

# 記 録

文書番号	SCJ 第 20 期 200822-20530700-008
委員会等名	日本学術会議 情報学委員会 E-サイエンス分科会
標 題	日本における E-サイエンスの 推進に関する諸課題
作成日	平成 20 年（2008 年） 8 月 2 2 日

※本資料は、日本学術会議会則第二条に定める意思の表出ではない。掲載されたデータ等には、確認を要するものが含まれる可能性がある。

## 目次

1. はじめに .....	1
2. 新たな科学の方法論としてのE-サイエンス .....	1
3. E-サイエンスの具体的形態 .....	2
4. E-サイエンス推進の諸課題 .....	4
4.1 E-サイエンス推進に必要な学術情報基盤整備 .....	6
4.2 オープン・アクセス(OA)に関する国レベルの方針提示 .....	9
4.3 サイエンスデータベース、学術コンテンツなどの整備・拡充に関する 国レベルの統合的な体制作り .....	11
4.4 E-サイエンスに関わる人材育成 .....	14
4.5 E-サイエンスを推進するための情報学分野としての技術課題の発掘 .....	15
5. おわりに .....	16
参考資料 .....	16

## 情報学委員会 E-サイエンス分科会委員名簿

委員長	西尾 章治郎	大阪大学 理事・副学長
副委員長	小柳 義夫	工学院大学 情報学部長
幹事	関口 智嗣	産業技術総合研究所 グリッド研究センター長
幹事	松岡 聡	東京工業大学 教授
	青山 友紀	慶應義塾大学 教授
	阿草 清滋	名古屋大学 教授
	安達 淳	国立情報学研究所 教授
	荒木 啓二郎	九州大学 教授
	市川 晴久	電気通信大学 教授
	喜連川 優	東京大学 教授
	坂内 正夫	国立情報学研究所 所長
	下條 真司	情報通信研究機構 上席研究員
	田中 譲	北海道大学 教授
	近山 隆	東京大学 教授
	辻井 潤一	東京大学 教授
	三宅 なほみ	中京大学 教授

## 1. はじめに

次世代の学術研究の方法論として、高速ネットワーク上で、研究リソースや、成果としてのソフトウェアや論文・データベース等、さらには、研究プロセス自身を連携させる「E-サイエンス」が重要視されている。また、その遂行のためには高度な学術情報基盤の形成が緊急の課題となっている。そのような状況のなかで、日本学術会議情報学委員会E-サイエンス分科会は、世界各所で開始されているE-サイエンスに関わる諸活動を調査・俯瞰し、学術分野全体としての推進方策のあり方について検討し、提言をしていくことをその活動の主な趣旨として設置された。

第20期における活動として、このような本分科会設置の趣旨に添ったテーマを設定し、本分科会委員、あるいはオブザーバーからの話題提供、調査および現状報告などをもとに、日本におけるE-サイエンスの推進に関する諸課題についての議論を行ってきた。本稿は、これらの課題に関する討議の結果をまとめ、「記録」として提示するものである。そのことは、E-サイエンス分野における諸課題を明確にし、将来、「提言」などによって本分科会の考えを広く世に問う準備として重要であると考えている。

本稿によって、E-サイエンスの振興に向けた具体的な課題や、その推進のためのステップを提示することは、新たな学術の方法論を構築していく上で日本学術会議が果たすべき重要事項であり、このことは日本の学術のレベルを高め、長期的な発展をするために大きく資するものであると確信する。

## 2. 新たな科学の方法論としてのE-サイエンス

科学の方法論については、歴史的には、経験的事実を対象とした学問としての「経験科学」(実証的諸科学を含む)が自然現象を解明する科学として古くから芽生えた。その後、ニュートンの法則、マクスウェルの方程式をはじめとする厳然とした理論を中心とした「理論科学」が17世紀に発展し、近代科学の扉が開くことになる。

その後、20世紀に入り、ここ数十年来のコンピュータの発展とともに、複雑な自然および社会現象をシミュレーションにより予測・再現する科学の方法として「計算科学」が発展した。また、情報技術(IT)の急速な発展により、さまざまなセンサ類から獲得されるデータなど、科学者達にとってほとんど管理不可能な量のサイエンスデータが生成・収集されるに至り、このような状況のなかで「データセントリックサイエンス」と呼ばれる新たな科学の方法が生まれてきている。

さらに、計算科学、あるいはデータセントリックサイエンスを推進する過程で、科学者達が互いに連携をして大量のサイエンスデータを扱うための新しい手法の提案、あるいは新たな機能の開発等を行って、科学上の思いもかけない発見や新たな真理の探究を実現する可能性が出てきている。

そのような可能性を一層高めるために、高機能コンピュータと高速ネットワークからなる高度なIT環境を駆使して、さまざまな科学者コミュニティが、新たな科学の方法の創出のために実践を重ね、その手順を確立しつつある。その新たな科学の方法がE-サイエンスである。

E-サイエンスは、このように計算科学やデータセントリックサイエンスを包括しており、最先端のITによって研究者間の先進的な連携環境を実現し、多数の研究者達がデータを生成、共有、分析し、さらには協同作業をすることによって、新たな真理の探究を可能にするものである。特に、E-サイエンスは、実験、観察などによるデータ収集、シミュレーションなどの手法によるモデル構築とモデルによる検証、それらの結果を踏まえての高度な理論構築を統合した強力な枠組みと言える。

自然科学分野では、主には実験器具や観測機器からのデータを取得し、シミュレーション実験によってデータの生成を行い、ソフトウェアによって処理し、ファイル化、データベース化を行う。これらのサイエンスデータに対して、関連する研究者コミュニティでは、各人が連携しながらデータ検索、データマイニング技術等これまでに開発された研究手法や新しい分析手法を駆使し、その結果、新たな学問上の発見や理論構築、さらには結果の解釈に新境地を拓くような活動が展開されている。

また、例えば、人の認知行動や社会・文化的な行動を扱う融合科学分野では、言語運用や学習過程、社会・経済活動など人の行動に関して、最近、これまでは取得不可能だった詳細かつ継時的なデータが大量に蓄積されつつある。そのサイエンスデータの分析をこの分野の研究者コミュニティが共同で推進することによって、人の「知」の仕組みの解明に関する新たな研究が展開しつつある。

### 3. E-サイエンスの具体的な形態

ここで、E-サイエンス推進のための課題を探る意味でも、そのあるべき具体的な姿を探ってみる。ただし、実際はこれらの複合体的な形態になると考えられる。

#### (1) 研究用グリッド基盤（高速ネットワーク上での計算・データなどの大規模連携）

研究用のグリッド基盤は、大規模な資源としてはスーパーコンピュータから、小規模なものはキャンパス内に設置されたクラスタマシン群、さらには個人のパーソナルコンピュータ(PC)などを、高速ネットワークで連携させ、それら計算・データ・ネットワーク資源を共用することを実現した高度IT基盤である。そのIT基盤上でインターネットのように仮想化、平準化した高度なE-サイエンス向けサービスを実現することが可能であり、また、その基盤を下地にボトムアップ的に、学問分野ごとに容易にE-サイエンスサービスを構築することも可能である。連携する各々のコンピュータは、単独でも個別のE-サイエンスの課題のシミュレーションや高度なデータ処理を行う機能や性能を有する場合もあるが、そのままではその分野の一部の研究者の限定的な利用にとどまり、分野横断的な入出力データ共有による相乗的な効果や、多くの分野内外の研究メンバーの平準化したアクセスによる相互資源利用が困難となる。

逆に言えば、これらの高機能コンピュータやアプリケーション・ソフトウェア、データベースなどを効率的かつ効果的に運用し、所属機関の異なる離れた地にいる研究者にもインターネットを介して容易にかつ平準化された形でアクセス・利用でき、共同研究のベースとして活用できるようにするグリッド基盤は、E-サイエンスの具体的な形態の典型と言えよう。特に、国内におけるグリッド基盤システムの研究開発の代表的なプロジェクトとしては、2000年代に入ってから「超高速コンピュータ網形成プロジェクト(National Research Grid Initiative : NAREGI、通称ナレギ)」[参考資料1]が推進されてきた。

なお、このNAREGIプロジェクトの開始に先立つ平成13年(2001年)度から、日本原子力研究開発機構、理化学研究所をはじめとする6機関が参画して推進されたITBL(IT Based Laboratory) [参考資料2]プロジェクトは特筆に値する。本プロジェクトは、国内の一部の国立研究所系のスーパーコンピュータを国立情報学研究所(National Institute of Informatics: NII)のSuperSINETで接続して共用し、共同利用可能かつ実運用に供することが可能なスーパーコンピュータシステム網の構築を目指して平成17年(2005年)度まで推進された。ITBLは、ITを活用して分散した計算資源をシームレスに繋ぎ、仮想的な研究環境を実現する実験プロジェクトとして、その後のNAREGI等の本格的なグリッドシステムの研究開発に至るマイルストーンと位置づけられよう。

NAREGIは、平成15年(2003年)度から文部科学省の委託研究として5カ年計画で、国際標準に基づいた我が国のグリッドのミドルウェアの研究開発、システムの実証研究開発、評価を推進した産官学の連携プロジェクトである。当初はグリッドシステムのミドルウェアを研究開発するプロジェクトとしてスタートし、平成18年(2006年)度から平成19年(2007年)度まで「最先端・高性能汎用スーパーコンピュータの開発利用」プロジェクトの一環として推進された。国立情報学研究所内に「グリッド研究開発拠点」を設置し、ミドルウェアの研究開発を産官学共同の研究開発を推進するとともに、岡崎国立共同研究機構分子科学研究所内にある「計算科学研究開発拠点」と共同で、ナノサイエンス分野を中心としたグリッド技術を利用した計算科学アプリケーションの研究開発や、開発されたグリッド環境の検証を行った。産官学体制として、産業技術総合研究所、大阪大学、東京大学、九州大学、東京工業大学等の大型スーパーコンピュータを有する独立行政法人の研究所、大学などのメンバーが企業群と一体となって研究開発を行い、EGEE(Enabling Grids for E-Science) [参考資料3]やTeraGrid[参考資料4]等の欧米のグリッド基盤プロジェクトとも密な関係を保ち、グリッドの国際標準化団体であるOpen Grid Forum(OGF) [参考資料5]において多くの標準化活動に貢献した。

平成20年(2008年)度にはNAREGIミドルウェアVer. 1.0が公開され、現在、国立情報学研究所では、ソフトウェアメンテナンスやユーザサポートを継続し、さらにグリッドのオペレーションセンター(GOC)などの運用のための組織を設置して、大学や研究機関を横断する全国グリッド基盤の構築を継続的に推進している。

## (2) 特殊な実験リソースの高速ネットワーク上での共用

E-サイエンスの一つの有り様として、高性能な電子顕微鏡、特殊な化学実験装置、世界に類を見ない大型放射光施設(SPring-8) [参考資料6]などの高度な実験リソースを高速ネットワークで共用する形態が考えられる。例えば、SPring-8は世界最高性能の放射光を利用可能な大型の実験施設であり、国内外の研究者に広く開かれた共同利用施設として、高速ネットワークを介した利用も促進されており、物質科学、地球科学、生命科学、環境科学、産業利用などの分野で優れた研究成果をあげている。

## (3) 観測データのリアルタイム連携(センサネットワーク型)

天文関係、地震関係などの観測ネットワークから得られる超大量のリアルタイム・サイエンス・コンテンツを、高速インターネットを介して、遠隔地間で共有・処理し、新たな科学の発見に資する形態が考えられる。

例えば、米国のスローン・デジタル・スカイ・サーベイ(Sloan Digital Sky Survey: SDSS)の活動は典型的な例である[参考資料7]。SDSSの望遠鏡が設置されているアパッチポイント天文台は、アメリカの天体物理学研究機構(ARC)によって運営されているが、このサーベイでは全天の4分の1にわたって1億個以上の天体の位置と明るさを測定して詳しい地図を作ることを目指している。さらに、100万個の銀河とクエーサーに対しては、距離も測ることを目指している。天文学者は、ネットワークを介してこのサーベイデータを見ることによって、宇宙の進化に関してさまざまな理論を考えることが可能になり、異なった理論はそれぞれ異なった宇宙の大規模構造のパターンを予測する。これらの活動から、40年かかっても見つからなかった宇宙の構造が、わずか1年で発見されたという報告もなされている。我が国も東京大学宇宙線研究所や国立天文台をはじめとして、全国に点在した機関に所属する20名を越える研究者達が、

高解像度CCD撮像技術およびその画像処理技術などにおいて多大の貢献をしている。

天文関係以外に、湖沼の生態系に関する世界的な観測ネットワークも構築されており、草の根のネットワークを組織して継続的な活動を推進しているGLEON(Global Lake Ecological Observatory Network)は、よく知られている[参考資料8]。GLEONは、平成17年(2005年)に10カ国からのメンバーによって第1回の会合を開催したが、湖沼学者、ITの専門家およびエンジニアなど複数領域の研究者達が、その後も湖沼の生態系の観測ネットワークを持続的に運営し、世界規模の「湖沼の生態系の観測ネットワーク」を構築している。これらの観測地点からのサイエンス・コンテンツは、共通のWebポータルによって世界中の研究者達が簡単にアクセスできるように公開されている。

また、国内のグリッド技術の実用化研究の一例としてGEO Grid(Global Earth Observation Grid:地球観測グリッド)プロジェクト[参考資料9]が産業技術総合研究所を中心に推進されている。GEO Gridは、グリッド技術を用いて、地球観測衛星データの大規模アーカイブ・高度処理を行い、さらに各種観測データベースやGIS(Geographic Information Systems:地理情報システム)データと融合し、ユーザがこれらのデータを統合的に、しかも手軽に扱えることを目指したシステムである。

#### (4) サイエンスデータベース、学術コンテンツなどのデータ共用

現在、多様な分野におけるサイエンス・コンテンツのデータベースが大規模に整備されつつある。また、さまざまな学術コンテンツのデジタルアーカイブ化が急速に進み、その基盤システムの構築も推進されている。ただし、後述のように米国、英国などと比較した場合に、国内における進展具合が必ずしも良好でないことは一つの大きな課題である。このようなサイエンスデータベース、学術コンテンツなどのデータを共用化し、研究者が横断・網羅的に検索・解析可能な情報システムを提供することにより、学術の進展を助長し、新たな科学的真理の探究に資することは、E-サイエンスのあるべき姿の典型と言えよう。特に、学術コンテンツ等のデータ共用化は、理工系の諸分野のみならず、人文社会科学系の文献学あるいは博物学などが中心の学問分野に対しても大いに利するものと言えよう。

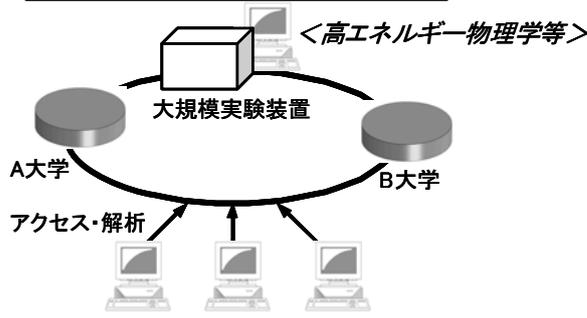
以上、E-サイエンスの具体的な形態の主要な例を4項目にわたって個別に述べてきたが、これらを統合的に連携させながら、大規模な国際連携科学として推進を図るところがE-サイエンスの特徴である。つまり、高速ネットワークを介して、コンピュータ、観測装置、データ、さらには人間(頭脳)を連携、共有しながら推進する科学が、E-サイエンスと言えよう。このような有り様について、グリッドシステムについて世界的なリーダーシップを発揮している米国アルゴンヌ国立研究所のIan Foster博士は、E-サイエンスを“Science as a Team Sports”と言及している。

E-サイエンスの具体的な形態のいずれも高速ネットワーク基盤が不可欠であり、その観点からE-サイエンスをサイバースペース上での科学の方法、つまり、「サイバーサイエンス」とも言うこともできる。ここで、国立情報学研究所が運営する高速ネットワークSuperSINETを利用した代表的な国内のE-サイエンスの成功事例を図1に示す。

## 4. E-サイエンス推進の諸課題

本記録の趣旨は、以上で述べてきた科学の新たな方法論として大きな可能性を有するE-サイエンスを推進する上での課題を提示し、今後の発展に資することである。その観点から以下では5課題を提示することにする。

### 広域分散データ解析システム

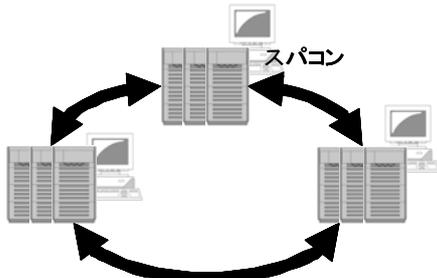


先端研究実験データのリアルタイム同時並行解析の実現

高エネルギー加速器研究機構で行われるBelle実験の大規模データを、東北大、東大、東工大、名大、阪大で解析し、「CP保存則の破れ」を検証。

### 大規模設計シミュレーション

<材料・ナノテクノロジー等>

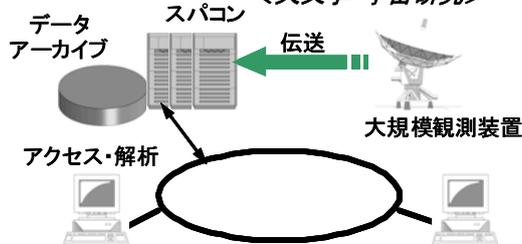


複数のスパコンをネットワークで結合し、材料設計等のための大規模シミュレーションを実施

東北大、東大、九大、分子科学研究所のスーパーコンピュータを直接連結し、超大規模シミュレーション計算を実行。水素吸収水化合物の構造の最適化や水素分子と水分子の結合状況の確定に成功。

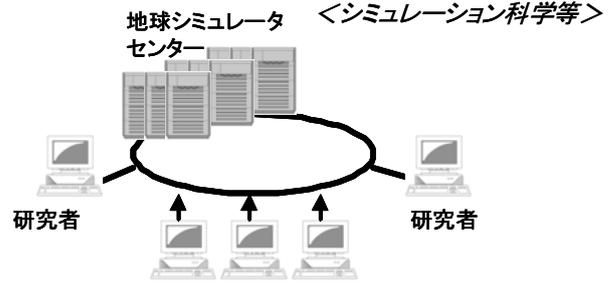
### 観測データの遠隔解析

<天文学・宇宙研究>



大規模な観測データを遠隔地から解析し研究に活用

### 連結階層シミュレーション マルチスケールシミュレーション

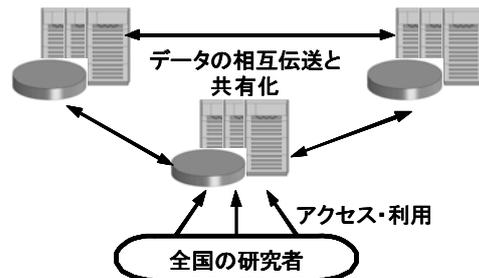


地球シミュレータセンターを高速ネットワークで利用し、新しいシミュレーション手法の研究開発を実施

地球シミュレータセンターを高速ネットワークで利用し、ヒートアイランドなどの都市現象の予測などの新しいシミュレーション手法を開発し、都市設計の提言などに利用。  
(海洋開発研究機構)

### 広域分散データベース整備と遠隔利用

<バイオインフォマティクス等>



研究成果データをネットワークで共有化し、全国の研究者に公開

国立遺伝学研究所で運営する日本DNAデータベースの情報を東大、京大で共有し、大量データを更新。米国、欧州と国際DNAデータベースの共同構築。

遠く離れた電波望遠鏡を超高速回線で結合し、これまでにない高い観測感度を達成する世界最長の光結合型VLBI(超長基線電波干渉計)の実現。

(国立天文台)

坂内正夫(国立情報学研究所長):「E-サイエンスに関する動向」, E-サイエンス分科会講演資料, p.16, 平成19年2月15日

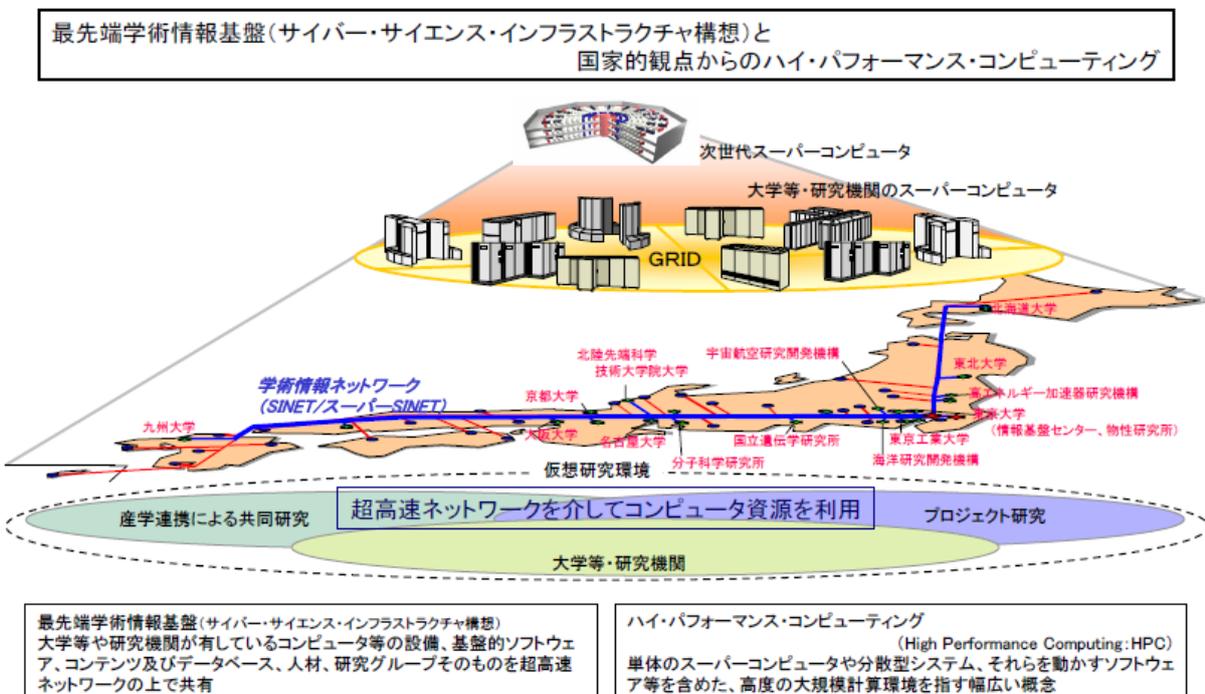
図 1 : SuperSINET を利用した E-サイエンスの具体的な形態

#### 4.1 E-サイエンス推進に必要な学術情報基盤整備

前章「3. E-サイエンスの具体的形態」の「(1) 研究用グリッド基盤（高速ネットワーク上での計算・データなどの大規模連携）」に示したように、E-サイエンスがサイバースペース上で展開される科学の形態であることを考えると、その基盤である「学術情報基盤整備（サイバー・サイエンス・インフラストラクチャ：CSI）」が最先端の情報通信技術をベースに高度な情報システムとして構築されているかどうかは、その科学の進展に大きな影響を及ぼす。

第3期の科学技術基本計画[参考資料10]には、「研究情報基盤は、研究活動に不可欠な、いわばライフラインとしての性格を有しており、特に、大型コンピュータや高速ネットワークなどは最先端の情報通信技術や国際動向に常に先行して整備していく。また、研究機関において不可欠な論文誌などの研究情報の体系的収集・保存、効果的発信並びに研究者・研究機関間の連携や協力を促進することにより、研究情報基盤の効果的かつ効率的な運用を進める。」と記されており、国としてもCSIの重要性に関する認識は高い。ところが、国立大学等における学内LANの整備が補正予算でなされてきた経緯があり、その予算の性格上、国立大学法人化後の運営費交付金に学内LANを整備する予算が積算されていない現状がある。このような憂慮すべき状況に対処していく意味も含めて、国内では、平成18年(2006年)3月23日付で文部科学省科学技術・学術審議会学術分科会研究環境基盤部会学術情報基盤作業部会が「学術情報基盤の今後の在り方について」（報告）[参考資料11]を提示している。

その報告書では、国公私立大学をはじめとする高等教育研究機関の学術情報基盤の整備、拡充に関する現状分析と今後の積極的な推進方策などが提示されている。特に、次世代スーパーコンピュータプロジェクトの推進と関連して、図2に示すような、将来のマルチペタフロップス級の次世代スーパーコンピュータを最上位層、現有ではテ



科学技術・学術審議会, 学術分科会, 研究環境基盤部会, 学術情報基盤作業部会:  
学術情報基盤の今後の在り方について (報告), pp. 51(附属資料7), 平成18年3月23日

図2: 学術情報基盤の階層的な構築

ラフロップス級ながら近い将来にはペタフロップス級になる各主要大学や研究機関のスーパーコンピュータを次の層に配置し、さらにその下位層に地域の中級レベルの大型コンピュータ群を配置し、統合的なグリッドシステムを構築することの重要性が記述されている。欧米においては、E-サイエンスが研究開発方法論の大変革をもたらすものとして非常に重要視されており、平成17年(2005年)以降、国家レベルの最先端のCSI構築の計画が開始されている。

まず、欧州では欧州連合(EU)がリーダーシップを発揮して、EU e-Infrastructure の構築の主要な事業の一環として、EGEE と DEISA (Distributed European Infrastructure for Supercomputing Applications) [参考資料 12] のもとで、E-サイエンス向けのグリッド基盤の構築が行われている。

EGEE は、CERN (欧州原子核研究機構、The World Wide Web 発祥の地としても名高い) が中心となり、大規模な Linux サーバーおよびストレージサーバ群をグリッド技術で仮想的に集約し、データ処理中心の大規模な計算環境を構築する世界最大の研究グリッドである。欧州を中心として、現状で「90 以上のサイト」、「10 万個以上の CPU 資源」、「10 ペタバイト以上のストレージ容量」を誇り、現在もその規模は拡大している。また、グリッドのミドルウェアとして gLite を開発し、多くの標準化に貢献していることでも有名である。

DEISA は欧州のスーパーコンピュータセンター群のコンソーシアムとして形成されているグリッドであり、10 カ所以上のセンター群がデータ共有や共通のジョブサブミッション、共通アカウント、さらには共通のアプリケーションサポート体制を形成している。今後は、EU が DEISA と並行して主導しているペタフロップス級のスーパーコンピュータ構築を目指す PRACE プロジェクト [参考資料 13] との連携が目されており、PRACE および各国と協力体制で構築予定の 4 機程度の次世代スーパーコンピュータを上位層として、現有のテラフロップス級の各国のスーパーコンピュータと共に統合的な仮想センターを形成するグリッドシステムを構築する計画が DEISA2 において進んでいる。

同様に、米国では、米国科学財団 (NSF) が中心となって、EGEE と類似した構造をも Open Science Grid (OSG) [参考資料 14]、ならびに DEISA と同様のセンター間の連携体制をもつ TeraGrid プロジェクトが推進されている。

近年、そのような欧米での急激な動きのなかで、国内においても科学技術における新たなブレークスルーを生むための情報基盤の構築が急務であることは、科学者間での一致した意見である。その一つの取組みとして、国立情報学研究所 (NII) を中心にした図 3 に示すような CSI 構想が推進されている。

そこでは、先に述べた SuperSINET、NAREGI 等の実運用した研究グリッド、さらに全国共通認証認可基盤を基幹システムとした E-サイエンス基盤である CSI 形成、さらには応用研究・開発が推進されている。特に、これらの基幹システム、サイエンスデータベース、学術コンテンツの複合 (連成) 体を構築することで、現在までに E-サイエンスの実運用としての成功例が得られており、先に図 1 に示した諸例は、それらの成功事例の一部である。また、CSI 構想において、後述のようにオープン・アクセス環境での次世代学術コンテンツ基盤の構築が重要であり、その具体的な事例として 57 大学と国立情報学研究所が展開している「機関リポジトリ」プロジェクトが推進されている。

以上のように、日本において E-サイエンスに関する先進的事例が着実に推進されていることを考えると、日本が E-サイエンス推進の国際的なリーダーの一つになり、他国



坂内正夫(国立情報学研究所長):「E-サイエンスに関する動向」, E-サイエンス分科会講演資料, p.8, 平成19年2月15日

図 3 : 日本における CSI 構想の例

と協調してより国際的で一般的なE-サイエンスインフラを構築していける可能性は十分にある。ただし、その推進のためには課題も多く、以下のような提起を行う。

### 課題 1

総合的な学術情報基盤(GSI)整備を怠った場合、国全体の研究活動が停滞することは必至であり、文部科学省のみならず他省庁も含めて、その構築に向けた掘り下げた議論が必要である。特に、E-サイエンスを強力に進めるには、国内学術諸機関(大学、研究所、研究センターなど)、関連省庁、国際関連機関、さらには諸々の国家プロジェクトを包括する強力な連携体制の構築が大変重要である。

- 最先端のCSIが必須な時代であり、国として総合的なGSI整備に関する将来ビジョンを策定することが必要である。具体的には、大学、研究所などの高度な教育研究諸機関が一体となり、日本学術会議が関与しつつ将来的なビジョン策定を行うことが重要である。
- CSI整備を財政面も含めて国として行うことが重要である。そのCSIの整備が、研究教育における効率化に大きな役割を果たし得ることから、投資効果は高い。また、公的なリソースとして、安全・安心の確保のためのライフラインとしてなど、さまざまな用途の使用が可能である。ただし、その整備にあたってはバラマキ型とならないようにすることが肝要であり、それぞれの設備等の投資効果をさまざまな観点から見極め、取捨選択や集中投資をすることへの配慮も重要である。

#### ■ 留意事項：

次世代スーパーコンピュータプロジェクトのように、国家プロジェクトとして E-

サイエンスの一翼をになう先鋭的な整備がなされつつある。それらの個別の設備の整備と関連しつつ、国全体として総合的、かつ網羅的に（例えば、大学の場合、国公立を問わず、また、大規模大学から小規模大学まで）OSI 整備を推進するための方策の導出が重要であり、有識者による高度なビジョン策定から現場の運営にいたるまでの総合的なプランニングが重要となる。

#### 4.2 オープン・アクセス(OA)に関する国レベルの方針提示

前章「3. E-サイエンスの具体的な形態」の「(2) 特殊な実験リソースの高速ネットワーク上での共用」、「(3) 観測データのリアルタイム連携（センサネットワーク型）」を推進する上で重要な課題がある。それは、さまざまな科学分野における実験データ、天文学の観測データなどのサイエンス・コンテンツのオープン化が実現しないことには、このような形態のE-サイエンスの発展が実現しないことである。また、前章の「(4) サイエンスデータベース、学術コンテンツなどのデータ共用」を実現するには、学術論文がオープン・アクセス（無料公開）可能になることが重要な鍵になっている。

まず、学術論文のオープン・アクセス運動については、国内では、先にも参照した「学術情報基盤の今後の在り方について」（報告）[参考資料 11]で論じられている。当報告書では、オープン・アクセス運動の理念を「無料で制約のない学術論文のオンライン利用を認める。」ことに集約している。通常、論文の生産者である研究者が、論文を出版しても直接経済的な利益を得ることがないにもかかわらず投稿するのは、「研究成果としての学術論文の内容は人類にとって共通の知的資産であり、その内容を必要とするすべての人に知ってもらいたい」からである、という強い認識が、この運動を展開する主たる根拠となっている。また、欧米の学術出版社の電子ジャーナルが年々高騰を続けるため、それに対抗する方策とも位置づけられている。

当報告書ではオープン・アクセスを実現する方法として、「セルフアーカイビング」と「オープン・アクセス雑誌刊行」の二つが提示されている。「セルフアーカイビング」の具体例として、「著者自身の Web サイト」、「研究分野別リポジトリ」、「機関リポジトリ」が記され、また、「オープン・アクセス雑誌刊行」の代表例として、最近、その存在が広く知られている PLoS (Public Library of Science) という団体の刊行する著者支払いモデルの雑誌が紹介されている。

最近では、「オープン・アクセス雑誌刊行」については、商業出版社をはじめ、学会の出版部門、大学出版会においても、非常に積極的に推進する意向をもっているところもあり、上記の諸例以外にも著者の選択により論文単位でオンライン公開を行う部分的なオープン・アクセス雑誌のモデルが種々提案されている。

このような学術論文のオープン・アクセス運動については、欧米では政府、研究助成機関等が直接的に学術情報流通に関与をはじめたことから、この運動に対して高い関心が集まるようになってきている。具体的には、米国では、2005 年 5 月から、国立衛生研究所 (NIH) が資金助成した研究の論文を、PubMed Central (PMC) [参考資料 15] に出版後 12 ヶ月以内に公開するように求め、2007 年 12 月には法制化されている。なお、PubMed Central は、米国国立医学図書館 (NLM) が運営するライフサイエンス分野のオープン・アクセスのデジタルアーカイブであり、電子ジャーナルが 200 誌以上閲覧可能である（タイトルによって、最新 1 ヶ月から 2 年間アクセスできないものもある）。一方、英国ではリサーチカウンシルズが、公的資金を助成した研究の成果である論文を機関リポジトリ等に掲載することを義務付ける提案を行っている。

それに対して国内では、論文情報の電子ジャーナル化が著しく遅延しており、国費

を投入した研究開発成果の無料公開は、強い政策提言がなされていないことが懸念されている[参考資料 16]。

我が国の特殊性として、欧米の学術情報源、すなわち電子ジャーナルを確保することが極めて重要である。現在、大学では欧米の学術出版社からライセンスする電子ジャーナルの価格の高騰が最も解決困難な問題であり、国レベルでの議論を必要としている。また、国際的には文献情報のオープン・アクセスへ向かう流れが強まりつつあり、我が国の学術界の対応も喫緊の課題である。

一方、我が国が生み出す学術文献情報も極めて重要である。我が国では商業学術出版活動はあまり大きくなく、主なステークホルダは学会と大学等である。我が国の学会の一般的状況としては、小さな学会が数多く存在し、独自に学術出版活動を行っているという現実がある。このため、出版活動が弱体で電子ジャーナル化への投資や編集人材の確保、そして電子ジャーナルビジネスの展開（サイトライセンシング等の契約による安定収入の確保など）が欧米の学術出版に比べて極めて遅れた経緯がある。これに対して、平成 15 年度から国立情報学研究所を中心に、我が国の英文雑誌の電子ジャーナルの国際展開を支援する SPARC Japan プロジェクトが進行しているが、日本語雑誌の電子ジャーナル化など今もって解消されていない課題も多い。

今後、学会における自立的かつ自律的な電子ジャーナル出版を活性化していくことが急務と考えられる。一方、人文社会科学を中心に大学紀要も重要な意義を持っている。これについては、大学と国立情報学研究所が協力して、後述する「機関リポジトリ (Institutional Repository)」構築事業のなかで電子化とオープン・アクセスによる公開が進んでいる。

そのような国内状況のなかで、オープン・アクセス運動の一環とも位置づけられる機関リポジトリ構築事業の活動が注目に値する。機関リポジトリとは、大学とその構成員が創造したデジタル資料の管理や発信を行うために、大学がそのコミュニティの構成員に提供する一連のサービスである。次世代学術コンテンツ基盤構築を実現するため、国立情報学研究所が大学と共同で確保すべき多様な学術コンテンツの中で、特に大学の教育研究活動の成果である学術情報の収集、組織化、保存および発信のための仕組みとして、この活動が行われた。国立情報学研究所では、これまでのコンテンツ関連事業の成果を継承、拡充させ、次世代学術コンテンツ基盤の整備に資するために、各大学における機関リポジトリの構築とその連携を支援してきた。平成 17 年度は 19 大学に、平成 18 年度からは公募により選定された 57 大学に、平成 19 年度からは 70 大学に、機関リポジトリ構築推進のための事業を委託し、平成 20 年 6 月には 83 機関が公開サービスを提供している。

今後は、学術論文、学術資料に加えて、観測データ、実験データなどのサイエンス・コンテンツのオープン化をいかに推進していくかが大きな課題であり、これらも大学等の機関リポジトリで公開されることが期待される。前章「3. 観測データのリアルタイム連携（センサネットワーク型）」に SDSS、GLEON の先駆的な活動の一端について述べたが、このようなサイエンス・コンテンツのオープン・アクセス化が、E-サイエンスの発展のために非常に重要である。しかし、サイエンス・コンテンツそのものが創造性に富む科学的な発見リソースであり、それをもとに学術論文の成果に結実していくことを考えると、観測データの段階で公開すべきか、未公開にしておくべきかの判断についてより慎重にならざるを得ず、難しい課題である。

## 課題 2

学術論文、学術資料については、国内の電子ジャーナル化が著しく遅延しており、その効率的かつ持続的な電子化の継続的な促進、および国費を投入した研究開発成果のオープン・アクセス化の促進が重要。その際、学会出版のコストモデルに十分留意し、学会の自律的な出版活動を阻害することのないような配慮が必要。その実現のためには、日本学術会議が中心となって、強い政策提言がなされるように働きかけをすることが必要。

一方、天文学の観測データ、素粒子分野における実験データなどのサイエンス・コンテンツのオープン化はE-サイエンスの発展のために重要。ただし、これらのデータを公開すべきか、未公開にしておくべきかは、コンテンツそのものが創造性に富む科学的な発見リソースであることを考えると慎重な判断を要する。そこで、公開する方向に誘導する何らかの権利保持に関する工夫が必要。

### ■ 留意事項：

オープン・アクセス化については、科学におけるオープン性という「建て前」と、激烈な競争を伴う創造的研究の推進という「本音」の狭間で、オープン化を促進する方策を今後どのようにして提案することが可能かの検討が必要。

## 4.3 サイエンスデータベース、学術コンテンツなどの整備・拡充に関する国レベルの統合的な体制作り

前章「3. E-サイエンスの具体的形態」の「(4) サイエンスデータベース、学術コンテンツなどのデータ共有」については、大規模な実験データ、画像データ、統計データあるいはシミュレーション結果データなどが典型例であるが、その共有のためには、個々のデータベースが持続的に維持され、適切なメタデータが付与されるような情報環境が実現されていなければならない。

従来から、我が国では、国立研究機関や大学の附置研究所が、研究コミュニティの要請に応じて持続的にサイエンスデータベースを形成してきた。この一例として、国立遺伝学研究所の作成する DDBJ (DNA Data Bank of Japan) [参考資料 17] がある。これは、DNA 塩基配列データベースを国際的な三拠点で構築するもので、我が国が一つの拠点となっている。さらに平成 19 年度から、国家プロジェクトとして分野別データベース拠点形成プロジェクトが開始された。現在、「ライフサイエンス統合データベース」と「地球環境データベース」について、当該分野の研究者グループが、各種のデータの組織化やメタデータの高度化、ユーザビリティの向上などに総合的に取り組んでいる。これらの例のようにサイエンスデータベースの形成には、国際的な観点、研究コミュニティを基盤とした拠点形成、そして持続性の維持が非常に重要である。

一方、サイエンスデータベースの作成支援のための競争的資金としては、文部科学省の科学研究費補助金（研究成果公開促進費）「データベース」が重要な役割を果たしてきた。すべての学術分野を対象とし、研究者グループが申請できるデータベース作成のための補助金として唯一継続的に運営されてきたことは、我が国として誇るべきことである。ただし、データベース形成活動は持続的に行うのが普通であり、高々数年の間の補助金では有用なデータベース形成まで至らない場合も多い。E-サイエンス振興のための資金として、その運用について再検討する必要がある。

サイエンスデータベースの共有には、その活用を補助するデータベースの整備・拡充が併せて行われることが肝要である。このようなデータベースの典型として、「文献情報」と「特許情報」があげられ、日常の教育研究活動にも必須の情報源になって

いる。最近、特許情報については、知的財産戦略本部「知財財産による競争力強化専門調査会」の審議報告書「オープン・イノベーションに対応した知財戦略の在り方について」(平成20年3月4日)[参考資料18]などで政策的な方向付けが配慮されている。特に現在では、特許庁が特許明細情報を電子化してオープンに提供し、商用サービス機関が付加価値を付けて一般に提供するという状況になっている。一方、文献情報については、前節「4.2 オープン・アクセス(OA)に関する国レベルの方針提示」で述べたように、我が国はいくつかの困難な問題を抱えている。

情報の電子化の急速な進展とともに、そのアーカイブと保存についても十分留意する必要がある。ここで「アーカイブ」とは、過去に遡る情報の電子化のことを広く指すものであるが、文献情報も紙媒体の遡及的電子化を進めるとともに、一方将来に向かっては、その保存についても災害等を配慮して複数拠点で確保するなどの配慮が必要である。

さらに、今後はWeb上に存在する電子情報すべてが学術活動に必須のものとなると考えられる。遡及的にもアーカイブ化することが重要と考えられるが、学術・科学技術分野を越えた広いジャンルの情報を含むものであるから、国レベルでの取組みが必要な課題といえる。

学問の分野は20世紀までに出揃い、21世紀は異なる学問分野の融合の時代だと言われている。今後、学術上の新たな発見、イノベーションを起こすには、異分野間の融合が必至であり、そのためには異分野間のサイエンスデータベース、学術コンテンツを駆使したE-サイエンスの推進は大きな意義を有する。その推進のためには、自己の研究テーマと統合可能な他の研究テーマ、研究者等を簡便に検索可能とするサービスが重要となる。科学技術振興機構(JST)が検討している「イノベーション創出に貢献可能な科学技術情報サービスを実現する上での条件」として、次のような事項があげられている。

#### (1) 情報の網羅性

インターネットの浸透、ITの進展に伴い利用可能なサイエンスデータベース、学術コンテンツが爆発的に増加しているが、それらの情報は多様な学問の分野を網羅的にカバーしているか。

#### (2) 情報提供機関の連携の推進

一つの中核的機関のみですべてを網羅するのには限界がある。収集する情報の学問領域、内容などに応じて、収集・提供するコンテンツを分担し、効率的に連携して情報提供サービスをしているか。

#### (3) 検索行動の多様化対応サービスの実現

必要な情報を探す、という利用局面の他に、関連情報から異質の情報に誘導するという機能を実現しているか。このような機能はE-サイエンスの推進のために重要な条件となり、そのためには大量のデータを俯瞰し、全体を把握する機能が求められる。そのために、情報加工、各種辞書類の整備を持続的に行うノウハウ、体制が必要である。

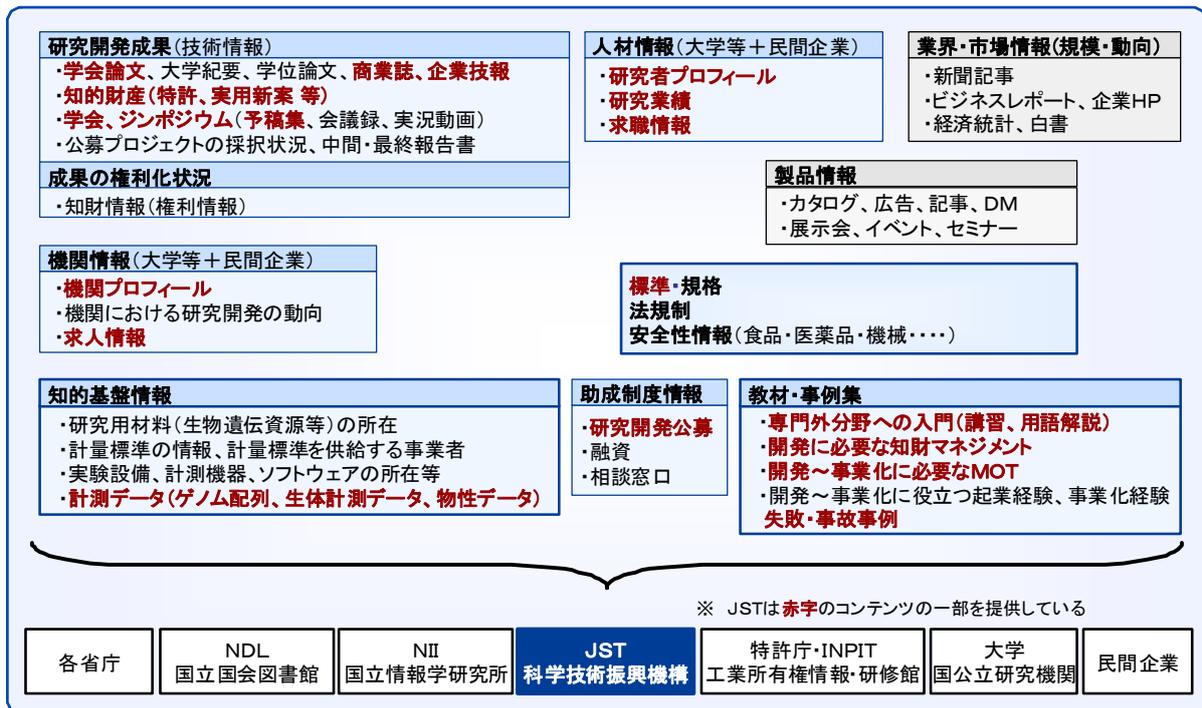
#### (4) 情報の公開と電子化

情報探索および解析をする上では、情報の公開と電子化が必要であるが、オープン・アクセス化以前の問題として電子化が実現されているか。海外と比べて遅延している情報の電子化を推進し、サイエンスデータベース、学術コンテンツの整備を促進するには、情報提供機関が重複を無くし、役割分担をもって推進することが重要である。

(5) ユーザビリティの高いサービスの提供

E-サイエンスの遂行において、サイエンスデータベース、学術コンテンツを利用する場合、自己が把握していない分野の探索が目的であることが多いが、これに対して正確な情報を迅速に把握可能とするユーザビリティの高いサービスを提供しているか。このユーザビリティには、操作性、可視化などの主要な機能に加えて、自国語化、正確性・信憑性など、トータルなサービスが求められる。

図 4 に科学技術振興機構 (JST) が情報の網羅性の観点から「科学技術情報の種類とその担い手」について自らの機関が網羅している種類を明確にするために調査した資料を示す。



JST検討資料より，平成20年6月

図 4：日本における情報の網羅性（ただし、確定版ではない）

以上のような課題の解決に向けて、図 4 に示した情報の種類に関して、国立国会図書館 (NDL)、国立情報学研究所 (NII)、科学技術振興機構 (JST) をはじめとする主要関連諸機関で連携・調整等を行い、研究成果の電子情報を生み出す大学や学会等の機関による自律的な活動を効率的に促すような仕組みを作っていくことが重要である。

**課題 3:**

サイエンスデータベース、学術コンテンツなどの整備・拡充は、国の学術研究の推進のためにますます重要。より効率的で網羅的な事業推進を可能とするために、国立国会図書館 (NDL)、国立情報学研究所 (NII)、科学技術振興機構 (JST) をはじめとする主要関連諸機関で連携・調整をしながら、関係各機関のミッションに沿った効率的な情報の電子化と相互運用性の向上についてより明確な棲み分けと連携を促進することが必要。

E-サイエンスは、異分野融合による新たな知見の創出に効果的であり、その推進に

においては、自己が把握していない分野の探索が目的であることが多く、検索行動の多様化対応サービスの実現、情報の公開と電子化、ユーザビリティの高いサービスの提供などが必須。

■ 留意事項:

主要関連諸機関で連携・調整、さらには関係各機関のミッションを尊重しつつ電子化情報の相互運用性についてより明確な棲み分けと連携の促進に関して、本分科会が中立的な立場から何らかの貢献ができないか検討し、方策を実行することは重要。

#### 4.4 E-サイエンスに関わる人材育成

E-サイエンスの振興には、計算科学的な「対象となる学問分野」の推進に「情報分野」の研究者が関わる形態が通常であるが、現状では、「対象となる学問分野」の研究者と「情報分野」の研究者が二つのグループに分かれてしまっている。E-サイエンスが特に強みを発揮可能な融合領域研究を推進する上では、他の分野のことも理解できる人材の育成が強く求められている。そこで、自身の研究領域に関する深い知見をもっていることと同時に、E-サイエンスを遂行するためには、その情報基盤に関する知識、あるいはそれを使いこなす技術が重要である。つまり、ITに精通しつつ、自身の研究分野についての知見も有する研究者をいかに育成するかが大きな課題である。

まず、最近、我が国で推進されている次世代スーパーコンピュータ開発プロジェクトでは、E-サイエンスに至る前段階の計算科学分野での人材育成に関する議論が盛んである。計算科学を推進する人材に必要な資質としては、コンピュータサイエンスに関する高度な知識に加えて、応用分野に関するある程度の背景的な知識をもっているか、あるいはそのような知識を効率的に学習する能力を備えていることであり、逆にそのような資質に欠ける場合は、計算科学を推進していく専門家とはなり得ない。例えば、ナノサイエンス分野あるいはバイオサイエンス分野の大規模なシミュレーション実験を行うには、対象分野の知識をある程度もち、あるいは学習しながら、その分野の専門家と連携して、大規模なシミュレーション計算を行い、しかもコンピュータサイエンスの見地からも高度・高品質なプログラム・アルゴリズムが記述できる人材が求められる。リーダーシップを発揮しながら計算科学を推進できる人材をいかに育てるか、特に、大学院生レベルでそのようなエキスパートを育てるためにどのようなコア科目を教えたら良いかが、現在大きな課題になっている。

E-サイエンスは、科学の方法論としてコンピュータサイエンスや計算科学の方法をベースにしてさらに拡張するものであり、この新たな科学の方法を推進する人材として、対象とする応用分野に関する計算科学的な高度な知識に加えて、さまざまなサイエンスデータベース、学術コンテンツを統合的に検索し、データ統合をしながら、新たな科学の真理を探究する資質が要請される。計算科学の場合と同様に、そのような能力を備えたエキスパートを育てるために、大学院生レベルでどのようなコア科目をベースにカリキュラムを組んだら良いかも大きな課題である。

さらに、E-サイエンスが、その科学の方法の要件の一つとして科学者間の協調的な研究推進をあげていることに呼応して、E-サイエンティストの教育にも、早い段階から協調的な学習環境を整え、学習者自身による問題発見、問題解決活動に基づく知識構成を誘発する新しいタイプの試みが有効であることが認識されはじめている。学習段階で自ら考え、知っていることをモデル化して現実に当てはめ、知識の異なる他者との相互吟味を経て自らの知識を構築して行く経験を十分に積むことは、コンピュータサイエンティストのみならず、E-サイエンスを駆使して科学を進化させようとする

研究者を養成する一つの強力な方法として検討に値し、そのような「E-サイエンス的な思考トレーニング」を科学における一つの根幹的な手法として、全分野で広めていくことは重要である。

#### 課題 4:

E-サイエンスの振興には、「対象となる学問分野」の研究推進に「情報分野」の研究者が関わっていく形態が通常であるが、現状では、「対象となる学問分野」の研究者と「情報分野」の研究者が二つのグループに分かれ、その間の交流が少ないことが一つの大きな問題であり、今後の情報分野全体の教育体制を含めて、より学際的な視点をもてる人材育成を行っていく体制が教育機関や研究所、さらには情報系のセンター群などで重要となる。

E-サイエンスが特に強みを発揮可能な融合領域研究を推進する上では、他の分野のことも広く理解できる人材、つまり、コンピュータサイエンスに精通しつつ、他の分野についての広い知見も有する研究者をいかに育成するかが大きな課題となる。

#### ■ 留意事項:

本分科会として、E-サイエンス分野を担う人材像をより明らかにすることが必要である。また、そのような人材を大学院レベルから養成する場合に、どのようなコア科目を教えるかの検討が重要である。

#### 4.5 E-サイエンスを推進するための情報学分野としての技術課題の発掘

他の科学の方法論、また、融合領域の研究を推進する場合に、往々にして情報分野の研究者は、他領域の研究の進展を支える存在として貢献してきた。しかし、例えば、真の融合領域研究を推進するためには、異分野の研究者とフィフティ・フィフティの関係、つまり、互惠の関係で研究を推進することが、情報分野の研究者が大きな成果を得るための必須の要件であると言える。

以上のことは、E-サイエンスの振興を図っていく過程においても同様のことが言えよう。つまり、情報分野の研究者が、E-サイエンスの推進過程において情報科学技術の新たな技術課題、研究テーマを開拓し、自身の学問分野の進展を大きく先導していくことが重要である。具体的には次のような課題が考えられる。

- (1) 世界をリードする先進的な情報ネットワーク、情報システム構築技術の提案。特に、次世代ネットワークあるいは新世代ネットワーク構築プロジェクト、グリッドプロジェクトなどとの連携のなかで、E-サイエンスと関連した情報通信技術に関する先駆的な研究開発が考えられる。
- (2) サイエンスデータベース、学術コンテンツの処理技術として新たな方法の提案。例えば、今後、センサネットワークから収集されるデータも含めると超大量のサイエンス・コンテンツがインターネット上で流通することになり、それらの種々雑多なコンテンツを収集・分析し、データベース化して蓄積し、検索要求に応え得るまでのデータ解析・加工技術の構築などを考えることができる。
- (3) 国家的事業である次世代スーパーコンピュータプロジェクトとの連携のなかで、E-サイエンスと関連した情報技術に関する先駆的な研究開発。
- (4) 情報分野自らが E-サイエンス的な研究手法の対象領域になることによって推進される研究テーマ群の推進。たとえば、全世界的規模で次世代ネットワークの実験を行う場合には、E-サイエンス的な組織間の連携や、大規模なシミュレーション実験、およびその結果に対するマイニング技術など、種々の情報科学技術によ

ってもたらされる手法が有益となる。

#### 課題 5:

E-サイエンスの推進過程において情報科学技術の新たな技術課題、研究テーマを開拓し、自身の学問分野の進展を大きく先導していくことが重要。

#### ■ 留意事項:

本分科会として、E-サイエンスの発展に本質的に関わる情報科学技術の新たな技術課題、研究テーマの掘り起こしが必要。これはE-サイエンス的な研究手法の対象領域になることによってそれ自身推進される情報学の研究テーマ群の推進も含む。

## 5. おわりに

本E-サイエンス分科会では、第20期において、欧米諸国のE-サイエンスに関わる諸活動を調査し、また、国内における学術情報基盤の整備状況や今後の整備計画などの報告を受け、これらの議論をさらに深め、本記録としてまとめた。本記録は、新たな学術の方法であるE-サイエンスに関して、その基盤となる情報インフラの世界規模での変革期という非常に重要な局面を迎えているなかで、今後、日本におけるその学術の方法のあり方についての「提言」などを行うための準備の一環と位置づけることができる。

国内においては、次世代スーパーコンピュータ開発プロジェクト、次世代あるいは新世代情報通信技術に関する開発プロジェクトが進行するなかで、学術情報基盤も大きく変容を遂げようとしており、そのような状況下で、E-サイエンスのあり様について日本学術会議の立場から積極的な発言をしていくことは非常に重要であると考えており、本分科会の活動がますます大きな意義をもってくるものと確信している。

## 参考資料

- [1] NAREGI: <http://www.naregi.org/>
- [2] ITBL: <http://www-riken.riken.go.jp/>
- [3] EGEE: <http://public.eu-egee.org/>
- [4] TeraGRID: <http://www.teragrid.org/>
- [5] Open Grid Forum: <http://www.ogf.org/>
- [6] 大型放射光施設 (SPring-8): <http://www.spring8.or.jp/ja/>
- [7] SDSS: <http://www.sdss.org/>
- [8] GLEON: <http://gleon.org/>
- [9] GEOGrid: [http://www.gtrc.aist.go.jp/project\\_geo](http://www.gtrc.aist.go.jp/project_geo)
- [10] 第3期科学技術基本計画:  
<http://www8.cao.go.jp/cstp/kihonkeikaku/kihon3.html>
- [11] 文部科学省科学技術・学術審議会学術分科会研究環境基盤部会学術情報基盤作業部会: 「学術情報基盤の今後の在り方について」(報告)(平成18年(2006年)3月23日)  
[http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/toushin/06041015.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/06041015.htm)
- [12] DEISA: <http://www.deisa.org/>
- [13] PRACE: <http://www.prace-project.eu/>
- [14] Open Science Grid: <http://www.opensciencegrid.org/>

- [15] PubMed Central: <http://www.pubmedcentral.nih.gov/>
- [16] 土屋俊:「日本の学術情報の電子化－絶望の現在と不安な将来－」(平成 19 年(2007 年)11 月 8 日: 2007 年図書館総合展講演資料)
- [17] DDBJ: <http://www.ddbj.nig.ac.jp/>
- [18] 知的財産戦略本部 知財財産による競争力強化専門調査会:「オープン・イノベーションに対応した知財戦略の在り方について」(平成 20 年(2008 年)3 月 4 日)  
<http://www.kantei.go.jp/jp/singi/titeki2/houkoku.html>