

IUGONETシステム報告

阿部修司[1], 梅村宜生[2], 小山幸伸[3], 谷田貝亜紀代[2], 新堀淳樹[4], 田中良昌[5], 上野悟[6], 八木学[7], IUGONETプロジェクトチーム

[1] 九州大・国際宇宙天気科学・教育センター、

[2] 名古屋大・太陽地球環境研究所、

[3] 京都大・理・地磁気センター、[4] 京都大・生存圏研究所、

[5] 国立極地研究所、[6] 京都大・理・附属天文台、

[7] 東北大・惑星プラズマ大気研究センター

IUGONET機関の所有する観測機器

アイスランド
オーロラ観測(2点)
地磁気観測(3点)

トロムソ
ISレーダー
流星レーダー
MFLレーダー

信楽MU
観測所

北海道HFレー
ダー(探査範囲)

太陽望遠鏡

赤道大気レーダー(EAR)

飯舘・女川観測所

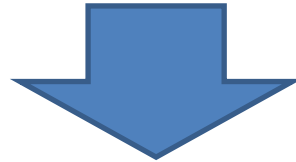
昭和基地
SuperDARNレーダー2台
MFLレーダー
オーロラ観測
地磁気多点観測点網
昭和・女川 ELF 同時観測

スバルバル: ISレーダー、
流星レーダー、オーロラ観測

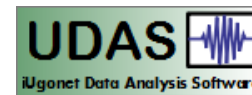


- MSTレーダー
- ▲ MF / 流星レーダー
- ◆ MAGDAS / 地磁気観測
- ★ FM-CWレーダー
- OMTI光学観測
- WDC / 地磁気観測所

- 自分の専門のデータ(例えば地磁気指数)はうまく使えるが、種類の違うデータ(例えば太陽撮像データ)を同時に使うための各情報を入手することが難しく、総合解析になかなか結びつかない
- 他の分野のデータに関する情報がなかなか得られないストレス



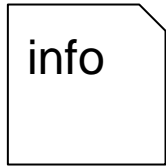
- このような問題を解決する手段として、IUGONETは**観測データのメタデータを共有するシステム**と**統合解析ツール**を開発



- 本講演では、**前者の観測データのメタデータ・データベース (IUGONET MDB)**の成果について紹介する

メタデータ・データベースの利用

- メタデータ・・・観測データそのものではなく、

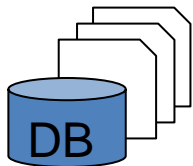


観測データに付帯した情報

e.g. 観測開始終了時刻、観測場所、
観測データの所在情報

→ メタデータを介して観測データへアクセス
させるとより便利

- メタデータ・データベース・・・**メタデータを収集し、かつ**



検索可能にしたもの

→ IUGONET MDBでは

様々な研究機関にある観測データのメタデータ

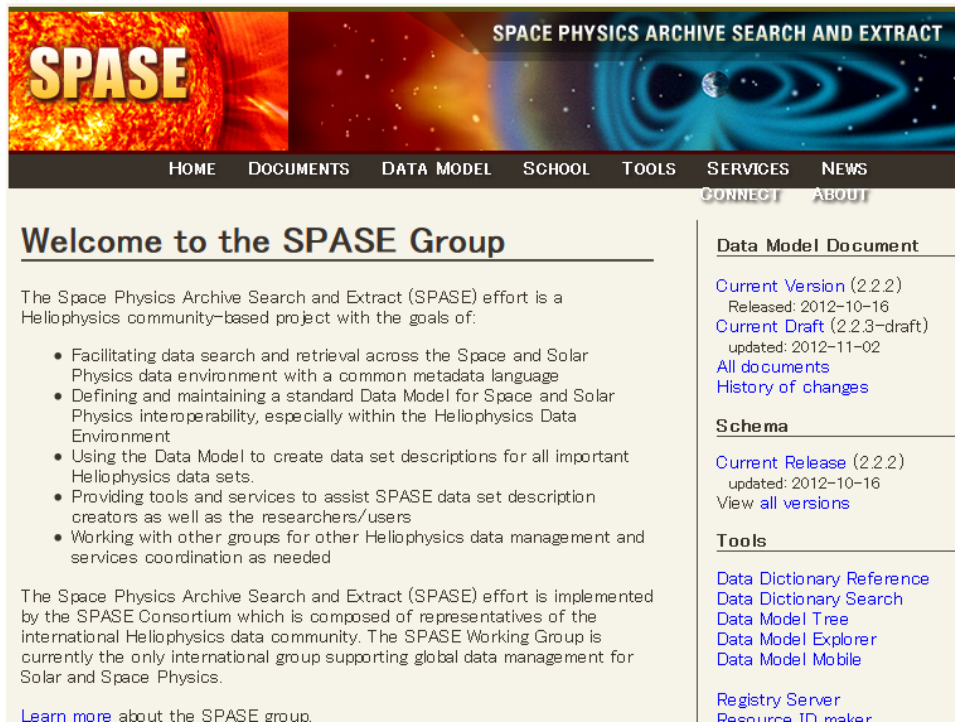
(いつ、どこで、どの物理量が同時観測されたのか?)を

1クエリーで検索可能

メタデータフォーマットの選定

IUGONET メタデータフォーマット = SPASE + 追加項目

(<http://www.spase-group.org>)



SPASE

SPACE PHYSICS ARCHIVE SEARCH AND EXTRACT

HOME DOCUMENTS DATA MODEL SCHOOL TOOLS SERVICES NEWS
CONNECT ABOUT

Welcome to the SPASE Group

The Space Physics Archive Search and Extract (SPASE) effort is a Heliophysics community-based project with the goals of:

- Facilitating data search and retrieval across the Space and Solar Physics data environment with a common metadata language
- Defining and maintaining a standard Data Model for Space and Solar Physics interoperability, especially within the Heliophysics Data Environment
- Using the Data Model to create data set descriptions for all important Heliophysics data sets.
- Providing tools and services to assist SPASE data set description creators as well as the researchers/users
- Working with other groups for other Heliophysics data management and services coordination as needed

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives of the international Heliophysics data community. The SPASE Working Group is currently the only international group supporting global data management for Solar and Space Physics.

[Learn more](#) about the SPASE group.

Data Model Document

[Current Version \(2.2.2\)](#)
Released: 2012-10-16

[Current Draft \(2.2.3-draft\)](#)
updated: 2012-11-02

[All documents](#)
[History of changes](#)

Schema

[Current Release \(2.2.2\)](#)
updated: 2012-10-16
View [all versions](#)

Tools

[Data Dictionary Reference](#)
[Data Dictionary Search](#)
[Data Model Tree](#)
[Data Model Explorer](#)
[Data Model Mobile](#)

[Registry Server](#)
[Resource ID maker](#)

SPASEとは?

metadata format developed by international consortium to comprehensively describe research resources regarding heliospheric and magnetospheric satellite observations

- closely related to STP and upper atmosphere researches (→ easy to use as a base format)
- new metadata elements & words appendable (→ customizable according to our data)
- widely-used in VxOs (→ possible metadata exchange in the future)

IUGONET's modifications

- 非デジタルアーカイブに関するワードの追加
- 太陽座標系に関するワードの追加
- 観測領域を表現するメタデータ要素の追加



- フリーソフトウェア **DSpace** をベースに構築

- <http://www.dspace.org/>
- 学術機関リポジトリで広く使用されているため、情報の入手が容易
- 少々のカスタマイズ (JSP、Servlet) で IUGONET メタデータを利用可能
- 採用しているメタデータのベースは SPASE (<http://www.spase-group.org/>)
- Dspace では直接 XML を扱えないため、フォーマット変換して格納 (プログラムを IUGONET にて作成 = git2dspace)



- 2012年3月より **運用**

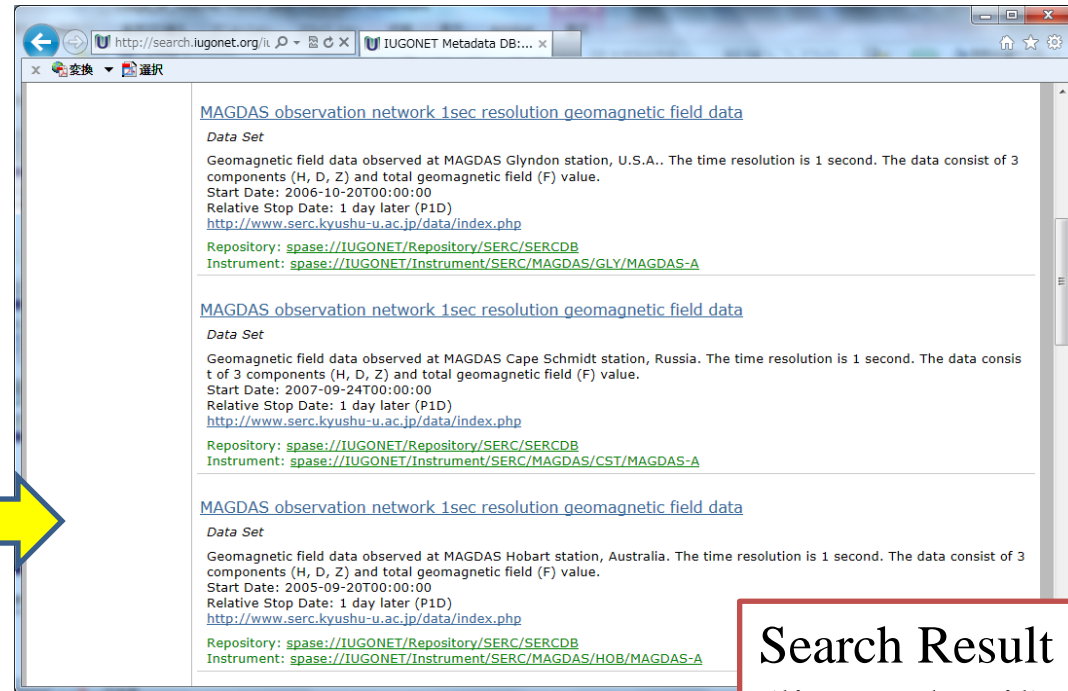
- β版公開時 (2011年5月) はサーバ1台のみ (iugonet1.stelab) の運用
- 2012年1月にメインサーバを移行 (iugonet7.icswse)。
- 旧メインサーバはサブとして使用し、現在、メインサブ2台の **冗長体制**
- **メタデータ総数は1000万件以上 (2014年8月15日調査、10679409)**

成果 : IUGONET MDB

<http://search.iugonet.org/iugonet/>



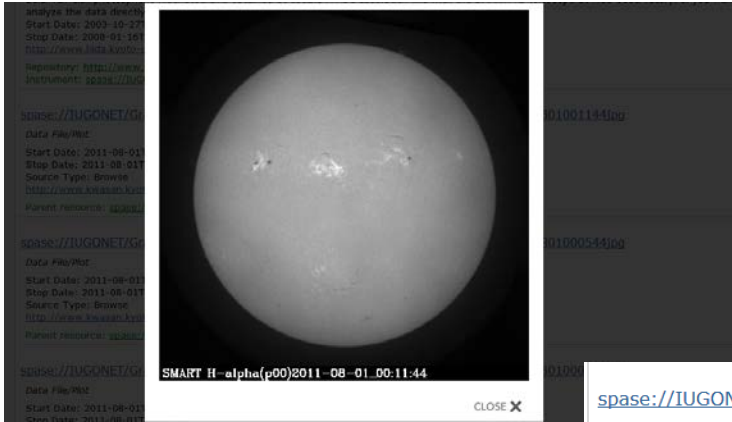
Freeword Search,
Data type Selection,
Visualized search, etc...



Search Result
(list or detail)

- Provide the service for cross-searching observational data distributed across the IUGONET institutions.
- Remarkable advancement in accessibility to the observational data and accelerate the interdisciplinary study.

1. Lightbox2を用いたDisplay データの表示



左記機能に加えて。。。



※JUDASの詳しい説明は
ソフトウェア発表にて

spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/dst198612_wdc

Data File/Plot
Start Date: 1986-12-01
Stop Date: 1986-12-31
Source Type: Data

<http://wdc-data.iugonet.org/data/hour/index/dst/1986/dst8612>

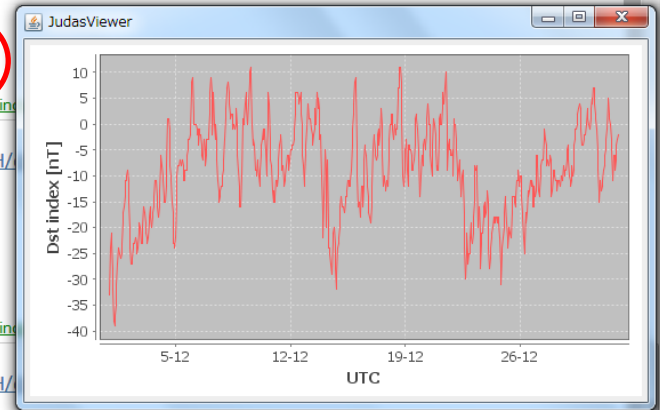
spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/

Data File/Plot
Start Date: 1958-02-01
Stop Date: 1958-02-28
Source Type: Data

<http://wdc-data.iugonet.org/data/hour/index/dst/1958/dst5802>

spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/

Data File/Plot
Start Date: 1958-01-01
Stop Date: 1958-01-31
Source Type: Data

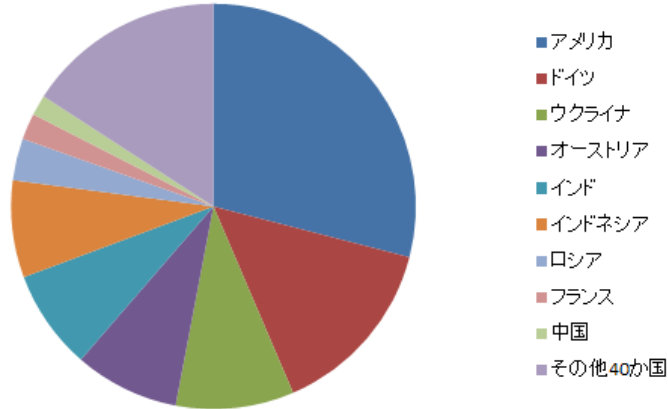


2. Java Web Start を用いたGranuleデータの表示

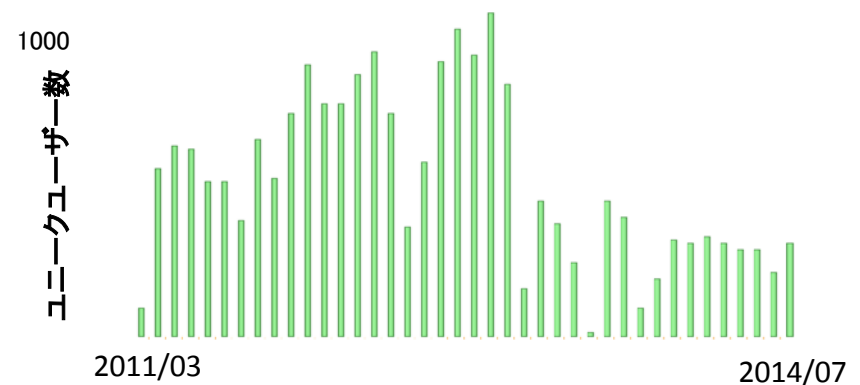
成果：MDBの利用状況

ユニークユーザー数の補足：
同一ユーザと思われる場合は
1月に何度アクセスしても1カウント

日本を除いた国別アクセス比率グラフ



月ごとのユニークユーザー数の推移



- ユニークユーザー数は**右肩上がり**で増加傾向していた
 - 2013年度末よりアクセス数の減少あり。現在はそこで安定
 - アメリカ・ヨーロッパ方面に加え、**アジア地域から**の多数の利用実績
 - インドネシア・オーストリア等からの高いアクセス数
- **国際アウトリーチの成果**

利用海外機関の例：

- ドイツ・ライプニッツ大気物理研究所(ドイツ)
- オーストリア科学アカデミー宇宙研究所(オーストリア)
- インド宇宙庁国立大気科学研究所(インド)
- インドネシア国立航空宇宙研究所(インドネシア)
- モスクワ大学(ロシア) 等

検索キーワード例

超高層物理分野の例	DST, MAGDAS, magnetometer, SUPERDARN, MU, EISCAT, MF Radar, smart, magnetogram 等
超高層物理分野以外の例	Jupiter, CO2, O3, rain, Sun Prominence, dellinger, adelic, ocean, Ca K, radiosonde, Wind Profiler 等

超高層物理分野のワードが多数を占めている一方で、隣接分野や他分野のワードも入力されており、**様々な分野のユーザから利用されていることが分かる。**

成果: OpenSearchの利用

OpenSearch: 検索結果を他のサイトから自由に利用できるようにするための仕組み

検索結果を他から利用する拡張に適應できるため、これを分散型MDBの構築に利用する

OpenSearch仕様は
<http://www.iugonet.org/opensearch.html>

http://search.iugonet.org/iugonet/opensearch/request?query=nipr_1sec_fmagsyoyo&ts=2010-01-01&te=2010-01-05&Granule=granule



**OpenSearchの結果
(ブラウザでの結果表示)**

成果：分散化MDB調査

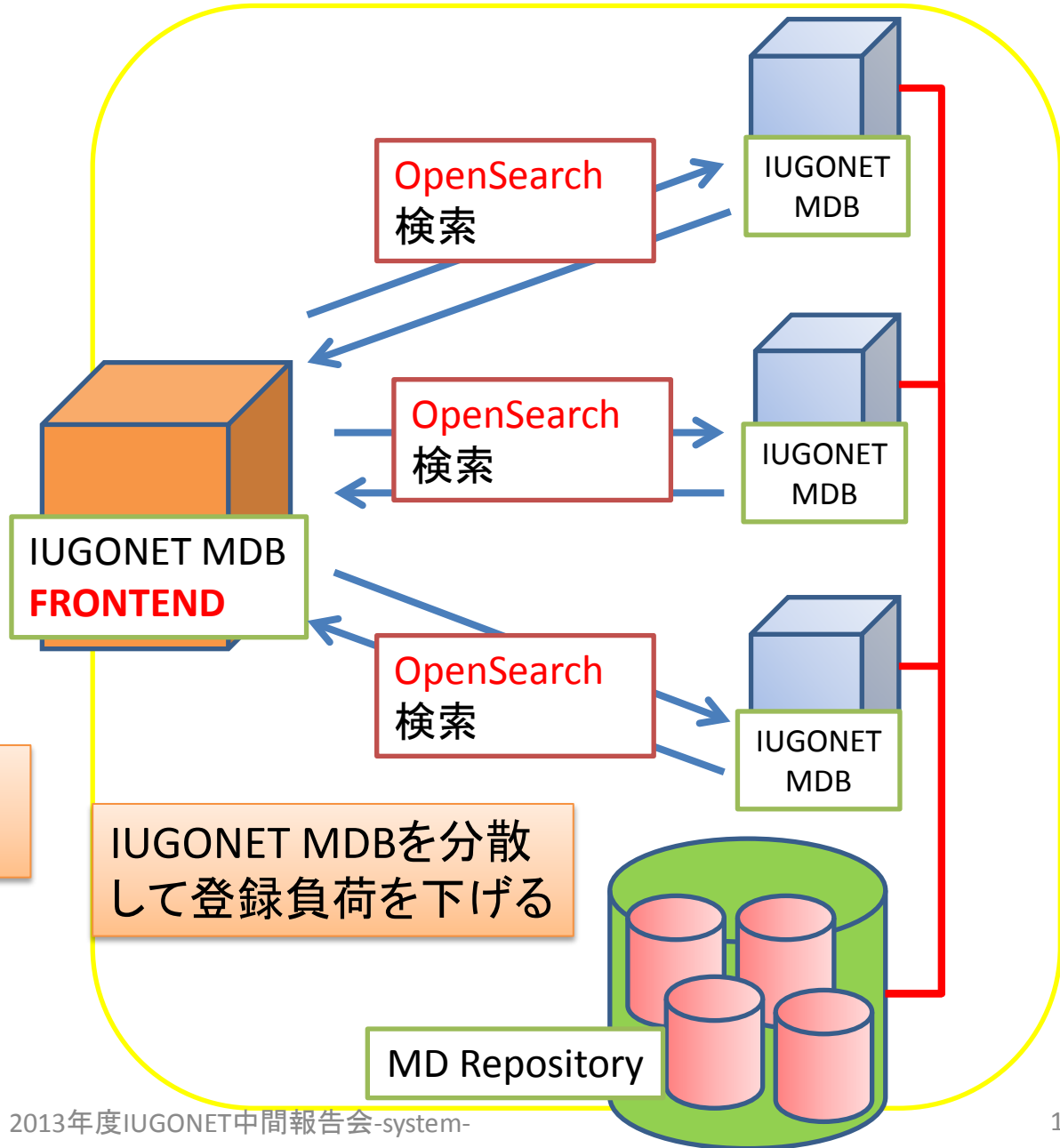


検索クエリ



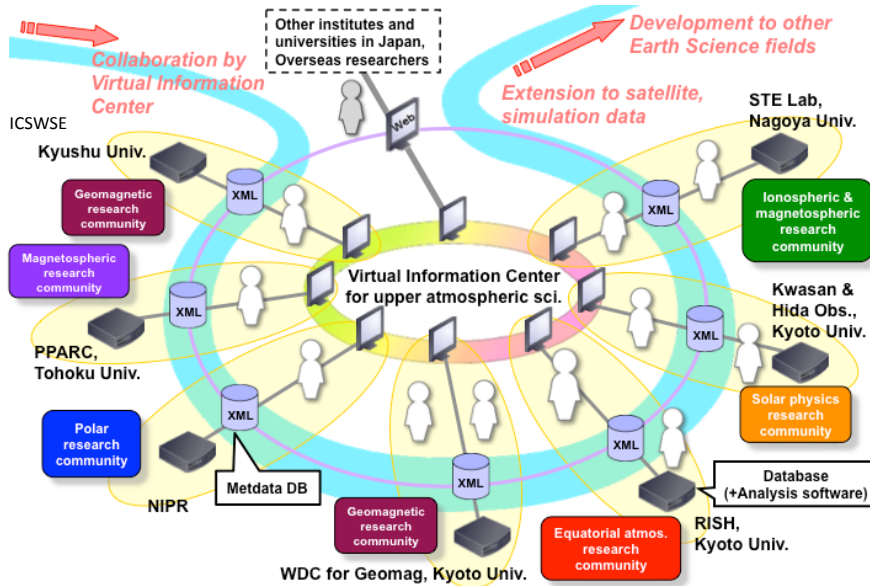
各サーバの結果をマージして返す

ユーザーは全サーバの情報を検索可能



成果：分散化MDB調査

分散化MDBの基礎は確立したが...



機関をまたぐ分散化では
各機関のネットワークのレスポンスが全体に影響(全サーバの結果を一度ソートしているため)等の問題が分かった

よって、現在は

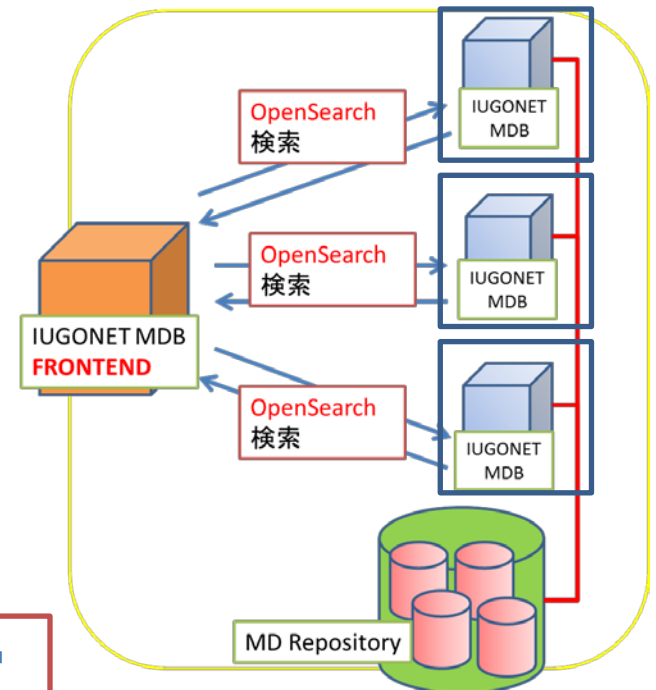
- ・実サーバとしては1台

だが

- ・内部にVMで各機関メタデータ用IUGONETサーバを持つことで処理の分散化を行なう

内部分散化の調査を進めた

実サーバ VM



成果：連想検索への取り組み

- ・専門分野外のメタデータ検索をする場合、**検索語句の選択が困難**
 - ・専門に近い分野であっても、正しく検索語句を選べるとは限らない
= **メタデータに記載されている語句でなければヒットしない**
- 等の解決に役立つ



Substorm

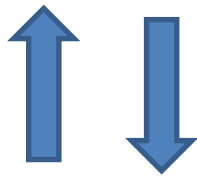


CME, AE,
Ionosphere...



国立情報学研究所で
開発された連想検索システム
GETAssoc の導入を検討

「Substorm」
を検索してみよう



メタデータヒットなし
関連用語は
CME, AE, ionosphere...



検索語句そのものがヒットしなくても、
関連する用語を検索語句候補として
表示することができる。

成果：連想検索への取り組み

どのような用語が連想されるのかは、連想検索システムに登録している辞書次第
 Googleスプレッドシートで管理した語句テーブルを辞書ファイルに変換するしくみを
 開発→辞書の共有、複数人による辞書作成

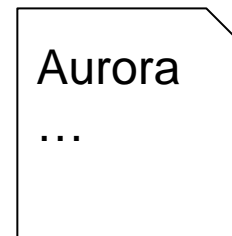
Google Spreadsheet

A	B
6 Thermosphere	The thermosphere is the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere.
7 Mesosphere	The mesosphere is the layer of the Earth's atmosphere that is directly above the stratosphere and directly below the thermosphere.
8 Stratosphere	The stratosphere is the second major layer of Earth's atmosphere, just above the troposphere, and below the mesosphere.
9 Troposphere	The troposphere is the lowest portion of Earth's atmosphere. It contains approximately 80% of the atmosphere's mass and 99% of its water vapor and aerosols.
10 Disturbance storm time index	The disturbance storm time (Dst, Kyoto Dst) index is a measure in the context of space weather.
11 K-index	The K-index quantifies disturbances in the horizontal component of earth's magnetic field with an integer in the range 0-9 with 1 being quiet and 9 or more indicating a geomagnetic storm.
12 Super Dual Auroral Radar Network	The Super Dual Auroral Radar Network (SuperDARN) is an international radar network for studying the upper atmosphere and ionosphere, comprising eleven radars in the northern hemisphere and seven in the southern hemisphere that operate in the High Frequency (HF) bands between 8.0 MHz (8.7m) and 22.0 MHz (14m).
13 Japan Meteorological Agency	The Japan Meteorological Agency, frequently abbreviated to JMA, is an agency of Ministry of Land, Infrastructure, Transport and Tourism.
14 International Association of Geomagnetism and Aeronomy	International Association of Geomagnetism and Aeronomy (IAGA) is the premier international scientific association promoting the study of terrestrial and planetary magnetism and space physics.
15 Committee on Data for Science and Technology	The Committee on Data for Science and Technology (CODATA) was established in 1966 as an interdisciplinary committee of the International Council for Science. It seeks to improve the compilation, critical evaluation, storage, and retrieval of data of importance to science and technology.
16 World Data Center International Council for Science	The World Data Center (WDC) system was created to archive and distribute data collected from the observational programmes of the 1957-1958 International Geophysical Year by the International Council of Science (ICSU).
17 International Council for Science	The International Council for Science (ICSU) is an international non-governmental organization devoted to international cooperation in the advancement of science.
18 Intermagnet	IntermagNet (International Real-time Magnetospheric Observatory Network) is a global network of observatories, monitoring the Earth's magnetic field.
19 CHAMP	CHallenging Minisatellite Payload (CHAMP) was a German satellite launched July 15, 2000 from Plesetsk, Russia and was used for atmospheric and ionospheric research, as well as other geoscientific applications, such as GPS radio occultation.
20 National Institute of Polar Research	National Institute of Polar Research is the Japanese research institute for Antarctica. The agency manages several research bases on the continent.
21 International Geomagnetic Reference Field	The International Geomagnetic Reference Field (IGRF) is a standard mathematical descriptor of the Earth's main magnetic field.
22 International Reference Ionosphere	International Reference Ionosphere (IRI) is a common permanent scientific project of the Committee on Space Research and the International Union of Radio Science started in 1960/62. It is the international standard for the terrestrial ionosphere since 1960.

語句テーブルのダウンロード
 辞書ファイルへの変換



辞書ファイル(.itb)



例：
 キーワード「substorm」での検索数0
 →「CME」「AE index」などが関連キーワード
 検索結果へのリンク作成、表示

Search Results

substorm

Time from: to

Data Set (Numerical Plot / Movie) Data File / Plot

Search produced no results.

Results/Page 10 | Sort items by Relevance

Did you mean:

- [CME](#)
- [AE index](#)
- [Ionosphere](#)
- [Intergovernmental Panel on Climate Change](#)
- [Exosphere](#)

成果：運用ツールの開発

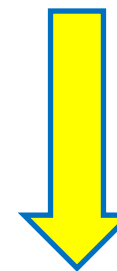
- **md_checker**: 各機関の作成したメタデータ内のSPASE要素他をチェックする
- **md_validator**: 各機関の作成したメタデータが正しいXMLとして記述されているかをチェックする
- **brokenlinkchecker**: 登録したメタデータに記載されている外部へのURLリンクと、外部機関の実データとが、正しく接続されているかを確認するツール。いわゆる「デッドリンク」を検出し、ユーザーがデータへ正しく到着できる手助けをする
- **g2d、git2dspace**: 各機関のメタデータを保存しているgitリポジトリからMDBへメタデータをインポートする

build.xml

- Github(<https://github.com/iugonet>)にて、MDB含む多数のプロジェクトを準備
 - IUGONET MDB
 - 連想検索システム
 - JavaベースのUDAS(解析ソフトウェア)
 - etc etc...

```

[osamu@osamu] ~/iugonet/build$ cat build.xml
<?xml version="1.0" encoding="UTF-8" ?>
<project name="iugonet" default="extract" ?>
  <property file="build.properties" ?>
    <target name="install_git" description="Install Git" ?>
      <get src="http://hg.ng.riken.go.jp/rikenforge-release/rikenforge-release-0.5.2-2.e15.r1.1398.mn" dest="/tmp" useIntermediate="true" ?>
        <exec executable="sudo" ?>
          <arg value="rm" ?>
            <arg value="-rf" ?>
              <arg value="/tmp" ?>
            </arg>
          <exec executable="tar" ?>
            <arg value="xzf" ?>
              <arg value="rikenforge-release-0.5.2-2.e15.r1.1398.mn" ?>
            </arg>
          </exec>
          <exec executable="sudo" ?>
            <arg value="cp" ?>
              <arg value="install" ?>
              <arg value="git" ?>
            </arg>
          </exec>
          <exec executable="git" ?>
            <arg value="config" ?>
            <arg value="--global" ?>
            <arg value="--user.name" ?>
            <arg value="${GIT_USER_NAME}" ?>
          </exec>
        </target>
      </project>
    
```



1. Java
2. ant
3. build.xml

- メタデータスキーマは <http://www.iugonet.org/data/schema/> でも公開

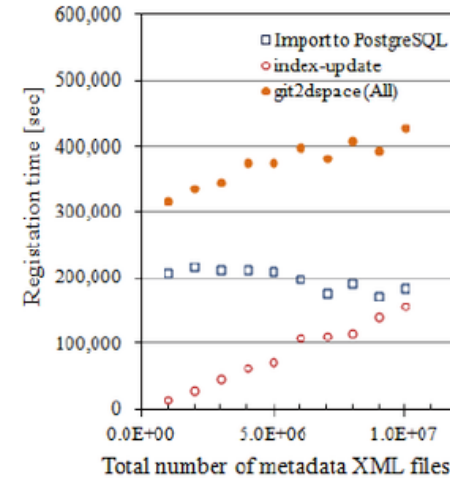


IUGONET PRODUCTS

- 積極的な成果発表
 - JpGU、SGEPSSではほぼ毎回発表
 - DEIM(データ工学と情報マネジメントに関するフォーラム)など他分野の研究集会への参加
- システムを主体とした論文の出版
 - 「超高層物理学分野の為のメタデータ・データベースの開発(小山他,宇宙科学情報解析論文誌,2012)」をはじめとした論文6編
 - 現在2編の論文をEPS CAWSES Special Issueに投稿中

メタデータ増加に伴う問題

- メタデータ増大に伴う各種パフォーマンス低下=登録から検索までに時間がかかる。メタデータの置き換え、削除処理を入れると所要時間が倍増
- 単一マシンでの運用限界=検索エンジン (Apache Lucene) での検索パフォーマンス低下



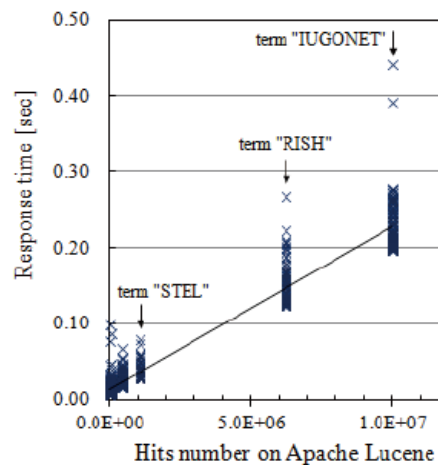
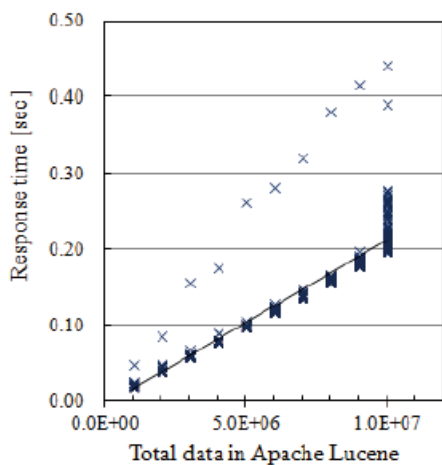
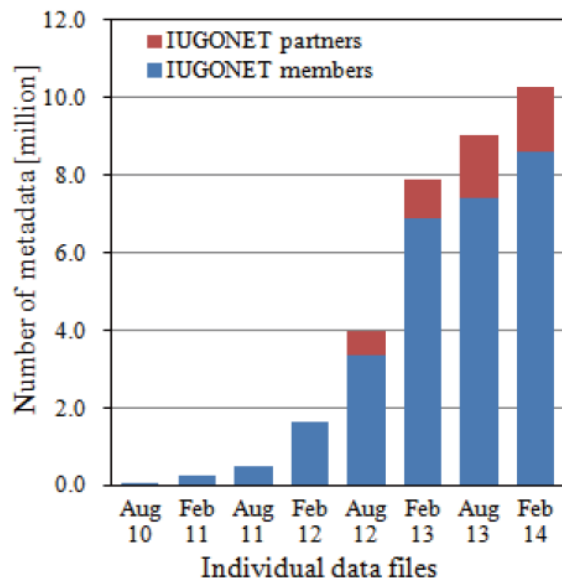
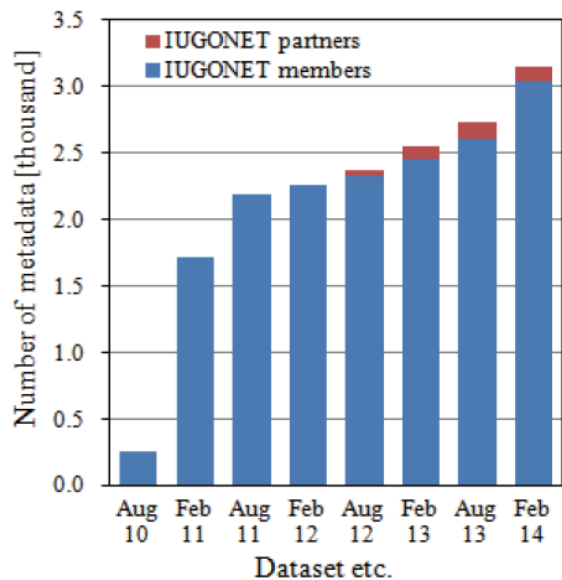
(Abe et al., submitted to EPS)

- MDB分散化による、既存システムを維持した大幅なパフォーマンス向上
 - 検索システムの分散化は大規模システムでは一般的
 - 例えば、Googleでは1クエリに対し数千台のマシンで処理
- OpenSearchによるメタデータのハーベスティングなしでの複数サーバの検索表示
 - 横断検索サーチエンジンの例: 国立国会図書館サーチ (NDL Search)
 - 他プロジェクトとのメタデータ交換が容易に

などを検討したが、解決には至っていない

システム的には例えばindexファイルの分散化などがあるが、より根本的な原因である「メタデータの粒度」についての議論も必要

検索エンジン(ApacheLucene)の応答時間



(Abe et al., submitted to EPS)

IUGONETでは、共通メタデータフォーマットによるメタデータ作成を助ける構造定義(XMLスキーマ)を公開

<http://www.iugonet.org/data/schema/>

メタデータ1.0.4リリースに伴うメタデータ一括更新にて

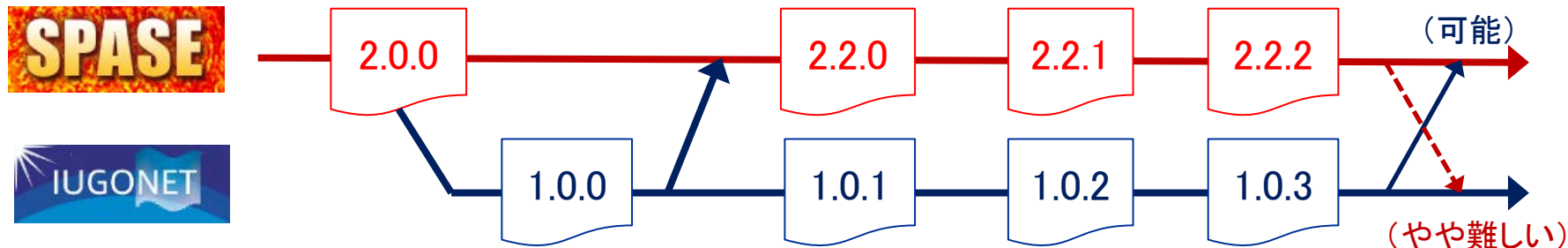
- ・XMLとしてinvalidのファイルが登録されている
- ・使いたい語句がスキーマ内に存在しない

などの問題が分かってきた

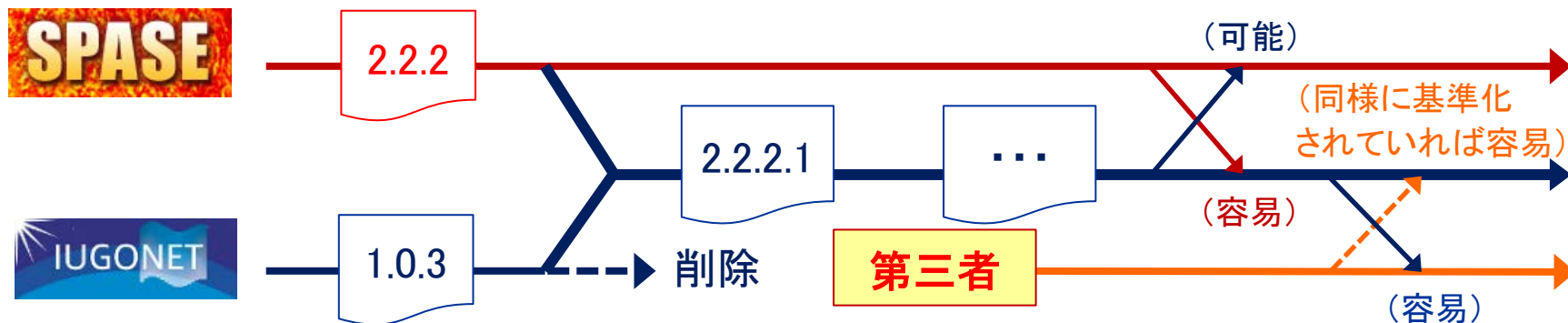
他方で、昨今のオープンデータに関する大きな流れに対応したメタデータスキーマ(iugonet-2_x_y_z.xsd)を検討中

- リファクタリング(IMPORT & REDEFINE)
- ORCID対応
- DOI対応(DataCite、JaLC、CrossRef、etc...)
- ライセンス情報の追加

これまでの更新ポリシー（名前空間をIUGONET独自のものに変更して派生）



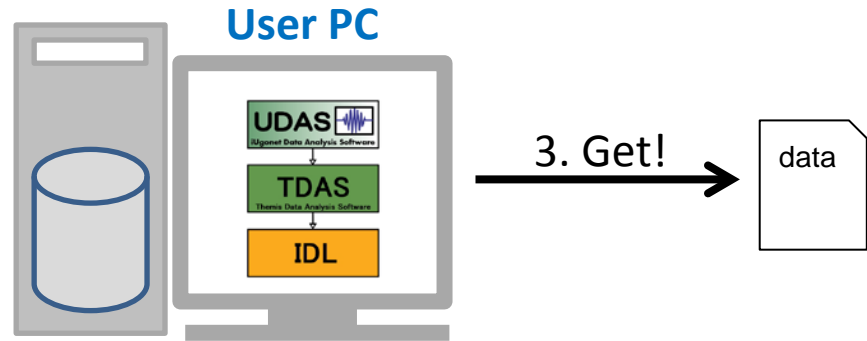
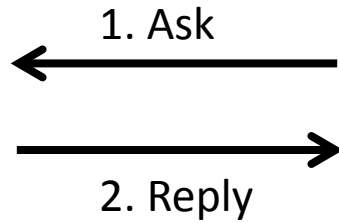
現在検討中の更新ポリシー（名前空間をSPASEのままとする）



- 利点**
- ・ SPASE側で拡張された内容を容易に取り込むことが可能
 - ・ IUGONETプロジェクトのXMLファイルであってもSPASEフォーマットとして扱うことができるため、第三者でもシステム処理が容易となる

XMLファイルの相互提供や相互交換が容易となり、流通の促進化につながる

Collaboration with our software tool(UDAS)



QR Reader



IUGONET Touch



次期プロジェクトに向けて

- IUGONETシステムの成果と問題点について紹介
- 成果
 - MDBの構築ノウハウ
 - 分野内及び分野横断交流の懸け橋
 - 各種サイエンスへの適用
- 課題
 - メタデータの粒度についての再考
 - スキーマアップデートに関する諸事
 - メタデータDB利用形態の充実
- 本研究集会の目的のひとつは、成果/課題についてオープンに話し合い、次期プロジェクトに繋ぐこと

複数の交流の場を設けています。

活発な議論をしましょう！



WELCOME! OSAIDQ