

3.15 統合データシステム研究開発室

室長 村山泰啓 ほか8名

科学データの国際的利活用促進と利用基盤技術の研究開発

【室概要】

近年、観測技術の向上や情報通信技術の進展に伴い、生み出されるデータの種類やサイズが爆発的に増大している。そのため、データが持つ貴重な知見が潜在的に増えつつも、それを十分に引き出せない状況がすでに起き始めている。加えて過去の地球観測データなど再現できない貴重な情報が散逸し、利用不能になるなどの問題も指摘されており、それらのデータの長期的な保存と利用体制の維持が国際的な課題になっている。これらの課題に対応するため、国際科学会議 (ICSU: International Council for Science) では、世界科学データシステム (WDS: World Data System) と呼ぶ枠組みを作り、国際的な連携体制構築に向けた取組が行われている。当室には WDS の事務局となる国際プログラムオフィス (IPO: International Program Office) が設置されており、ICSU と連携して WDS 関連事業を推進するとともに、国内関係機関の連携促進やとりまとめに向けた活動、科学データの取り扱いに関する研究開発を行っている。

【平成 26 年度の成果】

(1) WDS 関連活動

現状: WDS 事業としては、科学データの長期保存、分散管理されたデータの相互利用性の確保等を目指し、国際的な組織化やデータ出版 (data publication) の制度化、国際データカタログ事業等を進めている。平成 27 年 3 月末現在、科学データ保有機関 (国連・アメリカ航空宇宙局・中国科学院、各国大学・研究機関等の組織) や大手科学技術出版社等 89 機関が加盟している。国内では NICT 及び京都大学が加盟している。また RDA (Research Data Alliance) 等他の国際組織と MOU を結び、グローバルな連携を展開している (図 1)。

国際会議: ICSU の CODATA (Committee on

Data for Science and Technology) では、物理定数標準値の決定、国際的なデータポリシー策定等を行ってきたが、平成 26 年 11 月にニューデリー (インド) において、WDS-CODATA の合同国際会議 SciDataCon2014 を初めて開催した。そこでは、オープンリサーチデータに関する議論が活発に行われたが、NICT によるデータサイテーション利活用技術の講演に対して、座長・会場から特別の賛辞を送られた (3.5.4 参照)。

G8 と科学技術データ共有基本方針: 平成 25 年 G8 の科学研究データオープン化合意を受け、「国際的動向を踏まえたオープンサイエンスに関する検討会」(内閣府 総合科学技術・イノベーション会議) 及び G8 データ・インフラストラクチャ部会 (事務局: 欧州委員会) に当室室長が有識者委員として参加し、国内・国際的な科学技術データ共有基本方針等の最先端動向について情報・知見を提供した。

(2) 環境計測データネットワーク

NICT が保有する環境計測データについての管理、描画、ファイル提供を実現するための実験システムとして環境計測データネットワークを開発している (図 2)。中層大気等約 20 種類のデータを Web ページ経由で一般に公開するとともに、気象庁へも配信し、日々の気象予報で活用されている。

データの利便性の向上のため、旧 WDC (World Data Center) 機関 (国立極地研究所、京都大学等) との連携体制を構築し、保有データに付与された 100 万件以上のメタデータを、データカタログ相互交換実験に参加し、また国際的識別子である DOI (Digital Object Identifier) を研究データに付与する取組を進め

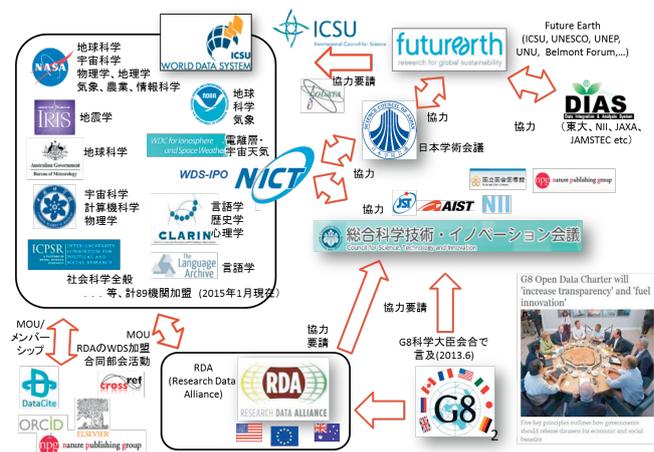


図 1 WDS と国内外機関の関係

た。国内の DOI 登録機関であるジャパンリンクセンターでのメタデータ仕様策定及び理化学研究所、東京大学地球環境情報融合プログラム、産業技術総合研究所、物質・材料研究機構等が参加する同センター実験プロジェクトにサブリーダー及び実験実施機関として参加するとともに、DOI 付与システムのプロトタイプ開発と基礎的な技術試験等を実施した。保有データ（10年間のアラスカ MF レーダー観測データ）を用いた大気波動現象の分析を行う等、長期科学データの利活用による成果創出と共に、ユーザー視点でのデータ管理のためのユースケース調査を並行して行った。

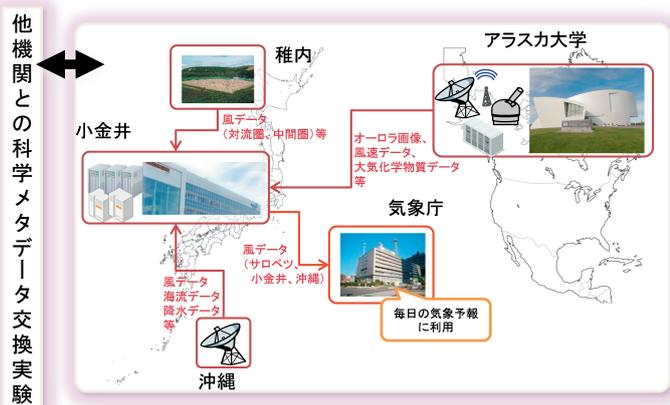


図2 環境計測データネットワーク

3
活動状況

(3) NICTサイエンスクラウド運用及び研究開発

NICTサイエンスクラウドでは研究機関及び民間企業等により30の研究プロジェクトが進められ、利用者数は月平均で700名を超え(図3)、50を超える査読付き学術論文が採録された。独自開発のデータ自動収集ツール(NICTY)も効率的にデータ収集を継続しており、3,000万ファイル(40TB)を超えるデータの自動収集を達成した(図4)。

ビッグデータ統合利活用のための次世代基盤技術による科学ビッグデータのオープンデータ技術として、帯域制御型高速データ伝送プロトコルを活用した多次元多階層型画像データベースによるマルチスケール可視化を進めた(図5)。多次元多階層画像データベースとして長期科学衛星観測データ(20年超)から10億を超えるデータ表示画像を作成し、マルチスケール可視化するためのツール(STARStouch)を開発した(図6)。



図3 サイエンスクラウドログイン数(平成24~26年)

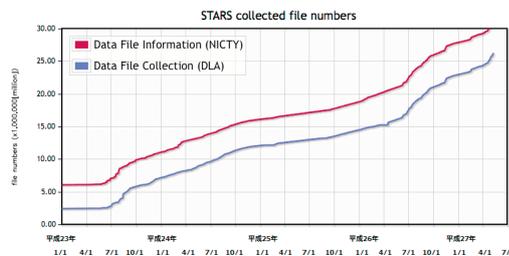


図4 科学データクローラによるデータ収集実績

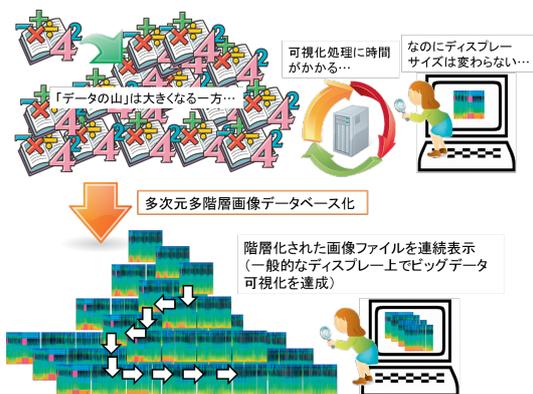


図5 多次元多階層画像データベース

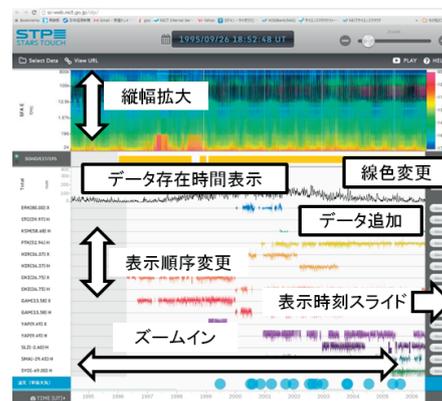


図6 STARStouch(科学衛星GEOTAIL版、<http://geotail.nict.go.jp/>)