e-研究インフラの海外動向

―これからの科学技術の 国際競争の分け目か?

国立情報学研究所 船守美穂

2017年12月15日

問題提起

- □ 欧米ではデジタル時代に鑑み、e-研究インフラ の重要性が強く認識され、整備が進められて いる。
- □ これに対して日本は、このような認識が薄く、 物理的研究インフラへの投資はなされている が、e-インフラについては手薄である。
- □ このままでは日本は、科学技術の国際競争に 遅れを取るのではないか?

目次

- 1. 欧州におけるe-研究インフラ整備の 動向
- 2. 日本におけるe-研究インフラ整備の 動向
- 3. 今後の展望一各分野の研究力を強 化するためのe-研究インフラを求めて

1

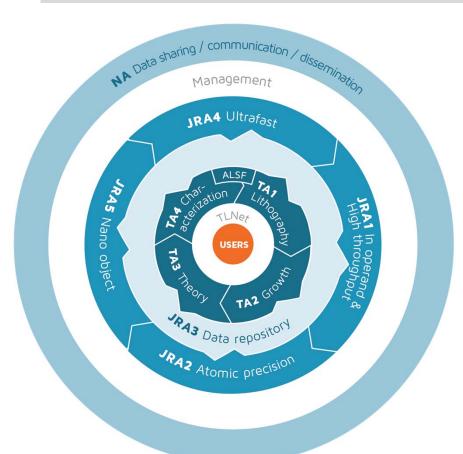
欧州におけるe-研究インフラ整備の動向

- > 事例紹介
- ▶ EU•各国政策

NFFA-Europe



nanoscience foundries & free analysis (1)



□ EU内の複数の測定機器により得られた測定値を一つのリポジトリに集め、データ統合、解析するツールを提供する。

(利用場面)

- → 一研究者が特定サンプルの計測のために、複数の施設を利用。データをNFFAのデータリポジトリに保存。
- → 研究グループがデータを共有、共同解析
- ▶ 他の研究者が、研究データを利用、解析

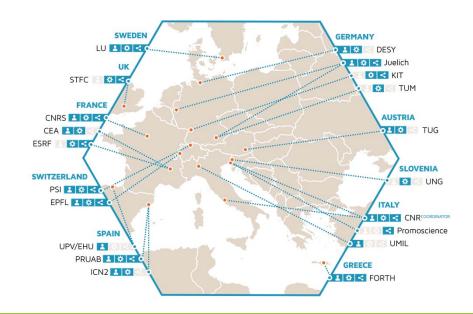
Industries Scientists & other Stakeholders

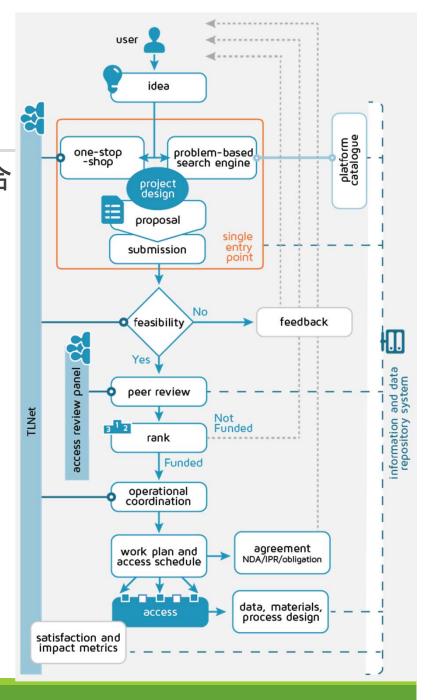
(X線回折、NMR、質量分析、電子顕微鏡、微細加工···)



NFFA-Europe (2)

- □ ユーザは、欧州に点在する測定・合成機器の利用申請を実験計画とともに提出。
- □ 採択となった場合、NFFAのプラットフォームも含めて利用可能となる。





European and Mediterranean Plant Protection Organization

Organisation Européenne et Méditerranéenne pour la Protection des Plantes

欧州•地中海植物保護機関(国際機関)



- □ 国際植物防疫条約(IPPC)に基づき、欧州・地中海地域51カ国について設置された、植物保護の協力・調和機関。
- □ 事務局は20名程度だが、研究 者は計4500名規模。
- □ 各観測点における植生や害虫、 天候等の情報をデータベース化。
- □ 各種基準や標準のガイドライン を作成。

〈データベース〉

- EPPO Global DB & EPPO codes
- EPPO Standards on Efficacy Evaluation of Plant Protection Products
- PQR EPPO database on quarantine pests
- EPPO database on Diagnostic expertise
- CAPRA (Computer Assisted Pest Risk Analysis)
- EPPO Pest Reporting System (restricted access)
- EPPO Plant Protection Thesaurus

〈EPPO標準〉

- EPPO Standards on plant protection products (3標準)
- EPPO Standards on phytosanitary measures (10標準)

Source:EPPO・Damian Griessinger氏(IT担当)より情報入手

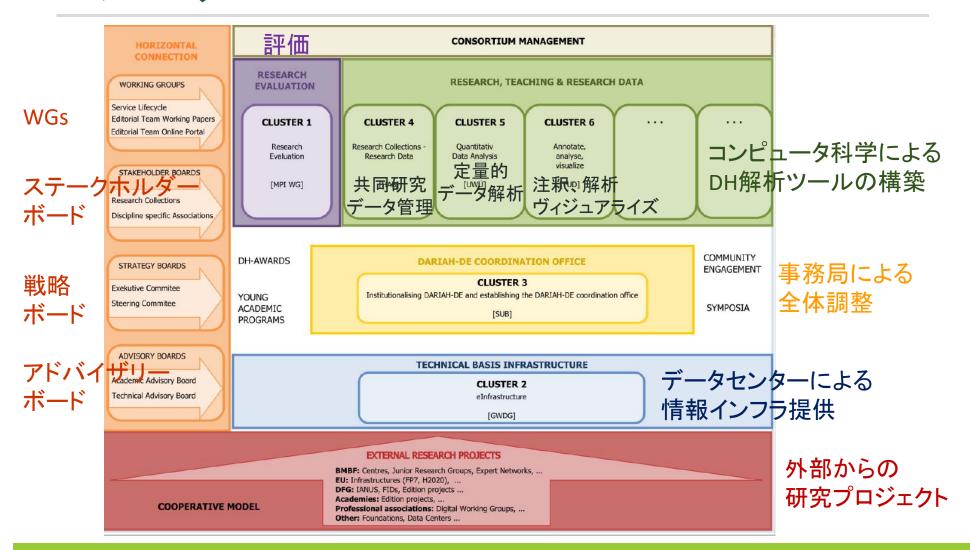
DARIAH

...Digital Research Infrastructure for the Arts and Humanities

□ EUのデジタル・ヒュマニティーズの取り組み

- > ESFRI (European Strategy Forum on Research Infrastructures) プロジェクトの一部
- ▶ 重点領域として、<u>長期助成</u>(数十年)がなされる。
- □ DARIAH-EUのもとに、各国DARIAHがある。
 - ➤ 各国DARIAHは、各国政府により助成される。
 - ドイツ: インフラ提供
 - フランス: 研究データ、メタデータcrosswalk
 - その他: DHツール、レポジトリ提供、WS実施、コンテンツ 収集

DARIAH-DE …ガバナンス



DARIAH-DE(ドイツ)

□ドイツ連邦政府による強い推進と助成

- ➤ KITを含む20機関程度が参加
 - 人文系の研究組織+情報科学+データセンター
 - DARIAHフルタイム約30名、関係者は計60名程度
- ▶ 現在、人文系で最も伸びている分野

■ DARIAHの主にインフラ整備面を担当

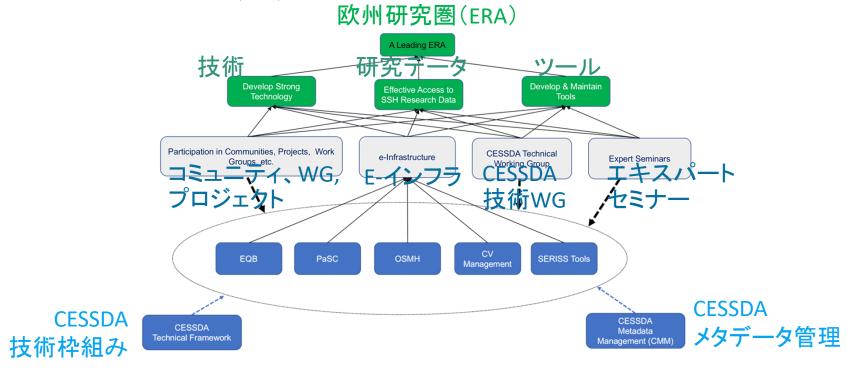
➤ 教育プログラムやWS実施、データ利用のためのサポート、 共同研究等を含む

□ドイツの特殊事情

- ▶ <u>州をまたがった共同研究の場合、各大学でデータを共有できない。</u>
 - ⇒EUレベルのリポジトリが必要

CESSDA...Consortium of European Social Science Data Archives

□欧州各国の社会科学系データアーカイブを連携させ、横断検索可能とするプロジェクトが進行中。 Technical Vision, Goals, Actions



Helmholtz Data Federation

...概要

- □ ヘルムホルツ協会が構築する、「研究データインフラ連合 (federated research data infrastructure)」
 - ▶ 協会内の複数の大型実験研究施設のデータが、統合的に管理される。
 - ▶ ドイツ国内で完全オープンにデータ提供。
- □ 欧州オープンサイエンスクラウド(EOSC)の一部として構築
 - ▶ドイツ分のEOSCとしてEUに提案する。
- □ HDFの柱
 - 1. データ管理と分析のための革新的ソフトウェア技術
 - 2. 優れた<u>ユーザサポート</u>と共同R&D
 - 3. 先進的ストレージと分析のハードウェア

データ利用希望者と共 同研究で、データ利用 の整備・便宜を図る。

Helmholtz Data Federation

...参加機関

Helmholtz Data Federation (HDF)

- Initial set of scientific cases identified by the six Helmholtz partners
- Reviewer statements:
 - "Sharing of infrastructure, services and datasets opens the way for improved cross-disciplinary working and re-use of data."
 - "...will be a first example of a national shared research data infrastructure in Europe."



極域氷床、海洋研究

素粒子物理

ゲノム研究、放射治療

バイオメド、植生

健康科学、原子核物理

気候、エネルギー、 宇宙物理

27 03.04.2017

SCC Steinbuch Centre for Computing

Helmholtz Data Federation ...メリット

□ データ利用者

- ▶ 様々な大型実験の研究データの利用が可能
- ▶ ユーザサポートもあり
- ▶ データ解析環境もあり

□ 参加研究機関

- ストレージを得られる
 - 各実験施設は歴史的に自身のストレージはあるが、年々、この維持・更新が 負担となってきている。
 - ストレージをKITに集約することがヘルムホルツ協会としての方針のため、各 実験研究施設でストレージを調達できなくなる。
- データ管理をKITに任せることができる
 - 各実験施設はデータ管理は手放し、個々の分野の仕事をしたい。
- ▶ <u>学際領域研究</u>等で外部資金獲得が可能となる

European Commission (2016)...Open Innovation, Open Science, Open to the World

Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process.



European Commission (2016)...European Open Science Cloud (EOSC)

■ EOSC aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market.

KEY FACTORS:

- New modes of scholarly communication
- Modern reward and recognition practices need to support data sharing and re-use.
- Core data experts need to be trained and their career perspective significantly improved.
- A real stimulus of multi-disciplinary collaboration requires specific measures in terms of review, funding and infrastructure.
- The transition from scientific insights towards innovation needs a dedicated support policy.



ESFRIにおけるe-インフラの重視(1)



Professor John Womersley

前ESFRI議長 John Womersley

A lot of the research infrastructures are really data infrastructures or networks that bring together datasets from existing facilities or large single-sited infrastructures that generate very, very large amounts of data. It is no longer possible or sensible to simply look at the facility without also considering the data handling.

The emerging "e" component of research infrastructure

- ESFRI: The European Strategic Forum for Research Infrastructures
 - ▶ 欧州域内において戦略的に整備・運営していくべき研究インフラについて欧州連合理事会に助言する、極めて影響力の大きい機関

欧州において戦略的に整備していくべき研究インフラについてESFRIが設計面で課す最低要件

□ 科学的卓越性

- ▶ 長期的科学プログラムが明確である。
- ▶ 科学コミュニティが確立している。
- ▶ 科学におけるリーダーシップが明確である。

□ 汎欧州としてのレレバンス

- ▶ 当該科学領域における汎欧州のアプローチが明確である。
- ▶ 対象となるユーザ・コミュニティが汎欧州である。
- ▶ 各国または国際的な施設が相互補完する、あるいはシナジー効果を有する。

□ 社会•経済的効果

▶ 社会的課題との関係性が明確にされ、経済的効果が予測されている。

□ e-二一ズ

- ▶ アクセス方針、セキュリティ対策を含む、e-インフラに対する要求のヴィジョンが出来ている。
- ▶ コミュニケーションネットワークや分散コンピューティング、HPC/HTCとの連携がある。

Strategy Report on Research Infrastructures ROADMAP 2018

ESFRIにおけるe-インフラの重視(2)

INVESTMENT STRATEGIES IN e-INFRASTRUCTURE (1)

ESFRI recommendation on coordination of Member States' investment strategies in e-infrastructures

- ESFRI aims to contribute to an effective and efficient approach to e-infrastructure and services (vertical and horizontal) for European science and its competitiveness in th global scene building on existing Research Infrastructures and electronic Infrastructures.
- ESFRI advises to establish urgently a *convergent policy of funding mechanisms* for e-Infrastructures at the various levels (institutional, regional, national, European). Such policy could include *support and financing of e-Infrastructures* for scientific users, providing incentives to researchers to generate *FAIR and reproducible (+R) data*, as well as the *development* of enabling e-tools/e-technologies and the *mainstreaming* of support actions addressing e-needs of all levels of intervention.
- ESFRI suggests to act as **STRATEGY FORUM OF FUNDERS** of the e-Infrastructuresfor European science as a key element of support of a coherent approach to policymaking on research infrastructure in Europe.
- ESFRI could be effective in facilitating the coordination of national and European efforts in e-Infrastructures for research and innovation, following the model that has been put in place for Research Infrastructures. The global dimension of e-Infrastructures shall also be addressed in a coherent way.



ESFRI

Strategy Report on Research Infrastructures ROADMAP **2018**

ESFRIにおけるe-インフラの重視(3)

INVESTMENT STRATEGIES IN e-INFRASTRUCTURE (2)

ESFRI recommendation on coordination of Member States' investment strategies in e-infrastructures

- ESFRI advises to proceed by *building on the EOSC HLEG vision* and by *strengthening the data*FAIR+R generator role of Research Infrastructures and the coordination of science and innovation communities also at broad international level as described by the e-IRG roadmap.3
- ESFRI advises that *urgent actions* must be taken to support the *training and hiring of e-infrastructure experts/scientists* and to expand the *data literacy* at all levels of education and innovation activities to enable the return from the investment in e-Infrastructures and Research Infrastructures and maximize societal benefits.
- ESFRI advises that *deploying strategic joint e-services*, including standardization and assisted open access, can be *most efficiently done at least at European level*.
- ESFRI recognizes that the capacity of distributed High Performance and High Throughput Computing, of communication networks for scientific data, and of transversal e-services, should **be strengthened by building on national resources** and adding an effective coordination and a possible additional funding such to realize a full European-scale impact.
- ESFRI recognizes that the funding approach to *HPC Tier-Zero facilities* bears challenges similar to more centralised large-scale Research Infrastructure facilities 4 with direct industrial policy implications and should be dealt with accordingly.





e-IRG...概要(1)

- □ e-IRG正式名称
 - > E-Infrastructure Reflection Group
- □ ミッション
 - ▶ 欧州域内の国・地域のe-インフラとサービスの統合を促進するための、戦略的グループ
 - ▶ 欧州のe-インフラが整合的・革新的・戦略的であるように、政策立案および統合的でサステイナブルなe-インフラサービスの開発を支援する。

e-IRG...概要(2)

□参加者

- > 各国の政策担当者(教育省、科技庁)
- ▶ E-インフラのサービスプロバイダ

欧州委員会(EC) ではなく、科学技 術担当として 各国調整

□ 活動内容

- ▶ 定期的に各国担当者と会議を開催し、その時々のe-インフラに関わる議題について議論。
 - EOSC、ESFRI、e-Infra Commons等
- ▶ 2年に一度程度のペースで、白書を出し、e-インフラの整備方針についてのガイドラインを提示する。
 - 強制力はないが、政策担当者がこれに基づいて自国の政策を誘導することができる。
- ▶ E-インフラに関するナレッジベースを整備・提供
- ➤ 近年、ESFRI(European Strategic Forum for Research Infrastructure) からの要請で、欧州域内のe-インフラのあり方について助言

e-IRG...e-Infrastructure Commons

e-Infrastructure Commons

- ➤ e-IRG Roadmap 2012にて提示
- ユーザが、e-インフラを組み合わせて使用するにあたり頻繁に、データの移動、保存、処理という作業に煩わされ、本務に集中できないことは問題。
- ▶ 欧州における知識・イノベーション・科学の創出のための単一のe-インフラ・コモンズの必要性を説く。
- ▶ 欧州における分散的な電子リソースを、費用対効果高く共有するための手段として提示。
- 変化し続ける研究や科学に対する要求に継続的に適合しながら、欧州域内の研究者がアクセスできるオープンでユーザフレンドリーな、各種リソースが統合されたエコシステム。
- ➤ このアイディアが、EOSCへと発展

e-IRGが選定した、e-研究インフラの色彩の強いESFRIプロジェクト

BBMRI: Biobanking and Biomolecular Resources Research Infrastructure (バイオバン キング) CLARIN: Common Language Resources and Technology Infrastructure(言語資源) CESSDA - Council of European Social Science Data Archives (社会科学データアーカ DARIAH - Digital Research Infrastructure for the Arts and Humanities (デジタルヒュマ ニティーズ) ECRIN: European Clinical Research Infrastructures Network(臨床研究) ELIXIR — European Life Sciences Infrastructure for Biological Information(生物情報) e-VLBI: Very Long Base Interferometry in Europe (VLBI) ESRF: European Synchrotron Radiation Facility(シンクロトロン放射光施設) EuroFEL: Free Electron Lasers of Europe (自由電子レーザー) KM3NET - The Cubic Kilometre Neutrino Telescope (ニュートリノ) LIFEWATCH – e-Science and technology infrastructure for biodiversity data and observatories(生物多様性) Preparing for SKA – Square Kilometre Array (SKA) European XFEL - European X-ray Free Electron Laser(X線自由電子レーザー)

e-IRGが選定した、特にデータ管理を必要としているESFRI対象プロジェクト

- BioMedBridges: Building data bridges between biological and medical infrastructures in Europe (生物一医学データ連携)
- CRISP Cluster of Research Infrastructures for Synergies in Physics (物理学)
- □ DASISH Data Service Infrastructure for the Social Sciences and Humanities(人文社会科学)
- ENVRI Common Operations of Environmental Research Infrastructures (環境研究)

各国のe-研究インフラに向けての イニシアティブ

□ 英国

- 英国研究会議協議会(RCUK)が「e-インフラ・ロードマップ」を発表。
 - さまざまなe-インフラを融合的に扱うことの重要性を指摘

□ ドイツ

- 独・情報インフラ・カウンシル(RfII)を2014.11に設置
 - 科学者ユーザコミュニティ」、「社会」、「e-インフラ運営主体」、「連邦政府、州政府」 の4つのグループからそれぞれ複数名がメンバー

□ オーストラリア

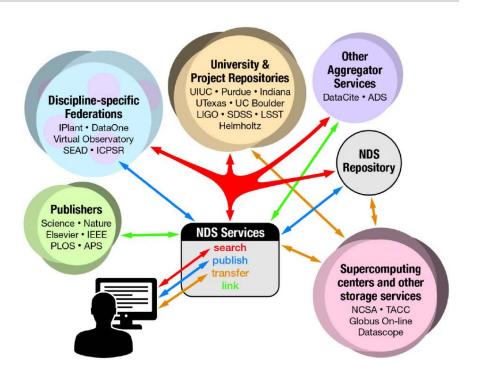
- ▶「e-リサーチ戦略と実施枠組み」が2006年に発表。これに基づき豪州国家 データサービス(ANDS)を形成。
- ▶ 豪州国家研究インフラ・ロードマップ(2016)で、「豪州研究データクラウド」 を構築し、データ集中、分野横断、グローバルな共同研究に対するニー ズに応える予定。

□ 米国

▶ 米・National Data Service (NDS)を設置

The National Data Service(米)

- □ 研究者にデータ発見、 利用、共有・公開を可能 とし、データと文献をリン クするサービス。
- □ NDSコンソーシアムによ るコミュニティ・ドリブンな 取り組み
 - 7大学、11データプロジェクト/フェデレーション、12情報基盤センター等、7出版社/学会等
 - > 200名体制



NDSの主要機能

- ▶ データ発見 ▶ データ共有・公開
- ▶ データ利用 → データと文献のリンク

2

日本における e-研究インフラ整備 の動向

日本のe-研究インフラに関わる取り組み事例

□ 取り組み事例がない訳ではない

- ▶ 学問分野別の学術データベースの整備
- 統合型e-研究インフラ(ex. データ統合・解析システム(DIAS))

□ 課題:

- ➤ 多くの場合、学術DBは当該分野の研究者により担われ、 研究者の評価につながらない上、近年、システム管理 負担が甚大となっている。
- ▶ 当該分野の研究プロジェクト予算の一部を割いて運営しているため、常に予算が少なく、かつ予算削減において最もしわ寄せを受けている。
- ➤ 結果として、DBは老朽化、DB管理者は疲弊しているこ とが多い。

日本のe-研究インフラに関わる政策(1)

- □ E-研究インフラ整備の視点が、政策にないわけではない。
 - ▶ 日本学術会議 E-サイエンス分科会(2008年)
 - ▶ 日本学術会議 国際サイエンスデータ分科会(2014年)
 - ▶ 日本学術会議 オープンサイエンスの取組に関する検討委員会(2016年) 等

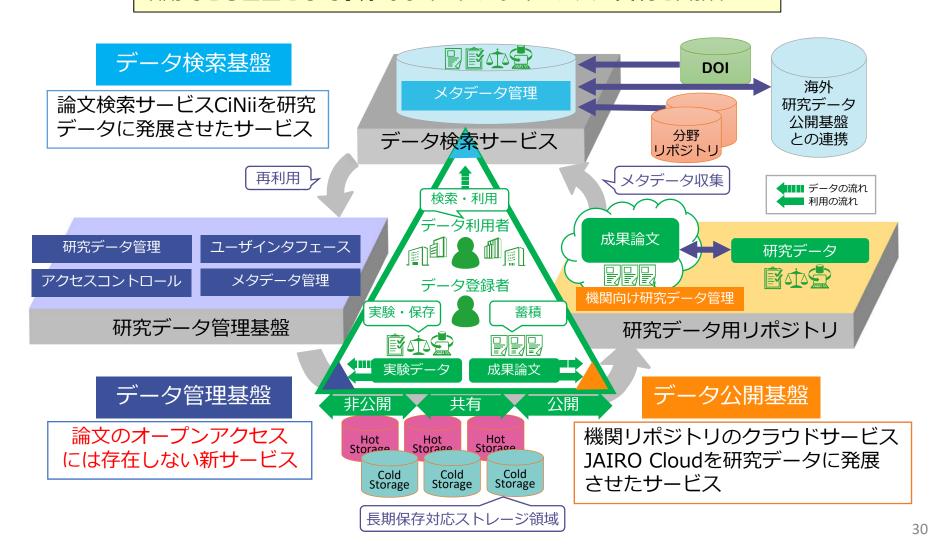
□ 課題:

- ▶ これら提言は海外動向を踏まえているため、同様の問題意 識で提言をとりまとめている。
- ▶ しかし、提言は抽象的で、実現した取り組みも漠然あるいは 小規模な場合が多い。
- ➤ 研究がデジタル時代に新たな次元のe-サイエンスに移行すると十分に信じていないため、取り組みに真剣みがうまれていない感がある。



オープンサイエンス対応 - 研究データ基盤

これまでに研究データの共有や公開の機会がなかった分野の研究者も 活用できる基盤として学際的なオープンサイエンスの実現を目指す



日本のe-研究インフラに関わる政策(2)

- □ E-研究インフラの整備が、物理的研究インフラの 効果的利用につながり、科学技術の国際競争力 につながるという認識が十分にできていない。
 - ▶ 日本学術会議が数年ごとに発表する「学術の大型研究計画に関するマスタープラン」において、e-研究インフラの併設整備の視点はない。
 - ➤ 第5期科学技術基本計画においても、「オープンサイエンス」や「オープンイノベーション」といった標語は並び、新たな取り組みとしての「Society 5.0」や「IoT」などの記述はあるが、既存の物理的な社会や学術を、具体的にどのように進化させるのかのイメージが出来ていない。

3

今後の展望:

各分野の研究力を強化するための e-研究インフラを求めて

オープンサイエンス推進の背景

説明責任 公的研究資金に おける義務化

データ再利用イノベーション

社会

アカデミック

研究の透明性 再現性

消極的

データ集中科学

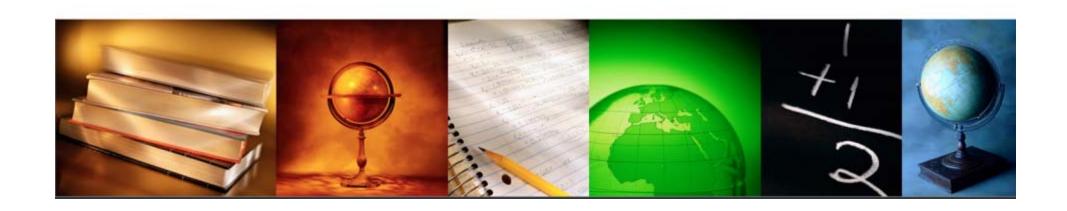
積極的

「オープンサイエンス推進」の課題

- □ これまでの「e-サイエンス」は、ICTにより学術を加速するといった、ストレートな理屈の概念であったが、
- □「オープンサイエンス」はこれに、説明責任や 研究公正、産業への活用など、社会的な要因 や社会への便益などの複合的な視点も絡まる ようになり、
- □アカデミアにとっては受け入れづらくなった。

The Fourth Paradigm: Data-Intensive Scientific Discovery

Tony Hey
Corporate Vice President
Microsoft External Research



A Digital Data Deluge in Research

Data collection

- Sensor networks, satellite surveys, high throughput laboratory instruments, observation devices, supercomputers, LHC ...
- Data processing, analysis, visualization
 - Legacy codes, workflows, data mining, indexing, searching, graphics ...

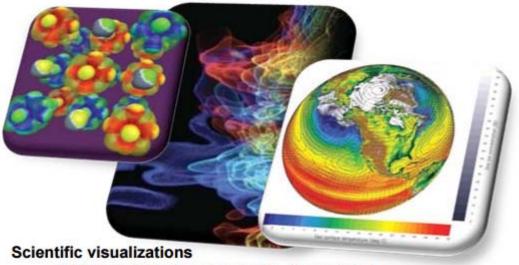
Archiving

 Digital repositories, libraries, preservation, ...



SensorMap

Functionality: Map navigation
Data: sensor-generated temperature, video camera feed, traffic feeds, etc.





NSF Cyberinfrastructure report, March 2007



Emergence of a Fourth Research Paradigm

- 1. Thousand years ago – Experimental Science
 - Description of natural phenomena

経験科学

- 2. Last few hundred years – Theoretical Science
 - Newton's Laws, Maxwell's Equations... 理論科学
- 3. Last few decades - Computational Science

データ集中科学

- 4. Today – Data-Intensive Science
 - Scientists overwhelmed with data sets from many different sources
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
 - eScience is the set of tools and technologies to support data federation and collaboration
 - For analysis and data mining
 - For data visualization and exploration
 - For scholarly communication and dissemination



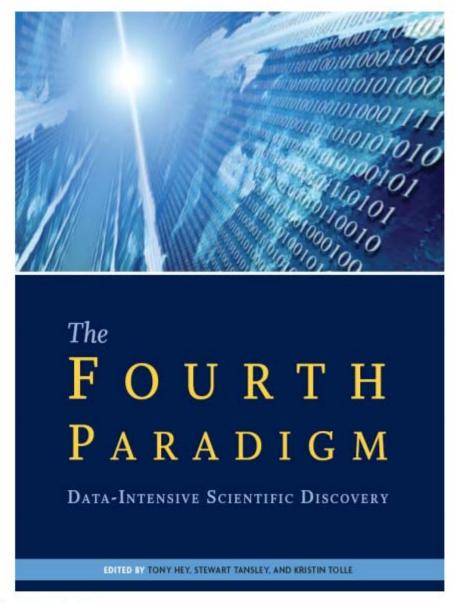
Science must move from data to information to knowledge

synthesized within the WorldWide Telescope service.



This work is licensed under a Creative Commons Attribution 3.0 United States License.

With thanks to Jim Gray





This work is licensed under a <u>Creative Commons</u>
Attribution 3.0 United States License.

e-研究インフラ構築に向けての 歴史的展開

- 1. 学術DBの整備や分野別解析ツールの開発
- 2. 高速・大規模コンピューティングの追求
- 3. 「e-サイエンス」や「eResearch」、「データ集中 科学」などの、ICTを用いた学術の加速
- 4. 「オープンサイエンス」などの、社会的要因や 社会への便益などの複合的な視点も絡まる 学術の推進

E-研究インフラが、研究力の強化につながるという認識の必要性

- □ 日本は、学術DBの整備や高速・大規模コンピューティング、そして物理的な研究インフラの整備には、重点的に投資してきたが、
- □ 「e-サイエンス」や「オープンサイエンス」の推進などの 抽象的な政策になって、どのような視点を学術において 求めるのかが、見えなくなってきている。
- □ 一方、欧米では、こうした抽象的標語のもとで、各学問分野における研究力を強化するための具体的な施策 (=e-研究インフラの整備)を行っている。
- □ 日本も、「研究力を強化する」という原点に政策面で立ち戻り、「e-研究インフラ」の整備および予算投下をした方が良いのではないか。