



IUGONET

Metadata DB for Upper Atmosphere

超高層大気長期変動の全球地上ネットワーク観測・研究
Inter-university Upper atmosphere Global Observation NETWORK

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

河野貴久*¹, 小山幸伸*², 堀智昭*¹, 阿部修司*³,
吉田大紀*², 林寛生*⁴, 新堀淳樹*⁴, 田中良昌*⁵,
鍵谷将人*⁶, 上野悟*⁷, 金田直樹*⁷, 田所裕康*⁵

*¹名古屋大・太陽地球環境研究所, *²京都大・理・附属地磁気世界資料解析センター,

*³九州大・宙空環境研究センター, *⁴京都大・生存圏研究所,

*⁵国立極地研究所・宙空圏研究グループ, *⁶東北大・惑星プラズマ大気研究センター,

*⁷京都大・理・附属天文台

超高層大気長期変動の全球地上観測・研究プロジェクト

IUGONET

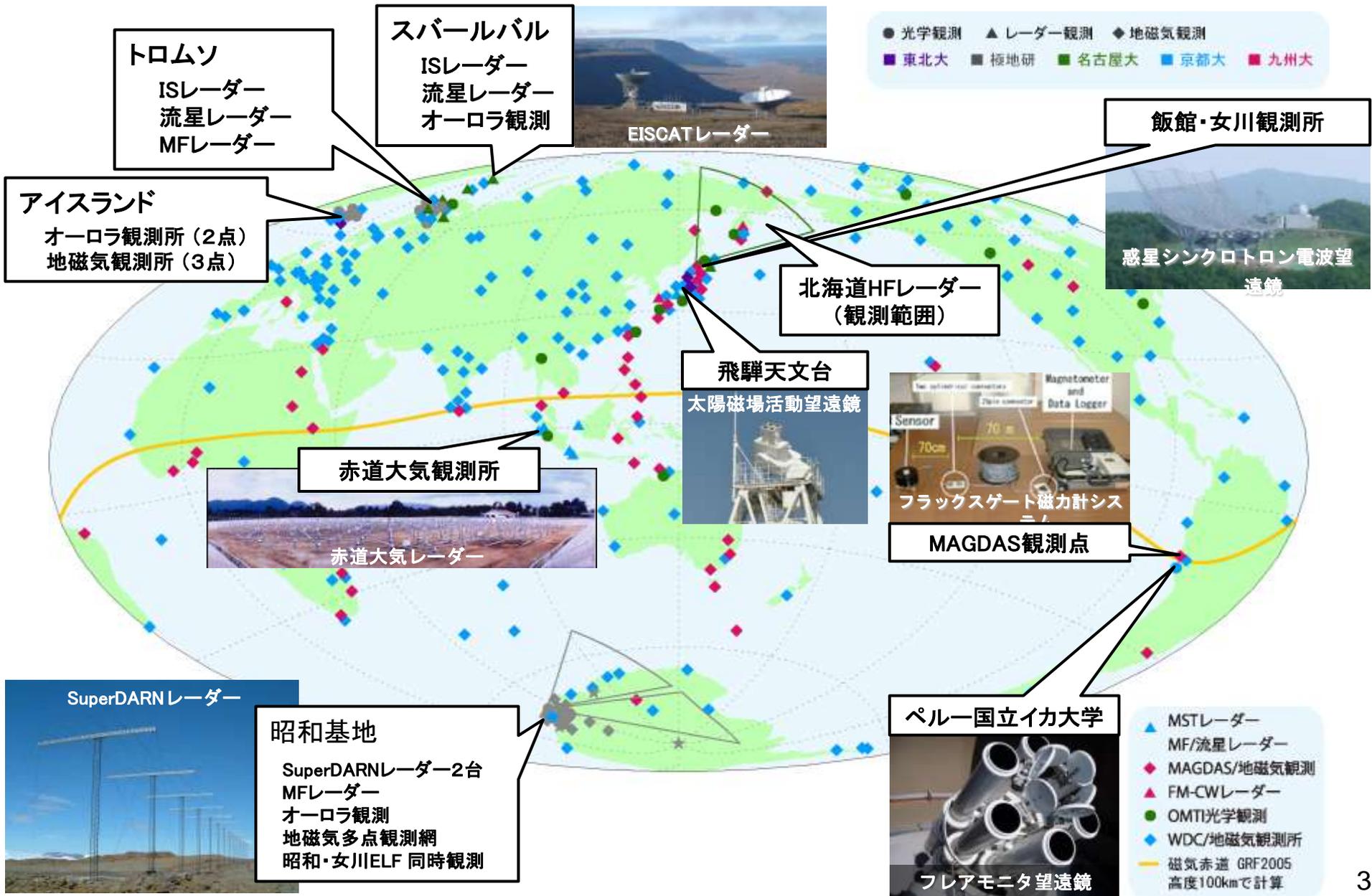
Inter-university Upper atmosphere Global Observation NETWORK

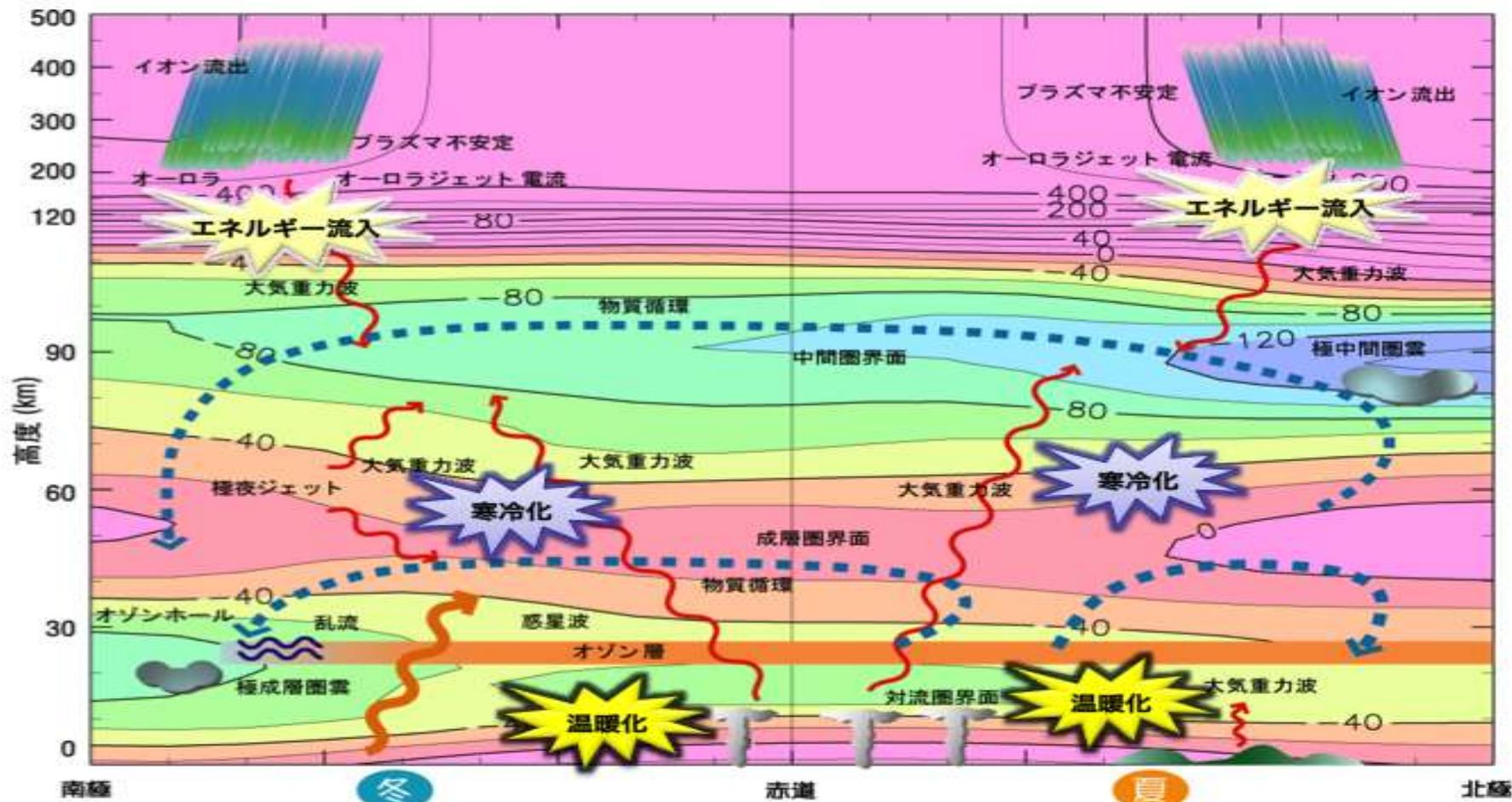
<http://www.iugonet.org/>

- ★ 大学や研究機関に分散する**超高層大気地上観測データを効率的に検索・取得、そして解析するための研究インフラを整備する**
 - IGY以来の国際共同観測事業で蓄積された観測データ（アナログ・デジタル）の流通
 - 分野をまたがる多様なデータの解析による超高層大気長期変動研究の進展

参加機関・組織

- 国立極地研究所
- 東北大学理学研究科附属惑星プラズマ・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
- 九州大学宙空環境研究センター

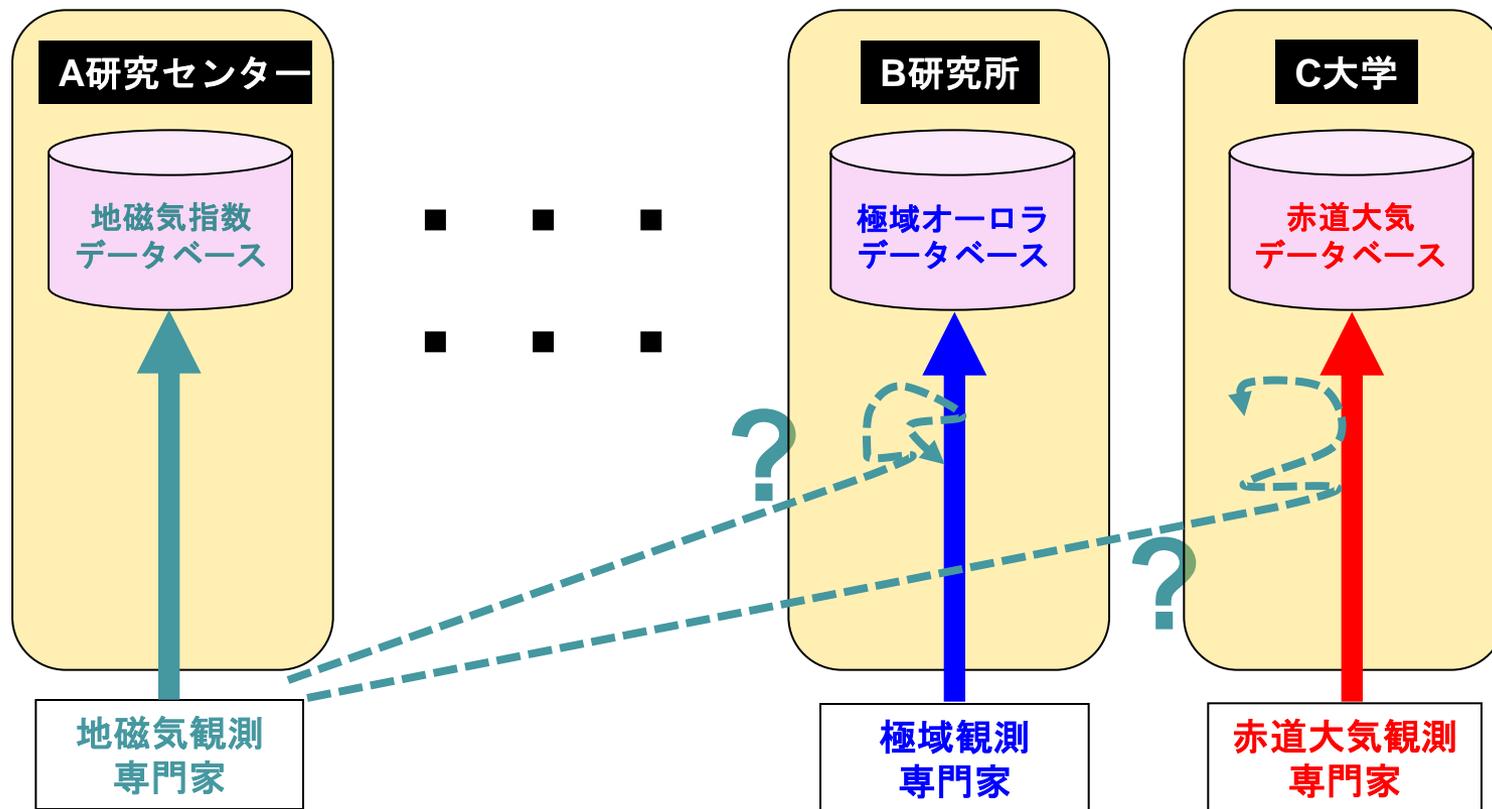




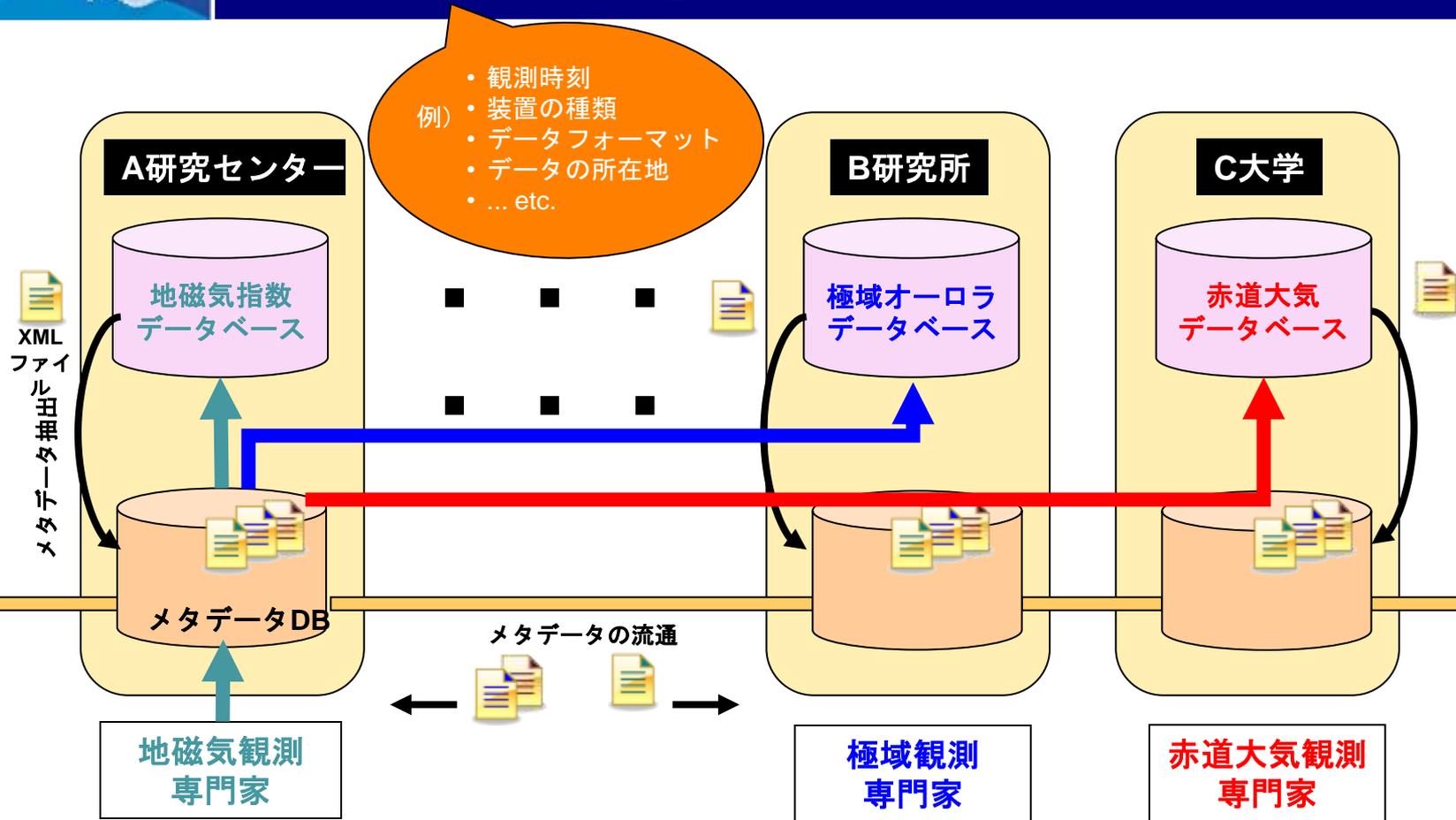
温室効果ガスの影響・対流圏の温暖化・超高層大気の寒冷化
極域現象・オーロラ帯へのエネルギー流入

極域から赤道域に設置した各種レーダー・磁力計・光学観測装置による
広い領域での風速・プラズマ・温度などの観測から
より正確な気候変動のシグナルを検出し、地球大気の大規模な振る舞いの定量化を図る

地球環境・気候変動の将来予測に貢献する。



- 複数の機関により多種類かつ大量のデータが観測・管理されている
- 他分野のデータへ辿り着く為の手がかりが無い・調査に時間が掛る



- 色々な機関の多種多様な観測データに関するメタデータ(データの所在情報など)を検索
- 実際の観測データは、各機関のデータベースが提供

- SPASEデータモデルをベースにした、IUGONET共通メタデータフォーマット(Ver.1)を策定した。これをもとにメタデータを作成中。
- DSpaceを利用して、メタデータ・データベースシステムを短期間で構築中。
- IUGONET各機関が提供する観測データを表示・解析するためのソフトウェアをIDL+TDASで開発中。

メタデータ検索システム

フリーワードによる検索

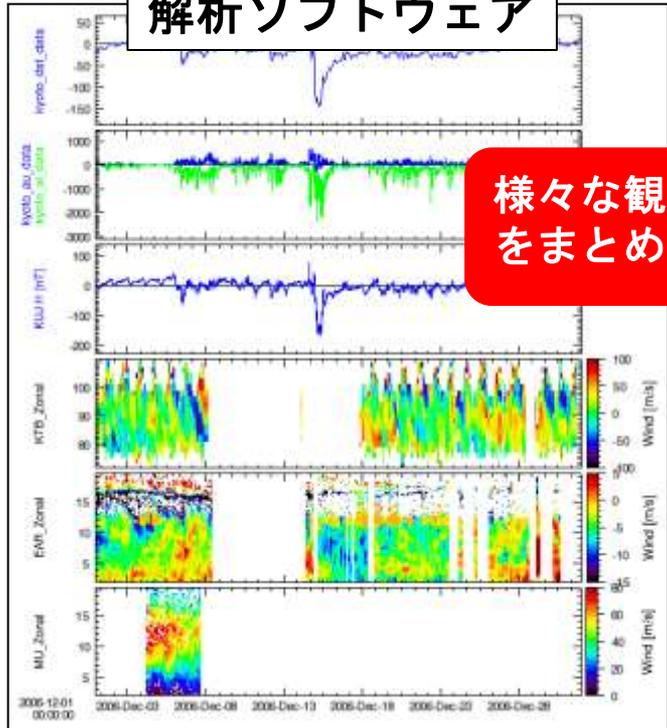
時間や場所による絞り込み

データタイプによる絞り込み



解析ソフトウェア

様々な観測データをまとめて描画



- リポジトリソフトDSpaceを採用
 - MITとHPが開発，オープンソースで公開
 - <http://www.dspace.org/>
 - v1.7.0: 2010-12
 - v1.6.2: 2010-06
 - v1.6.1: 2010-05
 - v1.6.0: 2010-03
 - Javaベースのシステム
 - Apache Tomcat, Apache Lucene, PostgreSQL
 - 学術機関リポジトリで多く利用されている



- IUGONET向けにカスタマイズをして利用
 - WEBインターフェース
 - メタデータ・フォーマット

1. 時間・地理空間の範囲検索

- 地球温暖化などの長期間にわたる自己相関
- 地球規模で起こっている現象の相互作用

2. 外部提供インターフェイス

- 可視化・解析ソフトからの利用
- 超高層大気分野以外のデータとの相互検索

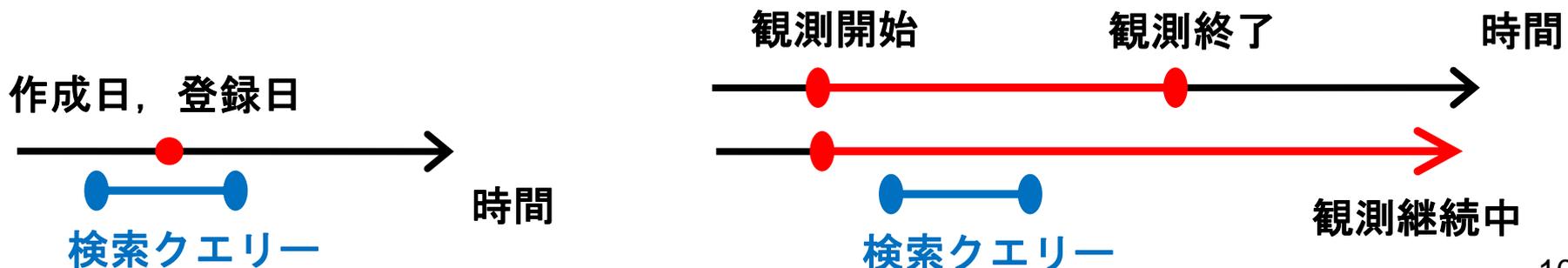
3. 管理(登録)のしやすさ

- プロジェクト終了後も研究者が管理できること
- 長期運用体制の確立
- 大量件数の登録・高頻度の登録

4. メタデータ登録速度

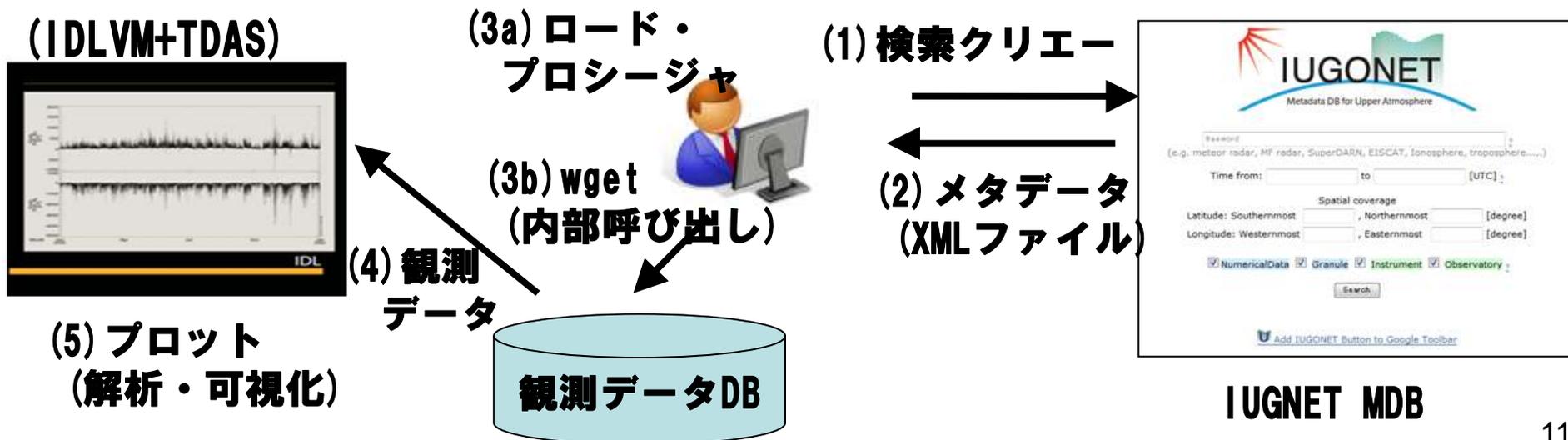
- 登録予定数は100万件以上
- 1度に10万件単位の登録の可能性がある

- DSpaceは標準メタデータフォーマットにDublin Core(DC)を採用
 - DCは書誌的な情報を記述するためのもの
 - たとえば, 作者やタイトル, 作成日
 - 作成日や登録日などの時間情報は, 点情報
- IUGONETで扱うデータは観測データであり開始時刻と終了時刻の間に情報を持つ
 - 検索エンジンLuceneのRange検索クエリを複数組み合わせることで検索を実現
 - 地理空間の緯度経度空間の検索も同様



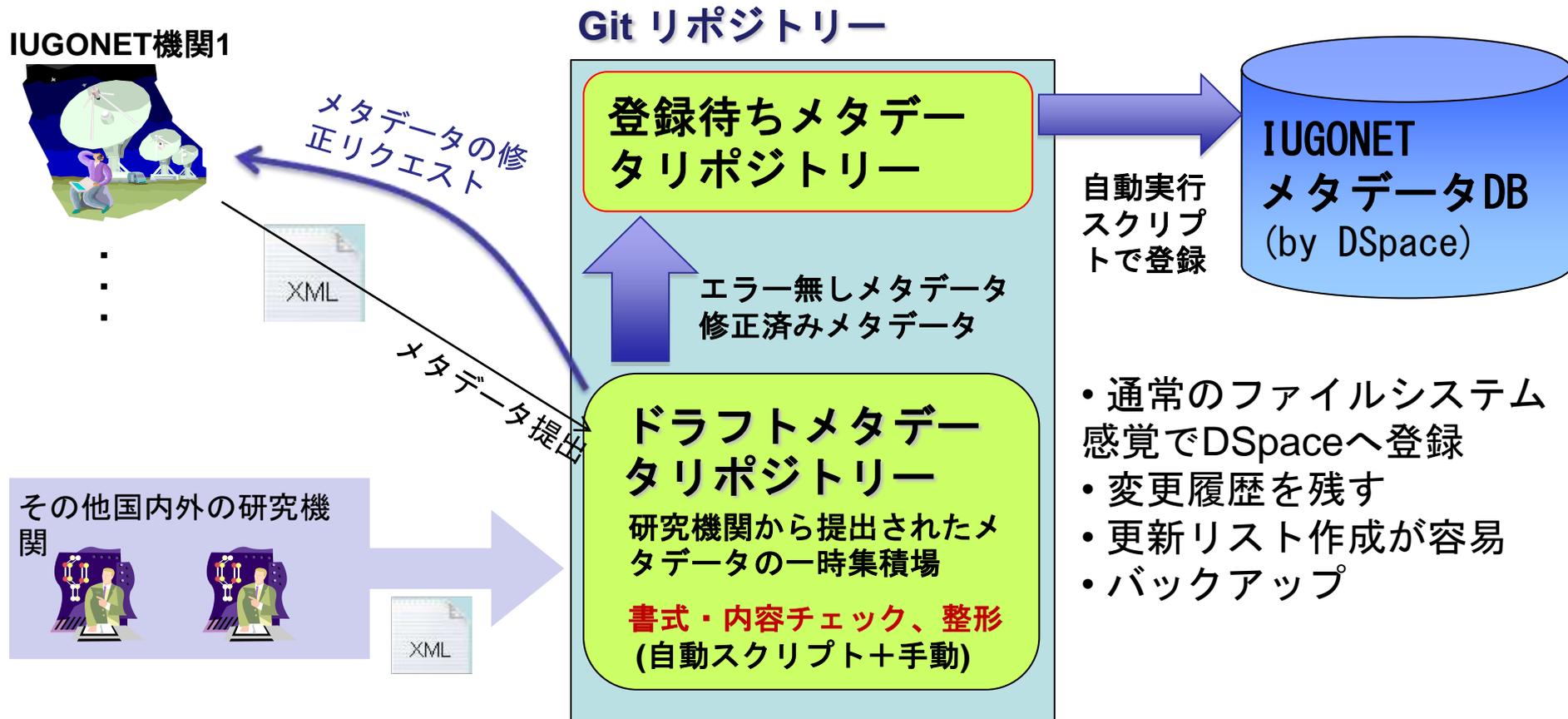
- DSpaceの外部提供インターフェース
 - メタデータ交換: OAI-PMH
 - 横断検索: OpenSearch, SRW/U(関連開発物)

- IUGONETの利用例
 - 解析ソフトウェア(観測データの所在情報等を取得)
 - 検索クエリー, 出力データのカスタマイズが必要



● メタデータ登録システムの開発

- IUGONET各機関で観測データからメタデータを抽出
- 提出されたメタデータのチェック, バージョン管理, データベース登録を行うシステム



➤ メタデータの登録

1. Gitリポジトリの履歴情報から登録リストを作成
 - 追加コミット分だけをチェック
 - 複数のコミット分を1つに集約
 - 同じファイルが追加そして削除された履歴は、無視
2. SPASEフォーマットからDSpaceフォーマットに変換
3. 削除, 追加, 更新の順番でDSpaceに登録
4. 登録ログとして, コミットIDとアイテムIDを保存
 - 次回の登録時に利用

Gitリポジトリ

メタデータ
リポジトリ

コミットID&アイテムID
リポジトリ

履歴情報から作成

- 削除リスト
- 追加リスト
- 更新リスト

DSpaceに登録

①

削除

②

追加

③

更新

登録ログ

4. メタデータ登録速度

- DSpaceのメタデータ登録速度の性能評価
 - DSpace 1.6.2と1.7.0を測定, 比較した
 - DSpaceは, IUGONETカスタマイズ版を使用
 - メタデータフォーマット(登録・検索対象)を追加
 - 1000件ずつ追加登録経過時間を測定

性能試験環境

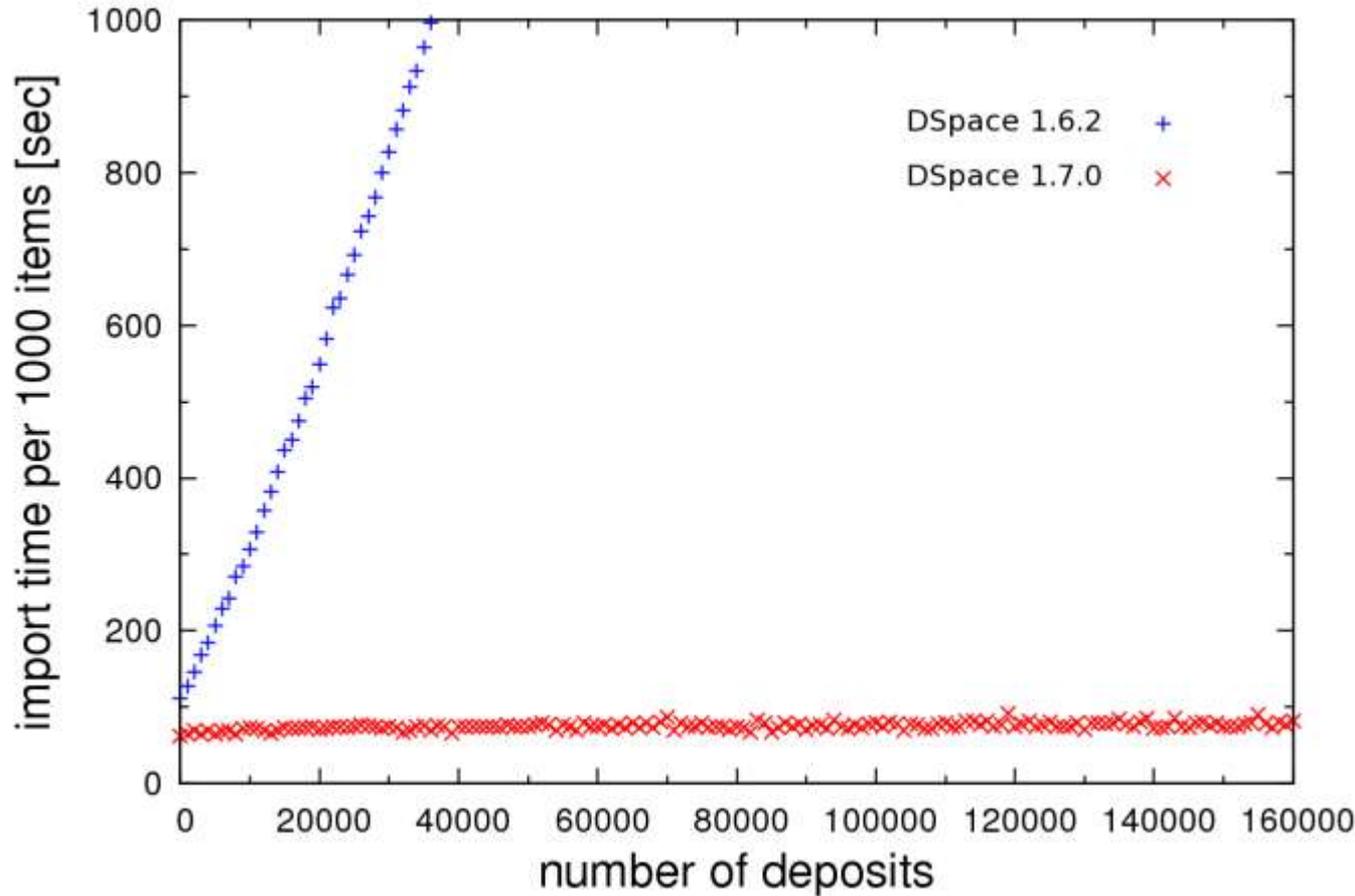
DSpace (検索エンジン)	1.6.2 (Lucene 2.3.0)	1.7.0 (Lucene 2.9.3)
CPU	Intel Xeon W5590 3.33GHz x 2	←
Memory	48GB	←
HDD	1TB(Software RAID1)	←
OS	CentOS 5.5 x86_64	←
データベース	PostgreSQL 8.4.5	←
Java	1.6.0_23-b05	←

DSpace 1.7.0は,
2010-12-17リリース

- 登録試験に用いたメタデータ
 - 観測データファイルに対応するメタデータ
 - 観測時間, 場所, データファイルのURLの情報
 - 登録要素数は少ないが, 現実に大量件数の登録が必要
 - 約20要素

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
- <Spase lang="en" xmlns="http://www.iugonet.org/data/schema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.iugonet.org/data/schema http://www.iugonet.org/data/schema/iugonet.xsd">
  <Version>1.0.0</Version>
- <Granule>
  <ResourceID>spase://IUGONET/Granule/STEL/MM210/CAN/fluxgate/PT1S_ergsc_cdf/mm210_1sec_can_19981130_cdf</ResourceID>
  <ReleaseDate>2010-09-01T00:00:00</ReleaseDate>
  <ParentID>spase://IUGONET/NumericalData/STEL/MM210/CAN/fluxgate/PT1S_ergsc_cdf</ParentID>
  <StartDate>1998-11-30T00:00:00</StartDate>
  <StopDate>1998-11-30T23:59:59</StopDate>
- <Source>
  <SourceType>Data</SourceType>
  <URL>http://gemsissc.stelab.nagoya-u.ac.jp/data/ergsc/ground/geomag/mm210/1sec/can/1998/mm210_1sec_can_19981130_v01.cdf</URL>
</Source>
- <SpatialCoverage>
- <CoordinateSystem>
  <CoordinateRepresentation>Spherical</CoordinateRepresentation>
  <CoordinateSystemName>GEO</CoordinateSystemName>
</CoordinateSystem>
  <NorthernmostLatitude>-35.30</NorthernmostLatitude>
  <SouthernmostLatitude>-35.30</SouthernmostLatitude>
  <EasternmostLongitude>149.00</EasternmostLongitude>
  <WesternmostLongitude>149.00</WesternmostLongitude>
  <Unit>degree</Unit>
</SpatialCoverage>
</Granule>
</Spase>
```

メタデータ登録速度試験の結果



追加で1000件登録したときの経過時間

累積が最終的な登録時間になる

- 1.7.0は、100万件まで登録したが登録時間はほぼ一定
- 1.6.2は、登録済み件数に比例して登録時間は増加
 - 線形増加すると仮定すると、100万件時には1000件の追加登録に約7時間掛る

- DSpaceを用いて地上観測データのメタデータ・データベースシステムを構築中である
 - 時間・空間の範囲検索を実装した
 - メタデータ登録システムを開発した
 - DSpaceのメタデータ登録速度の性能評価を行い、バージョン1.7.0を使えば大量件数時でも運用上問題ないことを確認した
 - 外部提供インターフェースの利用を検討
 - 現在、出力要素をカスタマイズ中である