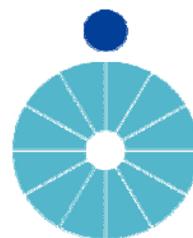


# ビッグデータに関する取組の 方向性について



文部科学省

MEXT

MINISTRY OF EDUCATION,  
CULTURE, SPORTS,  
SCIENCE AND TECHNOLOGY-JAPAN

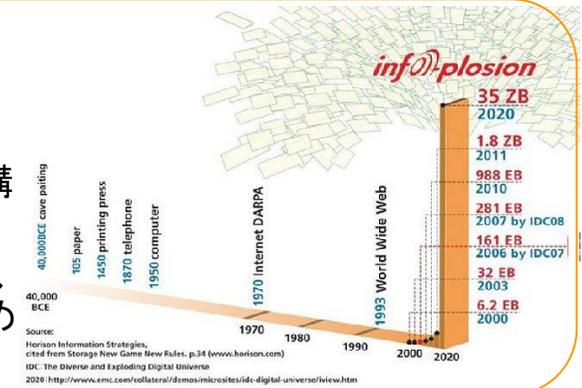
第4回ICT共通基盤技術検討WG

(平成24年6月28日(木)資料)

# ビッグデータを巡る状況

## 背景

- デジタル情報は、2020年までには約35ゼタバイト(2010年時の約35倍、DVDメディア約140億枚)へ拡大する見込みである。また、その膨大な情報(ビッグデータ)の中には、新たな知識や洞察が埋もれているといわれているが、その多くが整理・構造化されておらず、有効に活用できていない状況。
- 例えば、ライフサイエンス分野においては次世代シーケンサーの高性能化により、大量のデータが得られることになるが、大量データを活用して研究開発を進めるための情報分野からのさらなる対応が必須。



## 国内外の状況

- 米国では、平成24年3月29日にBigDataイニシアチブに関する公告を発表。このイニシアチブには6機関(NSF、NIH、DOD、DARPA、DOE、USGS)が総額2億ドル(185億円相当)を投資し、データへのアクセス、体系化、知見を集める技術を改善強化することが挙げられている。その他、欧州、アジアにおいても、ビッグデータに対する研究投資が行われている。
- 我が国は、各種センサー情報が発達していること、ハイパフォーマンスコンピューティング・自然言語処理などの領域において世界的に高い研究水準を保持していること、遺伝子情報など地域単位での研究が必要な必要が存在すること等から、大規模データの活用において、これらの強みが幅広い分野・領域に展開することで、我が国の科学技術の共通基盤の強化や産業競争力の強化が可能な環境である。

## さらなる研究が必要な項目

- 大規模データの蓄積・構造化・マイニング等のための研究開発
- 新たな洞察を導くデータの解析可視化技術の開発
- 文脈、意味・意図理解など高度なヒューマンインタラクションを実現するための研究開発
- シミュレーションとデータ科学の融合(データ同化)、先進数理的手法や機械学習アルゴリズムの開発
- アカデミッククラウド基盤構築のためのシステム研究
- アプリケーション側(バイオインフォマティクス、地球観測、防災等)と連携したプロジェクト研究 等

※異分野の研究者による研究チームがアプリケーションデータを共有し、異なる方向性から、ビッグデータに迫っていく研究アプローチ  
※異分野の研究者同士のインタラクションや協働が可能となるよう、機動的な研究チームを編成する研究体制  
※上記の取り組みを通じて、ビッグデータに関係する研究者に流動的なネットワークを生み出し、新たな人材育成スキームを構築

# アカデミッククラウドに関する検討会について

## 背景・概要

第4期科学技術基本計画においては、科学技術振興のための基盤として、「知識インフラ」の形成が盛り込まれている。情報化社会の進展に伴い、「情報爆発」と呼ばれるような大量のデータ・情報がインターネットを通じて流通している今日、様々な分野における知的活動の成果として生み出されている大量データを、効果的・効率的に収集、集約し、革新的な科学的手法により情報処理を行うことで、新たな知的価値を創造する、「データ科学」の重要性が認識されてきている。

「データ科学」は、第四の科学的手法として注目されている。

本年4月に有識者から成る「アカデミッククラウドに関する検討会」を文部科学省研究振興局に設置。本年夏の概算要求の時期までに議論し、中間的なまとめを行って、文部科学省の施策に反映させる。

## 主な検討事項

### (1) データベース等の連携

大学等の研究者が、全国に所在し、多分野にわたる様々なデータベースや研究資料を有機的に連携し、効果的かつ効率的に、研究に活用できるよう、①多分野にわたる異なるデータベース、リポジトリ等の連携、②その前提となるデータベース化、デジタル化・アーカイブ化の推進、③データ検索やデータマイニング等の情報処理システムとの連携について

### (2) システム環境の構築

上記(1)のような考え方で、有機的に連携するデータベース等を、研究者が手軽に一つのシステムとして利用でき、効果的・効率的な研究活動を行うことのできるシステム環境のあり方、データを取り扱う際の倫理規範、及び環境構築について

### (3) システムを利用して行うデータ科学の高度化に資する研究開発

上記(2)のシステム環境を活用して行うデータ科学の高度化に資する、情報検索、データマイニング、統計数理的手法等の 情報科学技術の研究開発について

## アカデミッククラウドに関する検討会 構成員

(主査)西尾 章治郎 大阪大学情報科学研究科教授  
(主査代理)喜連川 優 東京大学生産技術研究所教授  
相原 玲二 広島大学情報メディア教育研究センター長  
安達 淳 国立情報学研究所 学術基盤推進部長  
安達 文夫 国立歴史民俗博物館情報資料研究系教授  
北川 源四郎 情報・システム研究機構長  
五條堀 孝 国立遺伝学研究所副所長  
竹内 比呂也 千葉大学附属図書館長  
美濃 導彦 京都大学学術情報メディアセンター教授  
門田 博文 科学技術振興機構執行役  
安浦 寛人 九州大学理事  
山口 しのぶ 東京工業大学学術国際情報センター教授  
山名 早人 早稲田大学理工学術院教授

### ○米国との共同研究

#### 平野文部科学大臣とスレッシュNSF長官の会談において、ビッグデータ及び災害に係る研究協力について合意

MEXT and NSF Statement on Big Data and Disaster Research Collaboration  
From NSF Director Dr. Subra Suresh and MEXT Minister Mr. Hirofumi Hirano



(仮訳)

東日本大震災やそれに伴う津波、原子力発電所事故、またメキシコ湾原油流出事故、ハリケーン・カトリナ、地域的な干ばつ、洪水、火災に代表されるように、近年繰り返し壊滅的な自然・人的災害が起こっている。

このような災害や事故の経験から、基礎的な理解やイノベーションの創出のための基礎研究の推進が喫緊の課題であることが明らかである。ビッグデータによる革新は、適切な情報にリアルタイムでアクセスできることにより、これらの災害の影響を削減する可能性を有している。

平野博文文部科学大臣とスブラ・スレッシュ米国国立科学財団(NSF)長官は、平成24年6月5日に東京で会談を行い、日米間における災害研究協力の重要性を合意した。

日米各々の経験、専門知識を駆使することで、両社会の脆弱性を軽減し、社会のレジリアンス(回復力)を高め、日米相互に有益な効果をもたらさる。基本的合意が得られたのは、災害に対する堅牢性(ロバストネス)及び回復力(レジリアンス)の強化に関し、ビッグデータを通してコンピュータ科学、工学、社会科学、地球科学、といったように幅広い分野での研究協力支援である。

協力が期待できる研究分野の例は、以下の通りである。

- ・災害から得られた大量のデータを活用して、分析、モデリング、計算分析的な能力やハザード確率モデル等のアプリケーションの高度化を行う。
- ・情報技術のレジリアンスや応答力を改良し、即時意思決定に必須である、リアルタイムなデータセンシング、可視化、分析、予測を可能にする。
- ・レジリエントで持続可能な社会基盤と分散基盤網の実現のために、基礎知識とイノベーションを推進する。
- ・人、社会、経済、環境の各局面において、ビッグデータの獲得と、人的、社会的、グローバルスケールでの災害準備とその対応力についての理解を深める。
- ・緊急時の準備と対応に関し、多様な学問分野、エンドユーザからの入力、すべての情報源からの大量データ、それぞれの知見を統合する。

文部科学省とNSFは、平成24年末までに、より詳細な合意を発表することを目的として、実務レベルでの行動計画を進展させることに合意した。

# アカデミック分野におけるビッグデータに関する取組の方向性

研究開発

## ①ビッグデータに関する共通基盤技術の開発

- ・ビッグデータ処理の各段階(a.データ収集、b.蓄積・構造化、c.分析・処理、d.可視化)における基盤技術の開発(※基盤技術を開発する上では、実データが必須であり、以下の具体的分野やカウンターパートと連携・協業することによる研究開発を実施)
- ・アカデミッククラウド環境構築に関するシステム研究

### 分野間連携

#### ○アプリケーション側(ライフ、地球環境等)との協働によるプロジェクト研究

ライフ分野:

- ・効率的な相同性検索、ゲノム等情報から疾患の要因同定
- ・臓器等の数理モデルと実測の融合(データ同化)

地球観測分野:

- ・複数パラメータの処理技術、高度圧縮技術
- ・異なるデータ間の連携技術 等

### 国際連携

#### ○国際共同研究

- ・災害に対する堅牢性(ロバストネス)及び回復力(レジリエンス)の強化に関し、ビッグデータを通じた研究協力
- ・それぞれの国の災害経験及びその対処等を踏まえた共同研究

### 人材育成

## ②独立行政法人等における、ビッグデータ活用モデルの構築

- ・研究開発法人等が所有する専門データや死蔵されている膨大なデータの掘り起こしを行い、研究に活用できるよう、ルールを含め整備
- ・複数の研究機関が有するデータに関するデータベース連携の推進、これらのデータの民間等での利活用の推進
- ・データを活用・分析できる人材を育成

基盤構築

# ビッグデータに関する共通基盤技術の開発による新たな知の創造と社会的・科学的課題の解決（イメージ）

未来

- 新たな知の創造
- 科学技術イノベーション創出

- 社会的・科学的課題の解決  
（具体的な例）
  - ・消費エネルギー、CO<sub>2</sub>排気量削減
  - ・復興の高速化、必要な支援の見える化
  - ・病気や疾患の早期発見・予防
  - ・地震・津波等災害の予測・防災

ビッグデータの構造化・体系化、分野間のデータの融合

## ビッグデータに関する 共通基盤技術

人材  
育成

国際  
協力

グリーン

ライフ

復興

地球  
観測

防災

...

大量に散在する多様な学術情報、実世界データやセンシングデータ、WEB上のデータ等

独立行政法人等における  
ビッグデータ活用モデル構築

現在

分野