

## 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システム

### Automatic collection system of meta-data for Solar-Terrestrial Physics (STP) observation data via RSS1.0

# 村田 健史 [1]; 石倉 諭 [2]; 木村 映善 [3]; 久保 卓也 [2]; 山本 和憲 [2]; 篠原 育 [4]; 石井 守 [5]; 笠原 禎也 [6]

# Takeshi Murata[1]; Satoshi Ishikura[2]; Eizen Kimura[3]; Takuya Kubo[2]; Kazunori Yamamoto[2]; Iku Shinohara[4]; Mamoru Ishii[5]; Yoshiya Kasahara[6]

[1] 愛大・メディアセンター; [2] 愛媛大・理工; [3] 愛媛大 CITE; [4] 宇宙機構 / 宇宙研; [5] 情報通信研究機構; [6] 金沢大 [1] CITE, Ehime University; [2] Ehime Univ; [3] CITE, Ehime Univ.; [4] JAXA/ISAS; [5] NICT; [6] Kanazawa Univ.

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 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 または Virtual Organization) 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、宇宙を含めた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 (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ー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。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利用が始まっている RSS1.0 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。

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の項目については、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することができる。

図に、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行い、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。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。この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31万ファイルのメタデータの収集に成功した。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。

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。太陽地球系物理観測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。