共同研究課題

STRAS データベース(2006-2009 年度)

研究代表者: 村田　健史 (情報通信研究機構宇宙環境計測グループ　グループリーダー)
研究分担者: (所属・身分は分担者として最終年度のもの)
木村映善　　愛媛大学医学部　　准教授
篠原育　　宇宙航空研究開発機構宇宙科学研究所 准教授
笠原 禎也　　金沢大学 総合メディア基盤センター　教授
荻野龍樹　　名古屋大学太陽地球環境研究所　　助教
小嶋浩嗣　　京都大学生存圏研究所 准教授
家森俊彦　　京都大学地磁気世界資料解析センター 教授

1. 概要

地球科学の推進には、国際共同による地球観測とデータの自由な交換が不可欠である。1957 年から 1958 年に実施された国際地球観測年 IGY (International Geophysical Year) では、多岐にわたる地球観測が世界各国の協力で実施され、取得された大量の観測データは地球科学の推進に大きな役割を果たした。1957 年の国際地球観測年 IGY から 50 年後にあたる 2007 年から 2008 年にかけて、デジタル地球年 eGY をはじめとして、IPY、IHY、IYPE、GEOSS などの国際計画が進められており、地球観測データのデジタル化と統合的地球観測データ利用の議論が始まっている。わが国においても、文部科学省・地球観測推進部会が中心となり地球観測データ利用の推進について検討しており、並行して国土地理院の地球地図プロジェクト、産業総合研究所の GeoGRID プロジェクト、宇宙航空研究開発機構と海洋研究開発機構連合による海洋地球探査システムなどの巨大なプロジェクトが進められている。

これらの大規模観測データ利用において重要となるのが、高い透過性を持った分散データベースの利用である。自然科学観測分野では、現在、VO (Virtual Observatory) 技術が注目を集めている。VO は天文学で提案された分散データ共有の仕組みであり、世界各地の望遠鏡による大量の天文データを効率良く検索し、サイエンスを行える状態にまで一次処理された観測データ共有と利便性の高いデータ解析環境を目指している。VO 技術において最も重要なのがメタデータの共有と交換であり、天文 VO では OAI-PMH によりリポジトリとハーベスタを構築している。一方、本研究で対象とする太陽地球系物理 (Solar-Terrestrial Physics: STP) 分野は、天文 VO と比較してデータ公開機関が国内外に分散する傾向にあり、データ公開計画や公開ポリシが独自に行われる疎結合ネットワークである。このような疎結合ネットワークにおいて、OAI-PMH によりメタデータを収集するのは容易ではない。

2. 活動の結果・成果

そこで本研究では、疎結合下での太陽地球系物理観測データのメタデータ収集のため、近年インターネットサービスとして活発に利用が始まっている RSS1.0（RDF Site Summary）に注目した。システムではまず必要なメタデータを RSS1.0 で記述し、リポジトリとなる各データサイトに RSS 生成プログラムを組み込む。さらに RSS 収集エージェントがハーベスタとして機能することにより、メタデータ自動収集システムを構築することができる。

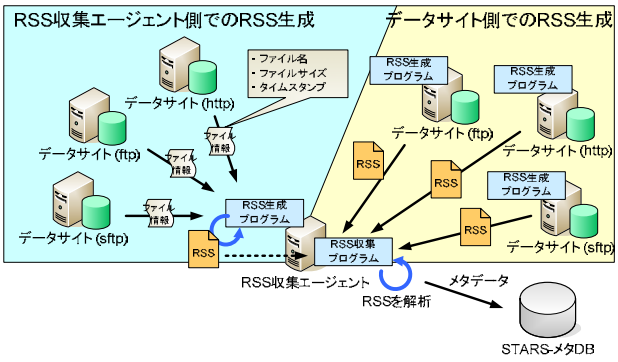

図１　RSS1.0 による観測データのメタデータ自動収集システム

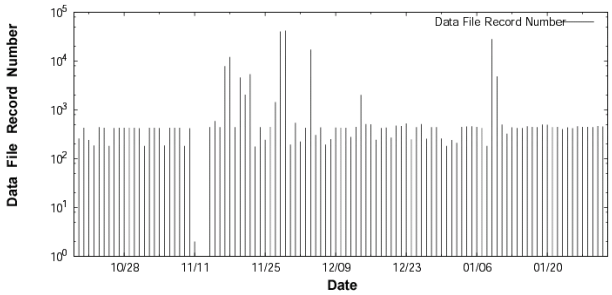

図２　試験運用によるメタデータ登録（2006 年 10 月～2007 年 2 月）：107 日間で約 30 万件のデータファイルの情報を自動登録

本システムで記述されるメタデータは、ファイル名やファイルのタイムスタンプなどの観測データファイル情報や観測日時情報などの観測データファイルのヘッダ情報と言った通常の Web では表わせない独自の項目を含んでいる。RSS1.0 で定義される語彙は基本的なものしかないが、RSS1.0 の特徴のひとつが XML 名前空間を利用した拡張性であり、名前空間を宣言することで Dublin Core モジュールや Syndication モジュールなどで定義される語彙を用いてメタデータを追加することができる。また、これらでは記述しきれない太陽地球系物理観測独自のメタデータ項目については、RDF の拡張性のひとつである独自モジュールを定義して記述できる。

図１に、本研究で構築したメタデータ自動収集システムを示す。システムでは、各データ公開サイトにおいて、RSS 生成プログラムを定期的に実行し、RSS ファイルを生成する。さらに RSS 収集プログラムを使った RSS 収集エージェントが RSS ファイルを定期的に収集し、必要な情報を抽出後にメタデータベースに登録する。一方、RSS 生成を許可しないデータ公開サイトについては、FTP や HTTP などのプロトコルによりデータサイトのファイル情報を収集し、RSS 収集エージェントサーバにおいて RSS ファイルを作成する。また、NASA/GSFC（ゴダード宇宙飛行センター）の NSSDC のように Web サービスによりメタデータを積極的に公開しているデータ公開サイトについては、Web サービスによりメタデータを収集後、RSS 収集エージェントサーバ上で RSS ファイルを作成することも可能である。

以上のように RSS1.0 を活用することで、疎結合ネットワークにおいて観測データファイルのメタデータ自動収集システムを実装した。自動収集システムを実装・運用したところ、4 つのデータサイトの約 30 万ファイルのメタデータの収集に成功した（図２）。本システムでメタデータを収集できなかったデータファイルはなく、本システムの有効性と実用性が確認された。

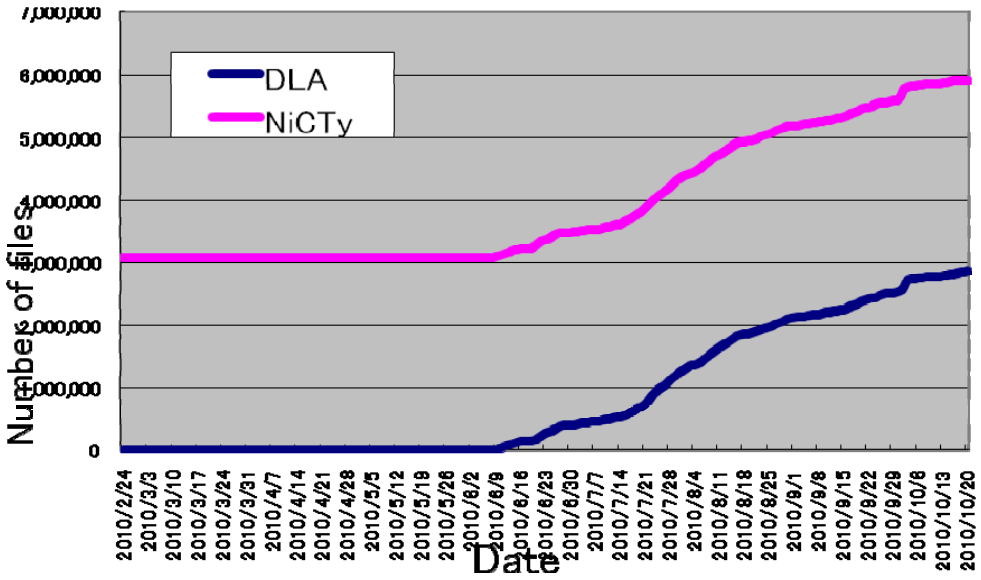

図 3　NICTY によるメタデータ情報自動収集結果および Download Agenet によるデータファイル自動収集結果

なお、本システムは新規追加される観測データファイルのメタデータ収集には有効であるが、データファイルが削除された場合に削除情報を収集することができない点に問題が残っている。これは、RSS ファイルが既存の観測データファイルのメタデータを元に作成されたものだからである。STP 分野では生成した観測データファイルが削除されることは頻繁にはないため実用上の問題はないが、より高い精度でのメタデータ収集ためには、今後の検討が必要である。本研究で実現したシステムは、太陽地球物理観測データに特化したものではなく、自己記述型の観測データファイルを公開する研究分野であればどのようなデータに対しても利用可能である。したがって、今後、疎結合ネットワーク下において Virtual Observatory を実現するためのメタデータ収集技術のひとつとして応用が期待される。

3.　データ収集・登録状況

本計画では、計画終了後も、データ収集システムを愛媛大学から情報通信研究機構に移し、自動収集システムを NICTY と名付けてメタデータの自動収集を行っている。さらに、NICTY メタデータベースを参照することで、HTTP や FTP、SFTP 等で公開されているデータファイルを自動収集するエージェント（STARS Download Agent）サーバを構築した。これにより、メタデータおよびデータファイルの長期自動収集を行っている。その結果を、図 3 に示す。2010 年 10 月現在で、600 万のメタデータファイル情報と 300 万のデータファイルのダウンロードに成功している。

なお、これらの自動収集されたデータは、NICT のサイエンスクラウド（OneSpaceNet）に格納される。OneSpaceNet は、科学研究専用のクラウドコンピューティングシステムであり、大規模分散ストレージも特長の一つである。大規模分散ストレージは、可用性の高いミドルウェアである Gfarm により構築している。図 4 は、OneSpaceNet の大規模分散ストレージの可用性の高さを示している。

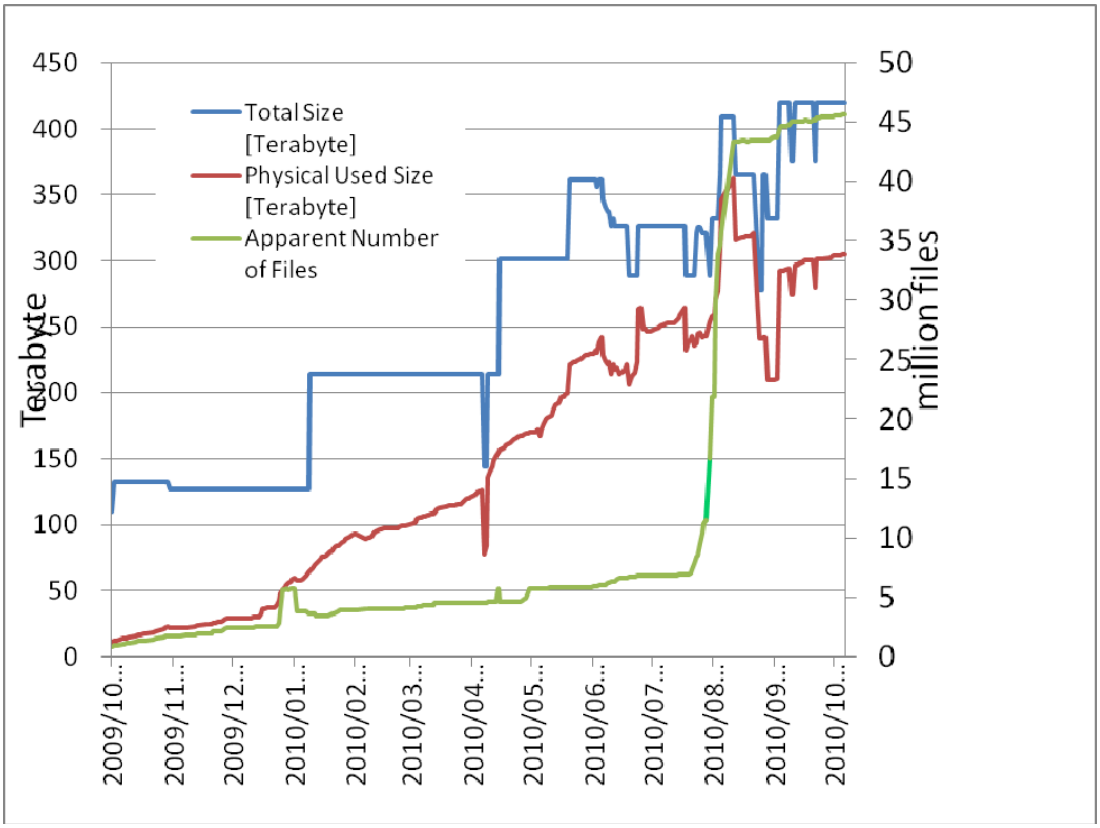

図 4　サイエンスクラウドストレージの可用性および利用状況

NICT では、これらにより収集されたデータをプロットするツールとして、STARS（科学衛星・地上観測データ解析参照システム）を開発している。STARS のデータプロットは多岐にわたる。STARS ドキュメントを、付録として添付する。

4.　関連論文（2006～2009）

2009 年：磁気圏プラズマ波動スペクトルにおける低域カットオフ周波数の自動検出法の開発、白石隆文、石坂圭吾、三宅壮聡、岡田敏美、村田健史、電子情報通信学会論文誌 B Vol.J92-B No.4 pp.794-802

2008 年： 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、松岡 大祐、村田 健史、藤田 茂、田中 高史、山本 和憲、木村 映善、可視化情報学会論文集、Vol.28, No.6, pp.38-46

2008 年：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、石倉 諭、村田 健史、久保 卓也、木村 映善、山本 和憲、篠原 育、電子情報通信学会論文誌(B)、Vol.J91-B, No.4, pp.499-509

2006 年：Development and Application of Geospace Environment Simulator for the Analysis of Spacecraft-Plasma Interactions, HIDEYUKI USUI, YOUHEI MIYAKE, MASAKI OKADA, YOSHIHARU OMURA, TOORU SUGIYAMA, TAKESHI MURATA, DAISUKE MATSUOKA, HIROKO UEDA, IEEE Transactions of Plasma Science Special Issue, Vol.34, No.5, pp.2094-2102

2006 年：RSS/RDF を利用した太陽地球系物理観測データのメタデータ配信の検討、木村映善、村田健史、情報処理学会論文誌、Vol.47, No.4, pp.1051-1062

5. 公開 web page のアドレス

本計画に関する情報は、すべて http://e-sw.nict.go.jp において公開されている。