

# 科学研究用クラウドシステム（NICTサイエンスクラウド）の提案

村田 健史<sup>\*1</sup> 渡邊 英伸<sup>\*1</sup> 鶴川 健太郎<sup>\*2</sup> 村永 和哉<sup>\*2</sup> 鈴木 豊<sup>\*2</sup> 磯田 総子<sup>\*3</sup>  
山本 和憲<sup>\*1</sup> 久保田 康文<sup>\*1</sup> 長妻 努<sup>\*1</sup> 坂口 歌織<sup>\*1</sup> 津川 卓也<sup>\*1</sup> 西岡 未知<sup>\*1</sup>  
建部 修見<sup>\*4</sup> 田中 昌宏<sup>\*4</sup> 深沢 圭一郎<sup>\*5</sup> 才田 聡子<sup>\*6</sup> 海老原 祐輔<sup>\*7</sup> 藤田 茂<sup>\*8</sup>  
木村 映善<sup>\*9</sup> 黒澤 隆<sup>\*10</sup> 村山 泰啓<sup>\*1</sup> 永井 亨<sup>\*11</sup> 水原 隆道<sup>\*12</sup>

## The NICT Science Cloud — A Proposal of Cloud System for Scientific Researches —

Ken T. MURATA<sup>\*1</sup> Hidenobu WATANABE<sup>\*1</sup> Kentaro UKAWA<sup>\*2</sup> Kazuya MURANAGA<sup>\*2</sup> Yutaka SUZUKI<sup>\*2</sup>  
Fusako ISODA<sup>\*3</sup> Kazunori YAMAMOTO<sup>\*1</sup> Yasufumi KUBOTA<sup>\*1</sup> Tsutomu NAGATSUMA<sup>\*1</sup> Kaori SAKAGUCHI<sup>\*1</sup>  
Takuya TSUGAWA<sup>\*1</sup> MIchi NISHIOKA<sup>\*1</sup> Mamoru ISHII<sup>\*1</sup> Shinichi WATARI<sup>\*1</sup> Osamu TATEBE<sup>\*4</sup>  
Masahiro TANAKA<sup>\*4</sup> Keiichiro FUKAZAWA<sup>\*5</sup> Satoko SAITA<sup>\*6</sup> Yusuke EBIHARA<sup>\*7</sup> Shigeru FUJITA<sup>\*8</sup>  
Eizen KIMURA<sup>\*9</sup> Takashi KUROSAWA<sup>\*10</sup> Yasuhiro MURAYAMA<sup>\*1</sup> Masao OGINO<sup>\*11</sup> Takamichi MIZUNAGA<sup>\*12</sup>

### Abstract

This paper is to propose a cloud system for science, which has been developed at NICT (National Institute of Information and Communications Technology), Japan. The NICT science cloud is an open cloud system for scientists who are going to carry out their informatics studies for their own science. The NICT science cloud is not for simple uses. Many functions are expected to the science cloud; such as data standardization, data collection and crawling, large and distributed data storage system, security and reliability, database and meta-database, data stewardship, long-term data preservation, data rescue and preservation, data mining, parallel processing, data publication and provision, semantic web, 3D and 4D visualization, out-reach and in-reach, and capacity buildings.

Keyword Science Cloud, Data-intensive Science, Informatics, Virtual Laboratory

### 概要

データ指向型科学は、実験科学、理論科学、数値シミュレーション科学に続く第4の科学研究パラダイムと言われている。サイエンスクラウドは、データ指向型科学の基盤環境として期待されているが、現在、サイエンスクラウドとは何か、サイエンスクラウドには何ができるかなどの議論が始まったばかりである。NICTサイエンスクラウドは、情報通信研究機構が2010年より構築を進めている科学研究専用のクラウドシステムである。本稿では、NICTサイエンスクラウドについて宇宙科学研究をはじめとした科学研究の利活用面から議論することで、サイエンスクラウドの定義を行うとともに、今後、サイエンスクラウドを活用した科学研究手法・環境構築についての筋道を示す。

キーワード サイエンスクラウド、データ指向型科学、インフォマティクス、仮想研究室

- 
- \*1 情報通信研究機構 (National Institute of Information and Communications Technology)
  - \*2 株式会社セック (Systems Engineering Consultants Co., Ltd.)
  - \*3 株式会社サイエンス・サービス (Science Service Co., Ltd.)
  - \*4 筑波大学計算科学研究センター (Center for Computational Sciences, University of Tsukuba)
  - \*5 九州大学情報基盤研究開発センター (Research Institute for Information Technology, Kyushu University)
  - \*6 統計数理研究所 (Institute of Statistical Mathematics)
  - \*7 京都市立大学生存圏研究所 (Research Institute for Sustainable Humanosphere, Kyoto University)
  - \*8 気象大学校 (Meteorological College)
  - \*9 愛媛大学医学部 (Department of Medical Informatics Ehime University)
  - \*10 株式会社日立ソリューションズ東日本 (Hitachi Solutions East Japan, Ltd.)
  - \*11 名古屋大学情報基盤センター (Information Technology Center, Nagoya University)
  - \*12 株式会社クレアリンクテクノロジー (CLEALINK TECHNOLOGY Co., Ltd.)

## 1. まえがき

これまでに発展してきた実験科学（第1の柱）、理論科学（第2の柱）、計算科学（第3の柱）に対して、データ指向型科学は科学的発見のための第4の柱であると Jim Gray は提唱した<sup>1)</sup>。筆者（村田）は、実験科学を支えるインフラストラクチャは実験装置や観測装置（宇宙科学研究では科学衛星による観測やロケット実験など）であり、計算科学を行う基盤となるのがスーパーコンピュータであるのに対して、データ指向型科学を支えるインフラとして提案するのが、科学研究専用のクラウドシステム（サイエンスクラウド）であると考えている<sup>2,3)</sup>。

サイエンスクラウドは、2008 年ごろにイリノイ大学によって提唱された<sup>4)</sup>。2010 年に第1回サイエンスクラウドワークショップ (ScienceCloud Workshop) が米国・シカゴにおいて開催された<sup>5)</sup>。その後、毎年1回の同ワークショップが開催され、サイエンスクラウドについての議論が進められてきた<sup>6,7)</sup>。これらのワークショップにおいてなされた議論は、クラウド基盤上での科学研究事例検討、クラウド環境および技術の有効性評価、クラウドシステムの耐障害性（fault tolerance）および信頼性、クラウド上でのデータ指向型研究の工程やツール、Map Reduce などのプログラミングモデルの利用、クラウドストレージ設計、クラウド内での I/O やデータ管理、クラウド内のワークフローやリソース管理、NoSQL などのクラウド技術の科学研究への利用、データストリーミングや動的アプリケーション、HPC 環境におけるクラウドコンセプトの応用、高機能並列分散ファイルシステムと仮想環境接続、クラウドセキュリティーに関する研究および実験などと多岐にわたる（表1）。

表1 第1回 ScienceCloud ワークショップでの議題一覧 (ScienceCloudWorkshop2010<sup>5)</sup>)

Scientific application cases studies on cloud infrastructure
Performance evaluation of cloud environments and technologies
Fault tolerance and reliability in cloud system
Data-intensive workloads and tools on clouds
Use of programming models such as Map-Reduce and its implementations
Storage cloud architectures
I/O and Data management in the cloud
Workflow and resource management in the cloud
Use of cloud technologies (e.g., NoSQL databases, etc) for scientific applications
Data streaming and dynamic applications on clouds
Application of cloud concepts in HPC environments
High performance parallel file systems and interconnects in virtual environments
Research and best practices in cloud security

これらの議論からわかることは、データ指向型科学研究のインフラストラクチャとしてサイエンスクラウドは重要であるという共通認識はあるものの、多くのクラウド技術は商用クラウドに共通したものであり、それらの技術を科学研究に流用しているにすぎないという点である。すなわち、技術的議論を通じての多くの参加者の興味や目的は、サイエンスクラウドとは何か、いわばサイエンスクラウドの定義を行う事である。サイエンスクラウドが、ビッグデータ科学のインフラストラクチャとして位置付けられることには誰も異論がないが、具体的に何をどのように問題解決するものであるかという事が明確ではない。

本稿ではサイエンスクラウドの個別の技術について議論するのではなく、利用面からのサイエンスクラウドを提案する。サイエンスクラウドを使った研究手法（研究スタイル）を規定することで、サイエンスクラウドにより期待される科学研究成果や、サイエンスクラウドに求められる個別の技術が明らかになるからである。

## 2. サイエンスクラウドとビッグデータ

### 2.1. サイエンスクラウドの事例

現在、欧米諸国を中心に様々なサイエンスクラウド（ScienceCloud）またはそれに準ずるシステムが提案または運用さ

れている（表 2）。サイエンスクラウドは、前述のとおりその定義が議論されている状況であり、これらのシステムを通じてその有効性や必要とされる技術が検討されている。また、表 2 を含む多くのシステムの中で、特定の研究テーマについてテストケースとしての成果が挙げられているものはあるが、複数の研究テーマで大きな成果を挙げた汎用性の高いサイエンスクラウドはまだ存在しない。

表 2 世界のサイエンスクラウドの動向

国名または組織名	クラウド名またはグループ名	URL	概要
CERN, CNR-IREA	HELIX NEBULA	<a href="http://helix-nebula.eu/">http://helix-nebula.eu/</a>	To create a multi-tenant Open Market Place for Science
INRIA Rennes (France)	Centre de recherche commn	<a href="http://www.inria.fr/centre/rennes">http://www.inria.fr/centre/rennes</a>	Scalable Storage for Sharing Application Data on Azure Clouds: TomusBlobs
	The Venus-C project	<a href="http://venusfeedback.codeplex.com/">http://venusfeedback.codeplex.com/</a>	
	SCIDP-ES	<a href="http://www.scidip-es.eu/">http://www.scidip-es.eu/</a>	Long-term data preservation for Earth Science
(Lavanya Ramakrishnan)	FRIEDA	<a href="http://frieda.lbl.gov/">http://frieda.lbl.gov/</a>	Flexible Robust Intelligent Elastic Data Management in Cloud Environments
	The Open Science Data Cloud (OSDC)		A petabyte-scale science cloud managed and operated by the Open Cloud Consortium (OCC)
	“Science Clouds”	<a href="http://scienceclouds.org/">http://scienceclouds.org/</a>	Science Clouds Blog

## 2.2. ビッグデータ処理の基盤技術

IT ビジネス分野の中心となりつつあるビッグデータという概念は、科学研究分野でも適用され始めている。1 節で議論した第 3 の研究手法である数値シミュレーションなどにおいて大規模計算の中核をなす技術の一つが HPC（High Performance Computing）である。HPC は計算指向型（compute-intensive）と言われており、個別のアプリケーション（プログラムや数値計算コード）の処理速度を最大化することが主目的となる。しかし、データ指向型（data-intensive）科学では、数値シミュレーションを含むあらゆる科学データを処理することが目的である。

一方、多様な計算機リソースを融合し、データ分散、並列データ処理やコンピュータとデータファイルをローカライズする工夫など、総合的なデータ処理環境をめざすのが MTC（Many-Task Computing）という考え方である。クラウド計算機リソースによる並列分散処理などがこれに該当する。

また、計算処理ではなくデータに視点を置いた考え方で提案されているのが DIC（Data-intensive Computing）である。Jim Gray が提言した第 4 のパラダイムであるデータ指向型科学では、数値シミュレーションを含むあらゆる科学データを処理することが目的である。クラウド内に配置された大規模科学データを解析するビッグデータ科学のためには、DIC 技術の向上が必須である。

特定パターンの商用サービスを実現することが多い民間クラウドと比較すると、複雑で特殊な目的実現型のサイエンスクラウドでは HPC はその一部であり、MTC 環境の実現が重要となる。

## 2.3. NICT サイエンスクラウド

1 節や 2.1 節で述べたように、現在欧米を中心にサイエンスクラウドに関する取り組みが行われているが、特定の研究目的に設計されたものや、民間のクラウドを模したシステムも多い。その結果、汎用性が高く多数の大規模プロジェクトで成果を達成したサイエンスクラウドはこれまでに存在しない。

本論文で議論する NICT サイエンスクラウドは、あらゆる科学分野でほとんどのデータがデジタル化されている現在、科学研究の第 1 の柱、第 2 の柱、第 3 の柱<sup>1)</sup>によるあらゆる科学データを収集・保存・解析（可視化）・公開データ処理するデータ指向型科学研究環境として設計された（図 1）。サイエンスクラウド構築においては、利用者（すなわち研究者）の視点が欠かせない。NICT サイエンスクラウドはユーザ指向型の設計を行うため、2010 年ごろからプロトタイプを構築してユーザ利用を進めると並行して、ユーザからのヒアリングを積極的に行った。

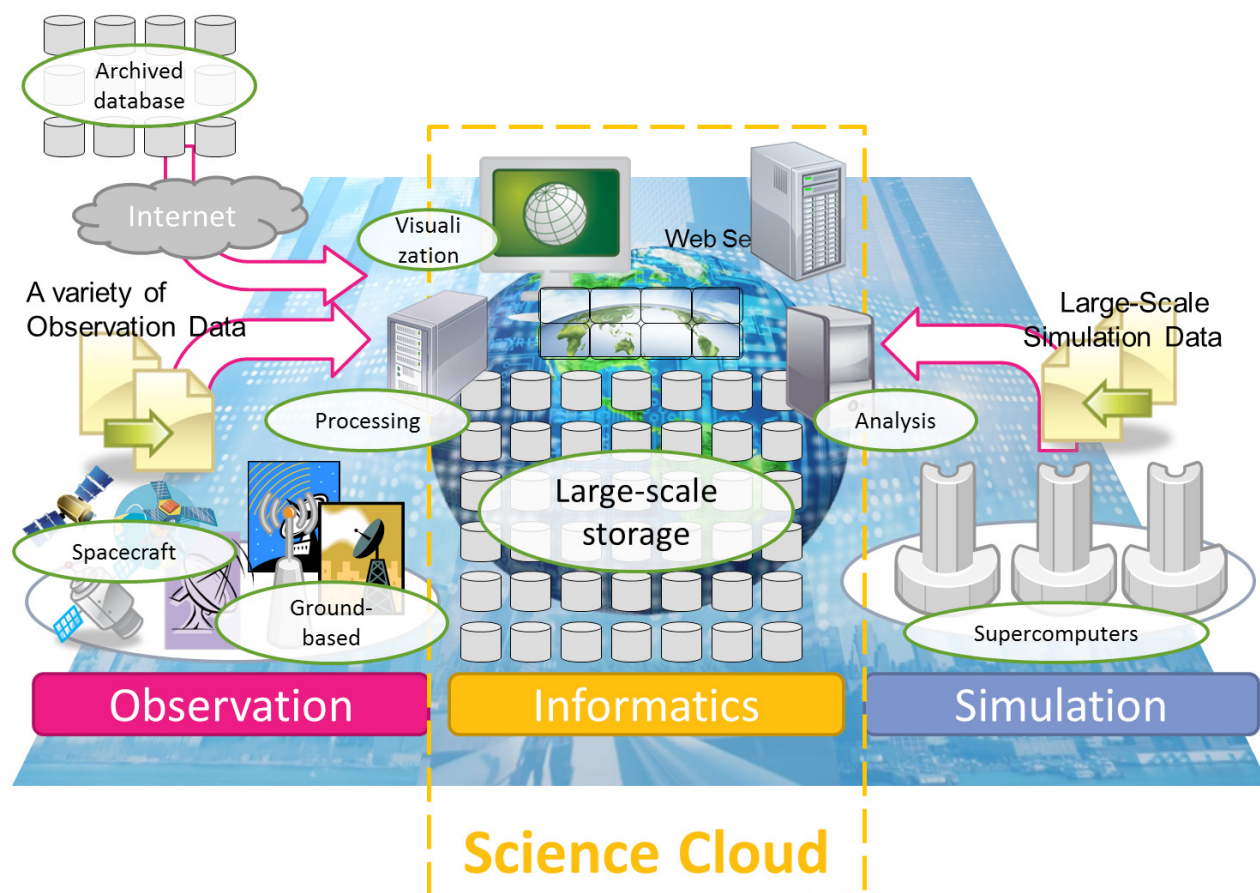


図 1 NICT サイエンスクラウド基本コンセプト

世界科学データシステム（WDS: World Data System）は、科学界の国連と呼ばれる国際科学会議（ICSU: International Council for Science）が実施しており、科学データに関する国際的取組の高度化を目指すプログラムである。情報通信研究機構（NICT）と ICSU は、本プログラムの実施主体となる国際プログラムオフィス（IPO: International Programme Office）に関する協力文書を 2011 年に締結した<sup>8)</sup>。本研究ではこの活動を背景として、広く科学研究者が要求または期待する要件を収集した。これまでに調査したサイエンスクラウドに求められる機能またはサービス要件を分析すると、クラウドサービスとしての機能とデータ指向型科学研究のインフラストラクチャの 2 つの側面があることが分かった。

クラウドサービス機能は、サイエンスクラウドの利便性に関わる機能である。民間クラウドと共通する機能であり、研究者がサーバ、ネットワーク、ユーザ管理等の管理から解放されることで研究の利便性が向上する側面である。民間クラウドと異なるのは、研究用のサービス（アプリケーション、ツール、システム等）が有効である点である。民間クラウドは、一般には計算機リソースや商用アプリケーションサービスの提供が主体となるが、サイエンスクラウドは研究に特化したサービスや研究用にカスタマイズできるツールを提供することで有効性が高まり、また商用クラウドとの差別化も行われると考えられる。NICT サイエンスクラウドの研究支援サービスについては、4 節において議論する。

データ指向型科学研究のインフラストラクチャ機能は、サイエンスクラウドの高度利用に関わる機能である。その中でも、特に大規模データ処理（ビッグデータ科学）と複合型データ処理が要求される機能の中心である。大規模データ処理では、単に大規模なストレージにデータを保存するだけでなく、それらのデータの収集、管理、公開と処理が期待される。また、複合型データ処理では、データ観測や数値計算、可視化やアウトリーチなど、サイエンスクラウド上においてすべてのプロセスを実現することが期待される。データ指向型科学研究事例については、5 節において議論する。

大型科学研究プロジェクト推進では、協調的研究環境としてのサイエンスクラウド利活用も重要である。近年、ほとんどの大型研究計画は国際プロジェクト化されており、高い成果を挙げるためには国際的協調研究環境が必須である。NICT サイエンスクラウドの重要な機能の一つが、国際的協調研究を行うための仮想研究室（バーチャルラボラトリー）機能（図 2）である。国際的協調研究環境を NICT サイエンスクラウド上に構築することにより、これまでよりもはるか



に短い期間で、多くの研究成果を生み出すことが期待される。バーチャルラボラトリーの実例については、5 節において議論する。

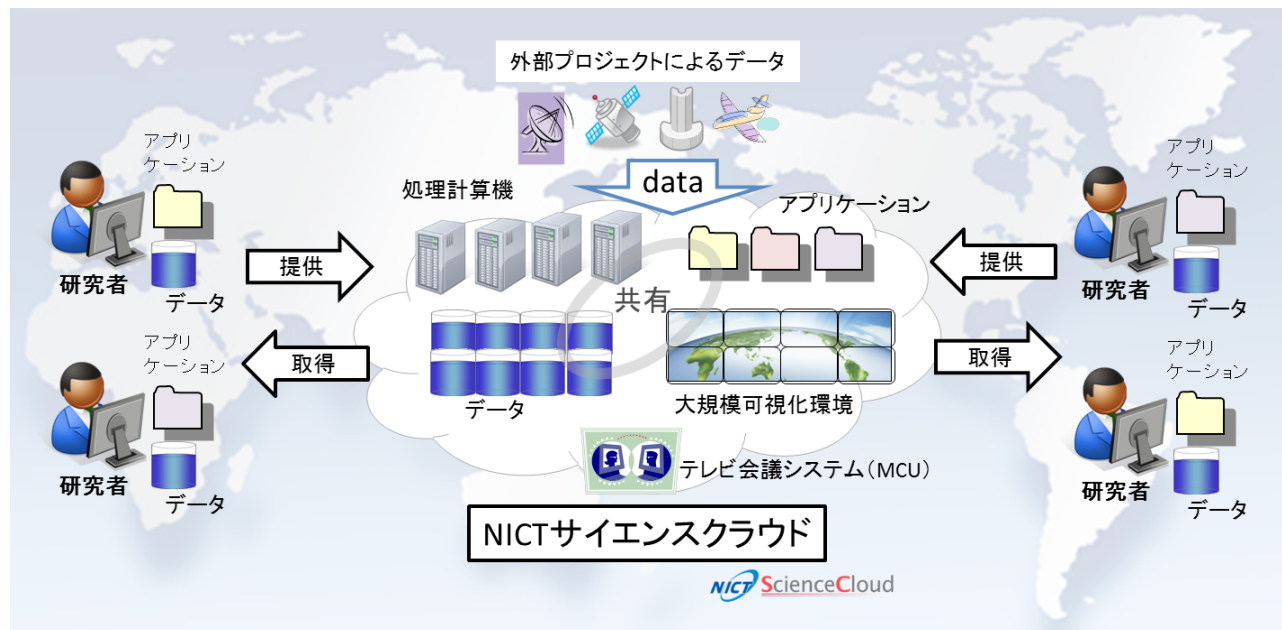


図2 仮想ラボラトリーによる協調的研究開発

### 3. NICTサイエンスクラウドのシステム概要

本節では、NICTサイエンスクラウドのシステム概要について議論する。NICTサイエンスクラウドは、広域分散型のクラウドシステムであり<sup>9)</sup>、2013年現在で国内5か所のDC（データセンター）をL2接続することで構成されている（図3）。インターコネクトには、情報通信機構の広域テストベッドネットワークであるJGN-X<sup>10)</sup>を用いている。

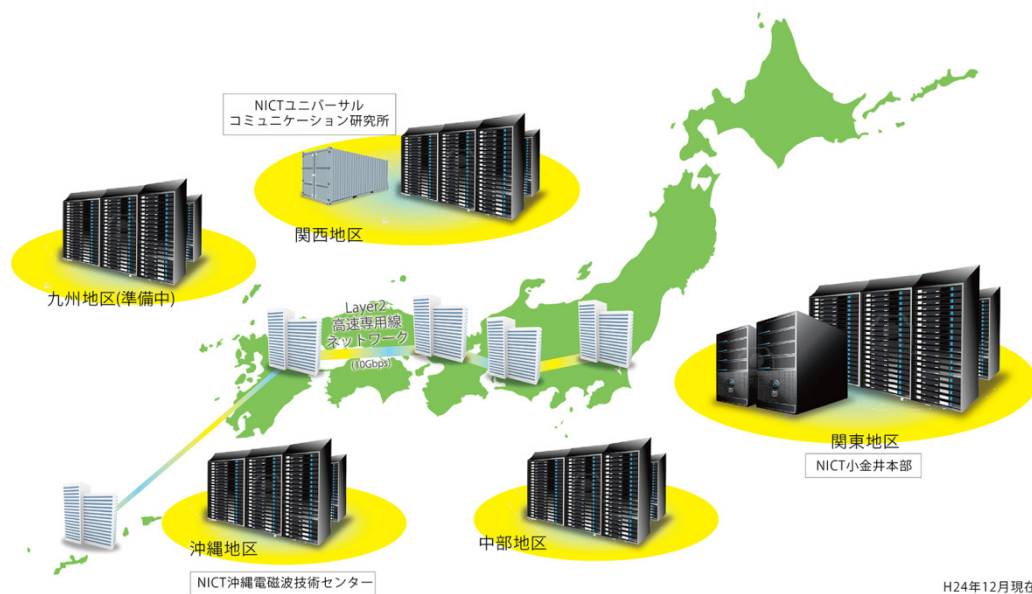


図3 NICTサイエンスクラウドデータセンター

図4に、NICTサイエンスクラウドのシステム構成図を示す。図4はクラウドリソース全体を表しており、すべてのクラウドユーザが利用できる安定で冗長化されたクラウド共通リソースと、ビッグデータ処理または定常データ処理を対象とするクラウド拡張リソースから構成される。NICTサイエンスクラウドは、図4に加えて、4節で述べる各種研究支

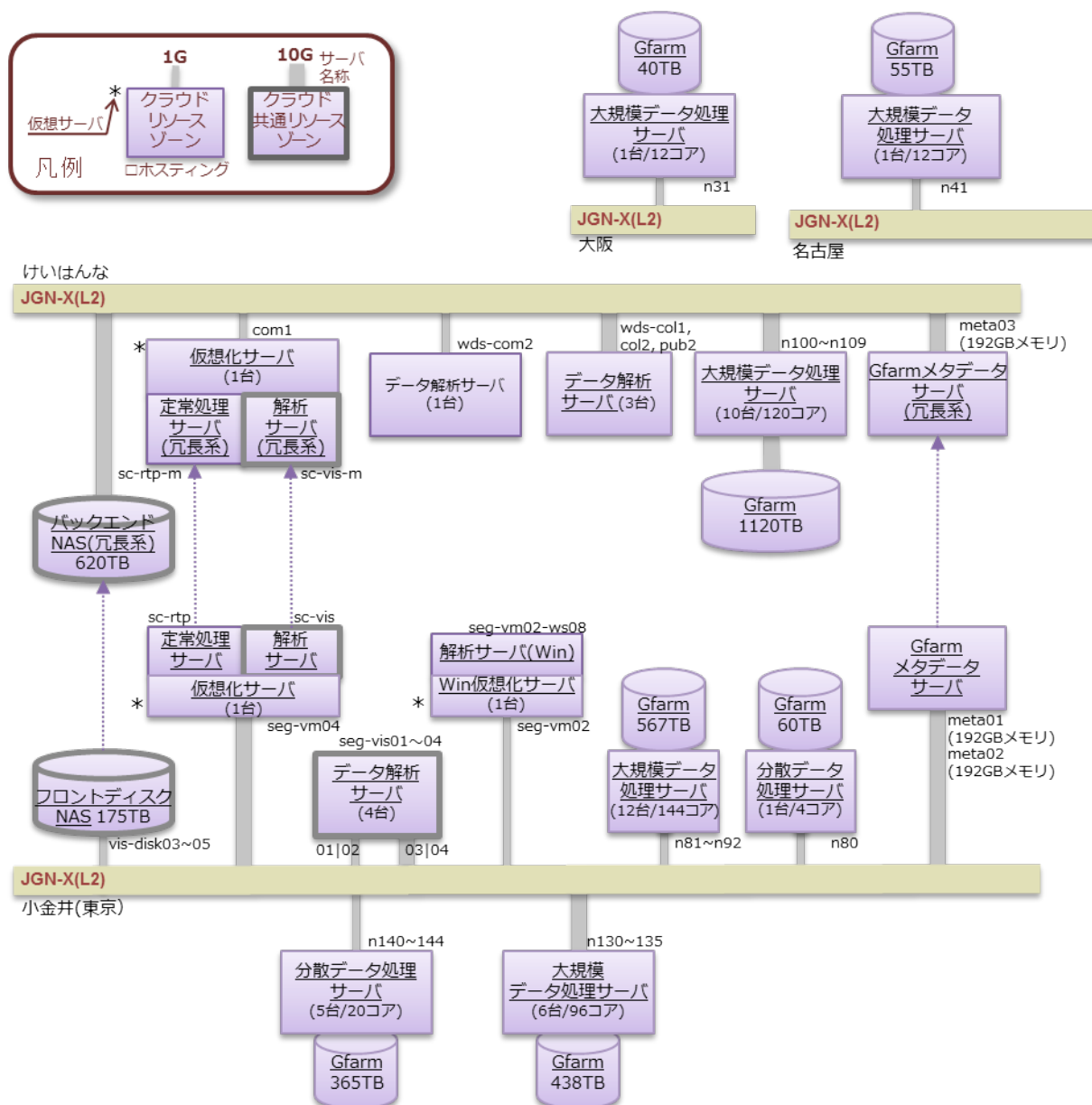


図4 NICTサイエンスクラウドシステム構成図

援サービスを行うサーバ群と4.6節に示す独自研究環境（ハウジング・ホスティングサーバ群）から構成されているが、本節ではクラウドリソースについてのみ議論する。

図5は、NICTサイエンスクラウドリソース利用を、クラウドリサービス利用と研究開発システム利用に分けて示している。多くの研究者は、クラウド利用を小規模な研究からスタートする。図6に示すクラウド的利用であり、商用クラウド利用と比較的似た利用である。クラウド的利用の場合には、安定性・冗長性に優れた共通リソースの利用が必須である。データの保存が多重化されているだけでなく、システム停止の頻度も小さいため、継続的な研究環境が保障されるからである。また、NICTサイエンスクラウド側がファイアウォール（IPSを含む）、ユーザ管理（LDAP）、データファイルのウィルススキャンなどを行うことで、研究者（ユーザ）の利便性は独自の研究環境と比較して高くなる。また、NICTサイエンスクラウドにライセンスされた民生品の研究用アプリケーションが利用できることも、サイエンスクラウドの利便性の一つである。これまでは、多くの研究者が個別にこのような研究環境を構築・運用してきたが、NICTサイエンスクラウドを用いることで構築・運用から解放され、研究に集中できることが期待される。

研究が進展すると、NICTサイエンスクラウド利用目的は、単純な利便の利用から、処理データの大規模化や特殊化、多目的化などに発展することが多い。また、4.6節で述べるように独自研究環境をサイエンスクラウドに移植し、クラウド

ドリソースを活用して目的を達成することもある。NICTサイエンスクラウドを活用した研究開発システム利用（図5）については、利活用する研究支援サービスを4節で、利活用事例を5節で述べる。

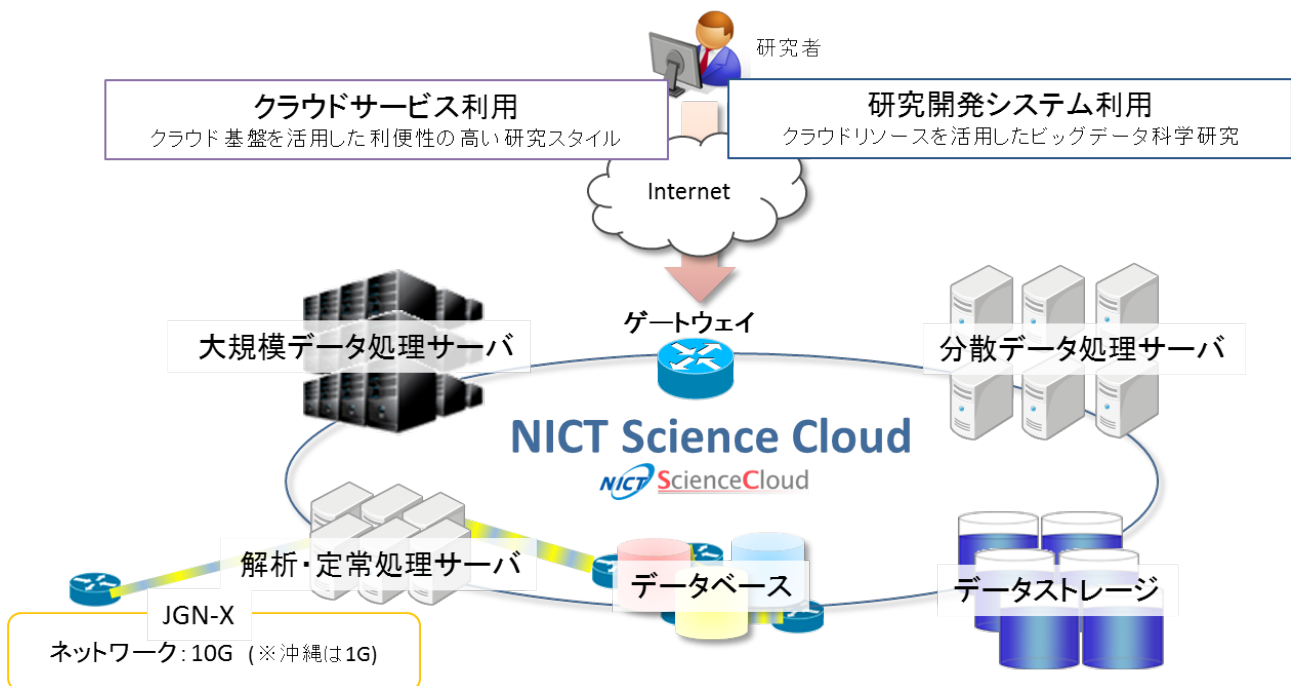


図5 NICTサイエンスクラウド利活用概要

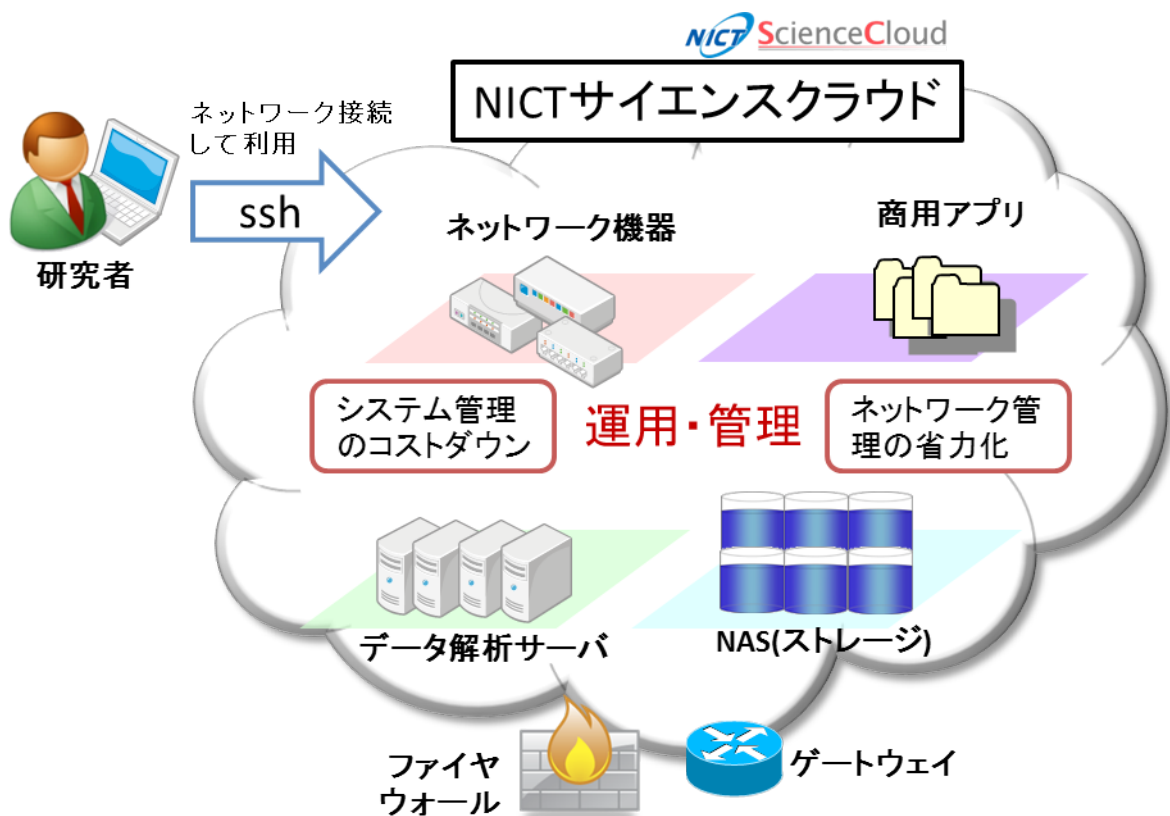


図6 クラウド研究環境利活用（クラウド的利用）

## 4. NICTサイエンスクラウドの研究支援サービス

### 4.1. クラウドエコシステム

商用クラウドでは、多くのソリューションが特定ベンダの独自技術と仕様に基づいており、その結果として囲い込み（ベンダロックイン）がユーザの利便性を損なうことが多かった。この状況を改善するため、ソフトウェア開発者、SIer、クラウド事業者、さらにクラウドユーザから構成されるステークホルダーによる、オープンソースのクラウド基盤ソフトウェアの利用と協業によるクラウドサービスやクラウドソリューションの提供という流れがある。オープンクラウドで市場を活性化し、共存共栄で成長するクラウドの生態系をつくり出そうという取り組みはオープンクラウドエコシステムと呼ばれている。

### 4.2. 科学研究用エコシステム

エコシステムは商用クラウドで提案された概念であるが、サイエンスクラウドでもエコシステムに取り組む事例もある<sup>11)</sup>。Helix Nebulaのエコシステムは、科学データから価値のある情報を抽出することを目的としており、これにより科学研究者と民間産業を結びつける試みである。

しかし、NICTサイエンスクラウドユーザからのヒアリングによると、クラウド基盤を使った新しいサイエンス（特にビッグデータ科学）を行う場合には研究テーマごとの特殊性が高く、エコシステムの実現は容易ではない。ただし、商用クラウドと比較すると専門性は高いが、図7に示すような科学研究分野で汎用性があるアプリケーション、ツールおよびシステムは可能である。NICTサイエンスクラウドではこのような汎用的な科学研究用アプリケーション・ツール・システム（科学研究用エコシステム）を開発し、公開してきた（表3）。本節では、それらの一部を紹介し、その有効性や汎用性について議論する。

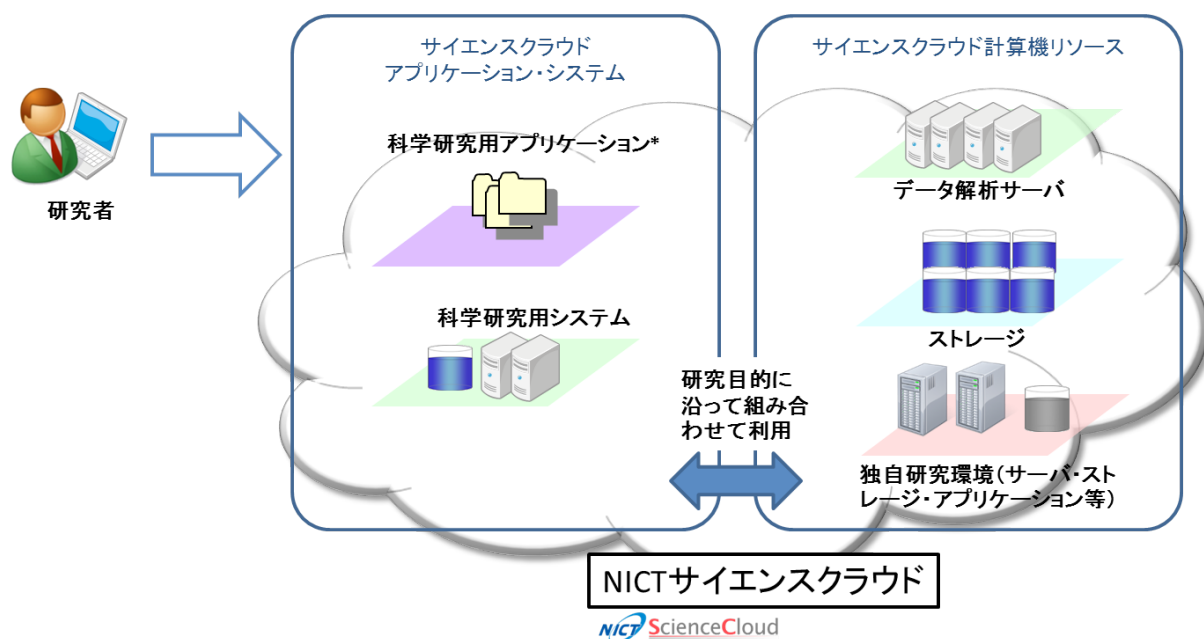


図7 科学研究用アプリケーション・システム利活用

### 4.3. 科学データファイルクローラ（自動収集）サービス（NICTY/DLA）

サイエンスクラウドにおいてデータ処理を行うために重要な機能の一つが、データ自動収集である。現在、科学研究データの多くはインターネット上で公開されているが、公開されるデータ量は多くの分野で増加している。図8は、NICTが収集している国内外の公開GPS（GNSS）受信機データ数の経年変化である。現在、5000を超えるGPS受信機の観測データが公開されており、公開サイトも20を超えている。このような公開データの多くは、各組織が独自のポリシーに従って公開している。データ更新の頻度、公開プロトコルはもちろんのこと、場合によってはデータ較正の変更のために公開サイト上の過去データファイルが連絡なく更新されることもある。また、サイト上で公開されるデータファイル形式、ファイル名命名方式、データ公開範囲やアクセス制限なども、公開サイトによって異なる場合がある。



表 3 NICT サイエンスクラウド科学研究用エコシステム利用例<sup>21)</sup>

科学研究用エコシステム名	概要
NICTY/DLA	インターネット上で公開されている科学データ（特に時系列観測データ）のメタ情報および実データファイルをクロールするツール。
バーチャルオーロラツール	地球および惑星磁気圏・電離圏の MHD シミュレーション（多流体系を含む）を 3 次元時系列可視化するためのツール。AVS Express（Linux および Windows）で動作する。
広域観測ネットワーク監視システム（WONM システム）	地球広域の観測網において観測所・観測装置からのデータ伝送を継続的に行うための監視システム。データ伝送サーバの監視、データ伝送状況の監視と再送、ネットワーク状態のモニタリングなどを行う。
WSDBank アプリケーション	サイエンスクラウド上にデータを保存または公開するための Web アプリケーション。（データファイルは NICT サイエンスクラウドのフロントエンドストレージに保存され、定期的にバックエンドストレージにバックアップされる。）
STARS（Solar-Terrestrial data Analysis and Reference System）	各種衛星観測・地上観測データを統合的に解析するツール。全対象データファイルをカプセル化したクラスライブラリ（SEDOC）のみを利用することもできる。

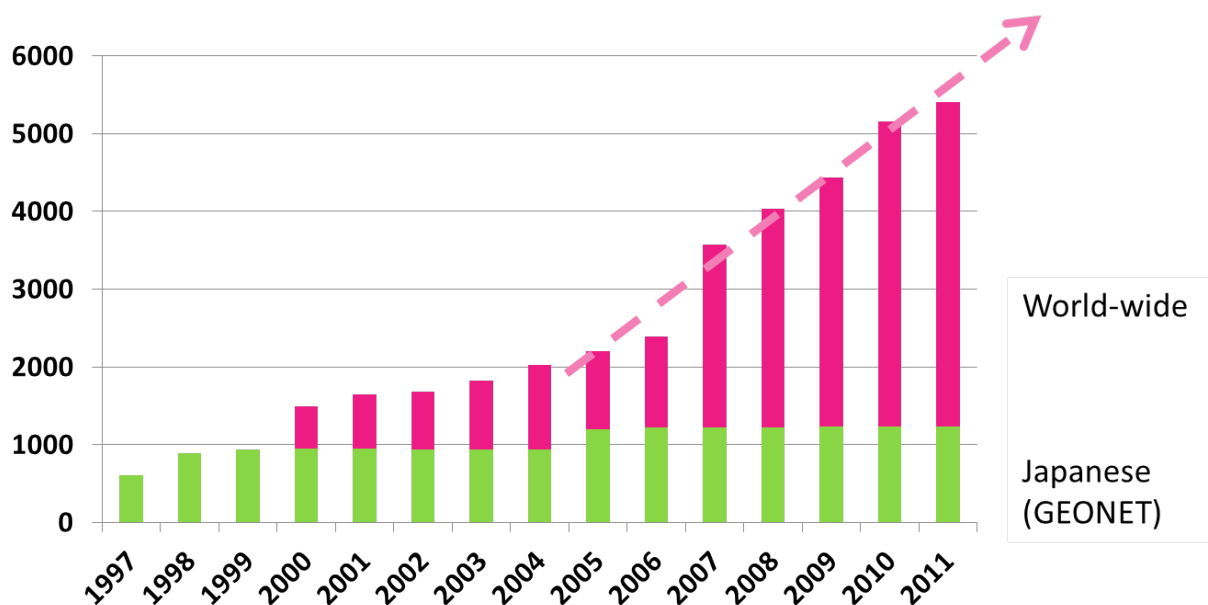


図 8 NICT が収集している国内外の公開 GPS 受信機データ数のトレンド

NICTY/DLA は、RSS によりメタ情報を記述し、FTP/HTTP などの標準的なプロトコルに対応することで、公開データの独自性や流動性を吸収できる<sup>12)</sup>。図 9 に NICTY/DLA（Download Agent）の機能概要を示す。NICTY では、あらかじめ対象となるデータの情報（ホスト名、ディレクトリ名、ファイル名フォーマットや公開ポリシーなど）を登録することで公開データのメタ情報をクロール・データベース (DB) 化し、これをもとにして各種プロトコルで公開されている実データファイルのダウンロード・DB 化を行う<sup>12)</sup>。

NICTY/DLA により独立して収集されたデータファイルは多様であり、多数のデータを解析するユーザがそれらのデータファイルから容易に情報を取り出すことができる環境が重要である。しかし、太陽地球系物理分野を含む関連する分野の科学データ記述形式を統一化することは容易ではない。NICT サイエンスクラウドでは、収集するデータのうち 400 種類のデータをカプセル化したクラスライブラリ (SeDOC クラス) を準備した<sup>12)</sup>。これによりユーザは、データ ID とデータ期間 (StartTime と EndTime) を指定するだけで、データファイル形式やデータの場所を意識することなく get 関数によりデータを容易に取得できる。SeDOC クラスは C++ 言語により記述されており、Java 言語からの呼び出しも可能である。

図 10 は、NICTY/DLA によるデータファイルの自動収集実績である。NICT サイエンスクラウドでは、NICTY/DLA をシステムとして実装し、定常運用を行っている。2013 年 4 月現在で、47 ホスト (428 データ) に対応した。これらのデータ対応において RSS に新たな定義を追加することで、順次データ対応を進めた。これまでに NICTY/DLA の収集対象と

なった 428 データにおいては、データポリシを RSS で吸収できなかった例はない。図 10 には、NICT サイエンスクラウドが NICTY/DLA により自動収集しデータベース化した科学研究データファイルのファイル数とデータサイズも示した。1200 万を超える観測データファイルと 7TB を超えるデータファイルが NICT サイエンスクラウド上に蓄積され、現在も最新の公開データが 6000 ～ 10000 ファイル／日で収集されている。

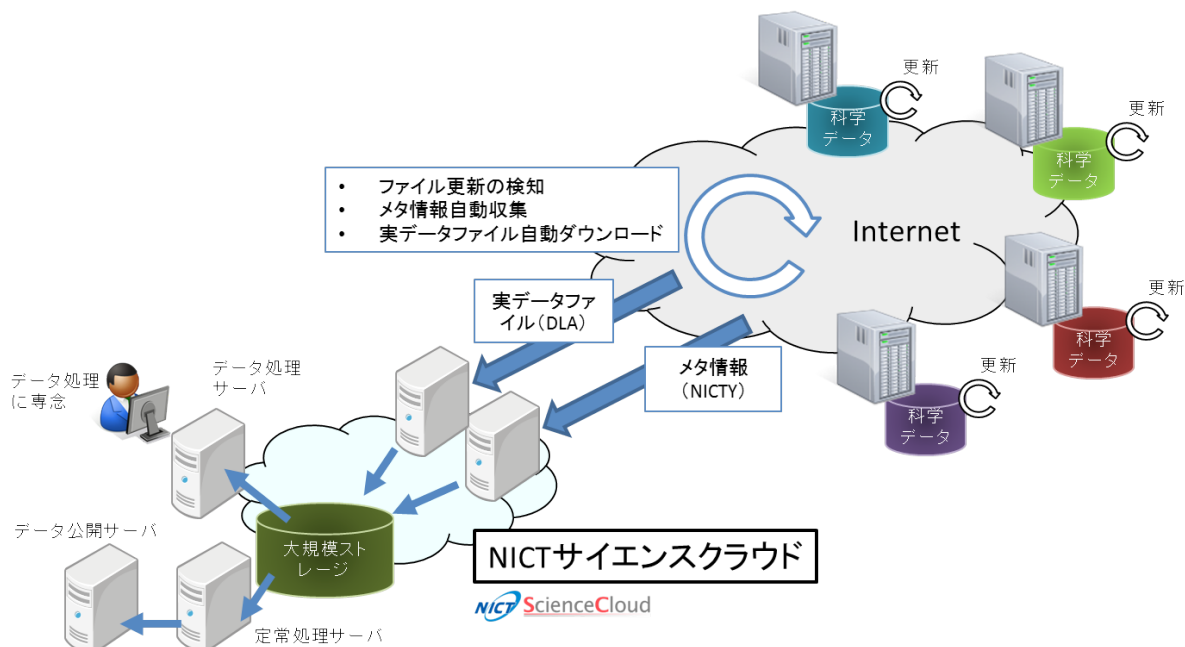
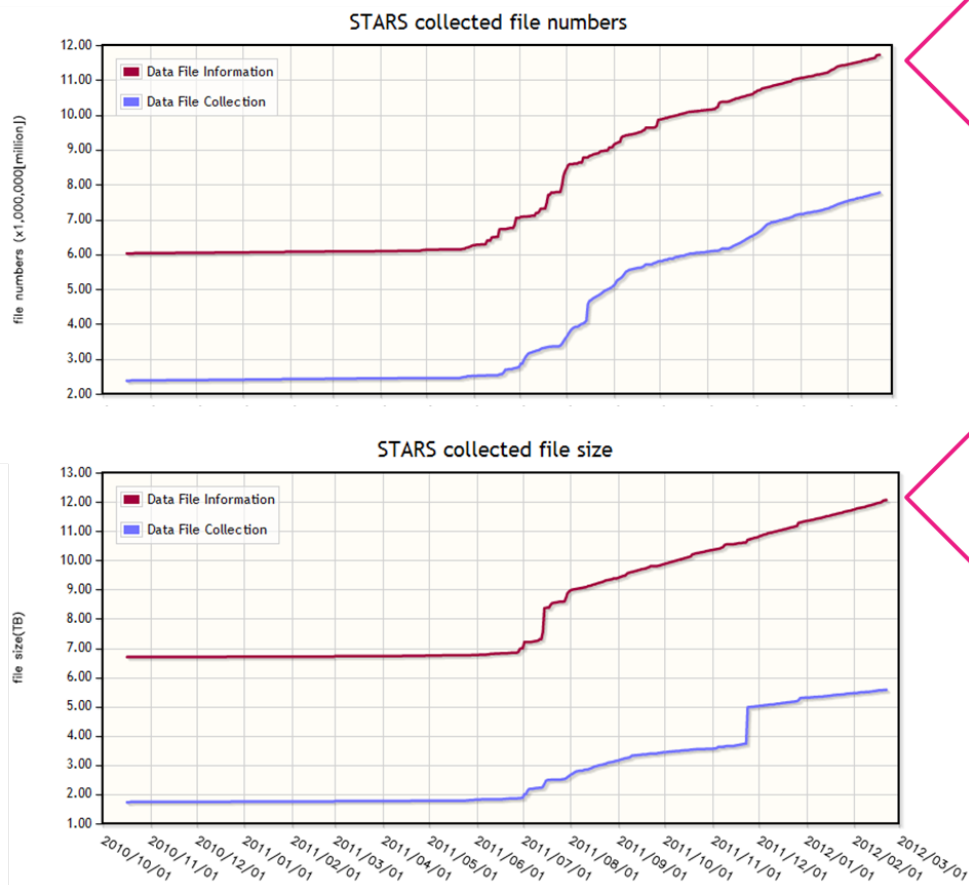


図 9 科学データファイルクローラ（自動収集）サービス（NICTY/DLA）の仕組み<sup>12)</sup>



More than 11  
million files

More than 12  
tera bytes

図 10 NICTY/DLA<sup>12)</sup> によるデータ自動収集実績（赤線が NICTY、青線が DLA）

#### 4.4. 広域観測ネットワーク監視システム

NICT サイエンスクラウドに示す通り、NICT サイエンスクラウドの基本コンセプトは、さまざまな形で公開されている科学データを収集し、格納し、解析することである（図 1）。地球・宇宙観測データについては、NICTY が収集対象とするようなインターネット上での公開データが対象となるが、研究機関が観測するデータを観測施設（Observatory）から直接伝送し、DB 化することもある。

現在、地球規模での地上観測ネットワーク網の試みが、様々な分野で進められている。これらの観測拠点数は、少ないものでは数か所であるが、多いものでは国内外で数千か所を超える大規模ネットワークもある。業務系・現業系の組織は別として、研究組織においてはこのような多数の観測拠点または観測装置を管理・運用することは容易ではない。一方で、地球規模での地球・宇宙環境・現象の理解には、このような大規模観測網が不可欠である。

特別な専用システムを除くと、一般研究機関での地球規模の観測システム運用やデータ収集には、専用ネットワーク回線ではなくインターネットが利用される。しかし、インターネットの物理層は 10Gbps を超える超高速なものから kbps オーダーで不安定な場合、また有線、無線などが混在している。また、インターネットでは QoS（Quality of Service）を確保することが難しく、ベストエフォートでのデータ伝送しか期待できない。これらの事情が、一般的な地球規模観測を困難にしている。たとえば、情報通信研究機構（NICT）の地上観測網も例外ではない。NICT 宇宙環境インフォマティクス研究室では国内外を含めて 30 以上の拠点に観測所を直接的または間接的に運用しており、40 種類以上のデータが日々伝送されている。これらの観測拠点の観測機器およびデータ伝送の維持運営が大きな負担となりつつある。そこで、NICT サイエンスクラウドでは、広域観測拠点と NICT サイエンスクラウドを結ぶネットワークを仮想化し、運用を一元化するためのシステムとこれを管理運営する Web アプリケーション（広域観測ネットワーク監視システム）を開発した（図 11）。ここで使われるネットワーク技術は新しいものではないが、L3/L2（VPN）ネットワークを混在させ、データ再送制御とデータファイル整合性の確保、機器の遠隔監視、ルーティング制御、分散データベースとの連携などを組み合わせることで、悪条件化でも観測とデータ伝送を継続的に行う OCM（Observation Continuity Model）に基づいた設計となっている。さらに、これを小型サーバに組み込むことで、1 台 10 万円程度の安価なデータ伝送システムパッケージも実装した。

現在、NICT 宇宙環境インフォマティクス研究室では国内外 8 拠点にこのシステムを配置しており、拠点数はさらに増える予定である。室内実験として数 10 台のサーバでのこのシステムを試験運用し、悪条件ネットワーク環境でデータが

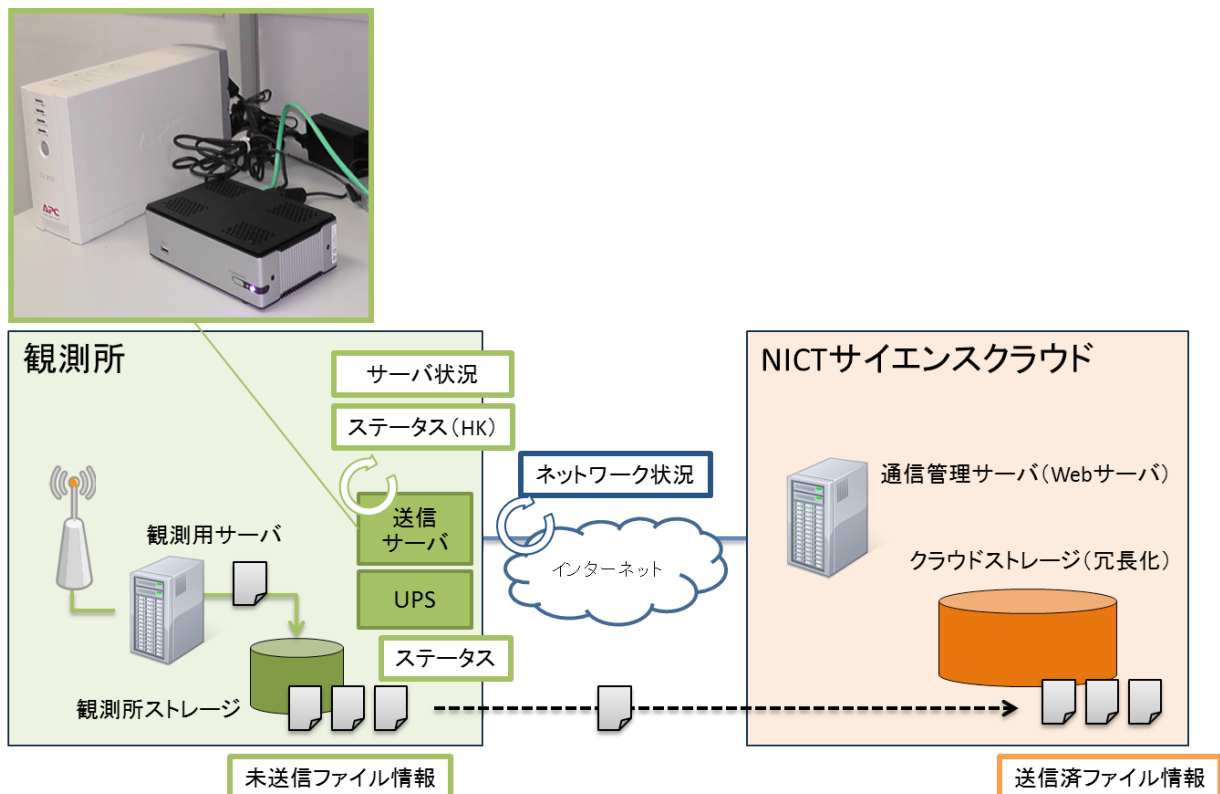


図 11 広域観測ネットワーク監視システム（写真はパッケージ化されたデータ送信サーバと UPS）

定常的に収集できることを実証している<sup>13)</sup>。なお、このシステムは、TCP/IP ベースであればネットワーク環境を選ばず、観測種類も選ばない汎用性の高いシステムである。

#### 4.5. 地球磁気圏 3 次元可視化ツール：バーチャルオーロラツール

筆者らの研究グループでは、宇宙・地球科学シミュレーションの可視化を積極的に進めており<sup>14,15,16,17,18,19)</sup>、ツールの標準化が有効であることを経験してきた。そこで、2008 年より地球近傍の宇宙空間（電離圏・磁気圏など）を対象とした 3 次元流体シミュレーションを時系列可視化するツールであるバーチャルオーロラツール（図 12）を開発している。バーチャルオーロラツールはサイバネット社の AVS Express<sup>20)</sup> により実装された、地球磁気圏物理学で必要となる標準的な可視化機能を網羅したツールである。バーチャルオーロラツールでは、対象となる時系列電磁流体（MHD）データを HDF5 形式とし、データファイル情報（メタ情報）を XML 形式記述された SPL（StarsProjectList）ファイル<sup>21)</sup>として表記することで、上記の可視化事例を含む様々な MHD シミュレーションコード（多流体型を含む）に対応しており、2013 年 4 月時点で、4 種類の Global MHD シミュレーションコードに対応している<sup>22)</sup>。

バーチャルオーロラツールは、AVS Express が対応している Windows と Linux の主要なディストリビューションで動作し、NICT サイエンスクラウド内のサーバ上であればライセンスを利用できる。

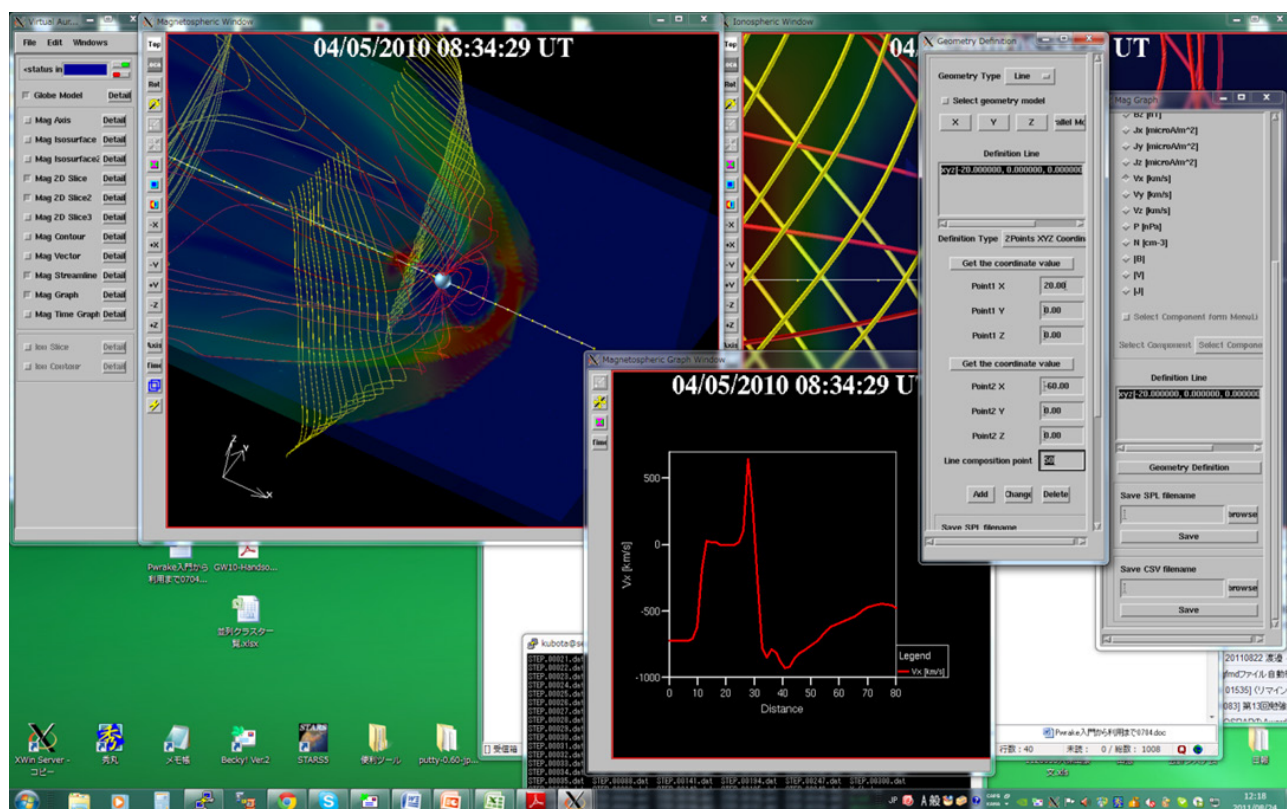


図 12 地球磁気圏 3 次元可視化ツール：バーチャルオーロラツール（図は Linux 版）

#### 4.6. 独自研究環境利用サービス（ハウジング・ホスティング）

アプリケーションやシステムといった科学研究用エコシステムとは別に、独自の研究環境を NICT サイエンスクラウド上で実現するためのハウジングやホスティングサービスも、サイエンスクラウドでは最も重要な機能の一つである。NICT サイエンスクラウドでは、特定の研究プロジェクトのためではなく汎用性の高いコンピュータリソースの提供を目指している。したがって、個別の研究テーマにおいて、NICT サイエンスクラウドのリソースで目的が達成できるとは限らない。そこで、NICT サイエンスクラウドでは、独自の研究環境を導入し、個別の研究目的に有効な機能のみを NICT サイエンスクラウドから利用する独自研究環境利用サービスを提供している（図 13）。

なお、独自研究環境利用サービスは商用のハウジングサービスとは (1) サーバ管理はユーザが行う、(2) NICT サイエンスクラウドのリソース（ファイアウォール／IPS やユーザ管理サービスを含む）を積極的に利用できるなどの点で異な



るため、NICTサイエンスクラウドではハウジングサービスと呼んでいない。ホスティングサービスは、商用クラウドとほぼ同じであるが、上記(2)のNICTサイエンスクラウドのリソースを積極的に利用する点が目的達成型のサイエンスクラウドの特徴である。

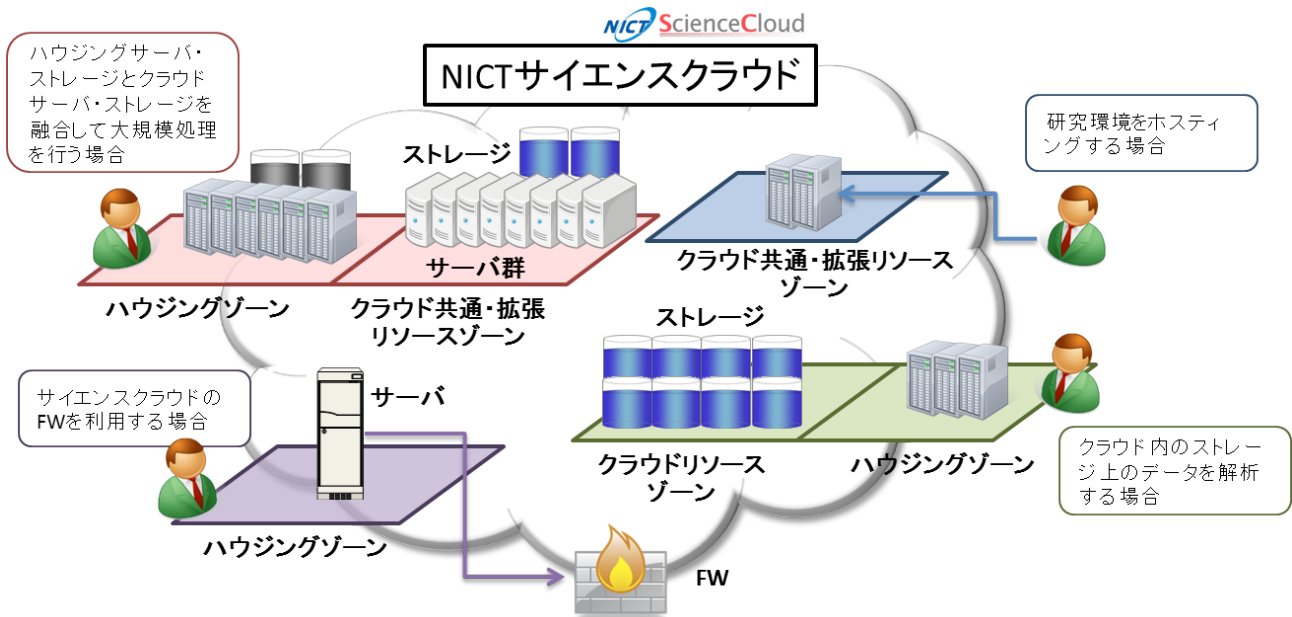


図 13 独自研究環境利用サービス（ハウジング・ホスティハウジング）

## 5. NICT サイエンスクラウド活用事例

前節までに、商用クラウドと比較したサイエンスクラウドの特徴や現状（2節）とNICTサイエンスクラウドのシステム（3節）および研究支援サービス（4節）について議論した。NICTサイエンスクラウドは、表2に示す各国のサイエンスクラウドと比較して、特に各研究テーマのためのカスタマイズ機能を重視している。本節では、宇宙科学研究の2つのテーマについて、NICTサイエンスクラウドの活用事例を示す。特に、前節までに議論したNICTサイエンスクラウドの各サービスをどのように導入・システム化するかについて重点的に述べる。

### 5.1. 放射線帯高エネルギー粒子予測システム

情報通信研究機構における放射線帯高エネルギー粒子予測システムをNICTサイエンスクラウドにおいて構築した例が図14である。放射線帯高エネルギー粒子予測システムでは、インターネット上から太陽風および静止軌道衛星データをNICTY/DLA（4.3節）により定常的に収集する（表4）。収集されたデータは、データ解析サーバ上で開発された放射線帯粒子予測アルゴリズム<sup>23)</sup>を用いて定常処理サーバ上で予測に用いられる。放射線帯粒子予測結果は、予測の精度確認（Verification）のための事後結果とともにWeb上で公開される。

今後、放射線帯高エネルギー粒子予測を高度化するためには、シミュレーションとの連動・融合が不可避である。本システムでは、NICTサイエンスクラウドが提供する大規模データ処理サーバを用いた放射線帯粒子計算と、NICTスーパーコンピュータ（Hitachi SR16000）によるGlobal MHDシミュレーション<sup>24)</sup>を融合させる計画である。これら3種類（衛星観測データ、粒子計算データ、MHDシミュレーションデータ）はすべてフロントエンドストレージに保存され、データ処理サーバで解析される。これは、図1に示すデータ指向型科学研究スタイルに一致している。

### 5.2. GPS全電子数（TEC）データ処理システム

図15は、NICTサイエンスクラウド上に構築されたGPS全電子数（TEC）データ処理システムの概要である。NICTでは、電離圏変動予測研究の一環として、世界中で公開されているGPSデータを収集し、独自のアルゴリズム<sup>25)</sup>によるGPS-TECデータの定常作成を行っている。GPSデータ自動収集については4.3節において述べたが、図15では収集したGPS

データ（表 5）をフロントエンドストレージに保存し、一部のデータ（日本上空の GEONET データ）については大規模処理サーバにおいて自動で TEC データに変換している。国内のデータについては、準リアルタイム（数時間程度の遅れ）で GPS-TEC 処理を行っている。生成された GPS-TEC データはマップ化され、様々な電離圏研究<sup>26,27,28)</sup>で活用されるほかに、NICT サイエンスクラウドの Web サーバ上で公開される。また、生成した GPS-TEC データは、地球科学・宇宙天気データを球面スクリーン上に立体的に表示する Dagik Earth<sup>29)</sup>でも利用され、展示イベントや出前授業などのアウトリーチに活用されている。NICT サイエンスクラウドの一部である可視化展示室でも閲覧可能である。

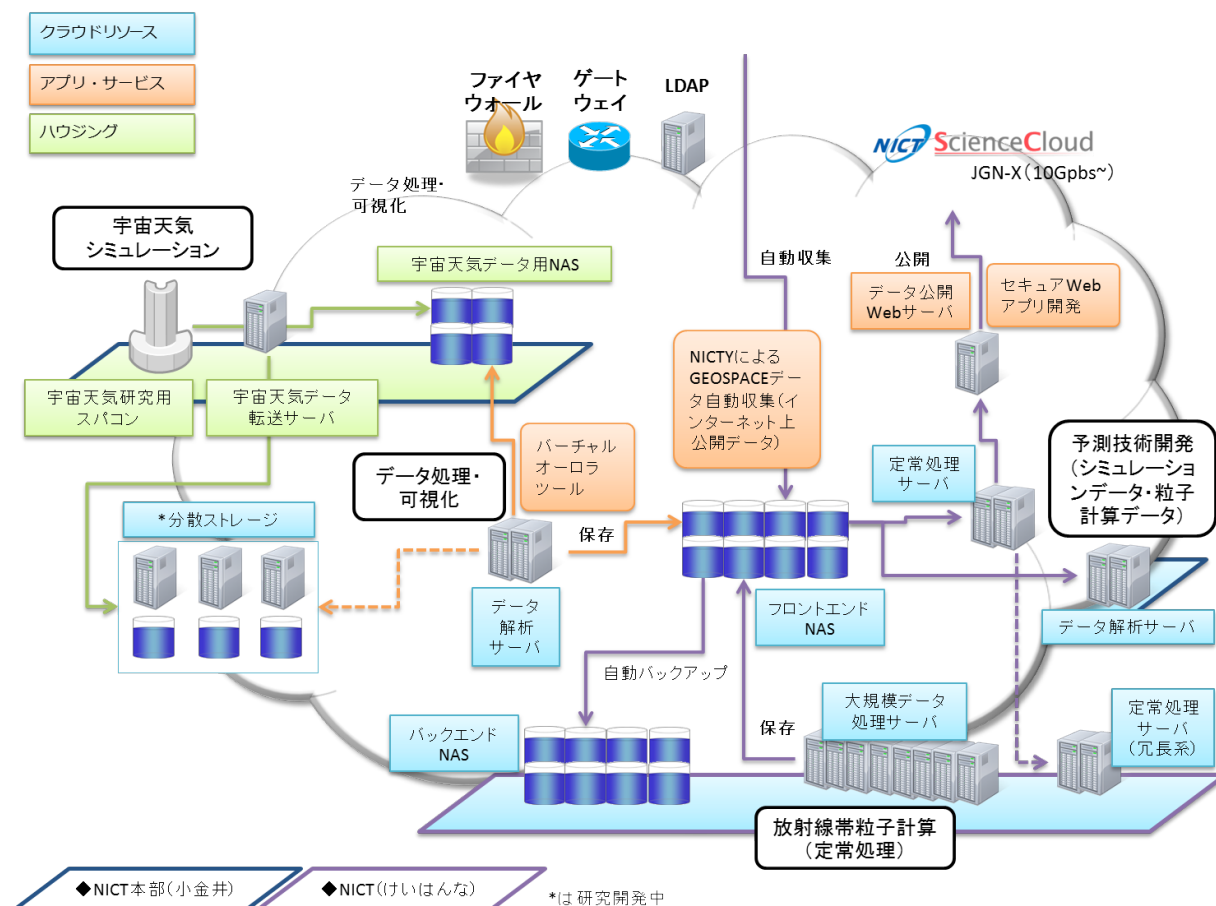
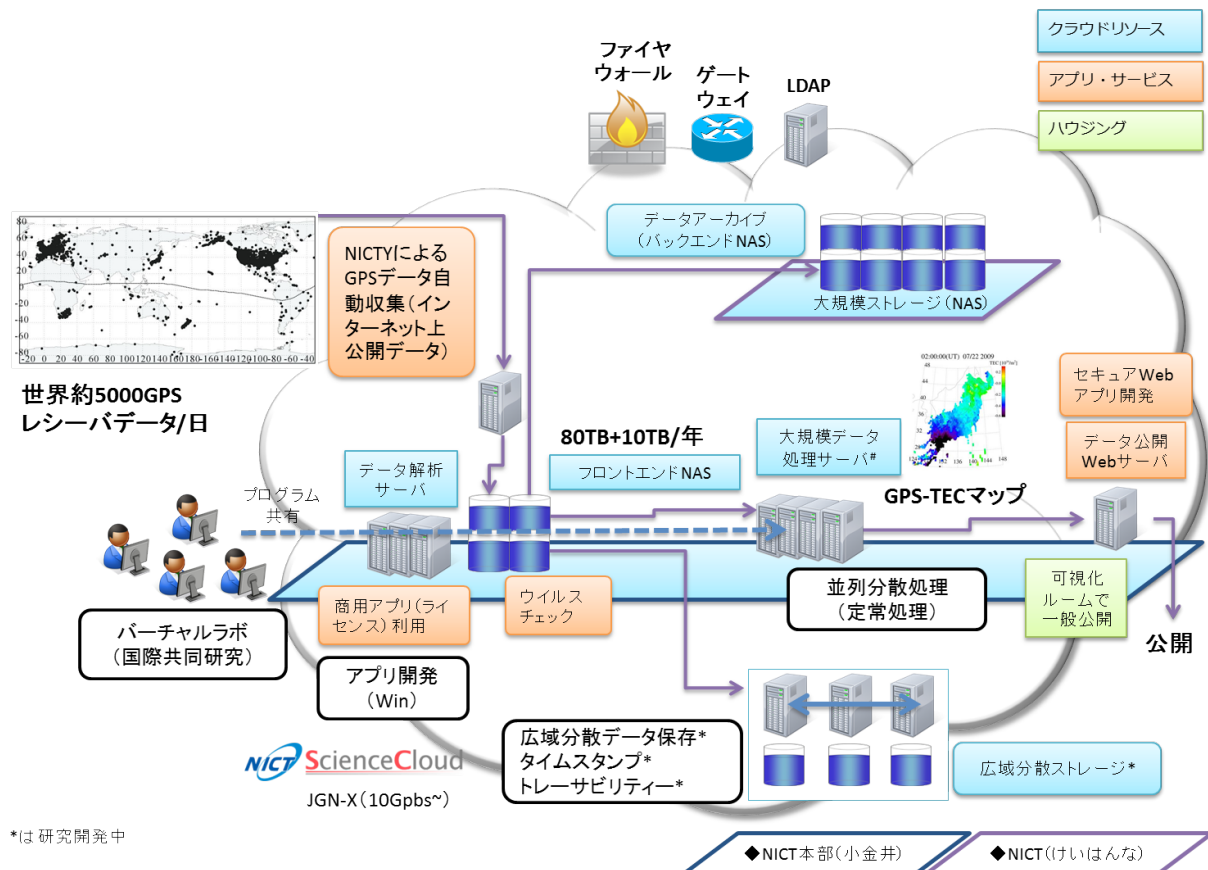


図 14 放射線帯高エネルギー粒子予測システムにおける NICT サイエンスクラウドシステム図  
(計画中や研究開発中の項目を含む)

表 4 NICTY/DLA による放射線帯高エネルギー粒子予測対象データ

データ名 (STARS メタデータベース)	データ説明	収集データ数
PARTICLE_5M/Realtime/GOES	静止軌道粒子データ / 5 分値 (GOES 衛星 / Secondary) 静止軌道粒子データ / 5 分値 (GOES 衛星 / Primary)	2 (ファイル/日)
FLUX/DOM/DRTS	静止軌道粒子データ (こだま衛星)	1 (ファイル/日)
SWEPAM_1M/Realtime/ACE	太陽風速度データ / 1 分値 (ACE 衛星) 太陽風密度データ / 1 分値 (ACE 衛星)	1 (ファイル/日)
SWEPAM_1H/Realtime/ACE	太陽風速度データ / 1 時間値 (ACE 衛星) 太陽風密度データ / 1 時間値 (ACE 衛星)	1 (ファイル/月)
MAG_1M/Realtime/ACE	太陽風磁場データ / 1 分値 (ACE 衛星)	1 (ファイル/日)
MAG_1H/Realtime/ACE	太陽風磁場データ / 1 時間値 (ACE 衛星)	1 (ファイル/月)
EPAM_5M/Realtime/ACE	太陽風粒子データ / 5 分値 (ACE 衛星)	1 (ファイル/日)
AE_Index_QL/GeomagneticIndex	地磁気データ (AE 指数)	5 (ファイル/月)
Dst_Index_QL/GeoMagneticIndex	地磁気データ (DST 指数)	1 (ファイル/月)



### 5.3. 考察

本節では、NICT サイエンスクラウドを活用した研究システム事例を2例紹介した。サイエンスクラウドを活用した科学研究テーマの多くは単一の研究機関では実行が困難であり、2.3節で述べた仮想研究室（バーチャルラボラトリー）による国際的な共同研究体制が必要となる（図2）。図15に、5.2節のGPS全電子数(TEC)データ処理システムでのバーチャルラボラトリーとしてのNICTサイエンスクラウド利用について示した。バーチャルラボラトリーを構築するにあたり、ファイアウォール、ゲートウェイ、ユーザ認証（LDAP）などはNICTサイエンスクラウドが用意するシステムやサーバを利用できることも、サイエンスクラウド上にシステム構築する利点の一つである。

図14や図15では、定常処理結果をWebで公開している。Webアプリケーションによる情報提供は、NICTサイエンスクラウドの運用面では重要である。近年、Webアプリケーションのセキュリティ対策は重要度が増す一方で専門性が高く、一般の科学研究者には対応が容易ではない。NICTサイエンスクラウドでは、セキュアWebアプリケーション開発手順書を作成し、NICTサイエンスクラウドWebで公開している<sup>22)</sup>。本手順書に沿ってWebアプリケーションを開発することで、セキュリティ対策を立てることができる。

2.2節においてHPC/HTC/MTCの概念を紹介した。放射線帯高エネルギー粒子予測システムでは、HPC環境としてのスーパーコンピュータを、独自研究環境利用（ハウジング）としてNICTサイエンスクラウドにネットワーク接続している。図14や図15の大規模データ処理サーバも、小型のHPCシステムである。一方で、図14や図15のシステム全体は、多種多様な計算機リソースを組み合わせることで目的となる速度でデータ処理を行うためMTC環境と言える。NICTサイエンスクラウドのリソースと独自研究環境（ハウジング）を組み合わせることにより、衛星・地上観測データ（Grayによる第1の研究手法）および数値シミュレーションデータ（Grayによる第3の研究手法）<sup>1)</sup>を融合したMTC型の研究環境を実現していることが分かる。

なお、各事例では研究内容の進展と並行してシステム利用の形態は変化する。本事例紹介は、2013年4月時点での事例であり、今後、発展・修正、新規追加、機能削除などが常時おこなわれることになる。

## 6. まとめ

宇宙科学研究分野は様々な科学研究分野の中でも研究規模が大きな分野である。宇宙科学の主要な研究手法である衛星観測および広域地上観測に対して、数値シミュレーションは観測を補完する手法であり、宇宙科学研究の二本柱と言ってもよい<sup>30)</sup>。これらの手法は、近年、予算の大型化に伴うメガサイエンスとなっている。それに伴い、出力されるデータも大規模化されている。宇宙科学では、他の科学研究分野同様に、観測とシミュレーションの融合が研究のブレークスルーとなることが期待されているが、二つのメガサイエンスの融合は20世紀後半に期待されたほどには進んでいない。これは、保存された大量のデータファイルから有益な情報を抽出する環境と手法が不十分であることを示唆している。

このような背景のもと提唱されているのが、第4のパラダイムであるデータ指向型科学研究である<sup>1)</sup>。一般にデータ指向型科学研究の基盤となるのがサイエンスクラウドであると考えられているが、サイエンスクラウドについては、技術的側面および利活用面からの議論が始まったばかりである。言い換えると、どのような方法でサイエンスクラウドがデータ指向型科学研究を実現するのかは明確ではない。本稿では、情報通信研究機構（NICT）で構築した科学研究用クラウドであるNICTサイエンスクラウドについて、「サイエンスクラウドとは何か」「サイエンスクラウドには何ができるか」という視点で議論を行った。

NICTサイエンスクラウドがデータ指向型科学研究に対して貢献できるのは、バーチャルラボラトリー構築を含むクラウドの利便性とビッグデータ処理を含む大規模データ科学研究環境の提供という2本柱の機能提供である。本稿では、特に前者に重点をおいて議論を行った。サイエンスクラウドにより真に大規模な研究プロジェクトを進めるためには、商用クラウドの様に計算機リソースを仮想化して提供するだけでは不十分である。セキュリティ対策やシステム冗長化はもちろんのこと、バーチャルラボラトリーを構築するためのユーザ管理、商用アプリケーションライセンスの提供、研究目的に特化したアプリケーションやシステムの提供、独自研究環境とクラウドリソースの融合などが総合的に機能して、初めてサイエンスクラウドとしての成果が期待できる。セキュリティについても、システムセキュリティの担保だけではなく、研究成果をWeb等で公開するためのセキュリティ対策のガイドラインなども必要である。

ビッグデータ科学の重要性が認識されつつある現在、今後も研究リソースを一か所に集約することでサイエンスクラウドを構築する研究機関が増えるであろう。しかし、単に大規模な計算環境を作っても、サイエンスクラウドとしては



機能しない。本論文は、真に実用的なサイエンスクラウド構築の助けになるだろう。

## 謝辞

本論文の研究は情報通信研究機構の NICT サイエンスクラウドおよび JGN-X を用いて行われました。バーチャルオーロラツールは情報・システム研究機構 新領域融合研究センターの育成融合研究課題として 2006 ～ 2008 年度に開発されたものを基盤としています。

## 参考文献

- 1) Hey, T., S. Tansley, and K. Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Washington, 2009.
- 2) 村田 健史, サイエンスクラウドは第四の研究基盤となるか?, 学術の動向, vol.17, no.6, 2012, pp.42-47.
- 3) 村田 健史, NICT サイエンスクラウドによる新しい科学研究の姿, 情報管理, Vol.55, No.8, 2012, pp.552-561.
- 4) “Science Clouds” Web site, <http://scienceclouds.org/>.
- 5) 2010 年第 1 回 ScienceCloud ワークショップ Web Site, <http://datasys.cs.iit.edu/events/ScienceCloud2010/>.
- 6) 2011 年第 2 回 ScienceCloud ワークショップ Web Site, <http://datasys.cs.iit.edu/events/ScienceCloud2011/>.
- 7) 2012 年第 3 回 ScienceCloud ワークショップ Web Site, <http://ceng.usc.edu/~simmhan/ScienceCloud2012/>.
- 8) ICSU/WDS Web サイト, <http://www.icsu-wds.org/organization/structure/ipo>.
- 9) Murata, K., T. Watari, S., Nagatsuma, T., Kunitake, M., Watanabe, H., Yamamoto, K., Kubota, Y., Kato, H., Tsugawa, T., Ukawa, K., Muranaga, K., Kimura, E., Tatebe, O., Fukazawa, K. and Murayama, Y., A Science Cloud for Data Intensive Sciences, Data Science Journal, Vol. 12, 2013, pp.WDS139-WDS146.
- 10) 尾家祐二, 超高速研究開発ネットワーク JGN II, 情報通信研究機構季報, Vol.51, Nos.3-4, 2005, pp.3-11.
- 11) Helix Nebula Web site, <http://helix-nebula.eu/>.
- 12) 石倉 諭, 村田 健史, 久保 卓也, 木村 映善, 山本 和憲, 篠原 育, RSS1.0 を利用した科学衛星・地上観測データのメタデータ自動収集, 電子情報通信学会論文誌 (B), Vol.J91-B (4), 2008, pp.499-509.
- 13) Watanabe, H., Yamamoto, K., Tsugawa, T., Nagatsuma, T., Watari, S., Murayama, Y. and Murata, K., T., An Integrated Management System of Multipoint Space Weather Observation, Data Science Journal, Vol.12, 2013, WDS2011-CS607.
- 14) 松岡大祐, 村田健史, 藤田茂, 田中高史, 山本和憲, 木村映善, Global MHD シミュレーションによる磁気フラックスロープの 3 次元構造解析, 可視化情報学会論文集, Vol.28, No.6, 2008, pp.38-46.
- 15) 松岡大祐, 村田健史, 藤田茂, 田中高史, 山本和憲, 大野暢亮, 3 次元可視化とビジュアルデータマイニング, 情報通信研究機構季報, Vol.55, Nos.1-4, 2009, pp.459-469.
- 16) Murata, K., T., Yamamoto, K., Matsuoka, D., Matsumoto, H., Okada, M., Mukai, T., Sigwarth, J., B., Fujita, S., Tanaka, T., Yumoto, K., Ogino, T., Shiokawa, K., Tsyganenko, N., A., Green, J., L. and Nagai, T., Development of the Virtual Earth's Magnetosphere System (VEMS), Advances in Polar Upper Atmosphere Research, No.19, pp.135-151, 2005.
- 17) Saita, S., Kadokura, A., Sato, N., Fujita, S., Tanaka, T., Ebihara, Y., Ohtani, S., Ueno, G., Murata, T., K., Matsuoka, D., Kitamoto, A., Higuchi, T., Displacement of conjugate points during a substorm in a global magnetohydrodynamic simulation, Journal of Geophysical Research-Space Physics, Vol.116 (A6), 2011, A06213.
- 18) Kubota, Y., Yamamoto, K., Fukazawa, K., and Murata, K., T., Visualization of the Flux Rope Generation Process Using Large Quantities of MHD Simulation Data, Data Science Journal, Vol.12, 2013, pp.WDS134-WDS138.
- 19) Fujita, S., Yamagishi, H., Murata, K., T., Den, M. and Tanaka, T., A numerical simulation of a negative solar wind impulse: Revisited, Journal of Geophysical Research, vol.117, 2012, A09219.
- 20) Murata, K., T., Yamamoto, K., Matsuoka, D., Matsumoto, H., Okada, M., Mukai, T., Sigwarth, J., B., Fujita, S., Tanaka, T., Yumoto, K., Ogino, T., Shiokawa, K., Tsyganenko, N., A., Green, J., L. and Nagai, T., Development of the Virtual Earth's Magnetosphere System (VEMS), Advances in Polar Upper Atmosphere Research, No.19, 2005, pp.135-151.

- 21) Murata, K., T., Yahara, H. and Toyota, K., Software design via object-oriented methodology and network database for solar-terrestrial observation data, Database system, 123-5, database, 2001, pp.31-36.
- 22) NICT サイエンスクラウド Web site, <http://sc-web.nict.go.jp>.
- 23) Sakaguchi, K., Miyoshi, Y., Saito, S., Nagatsuma, T., Seki, K. and Murata, K., T., Relativistic electron flux forecast at geostationary orbit using Kalman filter based on multivariate autoregressive model, Space Weather, Vol.11 (2), 2013, pp.79-89.
- 24) Yasubumi Kubota, Ryuho Kataoka, Mitsue Den, Takashi Tanaka, and Tsutomu Nagatsuma, GlobalMHDsimulation of the magnetospheric response to large and sudden enhancement of the solar wind dynamic pressure, PEM05-07, JpGU2013.
- 25) Tsugawa, T., Observational Studies on Large-Scale Traveling Ionospheric Disturbances Using GPS Receiver Networks, Ph. D Thesis, Kyoto University, Kyoto, Japan, 2004.
- 26) Nishioka, M., Otsuka, Y., Shiokawa, K., Tsugawa, T., Effendy, Supnithi, P., Nagatsuma, T. and Murata, K., T., On Post-Midnight Field-Aligned Irregularities Observed with a 30.8-MHz Radar at a Low Latitude: Comparison with F-Layer Altitude near the Geomagnetic Equator, Journal of Geophysical Research, Vol.117, 2012, A08337.
- 27) Tsugawa, T., Saito, A., Otsuka, A., Nishioka, M., Maruyama, T., Kato, H., Nagatsuma, T. and Murata, K., T., Ionospheric disturbances detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake, Earth Planets Space, Vol.63 (7), 2011, pp.875-879.
- 28) Otsuka, Y., Suzuki, K., Nakagawa, S., Nishioka, M., Shiokawa, K. and Tsugawa, T., GPS Observations of Medium-Scale Traveling Ionospheric Disturbances over Europe, Annales Geophysicae, 31, 163-172, doi:10.5194/angeo-31-163-2013, 2013.
- 29) Dagik Web サイト, <http://www.dagik.net>.
- 30) 松本 紘, 宇宙開拓とコンピュータ, 情報フロンティアシリーズ, 情報処理学会編, 共立出版, 1996.