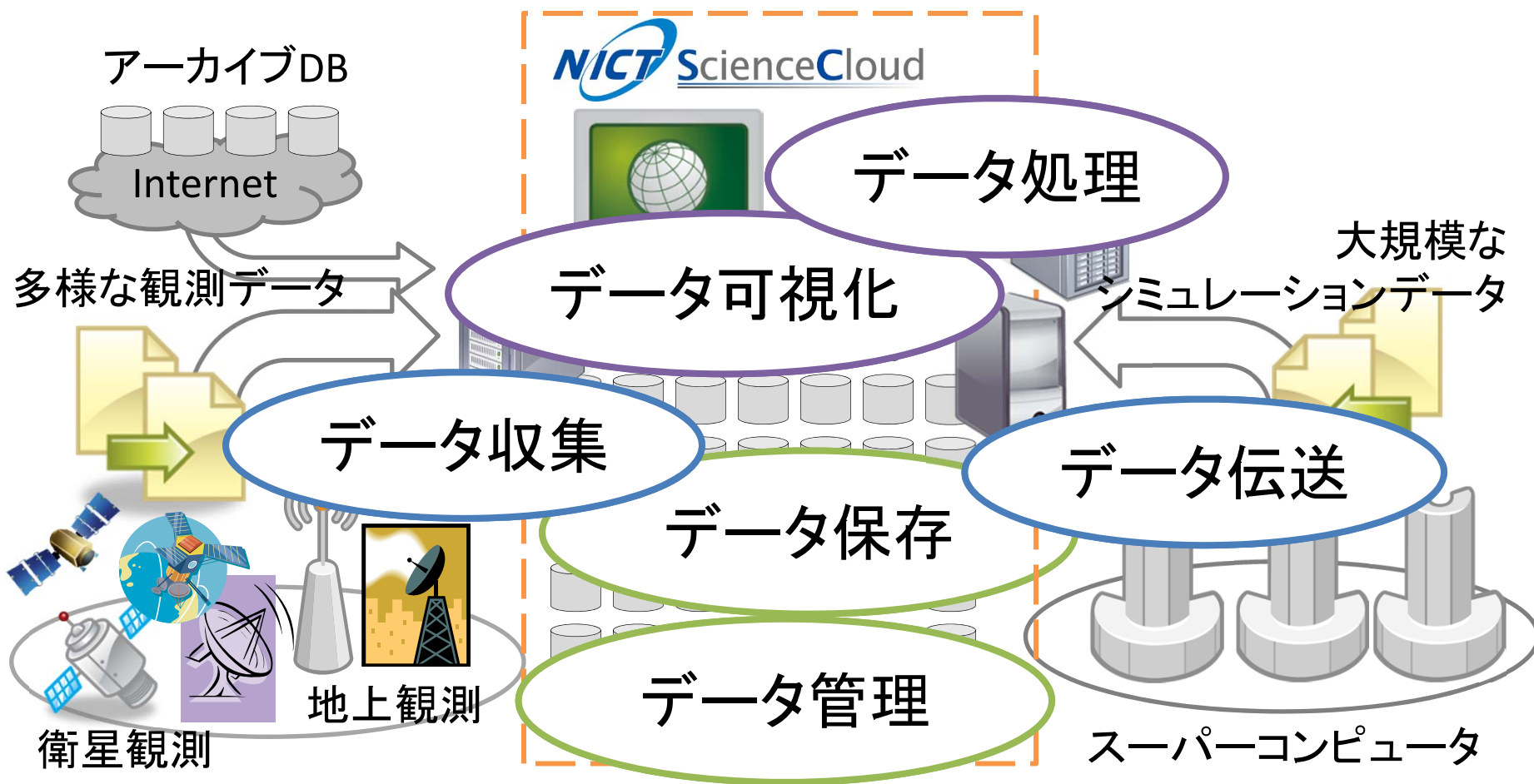


NICTサイエンスクラウドの ビッグデータ処理技術開発と運用 H.25年度報告

村田 健史・渡邊 英伸・長屋 嘉明
情報通信研究機構

20150213 宇宙科学情報解析シンポジウム

NICTサイエンスクラウド システムコンセプト



NICTサイエンスクラウドの基盤技術開発

基盤技術(i) データ収集・伝送

グローバルデータ収集

- 世界中の観測拠点を監視する広域観測網監視システム
- インターネット上の科学データを自動収集・処理
- 異分野データ・ソーシャルデータを融合表示

遠隔高速ストレージ

- 遠隔地からのクラウドストレージの高速I/O技術

基盤技術(ii) データ保存・管理

クラウドストレージ

トレーサビリティ

- 広域分散ストレージのトレーサビリティシステム・タイムスタンプシステム開発と試験運用

セキュアWeb開発手順

- 研究者がWebアプリケーションをセキュアに開発する手順の確立
- H.25年度に6例のWebアプリで有効性を実証・運用

基盤技術(iii) データ処理・可視化

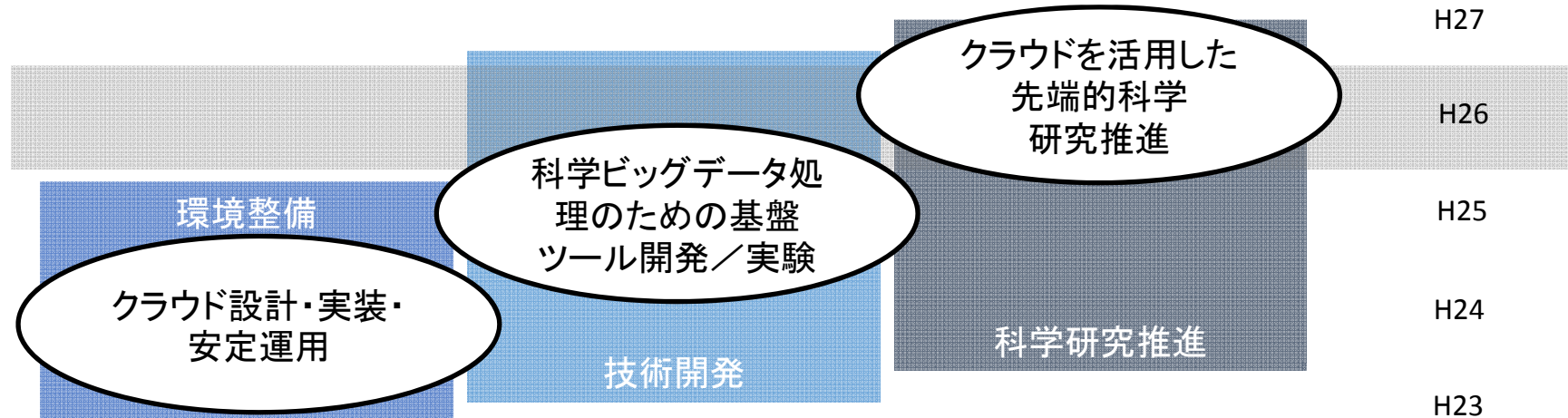
ビッグデータ並列処理

- 分散ストレージと分散処理の連携によるビッグデータ高速処理技術開発
- ビッグデータ処理システム開発開始・3次元レーダデータリアルタイム処理

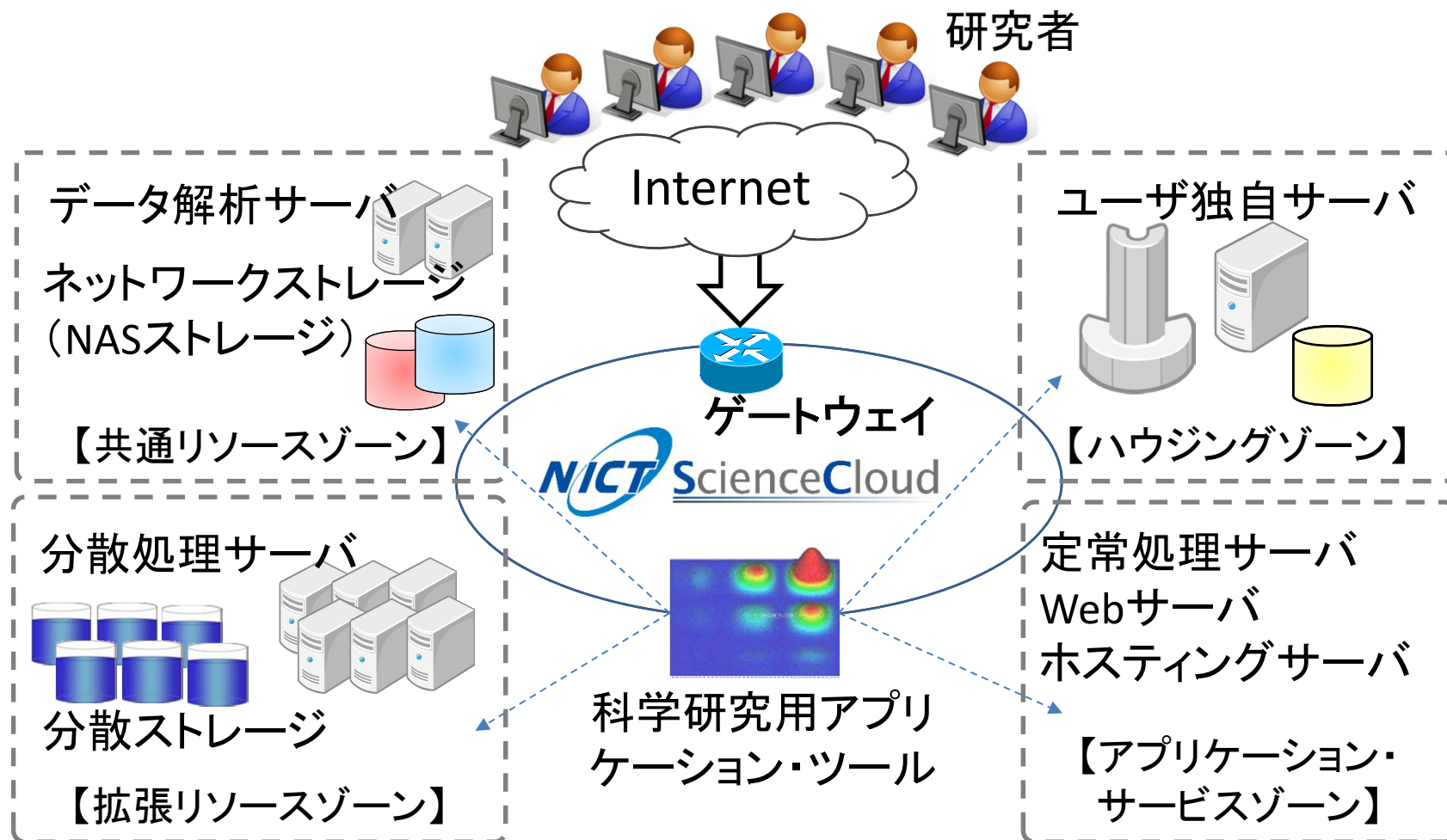
ビッグデータ可視化

- 宇宙天気・気象レーダの3次元可視化システム

個々の基盤技術をマッシュアップして初めて一つのシステム(アプリ)として機能する



NICTサイエンスクラウド リソース・サービス



申請プロジェクト一覧(H.25年度)

NICTプロジェクト比率 18/30

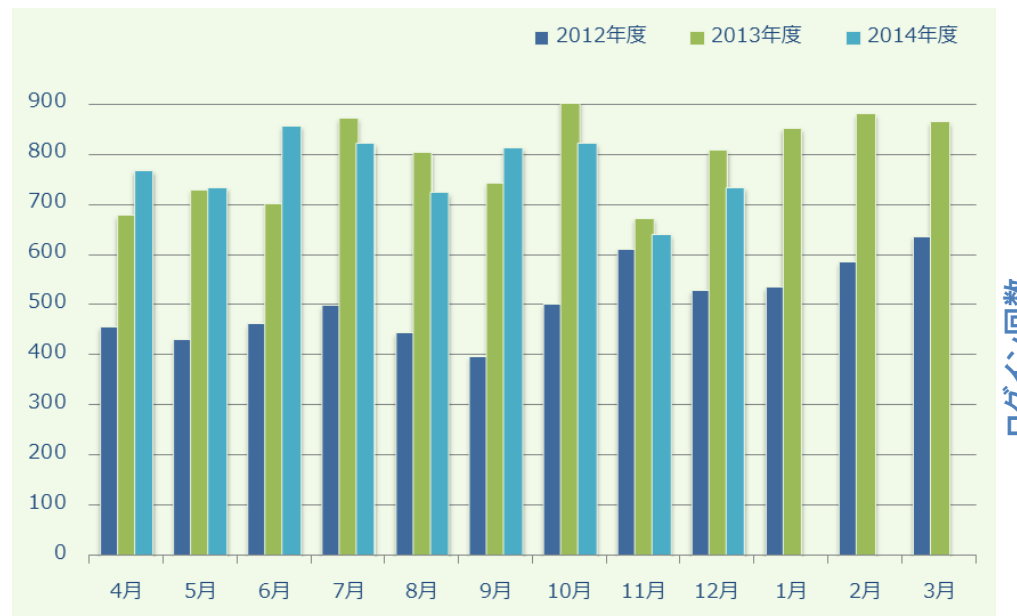
| PJ 番号 | プロジェクト名称 | 研究代表者 | 代表者所属 | 分担者数 (NCT外) |
|-----------|---|-----------|----------------------|----------------------|
| 1 | GNSS全電子数計測を利用した電離圏の研究 | 津川 卓也 | NICT宇宙環境インフォマティクス研究室 | 20(19) |
| 2 | 東南アジア域低緯度電離圏観測(SEALION) | 津川 卓也 | NICT宇宙環境インフォマティクス研究室 | 4(0) |
| 3 | 短波到来方向探査装置を利用した電離圏の研究 | 津川 卓也 | NICT宇宙環境インフォマティクス研究室 | 3(1) |
| 4 | 宇宙天気シミュレーション | 品川 裕之 | NICT宇宙環境インフォマティクス研究室 | 16(10) |
| 5 | 次世代宇宙天気情報処理の研究 | 亙 慎一 | NICT宇宙環境インフォマティクス研究室 | 2(0) |
| 6 | 太陽圏モデリング | 亙 慎一 | NICT宇宙環境インフォマティクス研究室 | 14(10) |
| 7 | 大気圏・電離圏長期シミュレーションデータの解析 | 陣 英克 | NICT宇宙環境インフォマティクス研究室 | 3(2) |
| 8 | 太陽風-磁気圏電離圏システム相互作用 | 久保田 康文 | NICT宇宙環境インフォマティクス研究室 | 2(2) |
| 9 | ジオスペース・放射線帯予測 | 長妻 努 | NICT宇宙環境インフォマティクス研究室 | 5(1) |
| 10 | 南極観測 | 長妻 努 | NICT宇宙環境インフォマティクス研究室 | 11(5) |
| 11 | Integrated Satellite Observaion SIMulator for a Coherent Doppler Lidar (ISOSIM-L)による衛星搭載ドップラーライダーのフィジビリティスタディ | 石井 昌憲 | NICTセンシング基盤研究室 | 5(4) |
| 12 | SMILES/GOSAT (プロジェクト申請未提出) | 笠井 康子 | NICTセンシング基盤研究室 | 45(37) |
| 13 | フェーズドアレイ気象レーダのデータ利用システム(気象レーダの3次元視覚化) | 佐藤 晋介 | NICTセンシングシステム研究室 | 9(7) |
| 14 | 生体電磁環境プロジェクト | 渡辺 聡一 | NICT電磁環境研究室 | 4(0) |
| 15 | 太陽圏の巨視的構造とダイナミックスの研究 | 鷺見 治一 | アラバマ大学 | 3(0) |
| 16 | 科学衛星搭載プラズマ波動観測器で得られた波形データの特徴解析 | 笠原 禎也 | 金沢大学総合メディア基盤センター | 2(2) |
| 17 | 惑星間空間磁場北向き時の磁気圏電離圏対流機構の解明 | 渡辺 正和 | 九州大学理学研究院地球惑星科学部門 | 5(3) |
| 18 | GNSS可降水量データベース | 藤田 実季子 | 独立行政法人海洋研究開発機構 | 1(1) |
| 19 | 静止軌道衛星帯電プラズマ環境の解析および予測の研究 | 中村 雅夫 | 大阪府立大学 | 1(1) |
| 20 | 地球磁気圏の形状と自由エネルギーに関する初期研究:大規模3次元電磁流体計算と観測を比較する方法の確立 | 齋藤 実穂 | 名古屋大学 | 1(1) |
| 21 | 気象分野におけるビッグデータ利活用技術の研究 | 大野 智生 | 気象庁観測部気象衛星課 | 2(2) |
| 22 | バーチャルオーロラツールを活用したデジタル磁気嵐現象の研究 | 海老原 祐輔 | 京大大学生存圏研究所 | 2(1) |
| 23 | SS-MIX標準ストレージのNoSQL実装と並列分散処理の検証 | 木村 映善 | 愛媛大学医学部 | 1(0) |
| 24 | NICTサイエンスクラウドを用いたゲノムデータ管理基盤に関する研究開発 | 原田 憲治 | (株)カイ研究開発部 | 3(0) |
| 25 | 社会インフラのメンテナンスに資するシミュレーションとセンシングデータの解析 | 中畑 和之 | 愛媛大学大学院理工学研究科 | 1(1) |
| 26 | Global MHDシミュレーションの大規模可視化によるプラズマダイナミクス | 深沢 圭一郎 | 九州大学 | 3(3) |
| 27 | SALMON (プロジェクト申請未提出) | 村山 泰啓 | NICT統合データシステム研究開発室 | 未 |
| 28 | 時系列データ表示アプリケーション(STARS touch)の開発※ | 村田 健史 | NICT統合データシステム研究開発室 | 5(5) |
| 29 | NICTサイエンスクラウドセキュリティ技術開発※ | 渡邊 英伸 | NICT統合データシステム研究開発室 | 5(5) |
| 30 | NICTサイエンスクラウド高速データ転送表示技術開発※ | 渡邊 英伸 | NICT統合データシステム研究開発室 | 6(6) |
| 合計 | | 30 | | 184 (129) |

※は技術開発プロジェクト。NICT外部利用者が多いプロジェクトはバーチャルラボとしての利用であると予想される。

クラウド利活用状況

サイエンスクラウド利用者数 (H.24年度~H.26年度)

ログイン回数は、サイエンスクラウドユーザが1ヶ月の間にゲートウェイサーバへログインした日数を示す。(1日に複数回ログインした場合は1回としてカウントする。)



データ収集保存実績

NICTサイエンスクラウドにおける科学データの収集・保存実績

| データ種別 | データ提供組織名 | 収集ツール | データファイル数 | 総データサイズ |
|------------------------|-----------------|--------------------|------------|---------|
| GNSS観測 | UNAVCO, 国土地理院など | NICTY+独自ツール | 23,506,753 | 9.6TB |
| SEALION | NICT | WONM | 60,787 | 1.6TB |
| GOSAT衛星 | JAXA, ECMWF | 独自ツール | 3,737,123 | 58.4TB |
| フェーズドアレイ 気象レーダー | 大阪大学(NICT) | 独自ツール | 2,358,677 | 217.1TB |
| SMILES | JAXA | 独自ツール | 564,253 | 1.4TB |
| 南極(電離層観測) | NICT | WONM | 280,859 | 424GB |
| 地磁気データ | NICT, 京都大学 | NICTY+独自ツール | 277 | 1.5TB |
| ひまわり衛星 | 気象庁 | NICTY+(HDD搬送) | 26,868 | 26.8GB |
| GEOTAIL衛星 | 京都大学 生存圏研究所 | NICTY | 29,840 | 204GB |
| GOES衛星 | NOAA | NICTY | 1,638 | 59MB |
| ACE衛星 | NOAA | NICTY | 759 | 203.5MB |
| こだま衛星 | JAXA | NICTY | 3,890 | 1.8GB |
| KAGUYA衛星 | 金沢大学 | -(伝送なし) | 3,227 | 221GB |
| 宇宙天気リアルタイム シミュレーション | NICT | NICTY (メタデータのみ) | 63 | 3.5TB |

成果リスト(学術論文・その他)

| | 学術論文(査読付き) | その他論文 |
|--------|------------|-------|
| H.24年度 | 25件 | 3件 |
| H.25年度 | 56件 | 7件 |

サイエンスクラウド運用の現状:プロジェクト一覧(2012-)

| #PJ | Category | Zone | Tool | Web |
|---------|----------|------|------|-----|
| #1(21) | O | CEAH | N | ● |
| #2(5) | O | C-A- | W | ● |
| #3(4) | O | C-A- | W | ● |
| #4(17) | S | C-AH | V | ● |
| #5(3) | S | C-A- | P | ● |
| #6(15) | O | C--- | - | |
| #7(4) | S | C-A- | - | ● |
| #8(3) | S | C-A- | V | |
| #9(6) | OS | CEA- | N | ● |
| #10(12) | O | ---H | W | |
| #11(6) | S | CEAH | - | |
| #12(46) | O | C--- | T | |
| #13(10) | O | C--- | WT | |
| #14(5) | H | CE-- | - | |
| #15(4) | S | C--- | - | |

| PJ | Category | Zone | Tool | Web |
|--------|----------|------|------|-----|
| #16(3) | O | C--- | T | |
| #17(6) | S | C--- | V | |
| #18(2) | O | C--- | - | |
| #19(2) | S | C--- | - | |
| #20(2) | S | C--- | - | |
| #21(3) | O | C--- | - | |
| #22(3) | S | C--- | - | |
| #23(2) | H | CE-- | - | |
| #24(4) | H | C--- | - | |
| #25(2) | S | C--- | - | |
| #26(4) | S | CE-- | VP | |
| #27(1) | O | -E-H | - | |
| #28(6) | I | CEA- | NWP | ● |
| #29(6) | I | C--- | - | ● |
| #30(7) | I | CE-- | P | |

(191)

Category: Observation, Simulation, Human, Informatics
太字は代表者がNICT所属

Space Science (18)

Earth Science (5)

Human Science (3)

Informatics (3)

NICTサイエンスクラウド 科学研究利用形態

サイエンスクラウド上で独自プラットフォームと独自アプリを構築

サイエンスクラウドのプラットフォーム上に独自アプリを構築

サイエンスクラウドが提供するアプリケーションを活用



独自アプリ

独自アプリ

SaaS

独自プラットフォーム

PaaS

PaaS

IaaS

IaaS

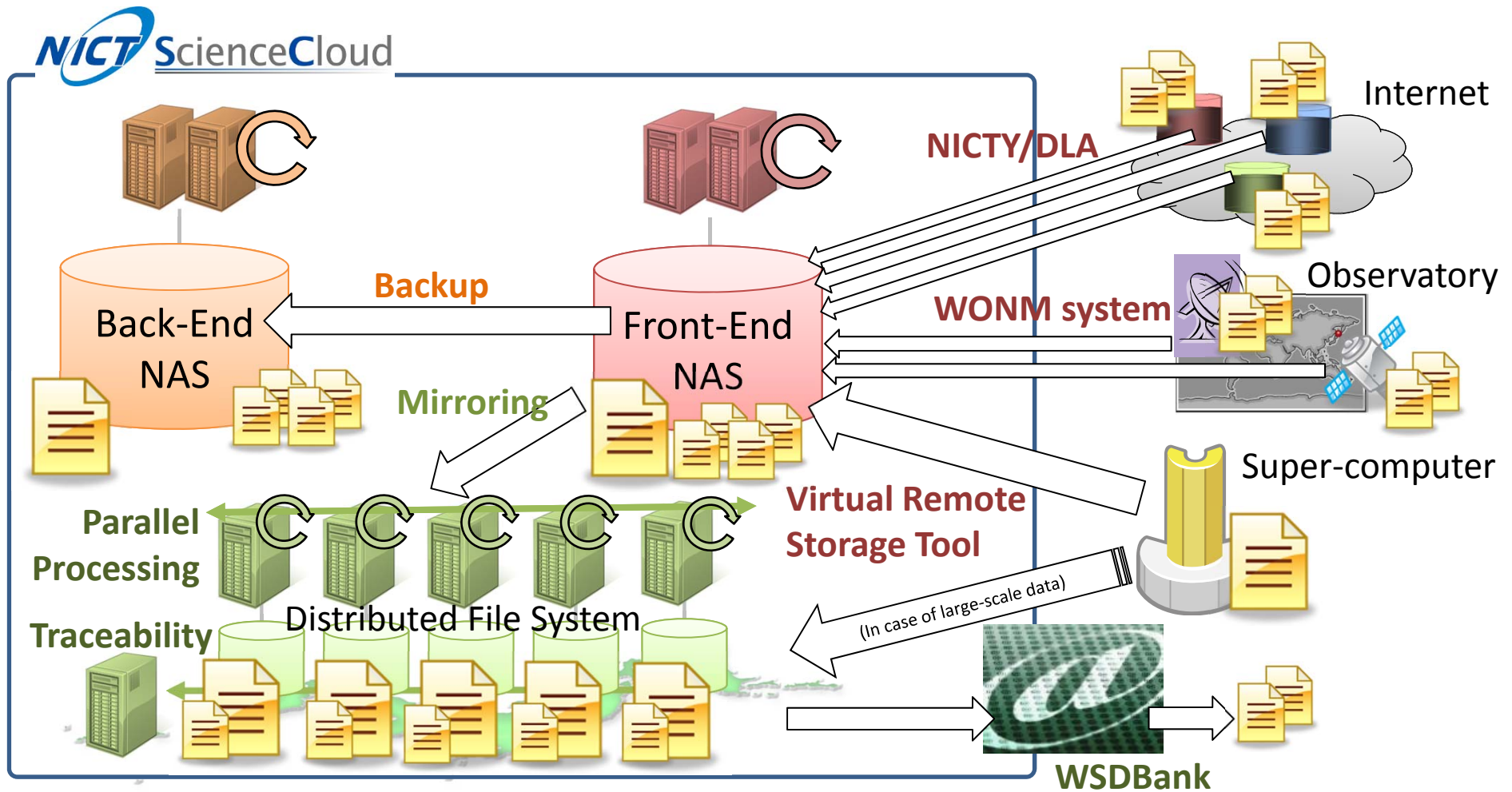
IaaS

NICTサイエンスクラウドオリジナルアプリ・ツール

| 目的 | アプリ名・ツール名 | 概要 |
|--------|---|---|
| データ収集 | NICTY/DLA | インターネットで公開されている科学データをクロールするツール。メタ情報自動収集ツール(NICTY)とデータファイルダウンロードエージェント(DLA)から構成。 |
| データ収集 | WONM(Wide-area Network Monitoring)システム | 広域観測網の観測所・観測拠点の観測システムを監視し、データ転送を自動的に行うツール。サーバツールとクライアントツールから構成されるが、クライアントツールをあらかじめセットアップした小型アプライアンスサーバを利用できる。 |
| データ伝送 | 遠隔高速ストレージシステム(High-bandwidth Virtual Remote Storage System) | 分散ファイルシステム(Gfarm)を仮想ストレージとして、遠隔地から高速データファイルの読み込み・書き出しを行うツール。クライアントサーバにセットアップすることで、APIとして利用できる。 |
| データ管理 | WSDBank(World Science Data Bank) | サイエンスクラウドのストレージ(NAS, 分散ファイルシステム)上のデータファイルにアクセスするためのWebアプリケーション。 |
| データ管理 | Gfarmトレーサビリティ | 分散ファイルシステム(Gfarm)上のデータファイルの履歴をファイル単位(インスタンス単位)で追跡するツール。管理者用。 |
| データ処理 | Pwrake | 複数の計算ノードでデータファイルを分散処理するための並列処理用タスクスケジューラ。NFSでもGfarmでも利用できるが、Gfarmと組み合わせローカルファイルに優先的にアクセスすることでI/O高速化を実現するアフィニティスケジューリングが可能となる。 |
| データ処理 | Torque/Maui | クラスタ計算環境で並列処理に適したタスクスケジューラ。リソース・マネージャ(Torque)とスケジューラ(Maui)から構成される。 |
| データ可視化 | バーチャルオーロラツール | Global MHDシミュレーションデータを可視化するツール。AVS Express/Devにより実装。 |
| データ可視化 | STICKER | フェーズドアレイ気象レーダデータとTwitter等のソーシャルデータを融合表示するWebアプリケーション。NICT情報利活用研で開発。 |
| データ可視化 | VDVGE | JAMSTECによるGoogle Earth用ボリュウムデータ表示ツール。 |
| データ可視化 | STARStouch | 異分野字形例つデータ融合表示ツール(Webアプリ)。GEOTAIL衛星版(公開済み)、ひまわり衛星データ版(近日公開)、読売新聞版(SOMATO)、フェーズドアレイレーダ版などを開発。 |

科学データ収集・転送・管理・保存・処理

Science Data File Crawling/Transfer, Preservation/Management and Processing

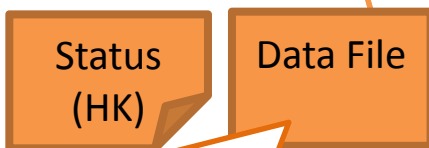
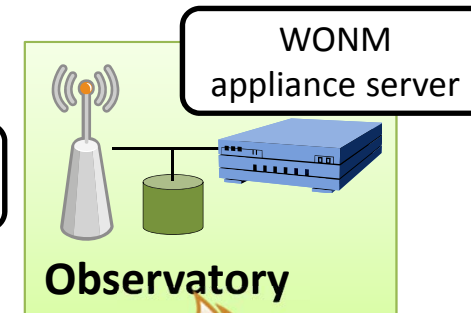
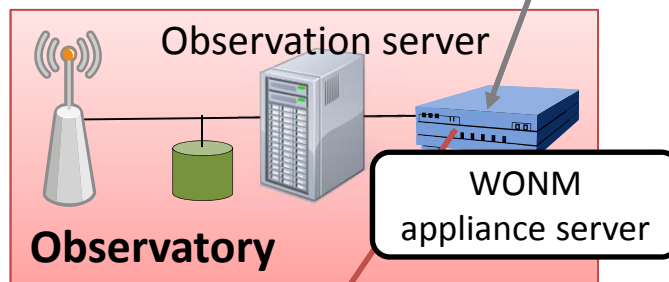
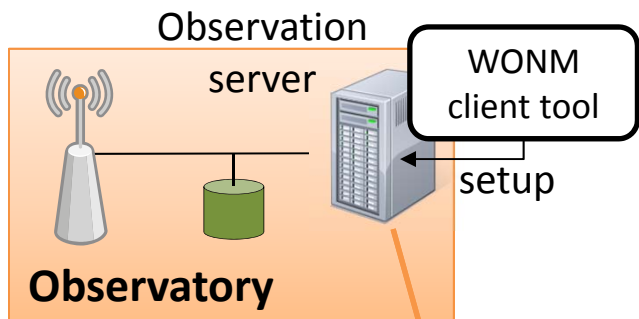


広域観測網監視システム

Wide-area Observation Monitoring System

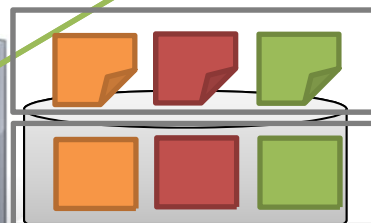
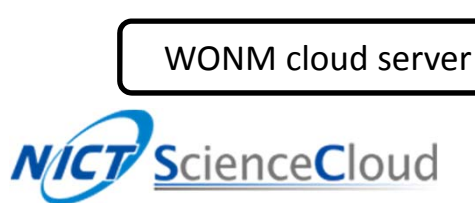


- ① Monitoring Server
- High tolerance
 - Automatic recovery
 - Redundancy



- ② Data Transfer
- Data Transfer Retry
 - High performance network band
 - Data consistency

- ③ Easy & Integrated Management
- Warning System
 - Monitoring System

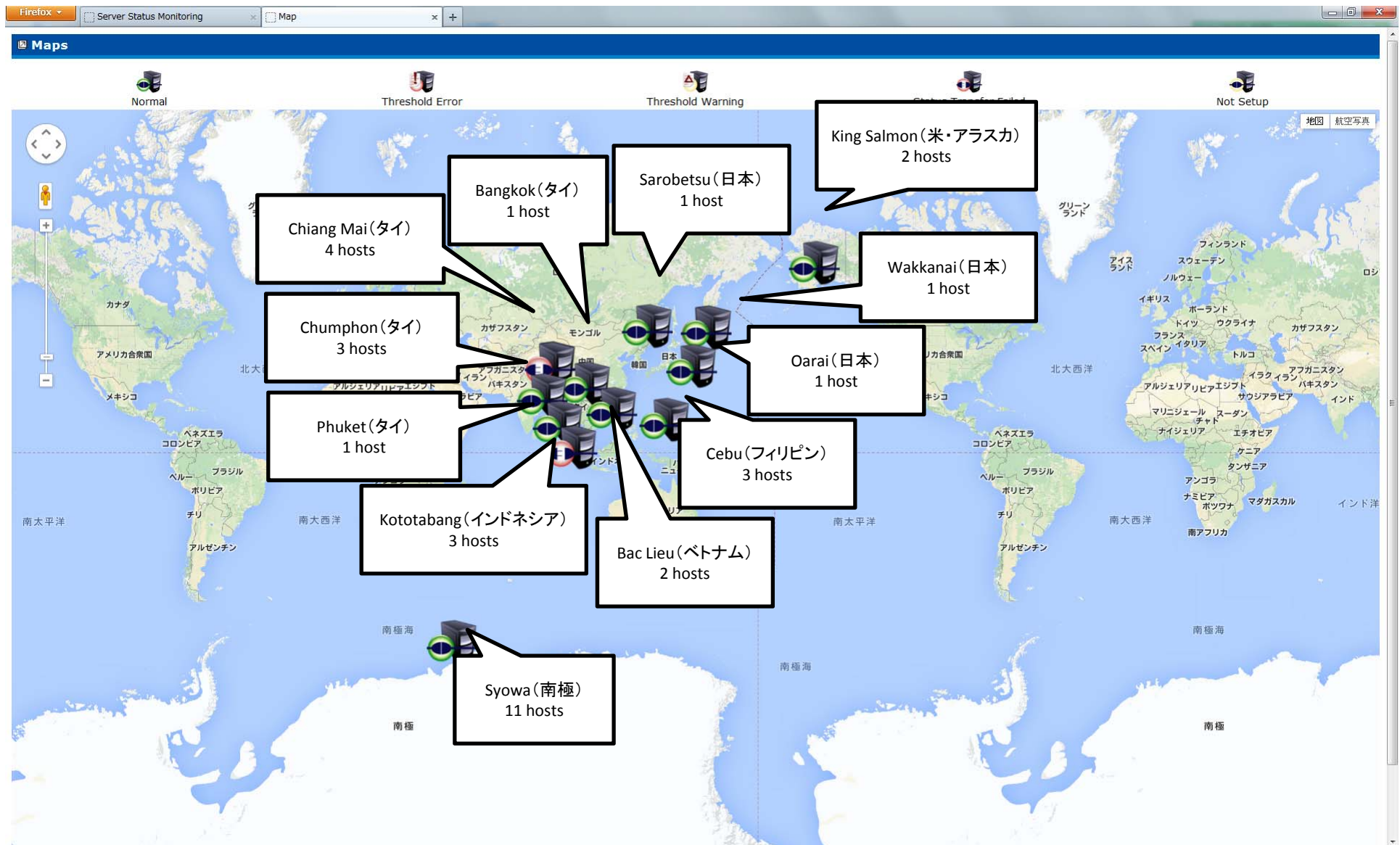


WONM Web

Data Analysis



WONMシステムによる観測所管理(宇宙環境のみ: ~H.26年度)



データ収集実績(～2014年2月)

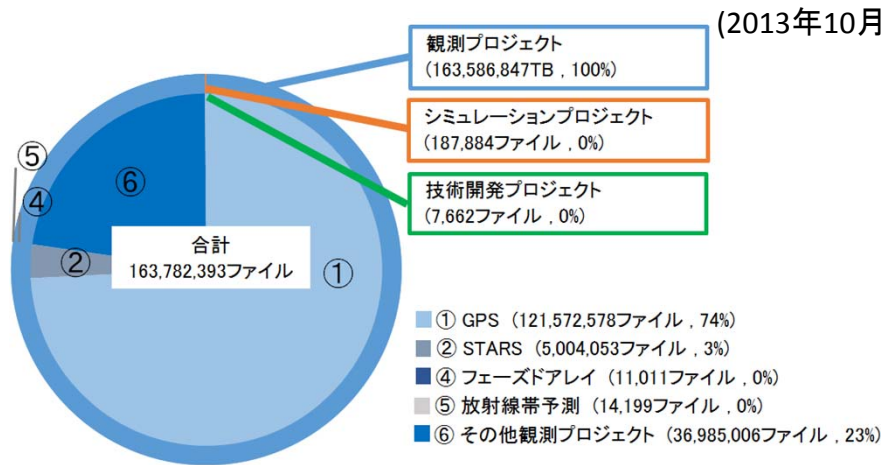
| Data | PJ# | Organization | Tool | Number of collected files | Total data size |
|---|-----|-------------------|-------|---------------------------|-----------------|
| GNSS observation (宇) | 1 | UNAVCO, GSI | NICTY | 23,506,753 | 9.6TB |
| SEALION (セ) | 2 | NICT | WONM | 60,787 | 1.6TB |
| GOSAT satellite (セ) | 12 | JAXA, ECMWF | * | 3,737,123 | 58.4TB |
| Phased-array meteorological data (セ) | 13 | Osaka Univ., NICT | * | 2,358,677 | 217.1TB |
| SMILES (セ) | 12 | JAXA | * | 564,253 | 1.4TB |
| Antarctica (ionosphere observation) (宇) | 10 | NICT | WONM | 280,859 | 424GB |
| Geo-magnetic data (宇) | 9 | NICT, Kyoto Univ. | NICTY | 277 | 1.5TB |
| GMS satellite | 21 | JMA | NICTY | 26,868 | 26.8GB |
| GEOTAIL satellite | 28 | Kyoto Univ. | NICTY | 29,840 | 204GB |
| GOES satellite (宇) | 9 | NOAA | NICTY | 1,638 | 59MB |
| ACE satellite (宇) | 9 | NOAA | NICTY | 759 | 203MB |
| KODAMA satellite (宇) | 9 | JAXA | NICTY | 3,890 | 1.8GB |
| KAGUYA satellite | 16 | Kanazawa Univ. | * | 3,227 | 221GB |
| Space Weather real-time simulation (宇) | 4 | NICT | NICTY | 63 | 3.5TB |

WONMシステム利用状況(2013年10月時点)

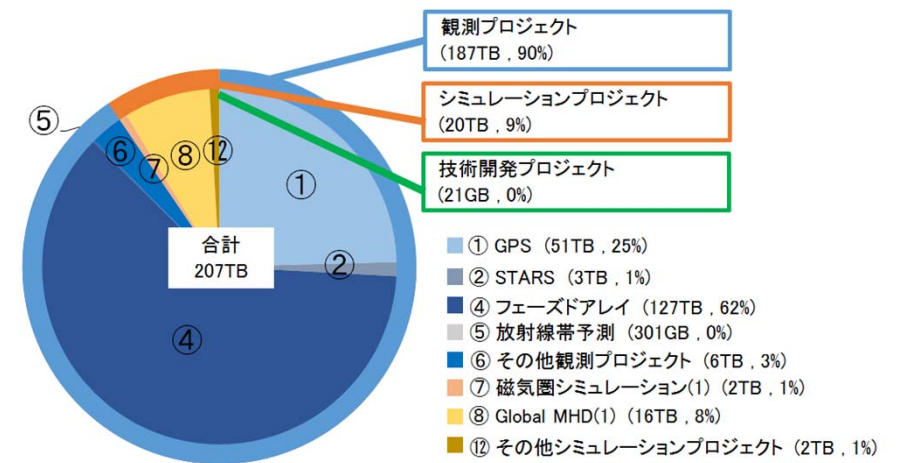
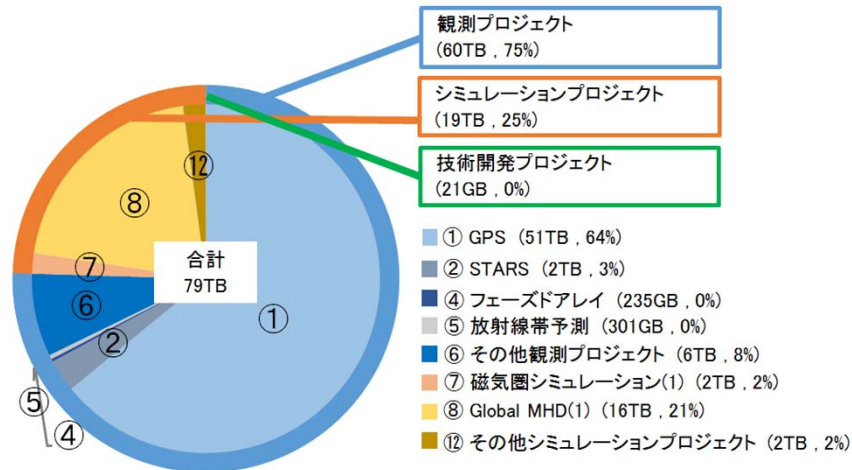
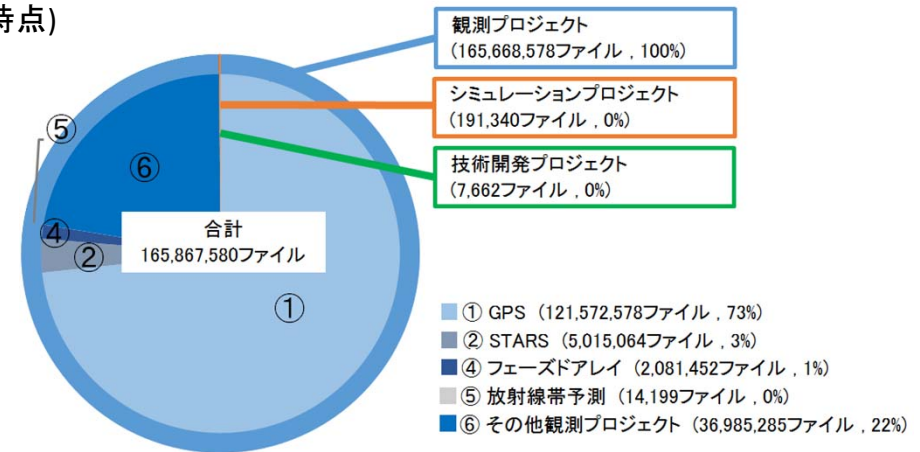
| プロジェクト | プロジェクト担当者 | 拠点 | サーバ数 | 備考 | |
|--------------------------|--|--|-------|------|--|
| Observation Network | 宇宙環境インフォマティクス研究室 山本和憲(kaz-y@ml.nict.go.jp) | 小金井 | 1 | テスト用 | |
| サイエンスクラウドサーバ 管理 | NICTサイエンスクラウド事務局 (osn-system@ml.nict.go.jp) | 小金井 | 96 | | |
| | | 沖縄 | 3 | | |
| | | 大阪 | 2 | | |
| | | 名古屋 | 2 | | |
| | | けいはんな | 15 | | |
| Observation Network Test | | 小金井 | 3 | テスト用 | |
| HIRAS | 宇宙環境インフォマティクス研究室 山本和憲(kaz-y@ml.nict.go.jp) | — | 0 | | |
| ISD-J | | 山川 | 1 | | |
| HF-TEP | | — | 0 | | |
| SEALION | | 宇宙環境インフォマティクス研究室 津川卓也(tsugawa@nict.go.jp) | チェンマイ | 3 | |
| | | | バンコク | 1 | |
| | チュンポン | | 3 | | |
| | プーケット | | 3 | | |
| | コトタバン | | 1 | | |
| | バクリウ | | 3 | | |
| | セブ | | 3 | | |
| HF Radar | 宇宙環境インフォマティクス研究室 長妻努(tnagatsu@nict.go.jp) | キングサーモン | 1 | | |
| Syowa Station | 宇宙環境インフォマティクス研究室 長妻努(tnagatsu@nict.go.jp) | 昭和基地 | 2 | | |
| | | 稚内 | 1 | | |
| | | サロベツ | 1 | | |
| Ishii Lab. | 仙台高専 石井誠四郎(ishii@sendai-nct.ac.jp) | 仙台 | 1 | | |
| Phased Array | センシングシステム研究室 佐藤晋介 | 小金井 | 1 | | |
| 合計 | | | 18 | 147 | |

NASストレージ(フロントエンド・バックエンド)

フロントエンドNASのプロジェクトごとの ファイル数(上)とディスク使用量(下)



バックエンドNASのプロジェクトごとの ファイル数(上)とディスク使用量(下)



NICT開発技術 信頼性機能(タイムスタンプ)

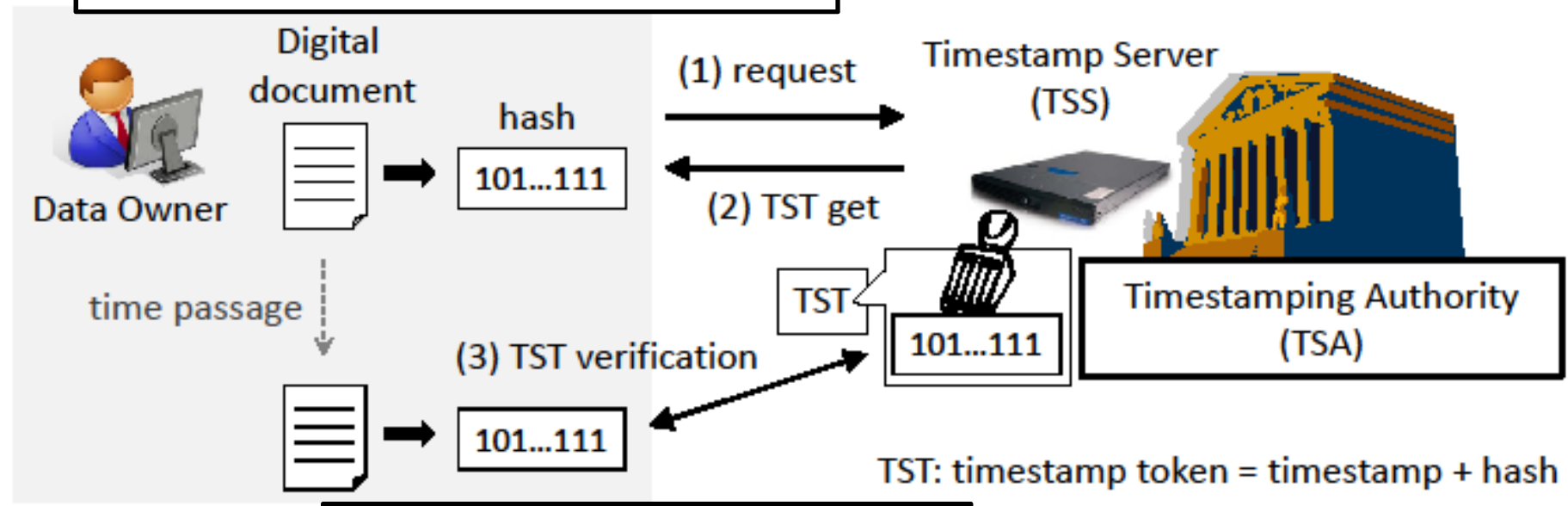
● 背景

- 時刻認証局(TSA)は日本では4社のみ(アマノ、セイコーソリューションズ、NTTデータ、北海道総合通信網)
- 現状では各社とも自社のTSAサーバでのタイムスタンプしか認めていないためクラウドの高速・大容量のデータベースには未対応
- 今後はこのようなクラウドに直結(内包)タイムスタンプ方式の実用化を視野に入れたビジネスモデルが必須である

● サイエンスクラウドの成果

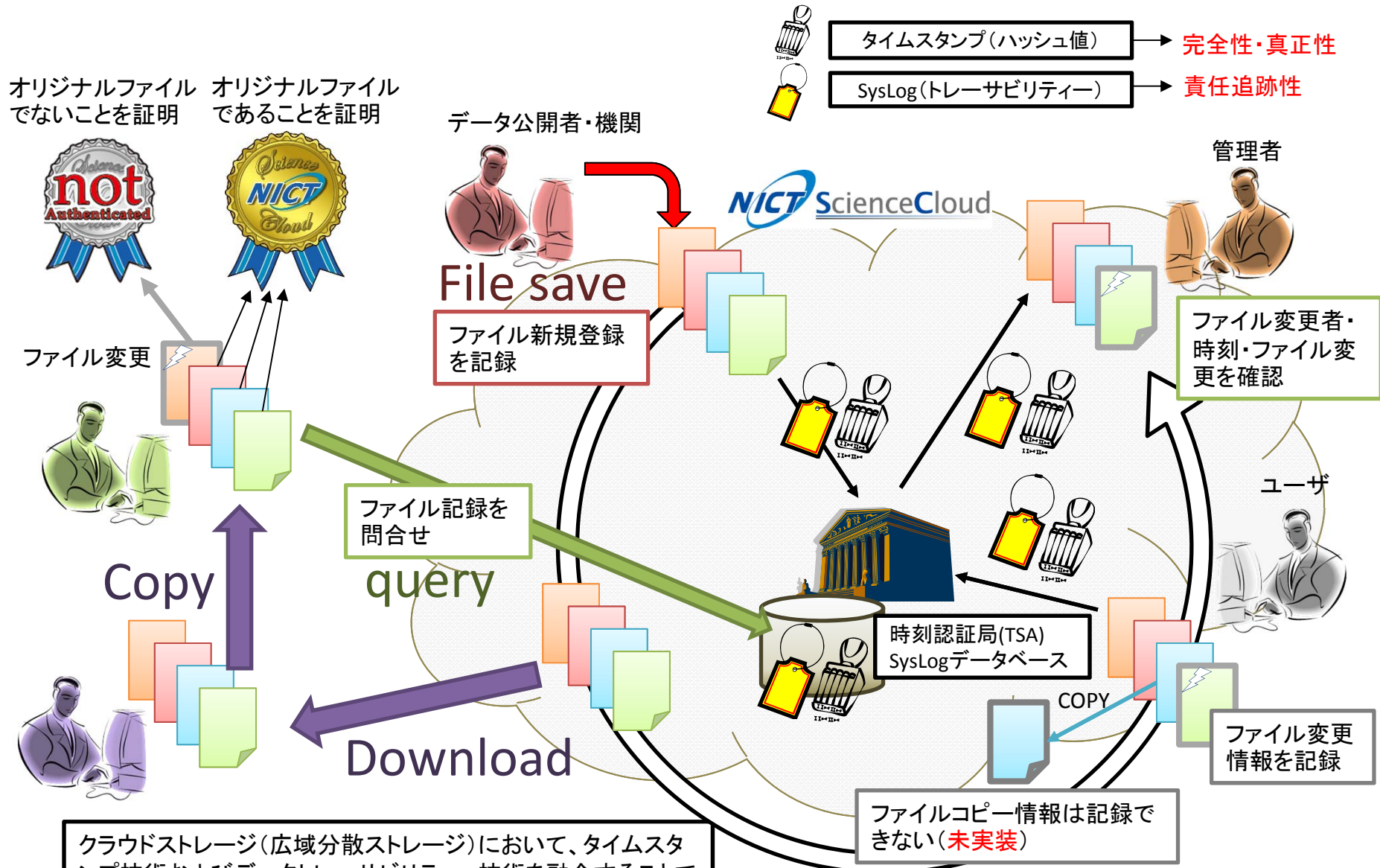
- 分散ストレージシステムとタイムスタンプ・サービスの協調機能によりデータトレーサビリティを実現
- 重要なデータの完全性と真正性を担保

ファイルをストレージに保存するだけで(ユーザは意識せずに)タイムスタンプ付与が可能となった。



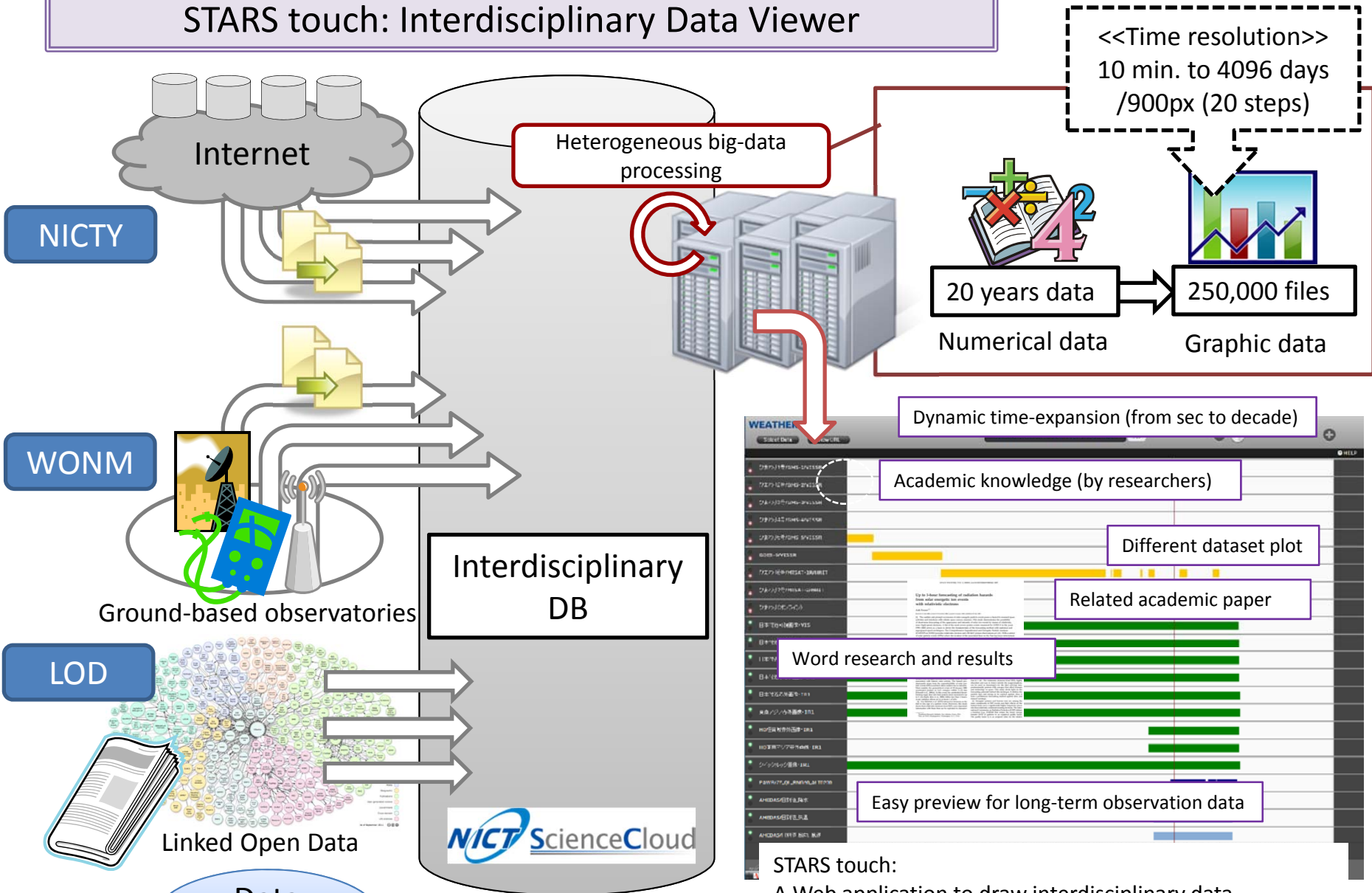
ファイルあたりタイムスタンプ付与(90ms)、タイムスタンプ検証(40ms)の現実性の高い性能を達成。

NICT開発技術 総合的信頼性認証機能



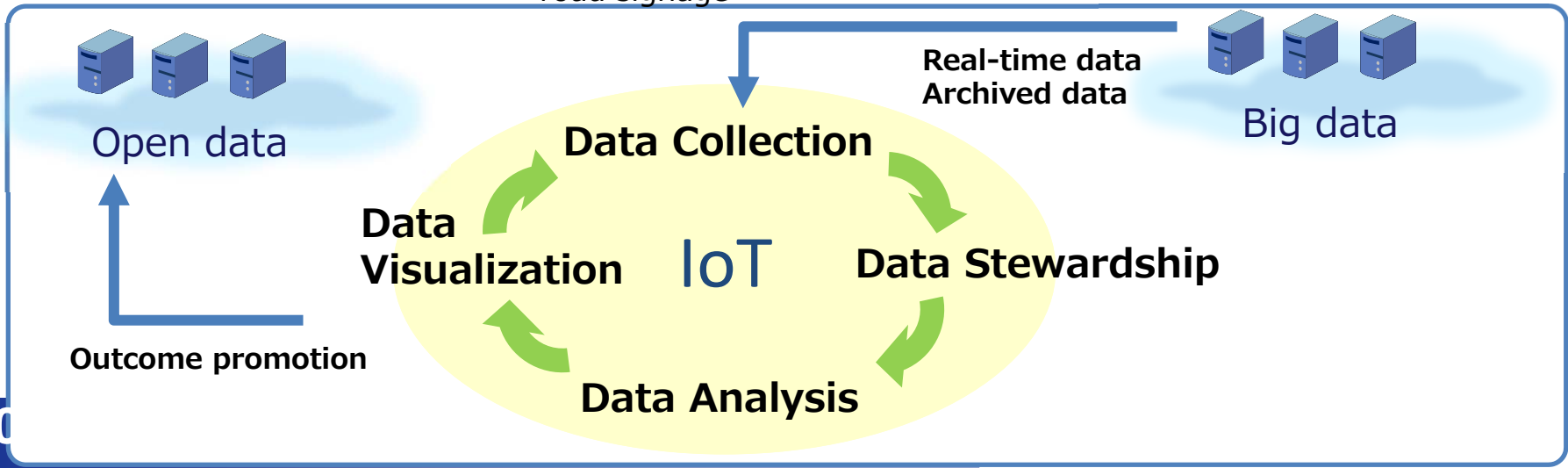
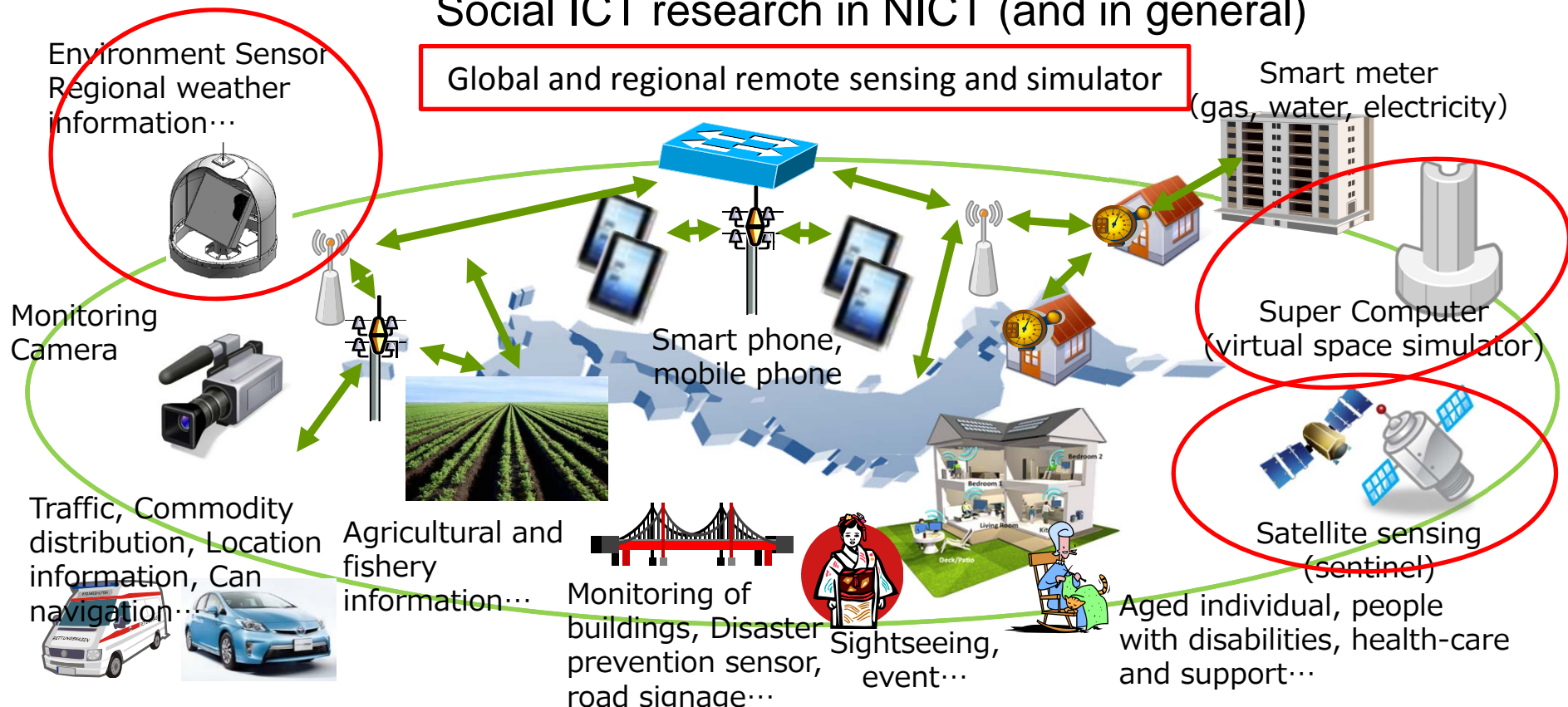
クラウドストレージ(広域分散ストレージ)において、タイムスタンプ技術およびデータトレーサビリティ技術を融合することで、データ完全性・真正性・責任追跡性認証システムを実現。

分野横断型時系列データプレビューア STARS touch: Interdisciplinary Data Viewer

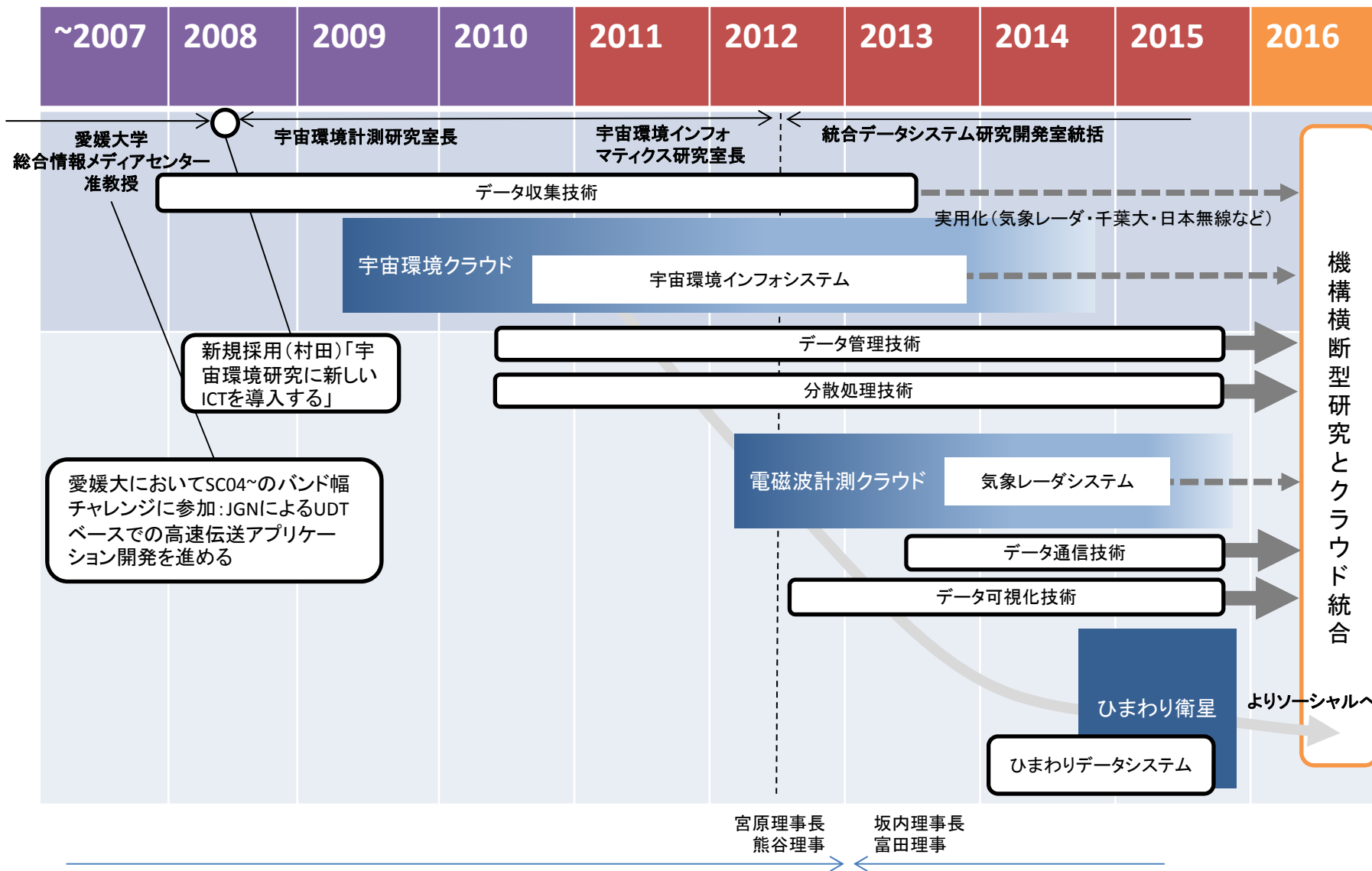


Social ICT research in NICT (and in general)

Global and regional remote sensing and simulator



サイエンスクラウドこれまでの経緯



ひまわり衛星データ転送・保存状況(2014年8月20日現在)

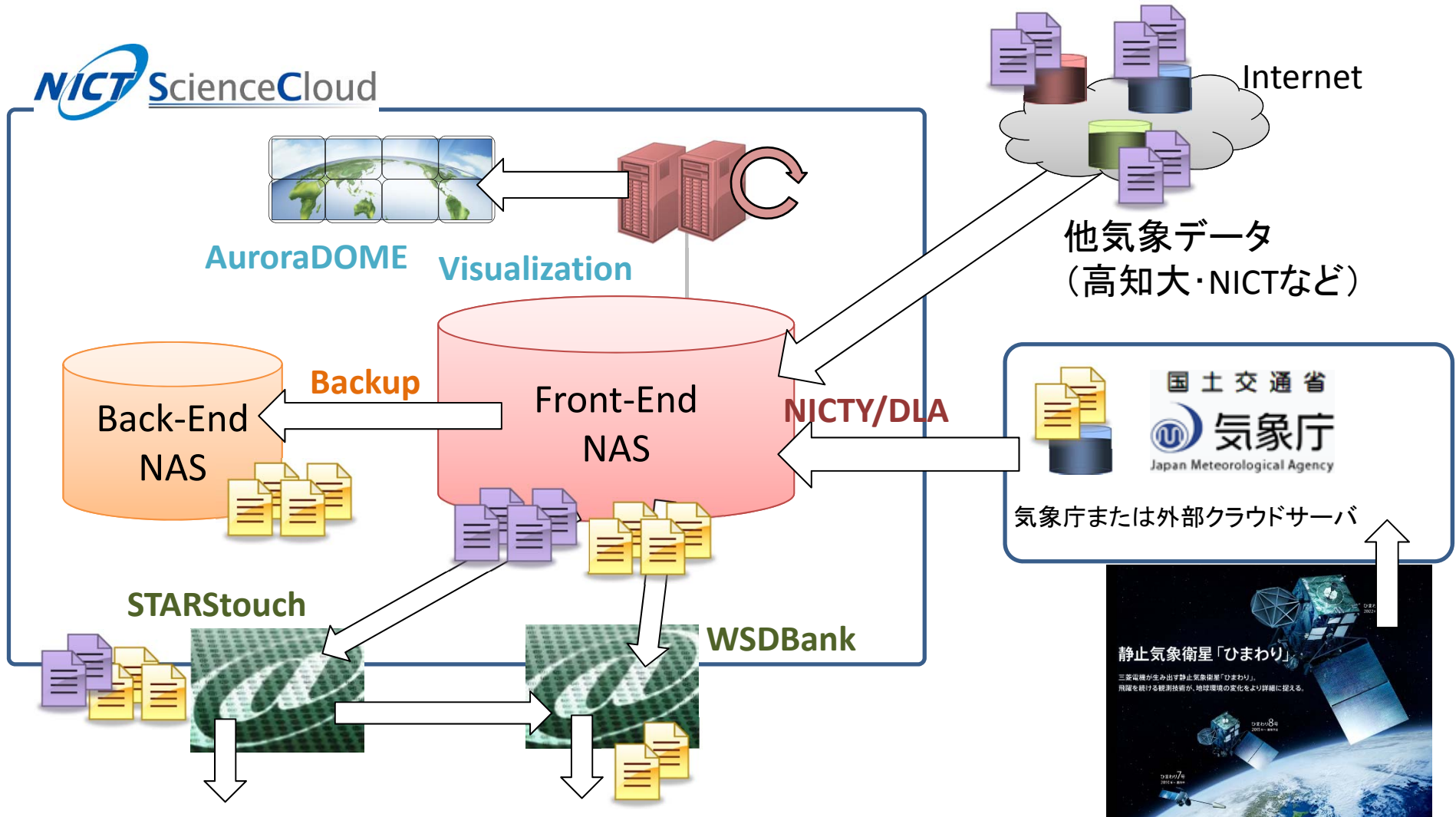
<http://sc-web.nict.go.jp/all-GMS/>

| 衛星名 | 期間 | データフォーマット | 伝送方法(*1) | 伝送状況 | データサイズ (現在: 2014/08/19) | データサイズ (最終) |
|----------------------|-------------|-----------|----------|---------|-------------------------------|--------------------------|
| ひまわり(初号機) | 1981年～1984年 | VISSR | オフライン | 完了 | 108GB | 108GB |
| ひまわり2号 | 1982年～1984年 | VISSR | オフライン | 完了 | 29GB | 29GB |
| ひまわり3号 | 1984年～1989年 | VISSR | オフライン | 完了 | 482GB | 482GB |
| ひまわり4号 | 1989年～1995年 | VISSR | オフライン | 完了 | 1.2TB | 1.2TB |
| ひまわり5号 | 1995年～2003年 | VISSR | オフライン | 完了 | 2.5TB | 2.5TB |
| GOES9号 | 2003年～2005年 | VISSR | オフライン | 完了 | 658GB | 658GB |
| ひまわり6号 | 2005年～2013年 | HRIT | オフライン | 完了 | 6.2TB | 6.2TB |
| ひまわり7号 | 2010年～2014年 | HRIT | オフライン | 完了 | 4.7TB | 4.7TB |
| ひまわり6号、 7号(オンライン) | 2013年10月～ | HRIT | オンライン | 取得中 | 1.2TB | 5.9TB |
| ひまわり8号 | 2015年～ | 標準データ形式 | -(オンライン) | 打ち上げ後予定 | — | 1.5PB |
| ひまわり9号 | 2022年～ | 標準データ形式 | -(オンライン) | 打ち上げ後予定 | — | 1.5PB |
| | | | 合計 | | 17.1TB | 21.6TB (ひまわり8, 9号を除く) |

22 (*)「ひまわり(初号機)」～「GOES9号」までは運用終了。「ひまわり8号、9号」は今後打ち上げ予定
 (*1) オフラインはHDD輸送による伝送、オンラインは、NICTYによる伝送



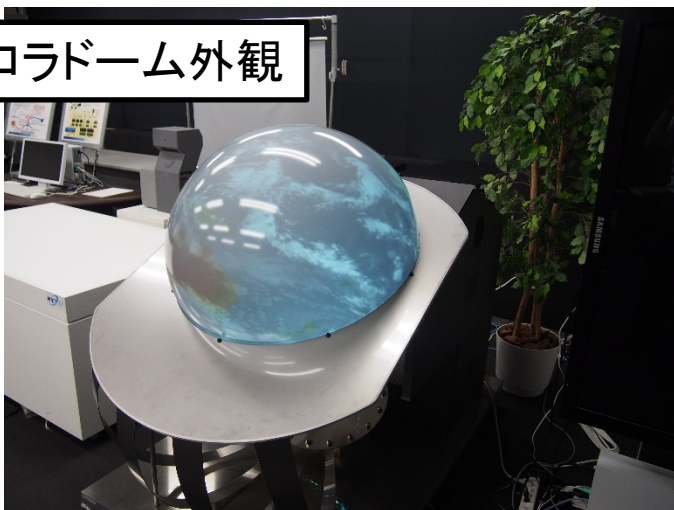
ひまわり衛星データフロー(計画・提案を含む)



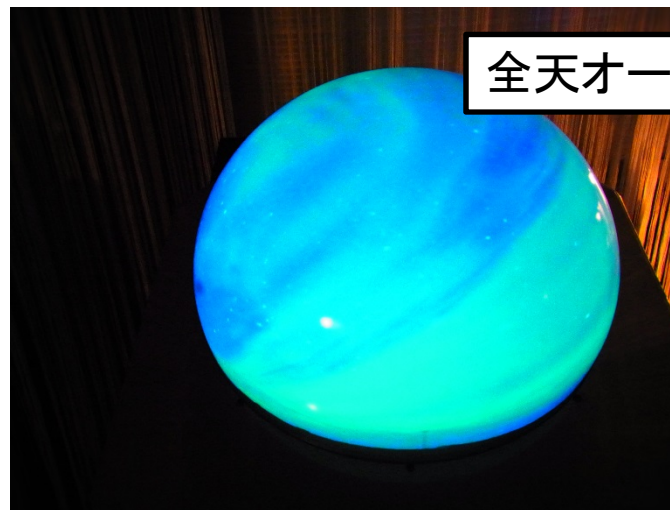
“AuroraDome (オーロラドーム)”

高知大学で準リアルタイムに可視化されるひまわり衛星データをドーム上に表示

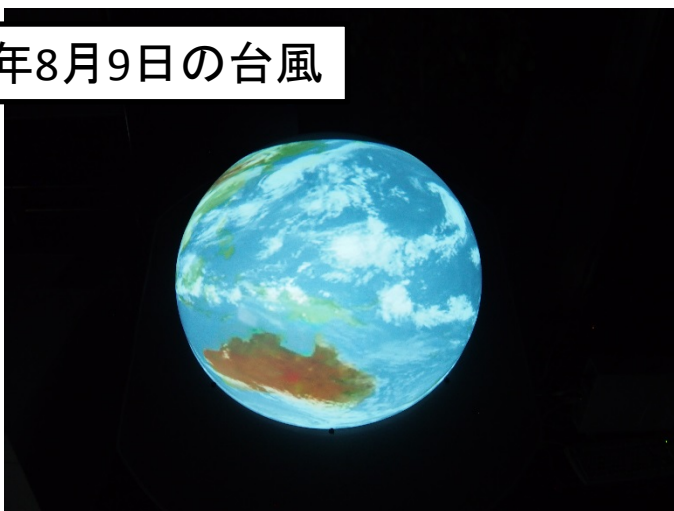
オーロラドーム外観



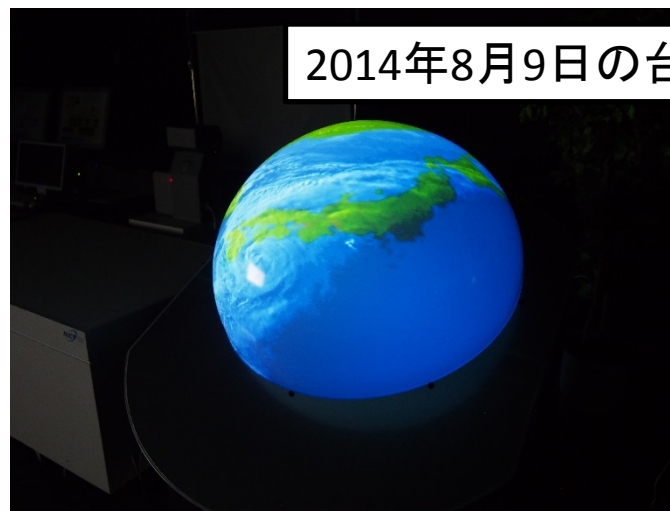
全天オーロラ画像



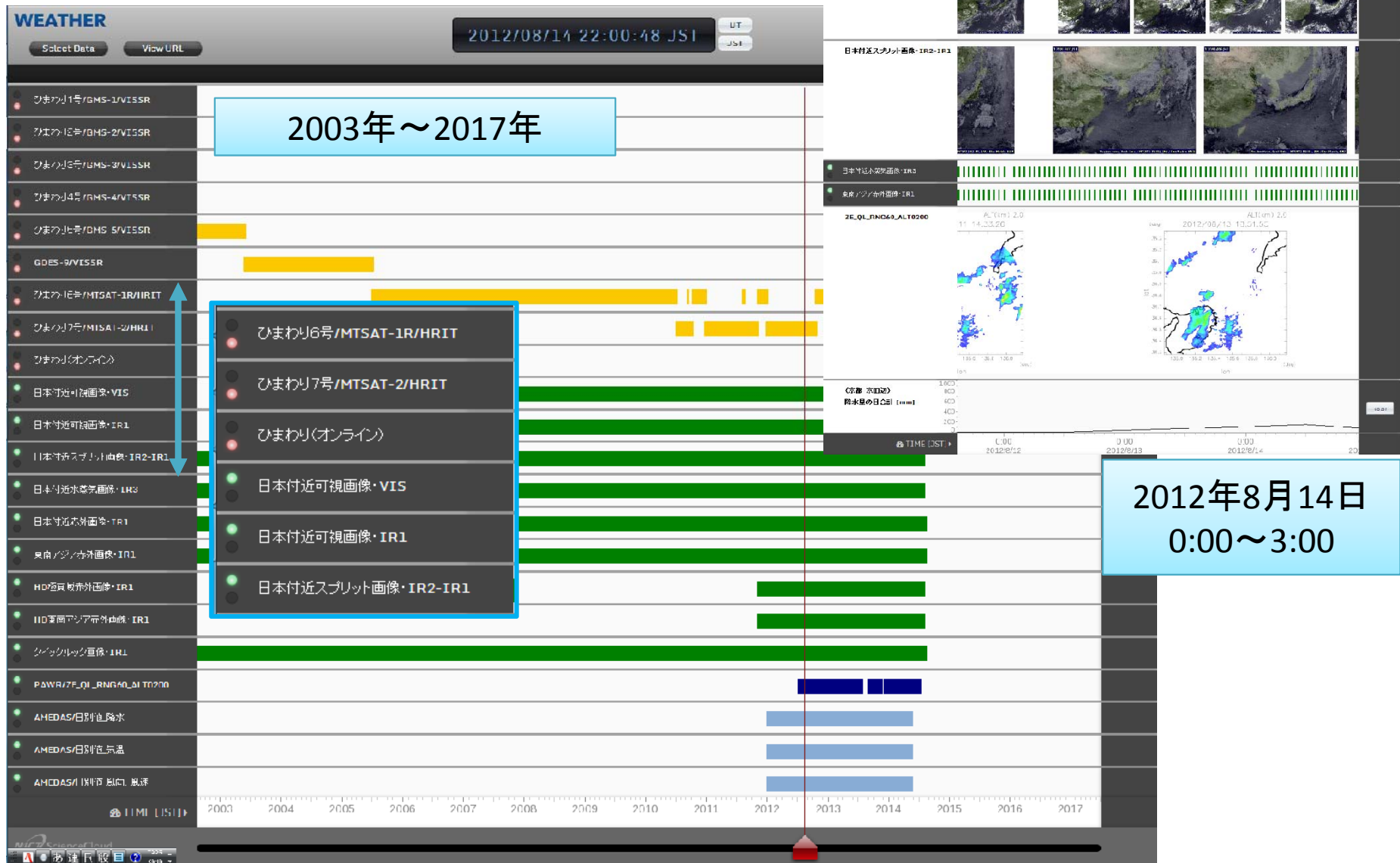
2014年8月9日の台風



2014年8月9日の台風(日本)



STARStouch



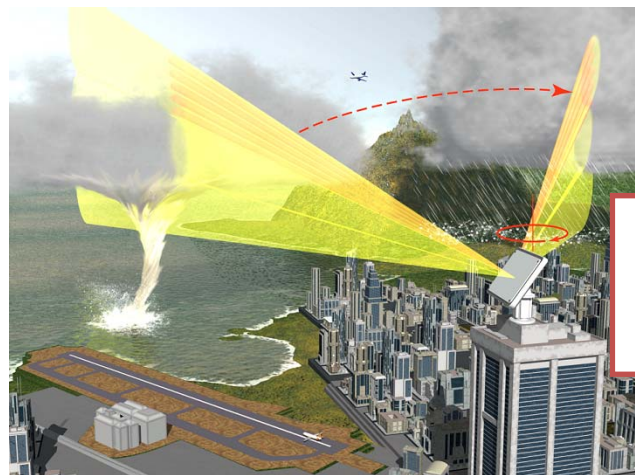
STARStouch→WSDBank

The screenshot shows the WEATHER interface with a list of satellite data sources on the left. A red circle highlights the first item, 'ひまわり1号/GMS-1/VISSR', with a red arrow pointing to a file browser window. The file browser window displays the WSDBank interface, including a table of files.

| Name | Size[B] | Timestamp (JST) |
|--------|---------|---------------------|
| 198103 | - | 2013/08/22 22:22:37 |
| 198104 | - | 2013/08/22 22:26:05 |
| 198105 | - | 2013/08/22 22:29:24 |
| 198106 | - | 2013/08/22 22:32:11 |
| 198107 | - | 2013/08/22 22:35:09 |
| 198108 | - | 2013/08/22 22:38:24 |
| 198109 | - | 2013/08/22 22:41:12 |
| 198110 | - | 2013/08/22 22:44:05 |
| 198111 | - | 2013/08/22 22:47:03 |
| 198112 | - | 2013/08/22 22:48:55 |
| 198401 | - | 2013/08/22 22:49:56 |

3D remote sensing via phased-array radar

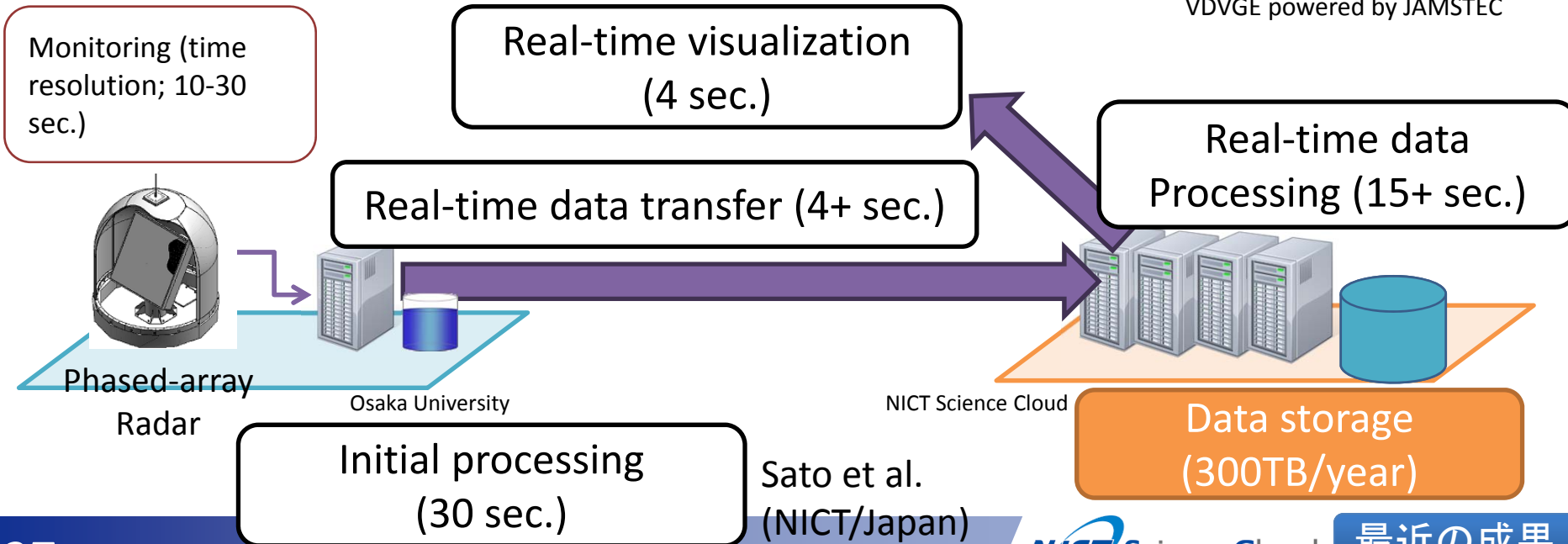
Data Storage and **real-time** data processing for “3D forecasting”



only
70 sec.

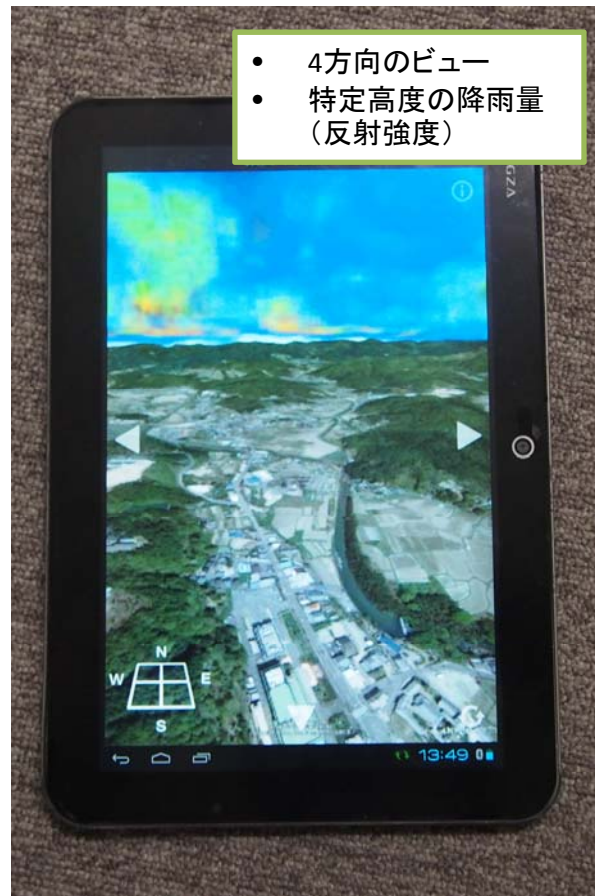


VDVGE powered by JAMSTEC



スマホ・タブレットアプリ試作

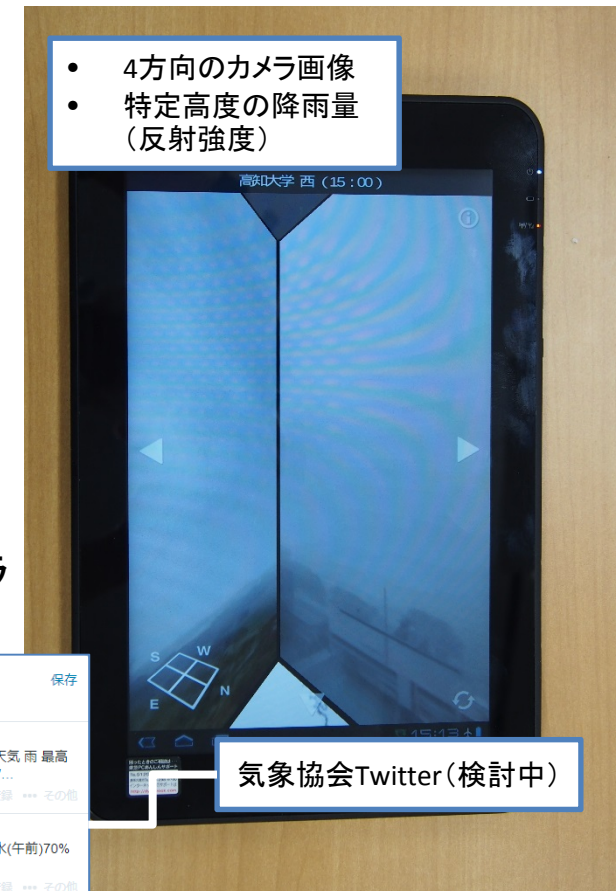
阪大フェーズドレイ版



- 4方向のビュー
- 特定高度の降雨量 (反射強度)

GPS機能により位置情報を取得→その場所からみた降雨

高知大学レーダ版



- 4方向のカメラ画像
- 特定高度の降雨量 (反射強度)

レーダーごとのカメラ画像とレーダ画像



気象協会Twitter (検討中)

“Visualization Gallery” of the Science Cloud

毎月平均5回程度のデモ



まとめ(独り言)

- NICTサイエンスクラウドは順調に成果を挙げている(と思います)
- 宇宙科学での利用が減って、地球科学や人間科学での利用、ビジネス利用が増えています
 - このままではサイエンスクラウドの宇宙科学利用は難しくなるかも
- 使ってくださいとは言いません、ノウハウを吸収してください