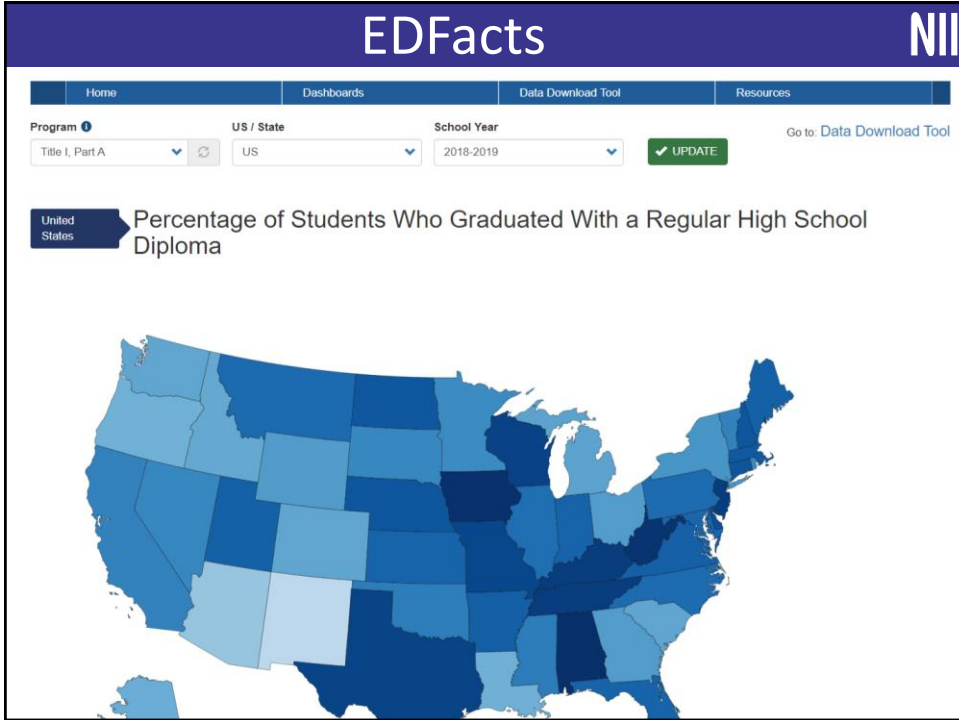
(備考) 1. 警察庁ホームページ「自殺者数」より作成。原数値。  
2. 令和2 (2020) 年分までは確定値。令和3 (2021) 年分は令和3 (2021) 年4月20日時点の暫定値。

男性微減 女性15%増 (19まで減少傾向)

## 次の国家を支える子供は？

### お話すること

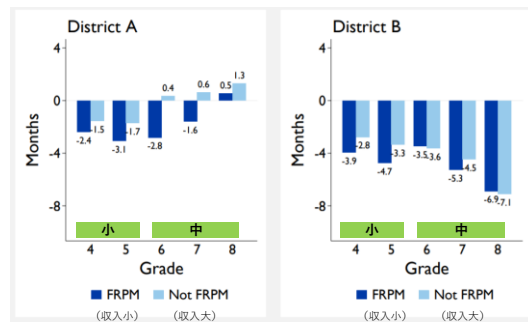
- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に



## NSLB: No students left Behind ESSA: Every Students Success ACT

## ジョージア州の分析事例：生徒の属性による違い（家庭の収入）

- 学区Aは、収入が低い世帯(紺色)の学力低下が大きい。収入の高い世帯(水色)は改善できている
- 学区Bは、収入に関わらず学力が低下している。小学校は収入によって低下度合いに差が見られる



小学生では、いずれの生徒も影響を受けているが、中学生では、収入が大きい生徒の学力改善が見られ格差が生じている

総じていずれの生徒も影響を受けている

FRPM (Free or Reduced-Price Meal Eligibility) :  
無料および割引価格で学校給食を提供される資格



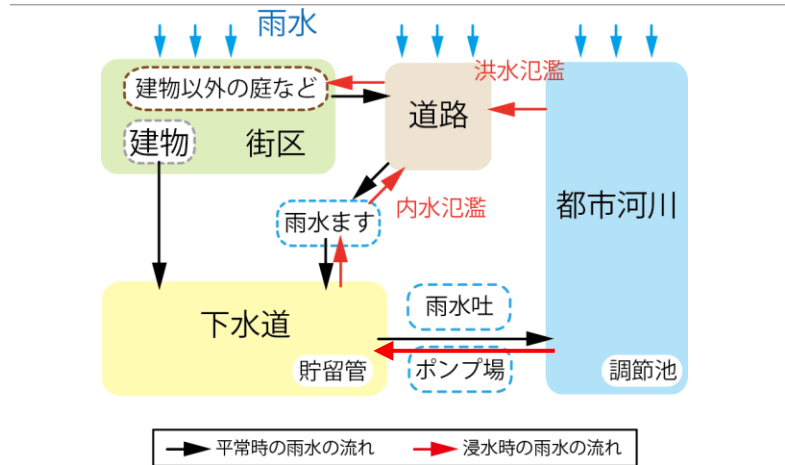
# 防災

NII

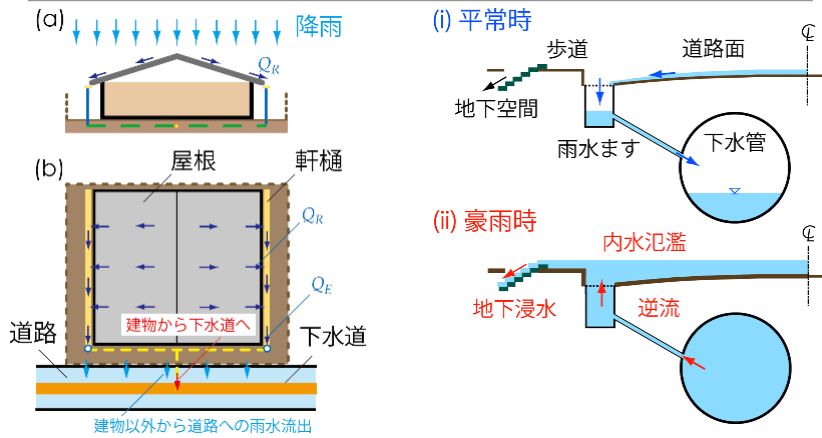
# DIAS

地球環境情報プラットフォーム

## 都市河川流域の水の流れの模式図

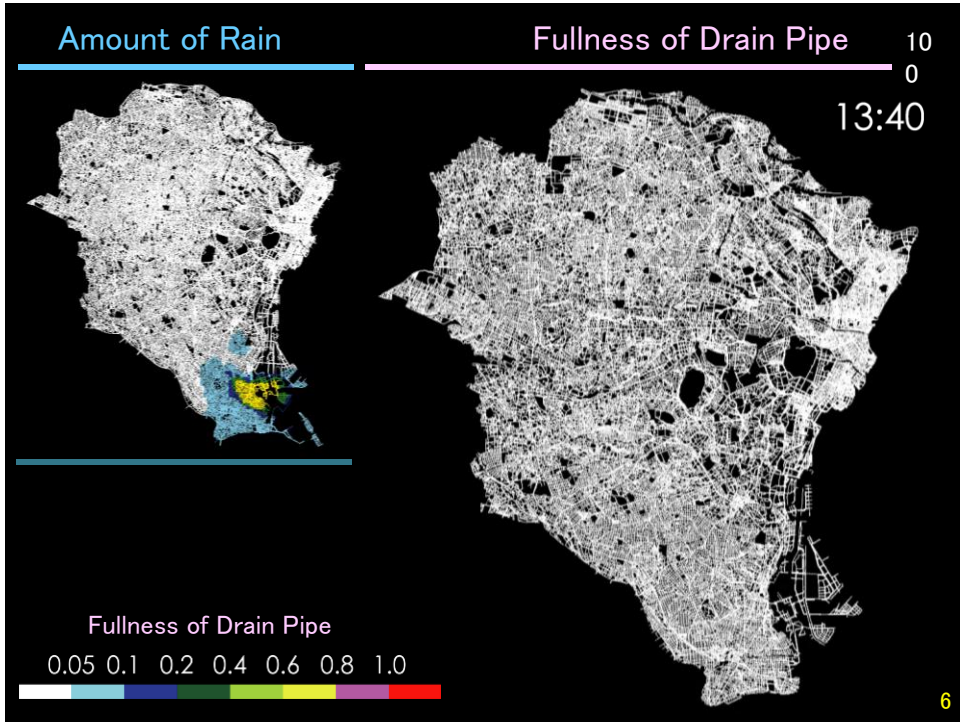


## 都市域内の雨水の流れの概略



街区内に降った雨水の流れ

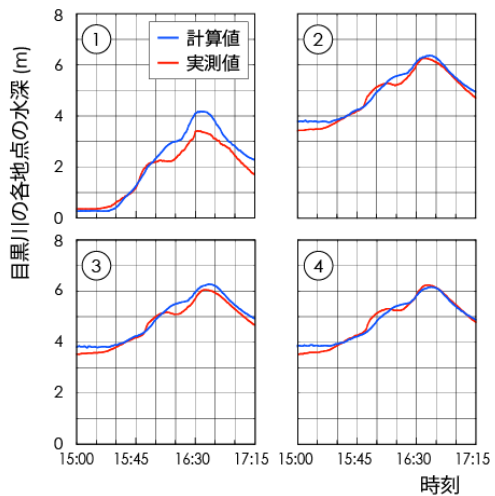
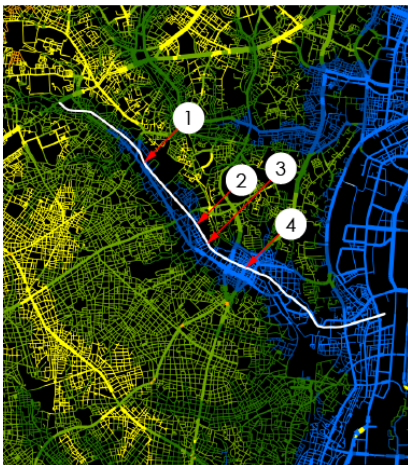
道路・雨水ます・下水道間の水の流れ



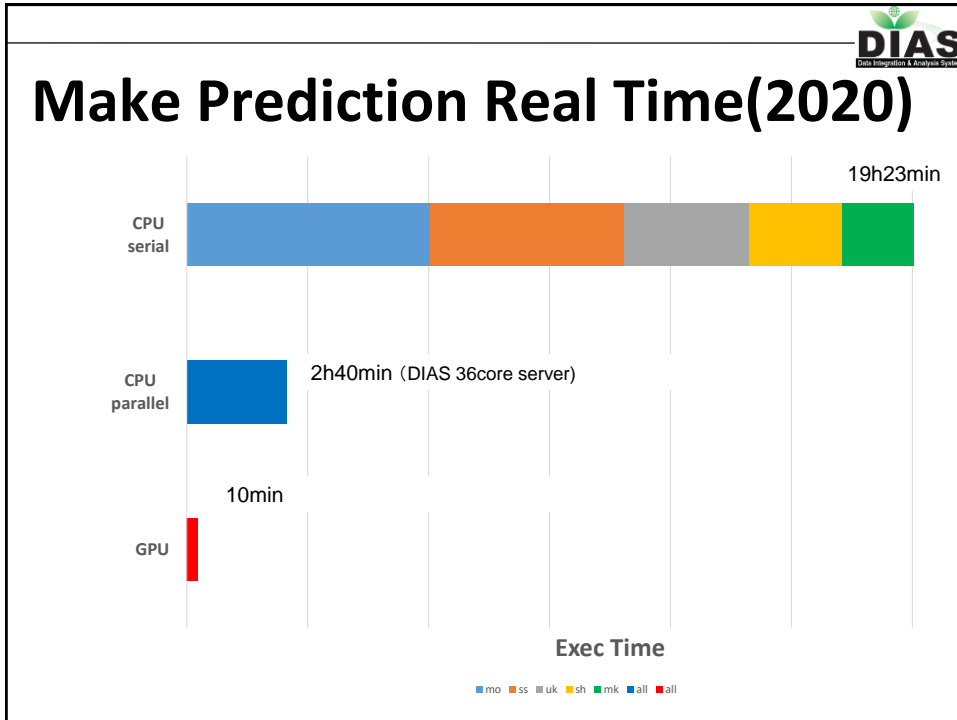
2. 予測手法の概要



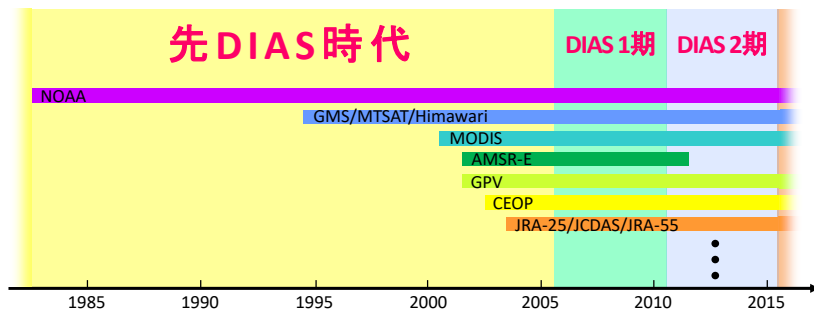
目黒川で計測された水深と本計算結果との比較



→ 非常に高精度での計算が可能



約40年：東大生研の地球環境データ  
リポジトリの開発の歴史  
DIAS：全体の開発期間の後半約3分の1



# 地球観測データ統合解析システム ≒ 100PB

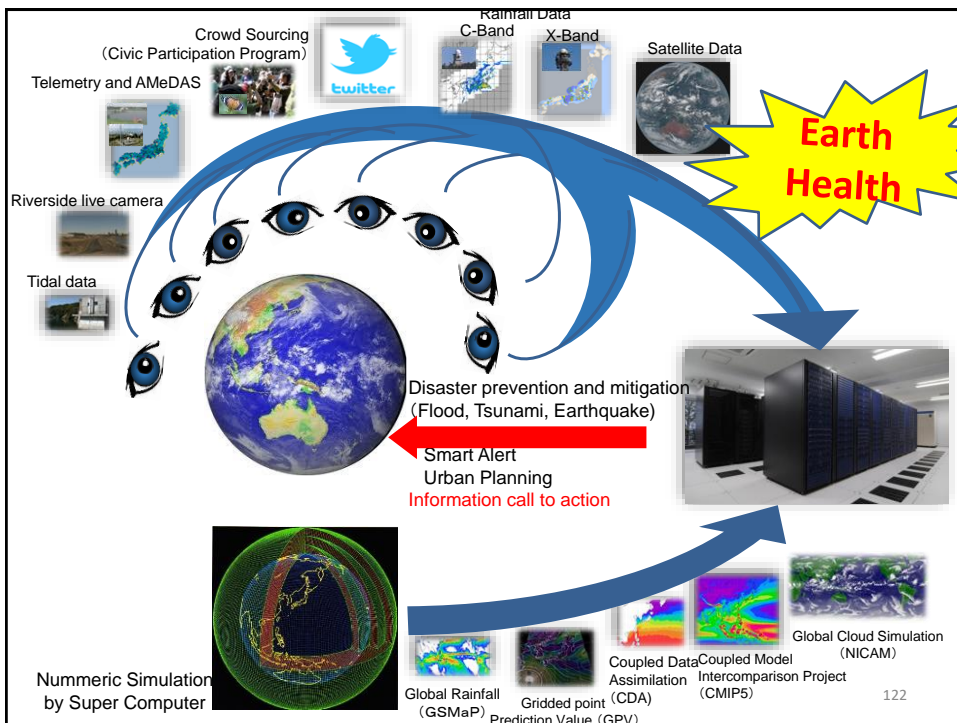


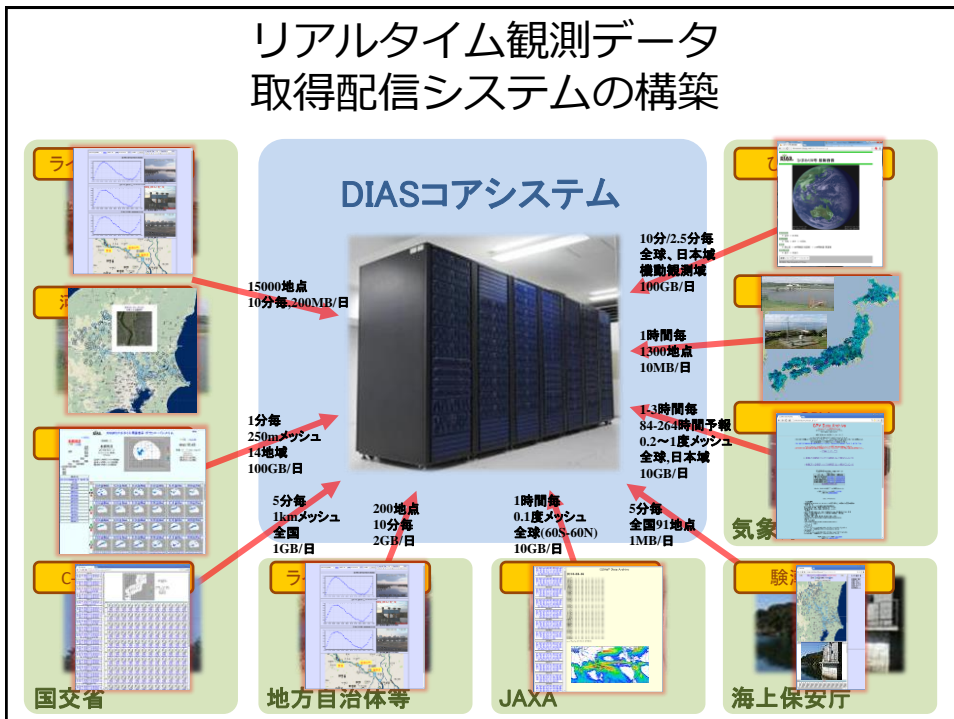
## DIASシステム構成



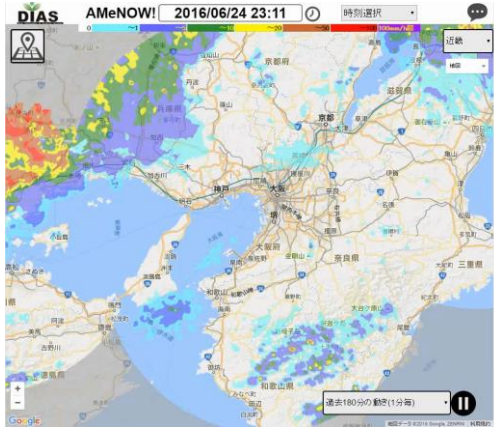
約100ペタバイト

Only one Weapon





## AMeNOW! リアルタイム降雨情報サービス



2015/10 公開  
降雨情報を地図上にリアルタイム表示  
PC,スマホ,タブ対応  
アカウント登録不要  
過去データ表示/動画表示/現在地表示  
/全国合成表示機能

<http://rain.diasjp.net/>

## 河川ライブカメラデータの活用

河川の増水・氾濫認識



## 河川の増水・氾濫認識

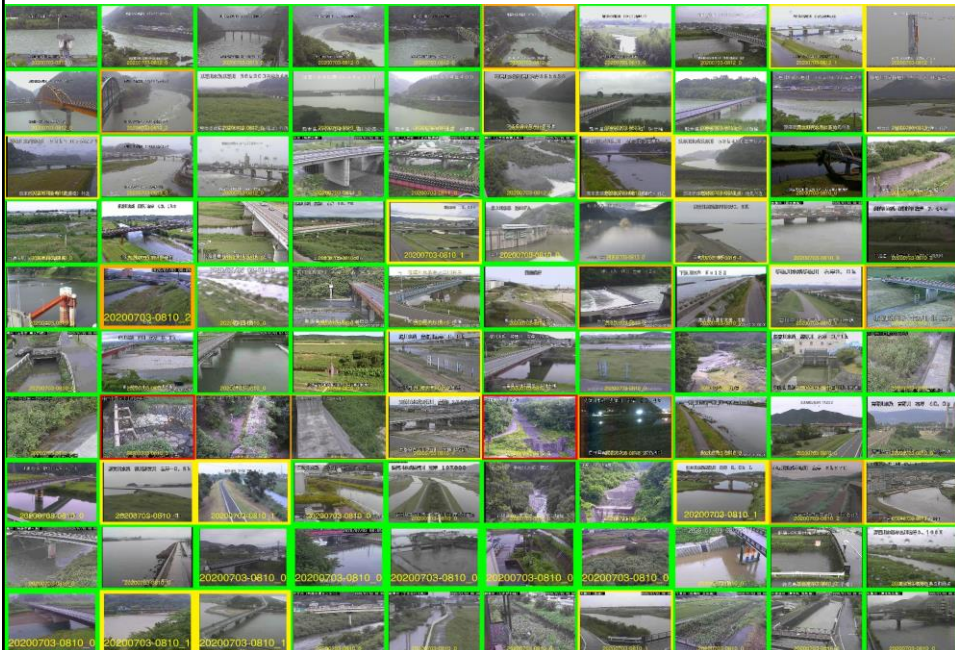
- 地球規模での気候変動の影響を受けて極端豪雨の頻度が高まり、河川の氾濫により流域に甚大な被害が多発
- DIASでアーカイブしている河川ライブカメラ画像にAI分析を適用して河川状態を自動認識し、事前に警報を出すなどして河川の氾濫被害を軽減することを目指す

球磨川の氾濫（熊本県人吉市）

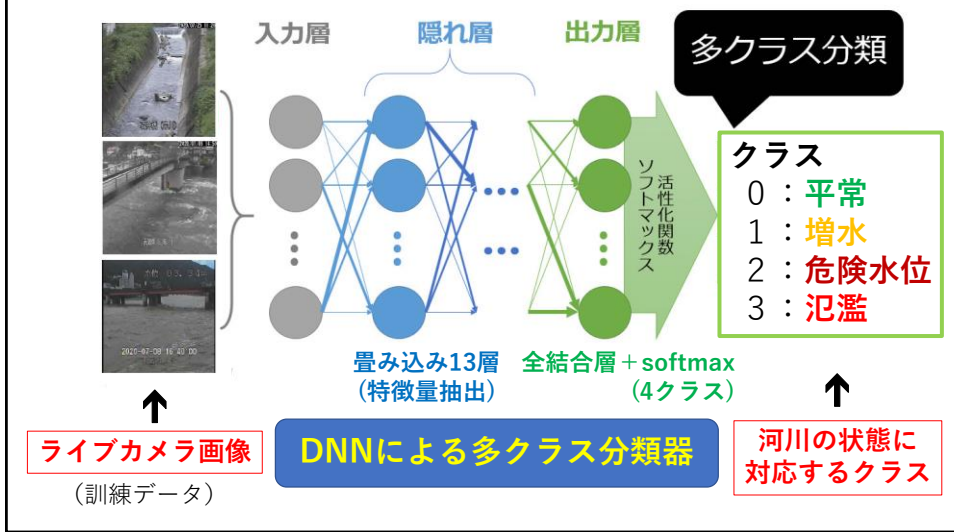


（毎日新聞 2020年7月4日）

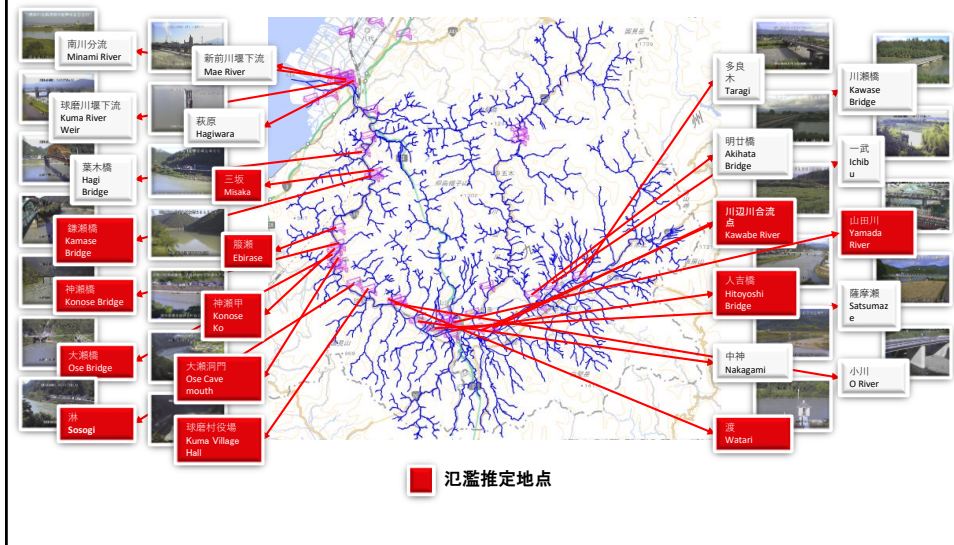
## 全国2718地点 河川ライブカメラ画像



# ライブカメラ画像を用いた氾濫認識 『多クラス分類問題』として定式化



## 球磨川水系ライブカメラ画像に 多クラス分類器を適用



# 深層学習を用いた氾濫画像の自動認識 (球磨川水系)

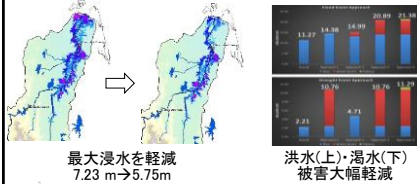


■ 平常(usual)    
 ■ 増水中(rising)    
 ■ 危険水位(critical)    
 ■ 氾濫(flood)

## スリランカ 気候変動・COVID-19下で 災害レジリエンス・持続可能性・包摂性を支える科学技術→質の高い成長14

灌漑局、気象局、災害管理センター、国家建築研究所、マハヴェリ開発庁からなる局長級プラットフォーム構築

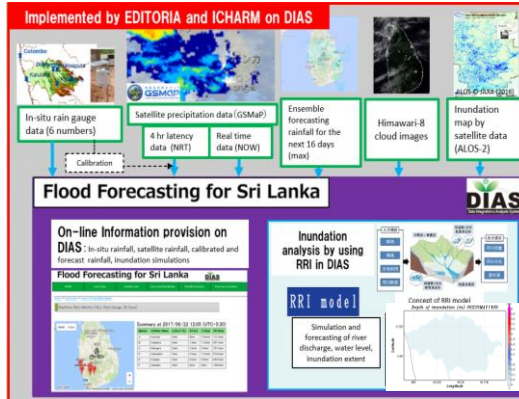
洪水被害軽減、農業生産拡大のマハヴェリ川の DIAS による効果的ダム運用基準を提案 (JICA 修士課程論文)



DIAS による知の統合オンラインシステム(OSS)構築 およびファシリテータ育成に具体的要請

- Flood Risk Mapping
- Flood Early Warning-
- Climate change adaptation planning-
- Economic damage assessment
- Contingency planning
- Capacity building for facilitators/stakeholders

2017年5月カル川大洪水時に DIAS による洪水予警報システムを2週間で開発→リアルタイム情報を提供し、洪水被害復旧に貢献



ヘルスケア

コロナAI 全滅！

# Covid禍でAIツール全滅！？

MIT  
Technology  
Review

Featured Topics Newsletters Events Podcasts

Sign In

Subscribe



AP

Artificial intelligence / Machine learning

**Hundreds of AI tools have been built to catch covid. None of them helped.**

Some have been used in hospitals, despite not being properly tested. But the pandemic could help make medical AI better.

by Will Douglas Heaven

July 30, 2021

The  
Alan Turing  
Institute



Data science and AI  
in the age of COVID-19

Reflections on the response of the UK's

結果：  
協力しないので  
少数の強いモデルではなく  
沢山の弱いモデル！

- Another problem Wynants, Driggs, and Mateen all identify is that most researchers rushed to *develop their own models, rather than working together* or improving existing ones. The result was that the collective effort of researchers around **the world produced *hundreds of mediocre* tools, rather than a handful of properly trained and tested ones.**

Key: Data Sharing

データ共有が肝！

そうしてみると、  
NII医療ビッグデータ研究セ  
ンタの取り組みは結構良  
かった？

Covid-19肺炎画像AI  
への挑戦

放射線学会、AMED、名大、NII


**順天堂大学**

---

**News & Information**
NO. 1  
2020年9月28日

---

**医療・健康**

新型コロナウイルスによる肺炎CT画像のAIによる解析手法開発  
 ～COVID-19胸部CT画像典型度分類において**83.3%**の精度を達成～


**順天堂大学**  
 Juntendo University


**名古屋大学**  
 NAGOYA UNIVERSITY


**NII**  
 国立情報学研究所  
 National Institute of Informatics

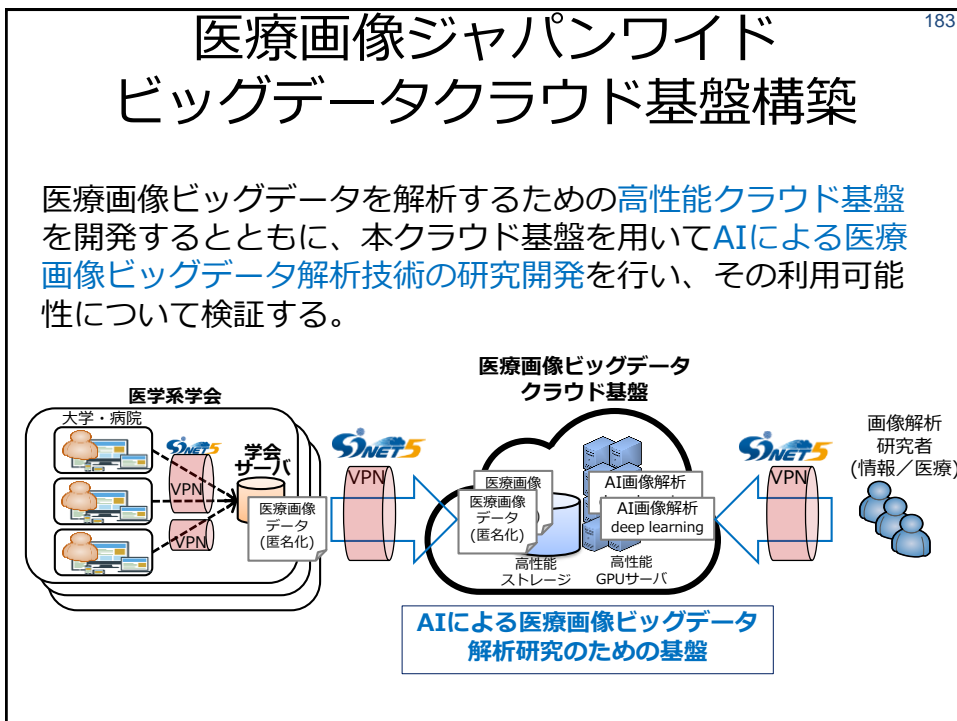

**JRS**  
 国立研究開発法人  
 日本医療研究開発機構

**概要**

順天堂大学は日本医学放射線学会所属の施設(東京大学、京都大学、大阪大学、九州大学、慶応義塾大学など)の研究代表機関として、国立大学法人東海国立大学機構・名古屋大学、国立情報学研究所などと共同で、新型コロナウイルス(COVID-19)肺炎CT画像をAIによって解析する手法を開発いたしました。人工知能技術(AI)を用いることで、胸部CT画像を入力すると、そのCT画像のCOVID-19肺炎典型度\*を判定できるようになっています。また、この解析を可能とするために、炎症などによって肺の形状がCT画像上で非常に識別しづらいような場合でも、AIが的確に肺の形状を推定できる手法も実現しています。今回開発された手法を用いたCOVID-19肺炎が疑わしい症例とそうでない症例の識別において、令和2年8月時点83.3%程度の典型度識別性能を達成しました。

本研究では、日本医学放射線学会と国立情報学研究所が中心となって進め、AMEDが支援するプロジェクト「臨床研究等ICT基盤構築・人工知能実装研究事業」「医療ビッグデータ利活用を促進するクラウド基盤・AI画像解析に関する研究」において構築・運用される画像情報収集プラットフォーム上に収集された、我が国におけるCOVID-19症例のデータベースが重要な役割を果たしています。ここでは国立情報学研究所が全国に展開する超高速学術情報ネットワークSINET5も大容量CT画像の収集とAI研究に大変役立っています。今後これらのシステムを活用して、さらなる精度向上が見込まれます。

\* ) 放射線診断専門医がCT画像を以下0-4つに分類した。1. COVID-19肺炎に典型的な所見を有するもの。2. 典型的とは異なる特徴的な所見で不確定なもの。3. 非典型的なもの。4. 肺炎の所見がないもの。これら2群に分けて、COVID-19肺炎典型度の高い疑わしい症例(1と2)とそうでない症例(3と4)とした。



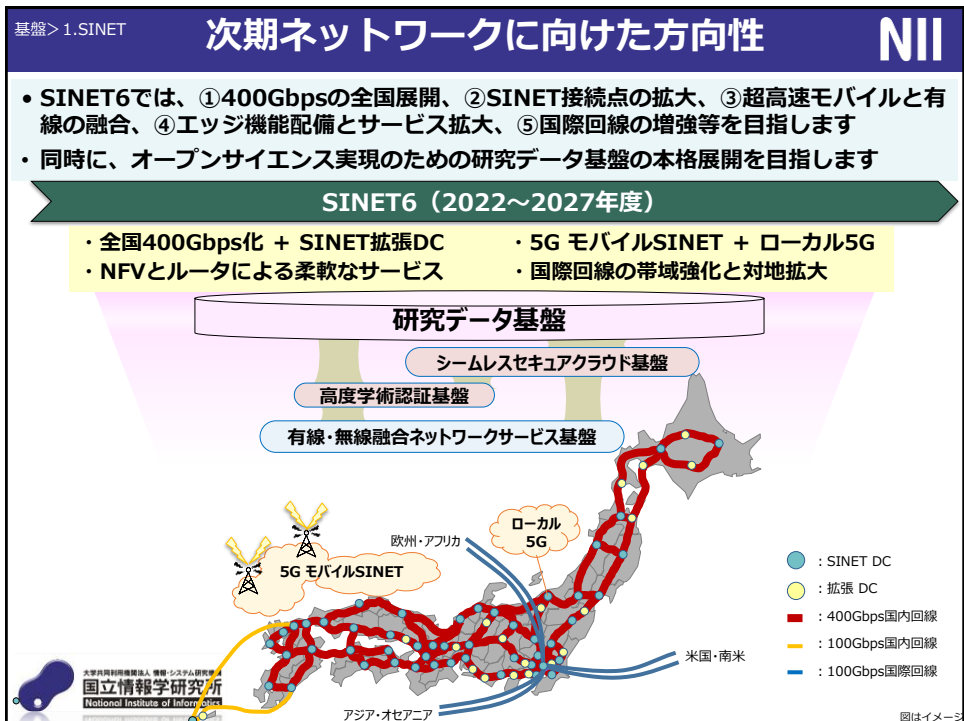
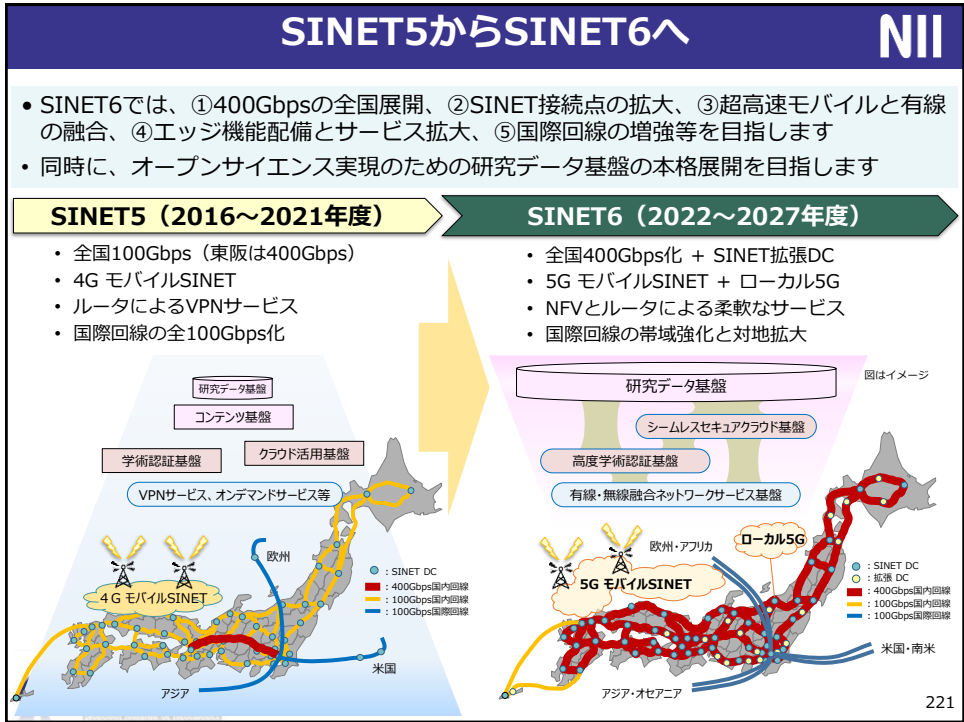


## お話すること

- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に

## お話すること

- ポスグレとの多くの接点
  - オープンソースの重要性
- NIIのご紹介（ちょっと）
- データの時代の到来
  - ゲームが変わった
- 事例紹介
  - 教育
  - 地球環境データ 100PB
  - ヘルスケア
- 社会との接点
- データガバナンスが肝となる時代に



## 計画の科学的意義

- 最先端の情報科学研究を幅広く実施し、次世代学術研究プラットフォームに反映
- 日本の学術研究全体を支える基盤として、ノーベル賞受賞を含む世界トップレベルの研究に寄与

### 次世代学術研究プラットフォーム

#### 研究データ基盤

新たな研究データ管理の責務に対応

① データガバナンス機能	⑤ キュレーション機能
② データプロビダンス機能	⑥ セキュア蓄積環境
③ コード付帯機能	⑦ 人材育成基盤
④ 秘匿解析機能	

#### ネットワーク基盤

高速化・高機能化

① 400Gbpsの全国展開と接続点の拡大
② 5Gと400Gbpsの融合
③ エッジ/NFVによるサービス拡大
④ 国際回線の増強

400G 超高速ネットワーク

225

## 次世代研究データ基盤: NII Research Data Cloud

2017-2021

2017年から開発開始  
⇒ 2021年から運用開始

2022-2027

データ管理基盤

データガバナンス機能  
データプロビダンス機能  
コード付帯機能  
秘匿解析機能  
キュレーション機能

データ検索基盤  
CiNii Research

データ公開基盤  
IAIR2 Cloud  
WEKO3

人材育成基盤  
セキュア蓄積環境

226

## 次世代研究データ基盤

- ① データガバナンス機能
- ② データプロビナンス機能
- ③ コード付帯機能
- ④ 秘匿解析機能
- ⑤ キュレーション機能
- ⑥ セキュア蓄積機能
- ⑦ 人材育成基盤

非公開

① Planning Experiment Collaboration Analysis Writing Curation Publication

データ管理基盤

公開

Preservation Discovery Idea

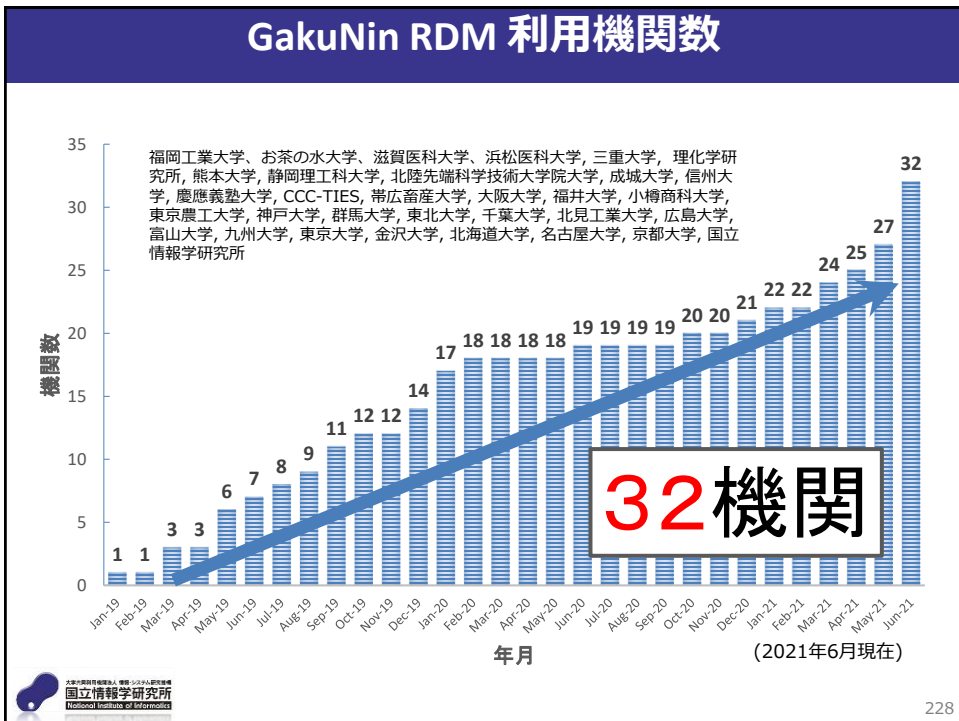
データ公開基盤

⑦

データ検索基盤

国立情報学研究所  
National Institute of Informatics

227



ご清聴ありがとうございました

