

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[†]

近年、太陽地球系物理の分野では、高度に発達した計測機器開発による観測範囲の拡大ならびに国際協調プロジェクトの立ち上がりを背景として、世界各地の観測データを蓄積したデータサイトからの観測データの横断的検索、分析を提供するシステムが要請されている。しかし、対象領域が多様性に富んでいるため、システム構成にも変化があり、Grid Computing や Web Service での統合も困難であることが明らかになっている。既存のシステムへの変化のインパクトを最小に抑え、かつ多様性の記述と柔軟な連携を実現するメタデータの記述を検討した。広くコンテンツ情報配信に利用されている RSS1.0 の簡素さと RSS1.0 が RDF/XML に基づいていることからオントロジ要素の記述に基づいた拡張ができることに注目し、RSS の独自拡張を施すことで既存のシステムと新しい流れの技術の親和的融合が実現できた。本論文では太陽地球系物理分野の観測データのメタデータの記述のコンテナとして RSS を利用するための検討結果について報告する。

Derivery Service of the Meta-data for Solar-terrestrial Observation Data with RSS/RDF

EIZEN KIMURA[†] and KEN T.MURATA[†]

In the Solar-Terrestrial Physics (STP) field, the Virtual Observatory (VO) project is difficult to realize because the data model of STP is wide-ranged and rich in diversity. To realize VO in STP, we have to develop the method that absorbs the gap between the data models in many data site. We have designed the light-weight meta-data delivery mechanism with RSS 1.0 which supports the extensibility from the RDF nature. We have designed the class and relationships of between Mission, Team, Group, User, and observatory data with OWL by the lesson of comparing the difference of models and web-services: CDAS and STARS. This RSS/RDF based meta-data proposes promising prospects which it shows capabilities of changing the specification without breaking forward application compatibility by making deduction from ontology information.

1. はじめに

自然科学分野では近年のネットワークの高速化と関連技術の発達の成果を基盤として eGY (The Electronic Geophysical Year)¹⁾ に代表されるように対象の観測、観測情報の蓄積、検索、データのダウンロード、解析といった新知見に至るまでの煩雑な事前の準備作業を簡略化し、科学の発展に寄与するようなシステムを志向する事例が始めている²⁾。観測データの所在を抽象化し、あたかも実物を観察しているかのように縦横無尽に観測データにアクセスできるシステムが仮想観測所：VO (Virtual Observatory) である。VO は、地球物理学分野や天文学分野で実際にシステ

ムの開発が進められ、一定の成功を収めている^{3),4)}。

しかし、太陽地球系物理領域ではいまだ VO を実現するには至っていない。この理由にはいくつか考えられる。最も大きな理由は、観測領域、対象と観測データの種類が多岐にわたり、天文学における観測データよりも複雑なメタ情報の記述、管理が要求されていることである。現在、STARS⁵⁾ のメタデータベースが管理している太陽地球系物理観測データは、観測データを蓄積しているデータサイトが 20 カ所であり、それらのデータサイトが従事している観測計画 (Mission) 数は 29、観測班 (Team) は 84、観測データファイル数は 1,884,901 (2005 年 9 月現在) になる。また、対象フォーマットは CDF, AEindices, DST, Kp 等をはじめとして、16 種類の観測データ形式がある。なかでも、CDF (the Common Data Format) はバイナリ形式の自己記述形式であり、CDF 内部で独自にデータ

[†] 愛媛大学総合情報メディアセンター

Center for Information Technology, Ehime University

格納様式を定義できるため、実際のデータフォーマットは数十種類にのぼる。

以上のように多様性に富む STP 分野のデータのメタデータを統一的に記述するために SPASE (Space Physics Archive Search and Extract ⁶⁾) による XML Schema を用いたメタデータ仕様策定が試みられている。しかし、メタデータ概念や定義の仕方に意見が分かれており、合意に至るまでの道のりが先行き長いといった状態である。

VO の成功には、(1) ユニバーサルなアクセシビリティの提供、(2) 容易に参加可能である、(3) データとモデルだけでなく、ツールと情報を提供する、(4) つねに進化し続けること、があげられている²⁾。(1)からは Grid Computing や Web Service といった SOA (Service Oriented Architecture) に基づいた提供が考えられるが、Grid Computing は (2) 実現の容易性の観点からは困難である。そして、(3) に対しては村田らの取り組み^{5),7),8)}があった。しかし、この研究は横断的検索のポータルサービスと入手したデータのプロットにとどまり、そして最後の (4) に対してもつねに変化に追従できる状態とはいえない。そこで我々は太陽地球系物理領域における VO を実現するために必要なメタデータの定義と配信環境の構築を試みることにした。

2. 太陽地球系物理領域の現状および STARS プロジェクト

2.1 プロジェクトの構成

STARS (Solar-Terrestrial data Analysis and Reference System) プロジェクト^{5),7)}とは、太陽地球系物理観測データの横断的検索、入手を実現するためのメタデータベース・プロジェクトである。このメタデータの対象である、STP (Solar-Terrestrial Physics) データのソースは人工衛星の観測データと地上観測データが中心である。STP 研究者はこれらのデータを総合的に解析することで STP 研究を進めている。たとえば、太陽面で爆発現象が起きて放出されるプラズマ放出の現象解析や太陽風による電子通信網の障害に対する予測等に使われている。

しかし、人工衛星を用いた地球観測のプロジェクトは莫大な予算が要求される。プロジェクトには多数の研究者や研究グループが参加し、打ち上げる人工衛星にそれぞれの研究グループが担当する観測機器を持ち寄って搭載することでコストをシェアするような研究計画が組まれる。このプロジェクトが Mission である。そして、それぞれの研究グループ (Team) の中で観

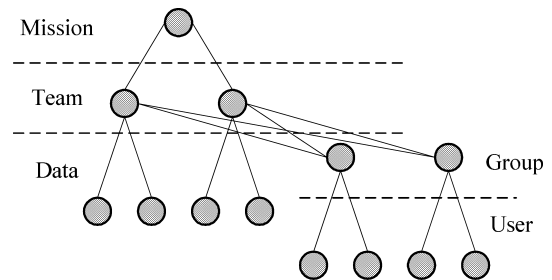


図1 STARSにおけるプロジェクトの構成

Fig.1 STARS project structure.

測機器の設計からデータ管理までを統括するのが PI (Principle Investigator) と呼ばれる責任者と、Co-I (Co-Investigator) と呼ばれる研究協力責任者である。PI や Co-I は特定の研究グループ (研究室, Group) に所属している。複数の研究室が1つまたは複数の観測機器の観測に携わり、共同研究の形で観測機器から観測された観測データをもとに研究を行う。

以上のことをまとめると、1つの Mission は1つの衛星に対応し、複数の観測班 (Team) によって構成されている。観測班 (Team) には複数の研究室 (Group) が属する。それぞれの研究室には、複数のデータ解析者 (User) が存在する。同様の関係が地上観測に対しても適用される。観測データのアクセスコントロールはこの研究組織階層に従って、Mission/Team/Group/User 単位で設定が可能である。一方、観測データの所属も n 分木の階層構造で表現される。人工衛星による科学観測においては1つの衛星 (Mission) に複数の観測班 (Team) が属している。そして各観測班は複数の観測データを蓄積するデータサイト (Data) を所有している。したがって、この構造は図1のような Mission を頂点とする n 分木の階層構造として表すことができる。この n 分木構造を RDBMS のテーブルにマッピングしたのが図2である。これら2つの階層的な帰属情報をもとに研究者の観測データのアクセス権限が決定される。

このようにアクセス権限、観測データ所属の関係が複雑な構造になっている理由は、(1) 複数の運用ポリシーが異なる研究室によって構成されていること、(2) 研究活動の利益保護のために、研究担当者が観測データへの優先的・排他的アクセス権限を一定期間有し、その後、随時公開されていくという時限的アクセス権限の設定を設ける必要があること、(3) 観測データが特定のデータサイトではなく、複数のデータサイトに分散して蓄積される場合があること、からきている。

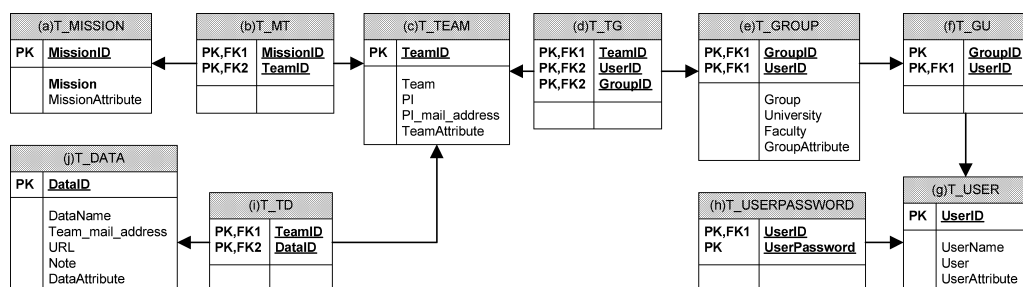


図 2 STP メタデータベースのテーブル仕様
Fig. 2 Table layout of STP meta-database.

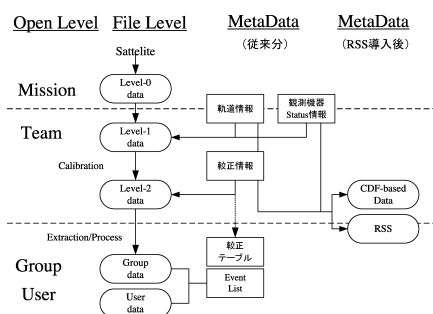


図 3 STP 観測データの処理フロー
Fig. 3 STP observation data workflow.

2.2 観測データの処理フローから閲覧まで

人工衛星にて観測されたデータは全観測情報を含む Level-0 データとして一括保存される (図 3)。そこから各 Team は各 Team の担当の観測データを Level-0 データより抽出し、軌道情報、機器ステータスを追加し、Level-1 データとして生成するが、この時点ではまだ研究データとしての価値はない。物理的に信頼できる数値として利用するために観測機器のバイアス等を修正するデータの較正作業が必要であり、この作業を経て生成されたものを Level-2 データと呼ぶ。この Level-2 データと最新の較正用テーブルが研究者に対して配布され、研究者らはこのデータをもとに解析を行う。解析結果はユーザ独自にデータセットを記述したイベントリストを添付する形で公開する場合もある。公開データフォーマットおよびファイルの命名が観測班によって異なり、ファイル形式のほとんどはアスキーデータ、画像フォーマット、バイナリデータフォーマットに分類される。そのため、横断検索を実現するためには観測データの PI/Co-I より承諾を得て、Level-2 に対応する情報を貰い、その情報からメタデータを書き起こして STARS メタデータベースに手動で登録する作業が必要である。

STP 研究者は STARS メタデータベースのフロントエンドである STARSdb を利用し、メタデータの

検索を通して複数のソースおよび時間軸の観測データを一括ダウンロードする。複数の観測班のデータサイトから、横断的にダウンロードした場合、データフォーマットが異なることがある。観測班ごとのデータのフォーマットの違いを吸収するために開発された SeDOC ライブラリを利用した STARSPlot を利用して複数のソースの観測データをプロットすることで地球磁気圏の多岐、広範囲にわたる現象を俯瞰することが可能である⁸⁾。

3. 太陽地球系物理領域に関するメタデータの現状

3.1 STARS の現状

STARS プロジェクト^{5),7)} では各研究機関に蓄積された太陽地球系物理の観測データに関する情報をメタデータとして蓄積することによって、横断的な検索を提供している。しかし、大元のデータサイトでは観測データは前節までに紹介したようにほとんどがプリミティブな形式や独自形式であり、観測に関するメタ情報は付随していない。また、命名規則を設けることで観測時刻やプロジェクト名をファイル名から類推できるものもあるが、情報量が不足しており、命名規則もデータサイトによって異なる。

そのため、STARS メタデータベースの管理者が各データサイトの管理者に了解をとり、メタデータ抽出に必要な情報を受け取り、該当ファイルの格納形式に従った分析をし、メタデータの抽出、付与をして、データベースに登録するという手間がかかる。メタデータベース運用者の負担が大きく問題となっており、メタデータの作成、配信、登録のプロセスを効率化することが望まれている。

3.2 IVOA における OAI-PMH によるメタデータ配信

IVOA (International Virtual Observatory Alliance) では天文観測データに関するメタデータを

OAI-PMH を利用して配信, 検索する試み⁹⁾ をすでに
行っており実用段階に至っている^{3), 4)}。しかしながら
OAI-PMH で使われている標準スキーマである OAI-
DC は図書のメタデータを扱うことを想定しており,
比較的制限されたデータセットの定義である。自然科
学分野への応用をするにはメタデータが不足するた
め, Dublin Core¹⁰⁾ では表記しきれないメタデータ
を独自スキーマとして定義し, 配信する試みがある¹¹⁾。
IVOIA で OAI-PMH を採用できたのは, データの種
類が比較的少ないためにデータの委細情報を記述する
必要がなく, メタデータを定義するスキーマの策定も
比較的シンプルなものにとどまったことが大きいと思
われる。STP 分野については後節で記述するように,
スキーマ策定そのものが難航している事情があるため,
OAI-PMH のように配信するメタデータのスキーマが
事前に決定されている必要がある仕組みでは実装が進
まない。そこで, スキーマが未確定の状態でも前方互
換性を保ちつつ, メタデータを配信することができる
仕組みを作る必要がある。前方互換性が問われる理由
は, STP 分野は運用ポリシーが異なる研究組織によ
って構成されているため, システムの導入の足並みを揃
えることが難しく, メタデータ配信, 受信システムに
おいてもつねに同じバージョンを利用することを前提
とすることができない。また, OAI-PMH を利用する
には REST アーキテクチャ¹²⁾ に準拠することが前提
であることも制約の 1 つである。セキュリティ確保の
ために, HTTP ではなく, SFTP での配布を希望し
ているデータサイトも存在するからである。

3.3 SPASE の現状

太陽地球系物理観測データのメタデータのために
XML Schema ベースでのメタ情報を策定する SPASE
コンソーシアム⁶⁾ があり, 現在も策定進行中である。
XML Schema を文書構造の定義に採用しており, 対
象範囲が広範にわたっているため, 構造の確定に時間
がかかっている。その結果 SPASE を前提とした応用
アプリケーションの開発, 検証も進まず研究活動の進
行が遅れている。

XML Schema はデータの記述の構造を定めるが, 同
時に, そのデータを扱うプログラムの構造も定めてし
まう。かつ, 構造の変化が生じるとそのスキーマを使
用しているアプリケーションの前方互換性がなくなる。
そのため, 将来にわたって改変が少ないようにより多
くの情報を収容しようとし, 仕様の先取りや拡張が繰
り返される結果, 策定が収束するのに数年かかる事例
が見られる¹³⁾。そして, 太陽地球系物理分野では, 大
規模かつ多岐にわたるシステム群が関わるため, 統一

されたデータ構造, およびインタフェースをすべてに
適用させることは現実的に不可能である。

この問題を解決するために, 仕様変更で前方互換性
がなくなる XML Schema でのメタデータ定義以外の
アプローチをとり, メタデータ仕様策定と研究開発が
並行して進行できるような方法が求められている。す
なわち, (1) 最低限のメタデータの情報交換をするこ
とから始め, (2) 運用を通して必要なメタデータを洗
い出していき, 随時メタデータ仕様を改編, 拡張して
いくという運用ができる方法が求められている。

この問題意識に対しては Semantic Web のオント
ロジーの適用を通して, 複数の異なる組織によって定義
されるデータ構造の多様性を吸収し, 相互の情報交換
を実現する手法が有効であると思われる。そして, 情
報交換を通して, 「最大公約数」として認知されるに
至ったメタ情報を SPASE コンソーシアムに必須メタ
情報として収録するように働きかけができることが期
待される。

4. モデル設計とサービス設計の相違

STARS では様々な観測計画のデータサイトのメタ
データを収集し横断的検索を提供しているが, 先述の
とおりメタデータはフォーマットが統一されていない
ため, 手作業の入力が主である。そこで NASA の運
営している CDAS (the Coordinated Data Analysis
System) の Web Service¹⁴⁾ を利用してメタデータ収
集の自動化を試みることで, メタデータの自動収集の
ための要件抽出を図った。図 4 に CDAS と STARS
におけるモデルと Web サービスを示す。CDAS では
観測ミッションのグループである Mission Group とそ
の観測データの Source (人工衛星) は, STARS では
Mission に相当し一本化されている。この違いはミ
ッションは人工衛星の打ち上げに付随して実施されるた
め, STARS ではミッションと人工衛星の関係をほぼ
同一視していることによる設計思想の違いからきてい
る。次に, CDAS では人工衛星に搭載された観測機器
(Instrument) とタイプ (Instrument Type) をミ
ッションに関連付けているが, STARS ではミッ
ションに従事するチーム (Team) が相当する。これは人
工衛星に搭載される様々な観測機器ごとに担当するチ
ームが 1 対 1 で割り当てられるため, 観測機器とチ
ームを同格に見なしているためである。最後に観測デ
ータのセット (DataSet/Data) とそれぞれのセットに
属する観測データの実体 (DataUrl/DataFile) はほ
ぼ同格である。STARS はこれらに加えてチームに所
属するグループ (Group), そして研究者 (User) 情

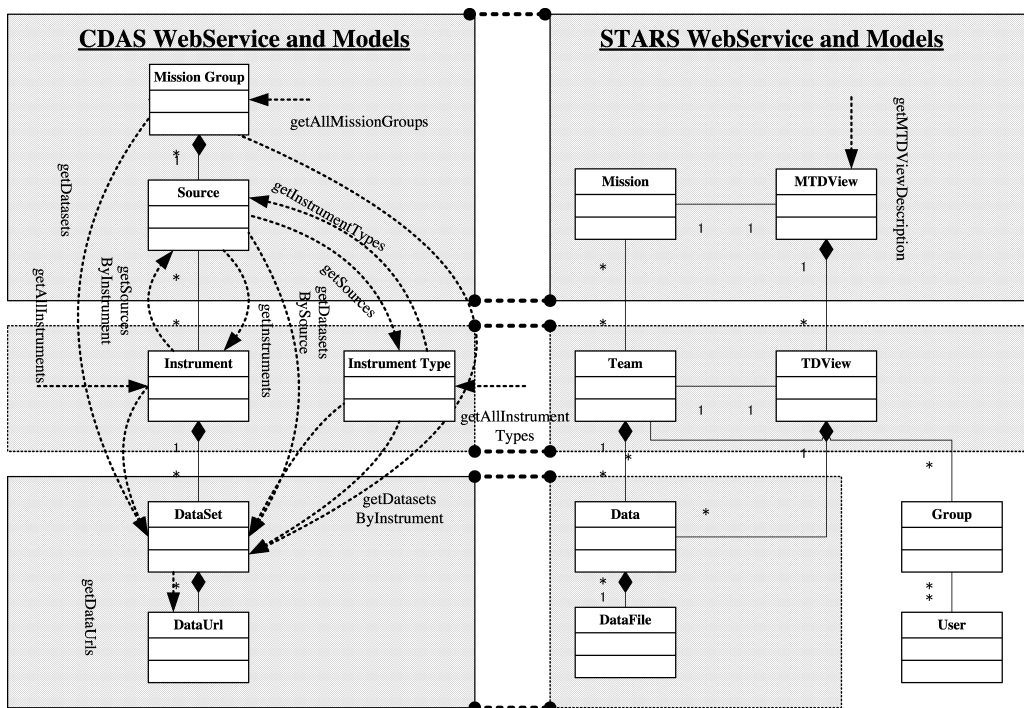


図 4 CDAS Web サービスと STARS Web サービスにおけるモデルとサービス

Fig. 4 The models of CDAS Web Service and STARS Web Service.

報も構成している。CDAS は「モノ」の視点からモデリングをしているのに対し、STARS は関与する「ヒト」の視点からモデリングをしているといえる。

以上のことから分かるように、CDAS と STARS では同一の対象を扱っているが、メタデータのモデリング手法が異なるため、STARS で CDAS のメタデータを収容するためには Mission Group の概念を収容するように STARS メタデータベースのテーブル拡張を行う必要がある。

さらに、メタデータを取得する Web サービスも前提となるデータモデルが違うために、取得メソッドの構成も異なる。図 4 に示すように、CDAS では一見メソッドが豊富ではあるが、実は各モデルの関連を導出するために十分な情報量を提供できていない。クラスの関連がどのように構成されているかが念頭にある研究者が自らの目的意識に沿ってデータを引き出せるように手順を Web サービス化しただけであり、構造・関連に対する前提知識を持たないエージェントやサービスとの連携は考慮されていない。一方、STARS は Mission/Team/Data の関連を記述する MTDView を手に入れることができる。ただし、これは STARS の構造に依存したものである。

このように同じ領域を扱っていてもモデリング手法

が違えば、モデルだけでなく提供するサービスも異なる。その結果、データの粒度やサービスの差異を吸収するために、前提知識となるモデルに関する情報をメタデータ収集エージェントにプログラミングしたり、メタデータを格納するデータベースのスキーマ設計を変更・拡張したりすることが必要となる。STP 分野では観測領域が多岐にわたるために、メタデータのモデリングやサービス定義も異なり、これらをすべて吸収、統合する試みは事実上、不可能である。

この問題を克服するには、統一スキーマを策定し採用することが考えられる。しかし、3.2 節、3.3 節で考察しているように、多岐にわたるメタデータを記述するスキーマおよび統一されたメタデータの提供サービスを策定するのは容易ではない。そこで、研究活動とメタデータ、サービスの仕様策定を並行して推進しつつ、メタデータの統一に至るまでの観測データモデルの差異を吸収していくための仕組みを構築していくことが要請される。

5. メタデータの形式の検討

以上のことから、(1) 現状のデータサイトを変更せずに簡単に実装できるものであり、(2) 太陽地球系物理の広範な概念をカバーできるメタ情報を定義でき、か

つ、(3) 自然科学観測技術の発展にあわせてメタデータの拡張もできるように、XML Schema 等によって構造が永続的に決定されるものではなく、RDF のように記述できる対象の拡大が後付でもできる柔軟性を持つメタデータの定義方法の採用、および (4) そのメタデータを扱うシステムの構築が必要であると判断した。また、膨大な観測データを迅速に配信できるように、軽量のコンテンツ情報記述用メタデータであることが不可欠である。

そこで軽量さと簡易な API で実現できることで爆発的な普及をみせている RSS に注目した。しかし、現状では RSS は RSS0.9¹⁵⁾、RSS0.91¹⁶⁾ および 0.92¹⁷⁾、RSS1.0¹⁸⁾、RSS2.0¹⁹⁾、Atom²⁰⁾ と乱立しており、バージョンの数字上では上位互換性がありながら、内実はまったく独自のものであり、非互換である。

3.3 節で論じたように、XML Schema によらない、柔軟に拡張可能なメタデータの記述方式としては Semantic Web で使われている RDF²¹⁾/OWL²²⁾ を利用することが考えられる。RSS1.0 (RDF Site Summary) は RSS-DEV によって策定され、RDF をベースとして拡張は XML 名前空間を利用し、かつ厳密な仕様が定義されている。そのため、この RSS1.0 に独自拡張部分を RDF/XML²³⁾ として埋め込むことで最低限のコンテンツ内容の記述に加えて RDF 記述による柔軟な拡張を実現できる。RSS1.0 は XML 文書であるため、RSS フィードをそのまま XML データベースに格納し、RDF ベースでの検索や情報の抽出といった二次的処理が容易となる。また、OSI トランスポート層に依存しないため、様々な手段で交換することが可能である。そのために、OAI-PMH のような REST アーキテクチャ準拠のシステムだけでなく、SOAP、XML-RPC、単純な HTTP や FTP 転送、SFTP といった様々な通信プロトコルでのメタデータの受け渡しが可能であり、既存のシステムへの適用が容易になる。そこで、観測データの配信情報のコンテナとして RSS1.0 を採用し、委細なメタデータを RDF/XML で記述することにした。

6. STARS-RSS の設計

6.1 全体設計

STARS に連動して使う RSS を STARS-RSS とし、位置付け、前述した RSS1.0 の RDF の拡張性を利用してモジュールを組み込むことでメタデータの拡張を行った。RSS 本体に対し、Dublin Core Module、Syndication Module と RSS の標準モジュールを必

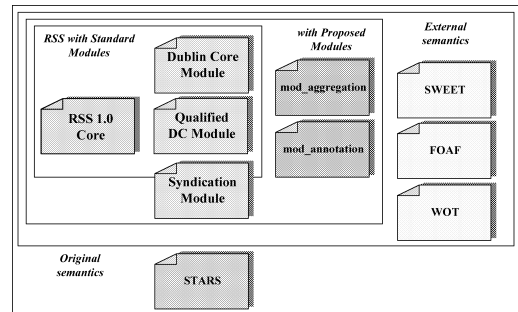


図 5 STARS-RSS のモジュール構成

Fig. 5 Modules in STARS-RSS.

須事項として取り込み、これらの標準モジュールでカバーできない部分を FOAF (Friend of a friend)²⁴⁾ や SWEET (Semantic Web for Earth and Environmental Terminology)²⁵⁾ といった既存のモジュールで定義を行った。さらに、これらが対応できない部分を本システムに関わる独自モジュールとして定義するアプローチをとった (図 5)。独自モジュールは OWL によりスキーマ設計を行った。

6.2 Dublin Core Module

DCMES (DC Metadata Element Set)¹⁰⁾ は基本的に書誌情報を前提に定義されたものである。作成者、責任者の定義は DC 精密化語彙を加えても数種類程度であり、観測プロジェクトの責任者、観測機器および採取されたデータの責任者、採取データからの加工を施した担当者、データのメタデータ作成者、データサイトの管理者といった多数の担当者が連携して 1 つのリソースに関わっているようなケースを記述できない。そこで、本システムでは通常の RSS の記述に倣って、dc:creator に RSS を作成した担当者の記述をするにとどめ、独自に定義したモジュールによって担当者を記述している。

6.3 Syndication Module

RSS を配信したシステムでは Ping²⁶⁾ や The Atom Publishing Protocol²⁷⁾ を利用して随時更新の通知を行う試みが行われているが、ファイアウォールやセキュリティ上の制限で双方向の通信ができないサーバも多数存在する。そういったサーバに対しては Pull 的に更新状況を確認する方法も提供することになるが、その更新頻度を Syndication Module で指定することで更新確認の照会トラフィックを抑制することができる。

6.4 RDF による拡張

6.4.1 SWEET

SWEET は、NASA が地球科学分野に関する用語を OWL で定義したものである。STARS では CDAWeb

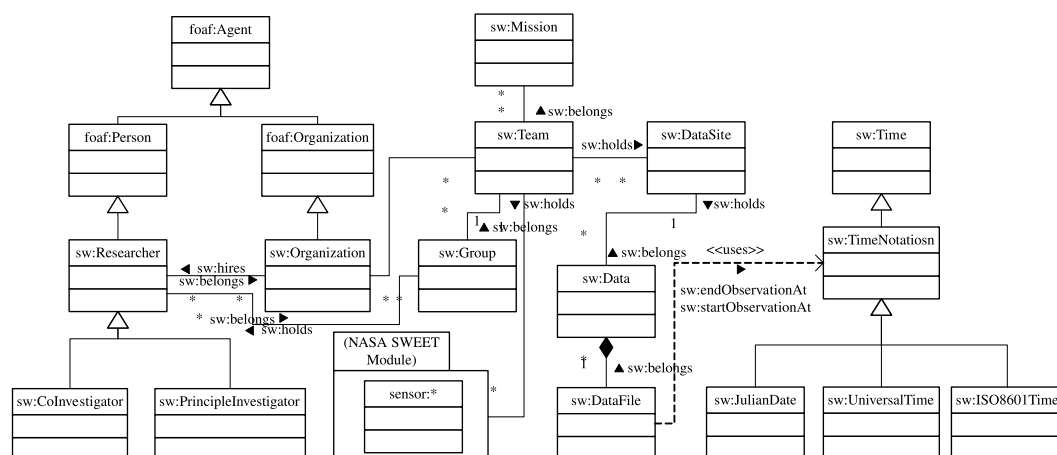


図 6 STARS モジュールの設計

Fig. 6 STARS module.

のメタデータも扱っており、ミッション構成の構造の違いを吸収する試みを行っている。そのため、NASA の Semantic Web への試みにあわせて、利用できるオントロジは本 RSS にも積極的に取り込んでいくことで将来の双方向の観測データのメタ情報の交換を容易にしていく布石を置く。今回は SWEET のモジュール群のうち sensor モジュールを利用して人工衛星に搭載するセンサの記述をし、CDAS における Instrument と STARS における Team のマッピングを試みた。

6.4.2 FOAF

STARS では多くの人がデータに関わっており、現状では人物を世界中で一意に識別するような URI として共通に認識され、IFP (Inverse Functional Property) として機能できるように枠組みは存在しない。しかし、人物を記述する試みで多くの支持を得ているリソース記述法を採用することで、他のメタデータへの互換性を極力確保する。FOAF が RDF に連動して人に関する情報を記述し、様々なメタデータの連係や人間関係のリンクを試みており、人物に関連するリソースとも連係することを考慮しているため、有用であると考えた。

そこで、本研究では、プロジェクトに關与している人物・組織を FOAF で記述することにした。foaf 名前空間で定義されているクラスを継承し、プロパティの拡張を行ったのが sw 名前空間における sw:Researcher (研究者クラス), sw:Organization (組織クラス)であり、かつ研究者クラスからプロジェクトリーダを表す PI, CoPI クラスを定義した(図 6)。研究者の所属情報は組織への帰属と観測班への帰属と様々な記述方法があるが、各クラス間に緩やかなプロパティ出現

回数の制約を加えることで記述に多様性をもたらすようにした。

6.5 独自モジュールの定義

本研究では、以上のように既存モジュールで適用できない場合、Schema Web²⁸⁾でOWLによって定義されたスキーマを探索し、それでも既存の語彙に見当たらないものを本プロジェクトに属する独自スキーマとして定義した(図6)。独自に語彙を定義した理由は、3.3節で論じたように、運用を通して、前方互換性を維持しつつ、メタデータの拡張を続けるためである。OWLでの定義を行ったのは、6.5.1項で考察されるように、時刻1つとっても、様々な表記や意味が変わるために、RDFでの単純な関係の記述では異なるデータモデル間のマッピングを行うための十分な情報を得ることができないためである。太陽地球系物理で扱われる語彙の関係を明示的に定義することで、異なる組織で定義されたメタデータに対する紐付けが可能になり、メタデータの仕様が変更になっても、それぞれのメタデータとの関連性の追跡ができる。現状のSTARSメタデータベースのリレーショナルテーブルとの親和性が高いように、Mission/Team/Dataクラスのスキーマ定義およびカージナリティをリレーション定義に沿って設計した。これを骨格として人間、組織、時刻情報といったモデリングのバリエーションをオントロジによる推論で収束させることを試みている。

たとえば、CDAS と STARS の両者間にまたがって、あるミッション、測定機器による特定期間の観測データの検索を試みる場合、Source (CDAS) と Mission (STARS)、Instrument (CDAS) と Team (STARS)、DataSet (CDAS) と Data (STARS) の関係は図 6

に示すように, `sw:holds/sw:belongs` でそれぞれ同じレベルでの対応付けがなされる. `DataSet` (CDAS) は `sw:JulianDate`, `DATA` (STARS) は `sw:ISO8601Time` によって観測時刻の表記方法が定義される. 観測開始・終了時間は `sw:startObservationAt`, `sw:endObservationAt` で記述される. そのため, 本方式に対応したエージェントでは, それぞれのデータ構造に関する前提知識や, データ変換処理を組み込むことなく, ミッションの検索で `Source/Mission`, 測定機器で `Instrument/Team`, そして観測時刻の種類と表記を適切に推論していき, 横断的な検索, 結果の取得が可能になる.

6.5.1 日付の解釈

XML Schema のデータタイプでは ISO8601 に準拠した日付の記述がなされている²⁹⁾. しかし, 太陽地球系分野では時刻は Julian Date で記述されることが多い. そこで XML Schema Datatype, `xsd:double` 型の精度を持つ時刻として定義した. また Julian Date 以外にも様々な時刻表記が考えられる. そのため, 時間を指し示す概念と時間の記述様式を分離した. `sw:Time` で STARS における時間の概念を表すクラスを, 時刻の表記クラスとして `TimeNotation` クラスを定義した. `TimeNotation` クラスのサブクラスで時刻の型について定義することで, 時刻情報の同定とその中の時刻の表記形式を判断できる. また, 観測データに関する時間は, 観測開始・終了時間, 公開開始時間, 更新時間と様々にあり, `DataFile` クラスからの様々なプロパティを定義することで表現している. 以上から時間の表記のパリエーションと, 様々な観測イベントに関連した時間を記述することが実現された.

6.5.2 公開鍵

メタデータおよびデータ配布にあたっては署名を行うことでファイルの認証, ダウンロード権限の認証を行う試みが行われている. しかし, データごとに公開鍵の公開と関連付けを行うのは煩雑である. そこで, 観測データのセット (`Data`) もしくは観測データサイト (`DataSite`) クラスのプロパティで `Wot Schema` (Web Of Trust RDF Ontology)³⁰⁾ を利用した公開鍵へのリンクと配布を試みている.

6.5.3 CDF ヘッダからの抽出

観測データは Level-2 フェーズにおいて生の観測データからメタデータを含む CDF (the Common Data Format)³¹⁾ 形式ファイルに変換され, 一般公開される. そこでこの CDF 形式の観測データからあらかじめ検索の対象となるようなメタデータを抽出し, RSS に組み込むことでメタデータの情報収集, 抽出の迅速

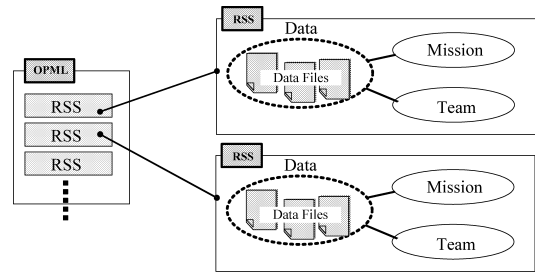


図 7 OPML によるデータサイトの RSS 集合の定義
Fig. 7 Description of RSS aggregation on data site with OPML.

化を図った.

6.6 OPML によるデータサイトの記述

OPML (Outline Processor Markup Language)³²⁾ は文書のアウトライン構造を記述するフォーマットであるが, RSS のアグリゲーションに利用されている. データサイトに設置する RSS は STARS の `Data` に対応し, 1 つの観測班に対して複数の所属する (図 7). 観測班とミッションは 1 対 1 の対応なので, データサイトにおいては観測班ごとに束ねるのが適度な粒度である. そこで観測班ごとの集約, もしくはデータサイトごとの集約に OPML を用いて記述した.

6.7 STARS-RSS のインスタンス生成

以上の設計をもとに生成した RSS ファイルの一例を可視化した結果を図 8 に示す. 図 8 の左上部分は RSS のコア部分である. RSS に対して図 8 右下に示すようにいくつかの名前空間を参照することでモジュールの組み込みを行い, メタデータの委細な記述を実現している. RSS の `item` 要素から各観測データへの情報にリンクしている. これら個々の `item` に対して RDF で `sw:isDataFile` によってデータファイルとしての定義を行い, 該当データファイルに関する観測情報 (観測開始・終了時間) の記述 (図 8 右上) を行っている.

RSS によってメタ情報が配信されている観測データを管理しているデータサイトについては RSS に `sw:describes` で観測班 (図 8 左下) およびホスティングしているデータサイト (図 8 右中央) に対する詳述を行っている. 観測班については, `sw:TeamAttribute`, `sw:TeamName` 等の `Team` そのものの情報のほか, その `Team` の所有しているデータ, 計測機器についても記述 (`sw:holdsData`, `sw:hasSensor`) している. これは, 初めてこのメタデータを受信した者が, この RSS に記述されている情報から, 2 章と 6 章とで論じているような STP 分野の関連するデータの構造を再構成できるための情報を提供するためである. すなわち, エージェントが RSS 経由でこの情報を受け取っ

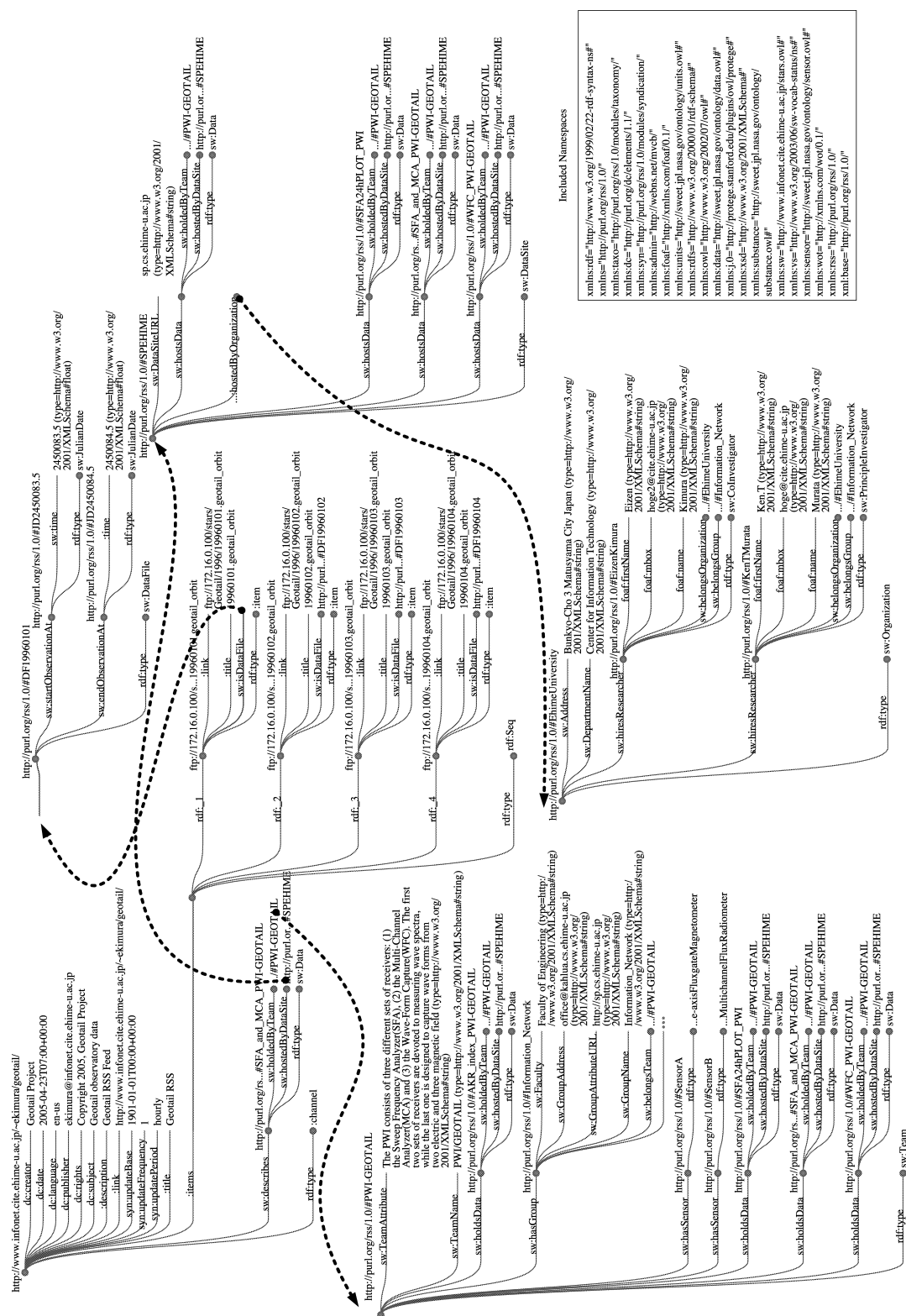


図 8 STARS-RSS の生成例
Fig.8 Sample instance of STARS-RSS.

て、データベースの構造を決定・構築し、格納ができるような粒度の情報提供を想定している。

そしてデータサイトを管理している組織および研究者について記述している(図8中央下)。6.4.2項で論じたように、現在は研究者を統一的に同定する手法はないため、擬似的IFPとしてFOAFを利用して研究者情報を記述している。

以上、一般に普及しているRSSリーダによって認識されるRSSの最小限の主要素から本プロジェクトによって運用されるメタデータ情報収集システムによって処理される広範な情報までが1つのRSSに共存することができた。このことは次章に述べるメタデータ配信と連携システムにみられるように様々なシステムの応用が実現できる素地となる。

7. 考 察

6章においてRSSを拡張することでRSSをメタデータのコンテナとして利用する試みについて述べた。現在、このRSSを利用した観測データのメタデータの配信と連携を実現するシステムの実装を進めている(図9)。

観測衛星の観測データは観測基地で非公開レベル(Level-0)のデータとして蓄積される(図9-(1))。蓄積されたデータはメタデータのない生の観測データであり、公開レベル(Level-2)においてCDFフォーマットのデータファイルに変換される際に基本的情報が付与される(図9-(2))。それに連動してCDFファ

イルに付与されたメタデータから本RSSに展開していく(図9-(3))。CDF化を試みるまではメタデータに関する情報は観測データとは別に配布され手動で登録していく必要があった(図3中央)が、Level-2ファイルのCDF変換のタイミングでRSSファイルにもメタデータを展開していくこと(図3右)でメタデータの自動配信を実現する。

RSSに展開された時点で速報的なデータの3次元可視化を行うサーバ等、更新情報をリアルタイムに要求するものに対してはRSS-Ping等を用いてRSSが更新されたことを通知する(図9-(4))。RSSを受信した可視化サーバは、RSS内の記述から最新の観測データをダウンロードし、可視化を行う。可視化が行われたものは3次元モデル閲覧用のファイルとして保存され、研究者の閲覧に供される。

RSS情報収集エージェントは(図9-(3))で生成されたRSS内のSyndicationモジュールによって記述された更新の頻度情報を獲得し、その頻度に応じて定期的に更新の照会を行い、更新がなされていればRSSのダウンロードを行う(図9-(6))。ダウンロードされたRSSはRDF文書としてRDFパーサによってメタデータの抽出が行われる。STARSメタデータベースへの登録に必要なメタデータを抽出した後、STARS-WS Webサービスを使って登録を行う(図9-(7))。

これらの一連の作業がミッション下それぞれのデータサイトで行われた場合、それぞれの観測データがRSSに記述されている更新頻度でSTARSメタデー

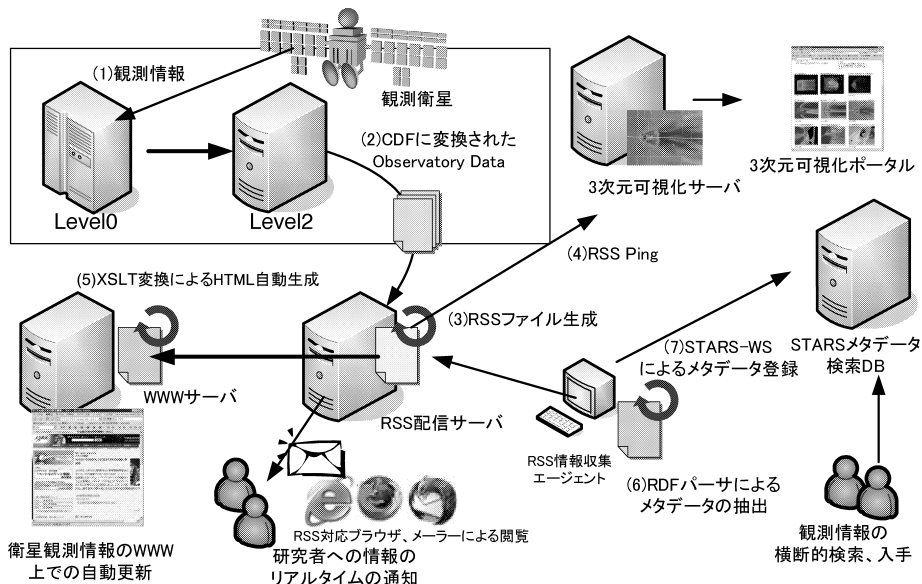


Fig. 9 The distribution and cooperation of the observatory data with STARS-RSS.

タに集積されていくので、研究者は比較的少ないタイムラグで観測データを横断的検索、入手することができるようになる。

また、これらシステム間のメタデータの配信、処理連係のほかに、本来の RSS の用途に沿った利用形態として、RSS 対応ブラウザ、メーラを利用している研究者への RSS フィードによる更新情報通知、および XSLT 変換等のプロセスを経て Web サイトへの更新情報の通知の利用があげられる（図 9-(5)）。

7.1 RSS/RDF のデータベースへの格納

現在はメタデータベースの根幹を RDBMS が構成しているため、今は RSS/RDF データからメタデータを抽出し、データベースに登録する作業を行っている。そのため、メタデータも従来の RDBMS のテーブル設計（図 2、図 6）に則って設計している。4 章で述べたようにデータモデルの差異を RDBMS に吸収させるために RDBMS のテーブルの拡張をするといった問題には対応できていない。しかし、メタデータの記述を RSS/RDF に移行させたことで、リポジトリの実装を RDBMS から XML データベースに移行することが可能になった。このことにより、将来においては、データモデルの差違の収束のためにデータベースやシステムの変更や開発に労力を割くことなく、セマンティクス情報を利用することでデータの横断検索、統合することが容易になる。現在は RDF リポジトリ実装のパフォーマンスの評価と従来のメタデータ・データベースにおける SQL および Web サービスから RDQL といった RDF のクエリ手法への橋渡しを検討中である。

航空宇宙圏の観測データにオントロジ情報を加味し、様々なデータのモデルのギャップを超えて統合する試みが報告されているが⁹³³⁾ 本システムのように RSS/RDF を利用することでイベントが発生したときにタイムリーな形でオントロジ情報を加味して伝播することは検討されていない。従来の手法では連携をするサービス間の最大公約数たる情報のみに限られるため、連携を続けていく過程で情報の粒度が低下していくか、もしくは当初から最大公約数を最大限にもっていくように巨大なスキーマの策定を試みて膠着状態に陥り、柔軟な研究の連携を困難にしていた。RSS/RDF は RDF による拡張ができるために、伝播の過程で付加的サービスを提供するものが独自に情報を挿入することが可能である。そのため、RSS の更新の頻度で情報の伝播および付加が可能になり、多様な連携の基盤として機能することが期待される。

8. おわりに

本研究では多様性のあるシステムとデータの差異を収束するためにコンテンツ配信に利用されている RSS1.0 を RDF/XML で記述したメタデータのコンテナとして採用した。その結果、メタデータの策定には今後も議論が続けられることになるが、それと並行して研究活動のために必要な観測データの情報配信を始めることが可能になった。かつメタデータの仕様拡張にともなう非互換性の問題を軽減することで、研究活動と研究活動支援技術の開発、現場への導入のラウンドトリップを短縮することが期待される。また RSS のデータの加工、取り出しの容易性によってメタデータの再利用による可視化や分析の連携への応用にも期待がかかる。現状では太陽地球系データの更新情報配信の範囲の議論にとどまっているが、今後の運用を通しての知見の積み重ねによって、メタデータの 2 次的利用の方法論の開発・効果の検証や、本手法の試みをさらに広い科学コミュニティでの汎用的なデザインパターンとして発展させていきたいと考えている。本研究の一部は、文部科学省科学研究費補助金（基盤 C（2）16500321）の支援を受けた。

参 考 文 献

- 1) eGY: The Electronic Geophysical Year (2005).
<http://www.egy.org/>
- 2) Thompson, B.: *Bridging the Digital Divide: eGY and Virtual Observatories*, Spring AGU Meeting, New Orleans, Louisiana (2005).
- 3) 大石雅寿, 白崎裕治, 田中昌宏ほか: パーチャル天文台を実現する分散データベース・計算資源の国際連携機構, DEWS2005 (2005).
<http://www.ieice.org/iss/de/DEWS/DEWS2005/procs/papers/4A-o3.pdf>
- 4) 本田敏志, 大石雅寿, 白崎裕治, 田中昌宏, 川野元聡, 水本好彦: 天文学連携データベースシステム（ヴァーチャル天文台）の開発 (2005).
<http://www.ieice.org/iss/de/DEWS/DEWS2005/procs/papers/1D-d2.pdf>
- 5) 村田健史: 国際太陽地球系物理観測の広域分散データベースシステム, 電子情報通信学会論文誌 (B), Vol.J86-B, No.7, pp.1331-1343 (2003).
- 6) Space Physics Archive Search and Exchange (SPASE) Consortium: Space Physics Archive Search and Extract (2005).
<http://www.igpp.ucla.edu/spase/>
- 7) 村田健史, 岡田雅樹, 阿部文雄, 荒木 徹, 松本 紘: 太陽地球系物理観測の分散メタデータベースの設計と評価, 情報処理学会論文誌, Vol.43, No.SIG 12(TOM16), pp.115-130 (2002).

- 8) 村田健史, スルディヤナ・ガーニ, 橋本弘蔵, 松本紘, 荻野龍樹: パーチャル地球磁気圏システムの提案, 電子情報通信学会論文誌 (B), Vol.J87-B, No.2, pp.309-313 (2004).
- 9) Plante, R.: An Evaluation of the Open Archives Initiative for VO Registries (2003). <http://www.us-vo.org/pubs/files/evaloai.html>
- 10) The Dublin Core Metadata Initiative: DCMI, Dublin Core Metadata Element Set, Version 1.1 (2004). <http://dublincore.org/documents/dces/>
- 11) de Sompel, H.V., et al.: Resource Harvesting within the OAI-PMH Framework, *D-Lib Magazine*, Vol.10, No.12 (2004).
- 12) Fielding, R.T and Taylor, R.N.: Principled design of the modern web architecture, *ACM Trans. Internet Technology*, Vol.2, No.2, pp.115-150 (2002).
- 13) Powers, S.: *Practical RDF*, O'Reilly (2003).
- 14) NASA: CDAS Web Services (2005). <http://cdaweb.gsfc.nasa.gov/WebServices/>
- 15) Libby, D.: RDF Site Summary (RSS) 0.9 official DTD (1999). <http://my.netscape.com/publish/formats/rss-0.9.dtd>
- 16) Libby, D.: RSS 0.91 Spec, revision 3 (1999). <http://my.netscape.com/publish/formats/rss-spec-0.91.html>
- 17) Winer, D.: RSS 0.92 (2005). <http://backend.userland.com/rss092>
- 18) Beget-Dov, G., et al.: RDF Site Summary (RSS) 1.0 (2001). <http://web.resource.org/rss/1.0/spec>
- 19) Center, B.: RSS 2.0 Specification (2002). <http://blogs.law.harvard.edu/tech/rss>
- 20) APE (Atom/Pie/Echo) Working Group: The Atom Syndication Format 0.3 (2003). <http://www.mnnot.net/drafts/draft-nottingham-atom-format-02.txt>
- 21) World Wide Web Consortium (W3C): Resource Description Framework (RDF) (2004). <http://www.w3.org/RDF/>
- 22) World Wide Web Consortium (W3C): Web Ontology Language (OWL) (2004). <http://www.w3.org/2004/OWL/>
- 23) Beckett, D., ed.: RDF/XML Syntax Specification (2004). <http://www.w3.org/TR/rdf-syntax-grammar/>
- 24) Brickley, D. and Miller, L.: FOAF Vocabulary Specification (2005). <http://rdfweb.org/foaf/>
- 25) NASA Earth Science Technology Office, Advanced Information Systems Technologies: Semantic Web for Earth and Environmental Terminology (SWEET) (2005). <http://sweet.jpl.nasa.gov/>
- 26) Six Apart: TrackBack Technical Specification (2004). http://www.sixapart.com/pronet/docs/trackback_spec
- 27) Gregorio, J., et al.: The Atom Publishing Protocol (2005). <http://www.ietf.org/internet-drafts/draft-ietf-atompub-protocol-04.txt>
- 28) Lindesay, V.: SchemaWeb (2005). <http://www.schemaweb.info/>
- 29) World Wide Web Consortium (W3C): XML Schema Part 2: Datatypes Second Edition (2004). <http://www.w3.org/TR/xmlschema-2/datatypes.html>
- 30) Brickley, D.: Web Of Trust RDF Ontology. <http://xmlns.com/wot/0.1/>
- 31) NASA/Goddard Space Flight Center: The Common Data Format (CDF) (2005). <http://cdf.gsfc.nasa.gov/>
- 32) UserLand Software: Outline Processor Markup Language (2000). <http://www.opml.org/spec>
- 33) Kogut, P. and Heflin, J.: Semantic Web Technologies for Aerospace, IEEE Aerospace Conference (2003).

(平成 17 年 7 月 12 日受付)

(平成 18 年 2 月 1 日採録)



木村 映善 (正会員)

昭和 47 年生。平成 15 年愛媛大学医学部附属病院医療情報部助手。平成 16 年愛媛大学総合情報メディアセンター助手。医療情報学, 福祉情報学, 宇宙情報工学の研究に従事。博士 (医学)。MEDINFO2001 若手研究者賞受賞。AMIA, ACM, IEEE Computer Society, 電子情報通信学会, JAMIA 各会員。



村田 健史 (正会員)

昭和 38 年生。平成 7 年愛媛大学工学部助手。平成 8 年愛媛大学工学部講師。平成 15 年愛媛大学工学部助教授。平成 15 年愛媛大学総合情報メディアセンター助教授。平成 17 年宇宙航空研究機構 (JAXA) 宇宙科学研究本部客員助教授。平成 17 年名古屋大学太陽地球環境研究所客員助教授。福祉情報学, 宇宙情報工学の研究に従事。博士 (工学)。地球電磁気・地球惑星圏学会, 電子情報通信学会, American Geophysical Union 学会各会員。