

超高層物理学のための分野横断型メタデータ・データベースの構築

梅村 宜生[†] 小山 幸伸 堀 智昭 阿部 修司 林 寛生

新堀 淳樹 田中 良昌 上野 悟 米田 瑞生 金田 直樹 元場 哲郎

[†]名古屋大学太陽地球環境研究所 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]umemura@stelab.nagoya-u.ac.jp

あらまし 超高層物理学分野における地上観測データは、太陽活動、地磁気、風速、オーロラなど多種に渡り、観測を行ったそれぞれの研究機関で分散管理されている。「超高層大気長期変動の全球地上ネットワーク観測・研究」(略称: IUGONET)プロジェクトでは、分散管理されている地上観測データへのアクセシビリティを向上させるために、地上観測データから抽出したメタデータを集約するメタデータ・データベースを構築している。また、メタデータと地上観測データ間の接続状況を確認するための外部 URL 接続状況検査プログラムを、本年度に新たに導入した。本論文では、メタデータ・データベースの概要と現状に加え、外部 URL 接続状況検査プログラムによる接続状況の確認方法と、その導入によりデッドリンクを解消した結果について述べる。

キーワード データベース, メタデータ, IUGONET, リポジトリ, 分野横断, デッドリンク, DSpace

Development of Cross-Field Metadata Database for Upper Atmosphere

Norio UMEMURA[†], Yukinobu KOYAMA, Tomoaki HORI, Shuji ABE, Hiroo HAYASHI,

Atsuki SHINBORI, Yoshimasa TANAKA, Satoru UeNo, Mizuki YONEDA, Naoki KANEDA

and Tetsuo MOTOKA

[†] Solar-Terrestrial Environment Laboratory, Nagoya University.

E-mail: [†]umemura@stelab.nagoya-u.ac.jp

1. はじめに

近年、太陽活動の活発化に伴い磁気嵐が発生し、人工衛星の電波障害を引き起こすなど、太陽と地球を取り巻く全球スケールで生じる物理現象とその影響に関する研究が注目を集めている。これらの現象は、太陽活動の活発化に伴い放出される電磁波や大量のプラズマと、我々が生活する大気圏の上層に位置する磁気圏から電離圏および熱圏に至る超高層大気が、複雑かつ相互的に作用するためと考えられている[1]。

上記の様に、全球スケールで生じる物理現象の相互的な変動を理解するには、太陽活動、地磁気、風速、オーロラなどの、多種の観測データを組み合わせる必要があると考えられている。しかしながら、日本における超高層物理学分野のこれまでの研究では、地上観測データは、観測を行ったそれぞれの研究機関から個別に公開されており、特定の分野の研究者にのみ、その存在が知られ、利用されることが殆どであった。また、アナログデータをも含む一部の地上観測データについては、公開されないまま記録メディアの中に埋も

れてしまう場合もあった。このことから、誰もが自由に多種多様な地上観測データを取得すること、そして、全球スケールで生じる物理現象を総合的に理解することは困難であった。

著者らは、平成 21 年度より 6 ヶ年計画として発足した「超高層大気長期変動の全球地上ネットワーク観測・研究」(IUGONET: Inter-university Upper atmosphere Global Observation NETwork)において、超高層物理学分野における地上観測データのメタデータ・データベース(メタデータ DB)を構築している[2][3]。IUGONET プロジェクトには、国立極地研究所研究教育系宇宙空間研究グループ、東北大学大学院理学研究科地球物理学専攻太陽惑星空間物理学講座および惑星プラズマ・大気研究センター、名古屋大学太陽地球環境研究所、京都大学生存圏研究所、京都大学理学研究科附属地磁気世界資料解析センター、京都大学理学研究科附属天文台、および、九州大学宇宙環境研究センターの、5 機関 7 組織が参画している。参画する各機関とその他の協力機関が連携し、それぞれの機関で個別に公開している地上観測データからメタデータを抽出してメタデ

ータ DB に集約することにより、分野を問わず、多種の地上観測データを横断的に検索・取得することができる仕組みである。また、IUGONET プロジェクトは、メタデータ DB を通じて得られた多種多様な地上観測データを組み合わせ、超高層大気領域で起こる複雑かつ相互的な物理変動の解明、特にアナログデータをも活用した超高層大気の長期変動についての研究も推進している。本論文では、IUGONET プロジェクトにおけるメタデータ DB の概要と現状について報告するとともに、本年度に新たに導入した、メタデータと各研究機関が公開している地上観測データが正しく接続されているかについて検査する外部 URL 接続状況検査プログラムと、その導入により接続異常を解消した結果について述べる。

2. メタデータ DB システムの概要

2.1. メタデータ・データフォーマット

IUGONET プロジェクトは、堀ら[4]の検討により、地上観測データのメタデータのベースとして、SPASE(Space Physics Archive Search and Extract)メタデータフォーマット[5]を採用した。SPASE メタデータフォーマットは、アメリカ航空宇宙局(NASA)、米国海洋大気庁(NOAA)、宇宙航空研究開発機構(JAXA)などの太陽地球系物理学の研究者で構成するコンソーシアムにより、同分野における観測データに関する記述を対象として、策定・管理されている。SPASE メタデータフォーマットは、SPASE データモデルに基づき、観測データファイル、観測機関、観測装置、観測データ担当者などの 12 種類のリソースタイプに分類される。実際の記述には XML を使い、異なる 2 種のリソースタイプ、例えば観測データファイルと観測機器などであっても、メタデータを一意に識別するリソース ID を参照することにより、関連付けることが可能である。さらに、IUGONET プロジェクトは、SPASE メタデータフォーマットのバージョン 2.0.0 に対し、各研究機関が保有する観測データ記述に適したものとなるように、(1)観測データ・プロットの開始時刻と終了時刻、観測範囲、(2)太陽観測に伴う固有座標系の要素を拡張したものを、IUGONET 共通メタデータフォーマットとして策定し、実際のメタデータ作成に用いている[4][6]。また、フォーマットの改良に際しては、SPASE コンソーシアムと協力して進めており、IUGONET プロジェクトで新たに加えたパラメータは、SPASE メタデータフォーマットのバージョン 2.2.0 として正式に採用された[4][6]。

2.2. メタデータ DB のミドルウェア

IUGONET プロジェクトは、小山ら[6][7]の検討によ

り、メタデータ DB のミドルウェアとして、DSpace を採用した。DSpace は、デジタルコンテンツに関する情報を記述したメタデータを、リポジトリとして管理し、インターネットに公開するために開発されている、オープンソースのリポジトリ・ソフトウェアである。DSpace は、登録されたメタデータを一意の URL で識別するハンドルシステム[8]を採用している。そのため、図書館の蔵書管理や、論文リポジトリデータベースなど、多くの機関で採用されている[9]。また、DSpace の標準仕様は、メタデータのベースの XML フォーマットに、題目、制作者、テーマなどのデジタルコンテンツに関する基本的な語彙群である Dublin Core[10]を適用しており、ブラウザ画面においても、これらのパラメータを検索キーとして実装している。これに対し、IUGONET プロジェクトは、超高層物理学分野のデータ特有のパラメータである、観測データ・プロットの開始時刻と終了時刻、観測範囲を検索キーとして利用するために、DSpace の標準仕様に対し、インターフェース層、アプリケーション層での改良を加えている。

2.3. メタデータの登録と公開の方法

メタデータの公開にあたっては、(1)メタデータを提供するメタデータ・プロバイダーが、地上観測データの情報を、XML 形式のメタデータとして作成し、(2)ssh プロトコルを用いてインターネット経由で、メタデータを、バージョン管理ソフトウェアである Git のリポジトリに登録し、(3)メタデータ DB において、メタデータを Git リポジトリから取り出し、スクリプトによる構文チェックを施し、(4)同じくメタデータ DB において、DSpace のデータストア層に登録する。なお、工程(3)、(4)は定期起動される。以上の工程を経た後に、利用者による、ウェブブラウザを使ったメタデータの

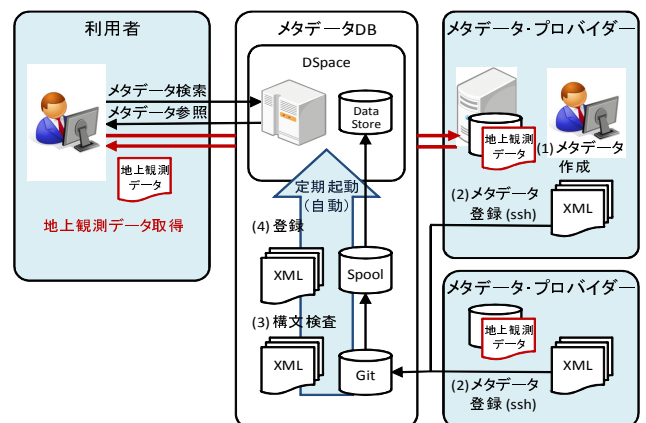


図 1. メタデータ・プロバイダーが作成・登録したメタデータは、メタデータ DB において定期的に構文チェックと DSpace のデータストア層への登録がされ、利用者による検索と参照が可能な状態になる。

検索と参照が可能になる。利用者は、メタデータ DB に対して検索を行い、検索結果として受け取ったメタデータに記載されているリンク URL から各研究機関の公開データサーバへ HTTP または HTTPS プロトコルにて接続することにより、地上観測データを取得することができる。また、メタデータの参照にはユーザ登録は不要であり、メタデータ・プロバイダーのみならず、誰でもメタデータを利用することができる。

2.4. 解析ソフトウェア UDAS からのデータ検索

IUGONET プロジェクトは、メタデータ DB の構築に加え、多種のデータを組み合わせた研究を推進するために、可視化・解析ソフトウェア UDAS(IUGONET Data Analysis Software)[11]も開発している。UDAS は、IDL(Interactive Data Language)で書かれた THEMIS 衛星ミッションの統合データ解析ツール TDAS[12]のプラグインソフトであり、各研究機関の地上観測データのフォーマット差異を吸収し、可視化と解析を容易にする。UDAS は、現在、開発中ではあるが、メタデータ DB で管理されている地上観測データの所在情報を取得するために、メタデータ DB と通信する。具体的には、観測時間範囲、観測装置、観測地、観測データ種別などを条件に設定し、メタデータ DB に対して検索要求を出す。この要求に対し、メタデータ DB は、OpenSearch を使い、メタデータの検索を受け付け、該当するメタデータを UDAS に応答する。このように、メタデータ DB は、ウェブブラウザのみならず、可視化・解析ソフトウェア UDAS からの検索要求にも対応している。

3. メタデータ DB の現状

3.1. メタデータの登録数と性能

河野ら[13]は、平成 23 年 3 月に、メタデータ DB に登録されるメタデータの総数は、将来的に 100 万件に到達すると報告したが、本稿の執筆時点(平成 23 年 12 月時点)において、既に 130 万件を超えている。この登録件数は、平成 23 年 7 月に、スペイン高等科学研究院(CISI)が発表した世界リポジトリ総合ランキング[14]にて、世界第 8 位、国内第 1 位を獲得した、京都大学学術情報リポジトリ・紅[15]の DSpace への登録件数の 8 万件を、大幅に上回っている。太陽地球系物理学の分野における典型的な地上観測データは 1 日あたり 1 ファイル作成され、それに 1 対 1 で関連付くメタデータが日々作成されるため、メタデータが急激に増加した。現在、メタデータ DB は DSpace のバージョン 1.7.0 を使っている。DSpace のバージョン 1.7.0 における、メタデータの単位数あたりの登録速度は、一定で維持されている[13]。

3.2. アナログデータのデジタル化

IUGONET プロジェクトは、今後のタスクのひとつとして、紙媒体やマイクロフィルムなどで保存されているアナログ形式の観測データをデジタル化して公開する作業を計画している。アナログデータをスキャンしてデジタル形式の画像ファイルを作成し、メタデータを付与することにより、観測データへのアクセスを容易にする予定である。現時点では、試験的な段階ではあるが、例えば、ミュンヘンに設置されていた、光学記録式磁場変化計で測定された地磁気変動の印画紙データのメタデータを登録している。今後も、各機関が保有する多くのアナログデータをデジタル化し、メタデータを登録していく予定である。

4. メタデータと地上観測データ間の接続状況の確認と対処

IUGONET のメタデータ DB は、地上観測データは保存せず、メタデータのみ保存している。そのため、仮に、

- 地上観測機器の不具合やメンテナンスにより、データの全部分、あるいは、一部分を欠損扱いとした場合、
- もしくは、各研究機関のデータサーバにおいて、ディレクトリの構成、あるいは、インターネットドメインが変更された場合、

メタデータ DB に登録したメタデータと、各研究機関が公開している地上観測データ間で、正しく接続されない事象、いわゆるデッドリンクが発生する。デッドリンクが発生する可能性は、時間経過ごとに、また、メタデータの登録数の増加とともに高まる。IUGONET プロジェクトは、デッドリンクの増加は、システムの運用や、データ解析の推進に影響すると判断し、メタデータと地上観測データが正しく接続されているかについて検査する外部 URL 接続状況検査プログラムを、本年度に新たに導入した。

4.1. 外部 URL 接続状況検査プログラムの概要

メタデータから各研究機関の地上観測データに接続するプロトコルは、HTTP または HTTPS である。接続先の外部のホスト数は、インターネット大手の検索サイトでは極めて多いのに対し、IUGONET プロジェクトのメタデータ DB では主に各研究機関に限られている。また、前者は、外部のホストが保有する全てのデータを管理していないのに対し、後者の IUGONET プロジェクトのメタデータ DB は、各研究機関が保有する地上観測データの全ての情報を管理している。こ

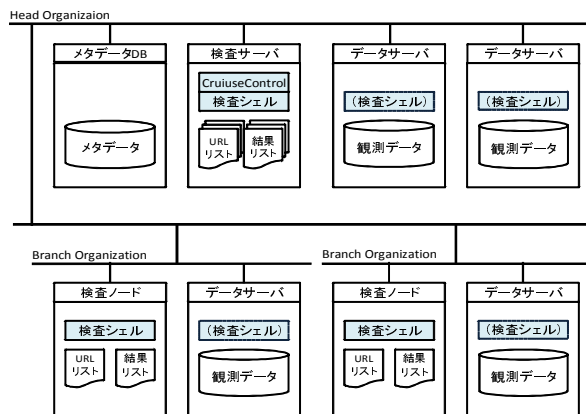


図 2. 代表機関には検査サーバを、それ以外の機関には検査ノードを設置し、各機関にて実施した接続状況の検査結果を代表機関の検査サーバに集めて管理する、集約・分散型を採用した。

の理由から、IUGONET プロジェクトは、1 台の検査サーバを使い、全ての観測データの接続状況について確認するよりも、まず、各研究機関にて、保有する観測データの接続状況について確認し、その結果を 1 台のサーバに集めて管理する、集約・分散型が好ましいと判断した。また、この構成は、1 台の代表サーバから HTTP(S) プロトコルを使った大量の接続をしなくてよい利点がある。外部 URL 接続状況検査プログラムの物理構成と論理構成を、図 2 に示す。図 2 のとおり、代表機関には結果を管理・集約するための検査サーバを、その他の各研究機関には検査を実施するための検査ノードを設置し、検査サーバを拠点として、各研究機関の検査ノードと接続した。検査プログラムは、検査を実行するシェルスクリプトと、検査対象の URL を記述した URL リストファイル、検査結果を記述する結果リストファイルにより成り立つ。Java などの、特定のコンパイラや動作環境を必要とするなどの制約を受けず、各研究機関で標準的に用いられている UNIX プラットフォームに容易に導入することができるよう、シェルスクリプトを採用した。また、シェルスクリプトは、可搬性の利点から、データサーバにおいても実行可能である(図 2 における点線枠)。また、地上観測データに対して HTTP(S) プロトコルを使って接続する方法と test コマンドを使って接続する方法の 2 つをシェルスクリプトに実装し、どちらかを選択する仕様にした。各研究機関において、仮に、HTTP(S) プロトコルを使うことによって発生するデータサーバの高負荷が懸念される場合には、シェルスクリプトの test コマンドによる、負荷の低減化を図る必要があるためである。また、結果を可視化するツールとして、CruiseControl を採用した。CruiseControl は、DSpace と同様、Java プラ

ットフォームで動作するツールであり、IUGONET プロジェクトでこれまで DSpace をカスタマイズした技術的経験や、JUnit 特有の関数を記述することにより結果を容易に可視化できるなど、IUGONET プロジェクトにおけるシステムの運用に容易に導入できるためである。

4.2. 検査プログラムの動作フロー

検査プログラムのシーケンスを図 3 に示す。まず、代表機関のメタデータ DB において、Java のプログラムを使い、データストア層に対して SQL を発行し、地上観測データの所在を示す URL を抽出する。このとき、各研究機関のインターネットドメインごとに URL をグループ化する。また、重複する URL は DISTINCT 句により、1 つとみなす。抽出した URL は、研究機関別の URL リストファイルに出力する。どの機関にも属さないインターネットドメインの URL は、1 つの URL リストファイルに集約する。また、URL リストファイルは、Git リポジトリに格納する。

次に、代表機関に設置した検査サーバと各研究機関に設置した検査ノードにおいて、シェルスクリプトを実行する。シェルスクリプトは、まず、Git の pull 命令により、メタデータ DB から URL リストを取得する。次に、予め設定した HTTP(S) プロトコルによる接続、またはシェルスクリプトの test コマンドによる接続のどちらかにて、URL リストファイルに記載された地上観測データへ接続する。接続できた場合は正常を示すコードを、接続できない、または、読み取り異常が発生した場合は異常を示すコード(HTTP/HTTPS プロトコルで接続する場合は HTTP ステータスコード)を、URL とともに、結果リストファイルに出力する。全ての URL に対して検査を終えると、最後に、再び Git を利用し、結果リストファイルをメタデータ DB の Git リポジトリに push する。

次に、代表機関のメタデータ DB において、Java プログラムを使い、各研究機関が push した結果リストファイルの内容を読み取る。結果リストファイルに正常を示すコードが示された場合は、メタデータと地上観測データ間の接続状況は正常と判定する。一方、異常を示すコードが示された場合は、接続状況は異常、すなわちデッドリンクが発生していると判定し、データストア層に対して SQL を発行し、その URL を含む全てのメタデータのリソース ID とファイル名を抽出する。抽出したメタデータの情報は、研究機関別の異常メタデータリストファイルに出力する。

最後に、異常メタデータリストファイルを、Git を利用して検査サーバに移し、CruiseControl にて、検査結果を可視化する。また、接続異常が発生している場

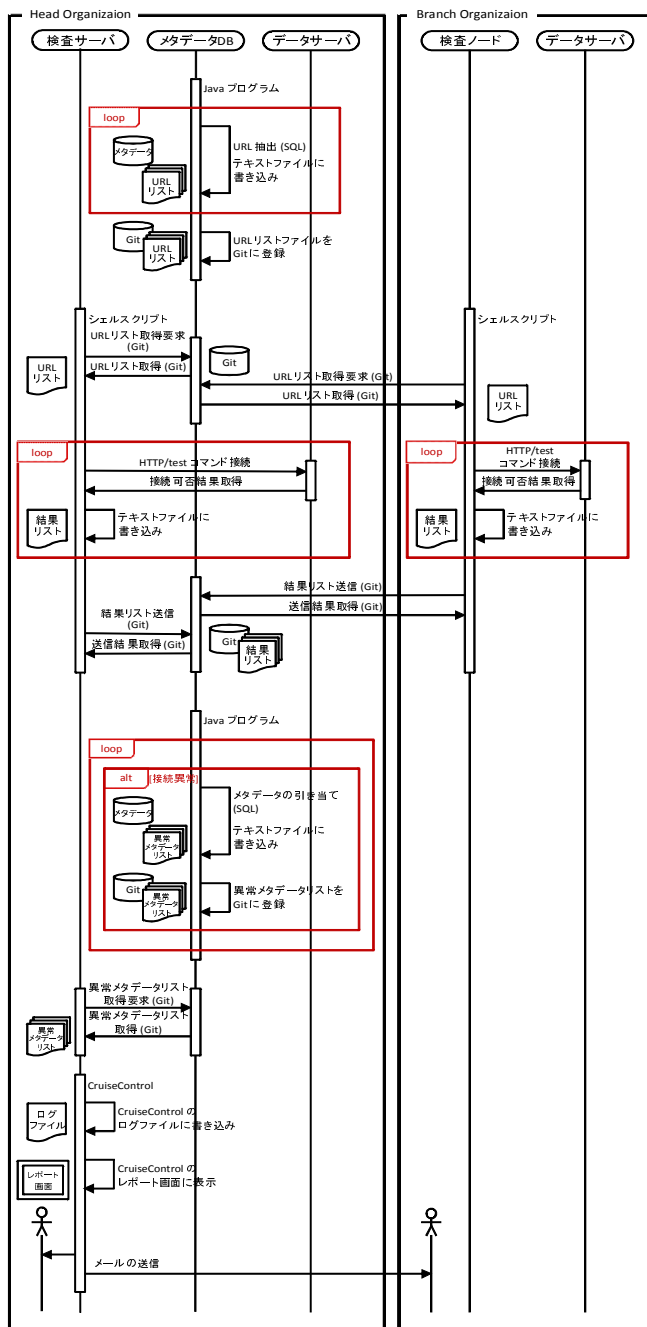


図 3. メタデータと地上観測データ間のデッドリンクを検出するためのシーケンス。

合は、対象の研究機関のメタデータ管理者に対して、異常メタデータリストファイルを電子メールに添付して送付し、接続関係の修正を促す。

4.3. 検査結果を踏まえた修正作業

メタデータと地上観測データ間に接続異常が確認された場合、メタデータ・プロバイダーは、

- メタデータを修正または削除し、Git リポジトリ

に再登録する、

- もしくは、各研究機関で公開している地上観測データの保存先を、メタデータが示す場所に変更する、

といういずれかの方法で接続異常を修正している。なお、IUGONET プロジェクトは、Git リポジトリを使ってメタデータの更新状況を管理しているため、DSpace のデータストア層を直接編集して接続関係を修正するという手法は採用していない。

IUGONET プロジェクトは、平成 23 年 12 月初旬に外部 URL 接続状況検査プログラムを導入し、これまでに登録したメタデータと地上観測データ間の全ての接続状況について検査した。接続異常数の推移を、表 1 に示す。表 1 のとおり、初回検査時の接続異常数は 1 万件以上であり、その発生確率は約 2.4 パーセントであった。これに対し、上述した接続異常解消のいずれかの対応を施した結果、本稿執筆時点(平成 23 年 12 月末)では、接続異常の発生確率は約 0.01 パーセントであった。今回の検査により、メタデータと地上観測データ間の接続状況について確認し、異常が発生していた接続関係を修正した。

表 1. 検査プログラムの導入による、メタデータと地上観測データ間の接続異常数の推移。

	初回検査時	修正後
接続異常数	17,689	102
ユニークな外部ヘリンク URL 数	750,415	750,415
	(2.3572%)	(0.0136%)

5. まとめ

IUGONET プロジェクトは、超高層大気領域で複雑かつ相互的に生じる物理現象を総合的に捉えるために、5 機関 7 組織が参画し、分野横断型メタデータ・データベース(メタデータ DB)を構築している。メタデータ DB は、各研究機関が保有する地上観測データのメタデータを集約している。カスタマイズした DSpace をメタデータ DB のミドルウェアとして、SPASE メタデータフォーマットに改良を加えた IUGONET 共通メタデータフォーマットをメタデータのベースとして用いている。メタデータ DB における現在のメタデータ登録数は 130 万件を超えている。メタデータと各研究機関の地上観測データ間の接続状況を検査するために、外部 URL 接続状況検査プログラムを新たに導入した。これまでに登録したメタデータと観測データ間の接続状況を検査し、異常が発生していた接続関係を修正した。このことから、外部 URL 接続状況検査プログラム

の導入は有用であることを確認した。

将来的には、外部 URL 接続状況検査プログラムを、IUGONET プロジェクト参画機関以外でも実行し、外部のネットワーク環境からも、地上観測データに正しく接続できることを確認する予定である。

6. 謝辞

大学間連携プロジェクト「超高層大気長期変動の全球地上ネットワーク観測・研究」は、文部科学省特別教育研究経費(研究推進)[平成 21 年度]および特別経費(プロジェクト分)[平成 22 年度から]の交付を受けて、平成 21 年度より 6 ヶ年計画で実施している事業である。

参 考 文 献

- [1] 柴田一成，上出洋介，総説 宇宙天気，京都大学学術出版会，2011.
- [2] 林寛生，小山幸伸，堀智昭，田中良昌，新堀淳樹，鍵谷将人，阿部修司，河野貴久，吉田大紀，上野悟，金田直樹，米田瑞生，田所裕康，元場哲郎，大学間連携プロジェクト「超高層大気長期変動の全球地上ネットワーク観測・研究」，宇宙科学情報解析論文誌，2011.
- [3] IUGONET Metadata Database
<http://search.iugonet.org/>
- [4] 堀智昭，鍵谷将人，田中良昌，林寛生，上野悟，吉田大紀，阿部修司，小山幸伸，河野貴久，金田直樹，新堀淳樹，田所裕康，米田瑞生，IUGONET 共通メタデータフォーマットの策定とメタデータ登録管理システムの開発，宇宙科学情報解析論文誌，2011.
- [5] SPASE メタデータフォーマット
<http://www.spase-group.org/data/>
- [6] 小山幸伸，河野貴久，林寛生，堀智昭，田中良昌，鍵谷将人，吉田大紀，上野悟，阿部修司，三好由純，金田直樹，能勢正仁，岡田雅樹，超高層物理学分野におけるメタデータ・データベースの構築，DEIM Forum 2010 F4-3, 2010.
- [7] 小山幸伸，河野貴久，堀智昭，阿部修司，吉田大紀，林寛生，田中良昌，新堀淳樹，上野悟，金田直樹，米田瑞生，元場哲郎，田所裕康，鍵谷将人，超高層物理学分野のメタデータ・データベースの開発，宇宙科学情報解析論文誌，2011.
- [8] Corporation for National Research Initiatives (CNRI)
<http://www.handle.net/>
- [9] DSpace Registry
<http://www.dspace.org/whos-using-dspace>
- [10] Dublin Core
<http://dublincore.org/>
- [11] 田中良昌，新堀淳樹，鍵谷将人，堀智昭，阿部修司，小山幸伸，林寛生，吉田大紀，河野貴久，上野悟，金田直樹，米田瑞生，田所裕康，元場哲郎，三好由純，関華奈子，宮下幸長，瀬川朋紀，IUGONET 解析ソフトウェアの開発，宇宙科学情報解析論文誌，2011.
- [12] TDAS (THEMIS Data Analysis Software suite)
<http://themis.ssl.berkeley.edu/software.shtml>
- [13] 河野貴久，小山幸伸，堀智昭，阿部修司，吉田大紀，林寛生，新堀淳樹，田中良昌，鍵谷将人，上

野悟，金田直樹，田所裕康，DSpace を用いた超高層物理学のためのメタデータ・データベースの構築，DEIM Forum 2011 C8-5, 2011.

- [14] Ranking Web of World Repositories

<http://repositories.webometrics.info/>

- [15] 京都大学学術情報リポジトリ・紅

<http://repository.kulib.kyoto-u.ac.jp/dspace/>