統合データベース（AMIDER）の紹介

PEDSC 特任研究員

小財正義

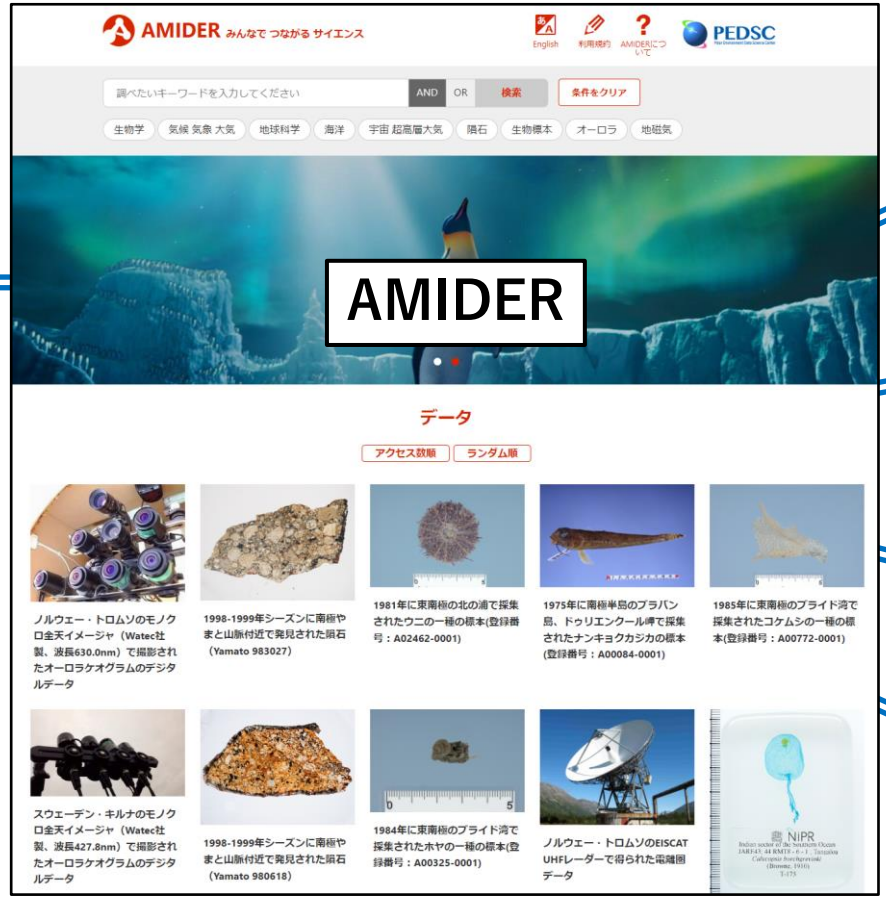
本研究は以下の支援を受けています。

- **文部科学省・国立情報学研究所**
  「AI等の活用を推進する研究データエコシステム構築事業　ユースケース創出事業」

AMIDERウェブサイトのコンセプト

- 研究データの“バーチャル博物館” or “プロモーションサイト” (観測データ, 標本データ, etc.)
- 研究データ：研究活動の直接のアウトプット⇒ダイナミックな研究活動の可視化
- 多様な科学分野へのOne-stop gateway
- オープンサイエンス・分野間連携への貢献

- ターゲット
- ✓ 専門分野外のユーザー
- ✓ 異分野との連携に興味を持つ研究者
- ✓ 教育関係者、学生



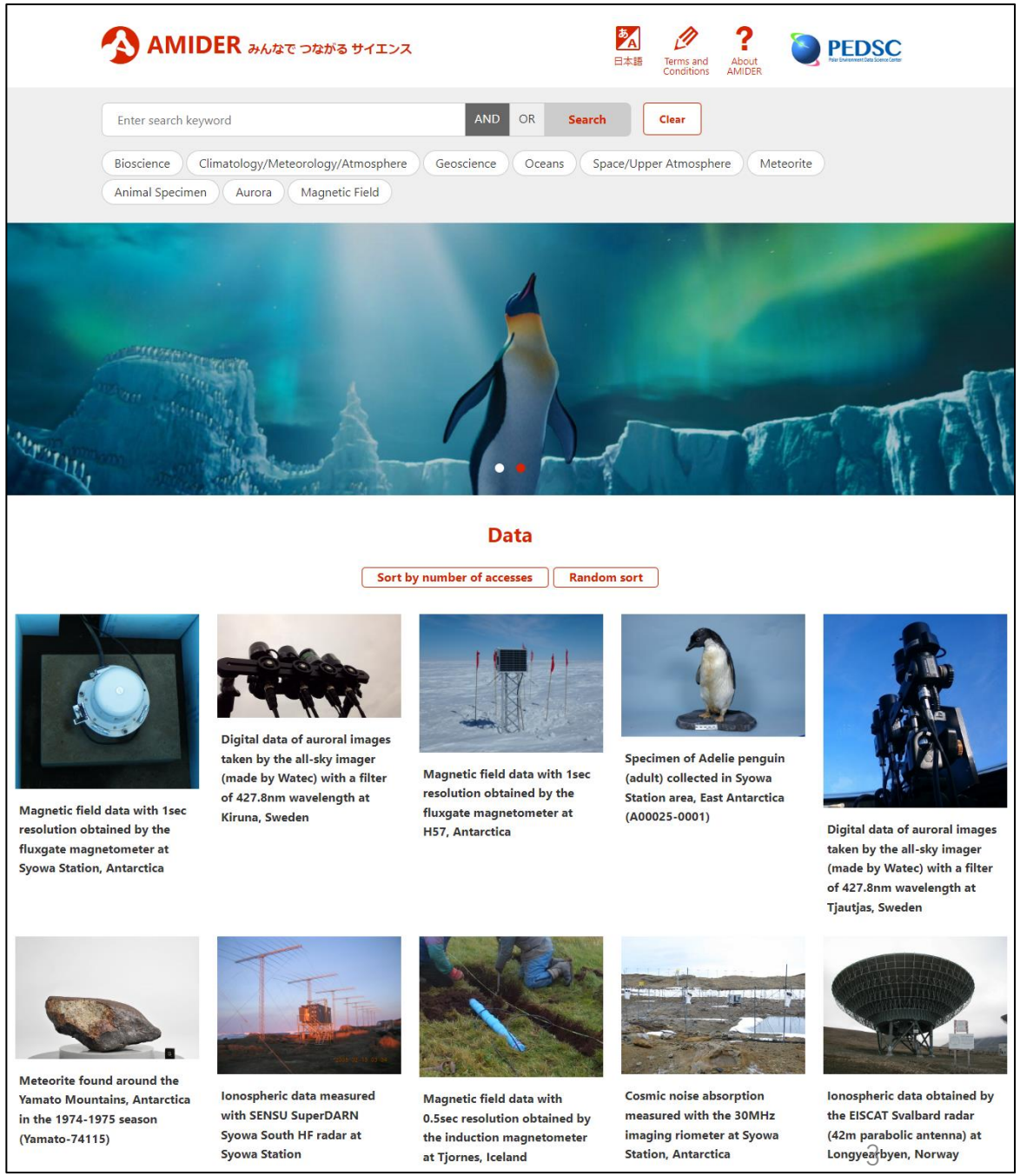
Space science

Geoscience

Bioscience

カタログ表示

- Webマーケティングで汎用的なデザイン
  - ✓ Amazon, YouTube,…
  - ✓ 研究データのプロモーションに活かせるヒント
  - ✓ サムネイル画像・スニペット
  - ✓ 多様なデータセットを一目で把握できる
- 研究データの“ブラウジング”自体を楽しめるように



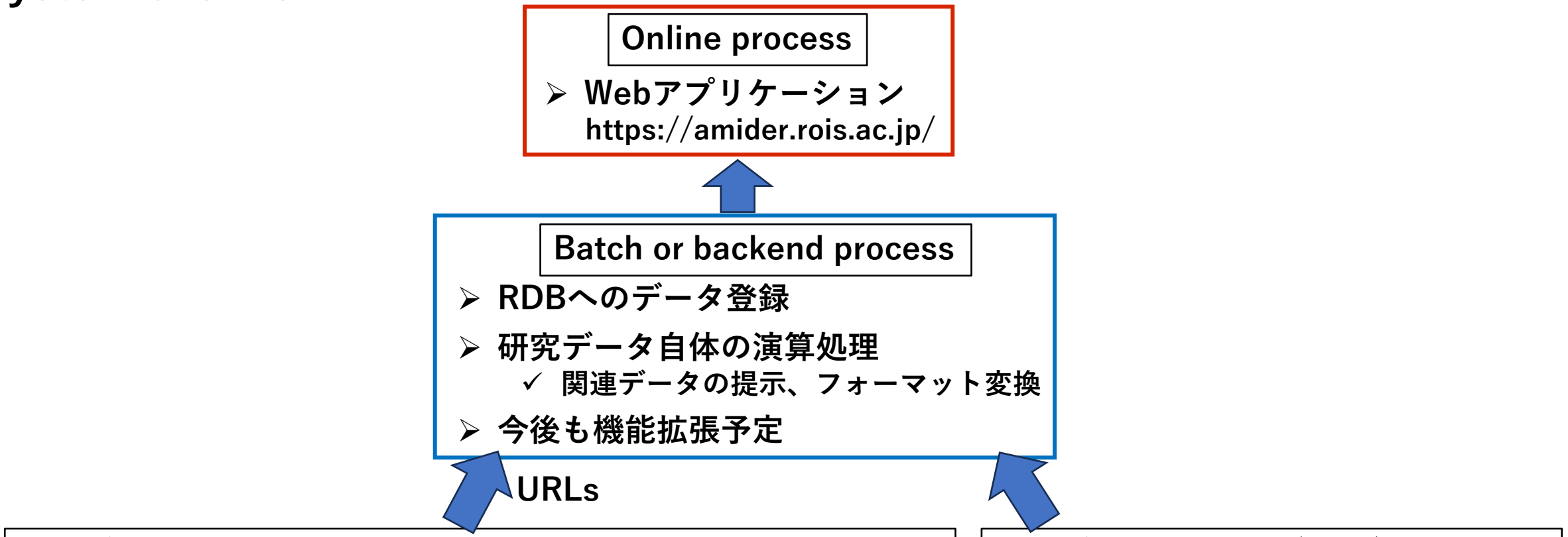
進捗状況

- 3月末頃の公開開始へ向けて準備中
  - ✓ 公開前に完了する作業/公開後のアップデートへ後回しする項目を仕分け
  - ✓ 第１次システムのフィックスへ向けたシステム改修
  - ✓ 各データセットの表示項目のチェック
  - ✓ サービス（データ受け入れなど）の運用方針の策定
  - ✓ AMIDERプロジェクト内・PEDSC内などでのレビュー

AMIDERウェブサイトのデモンストレーション

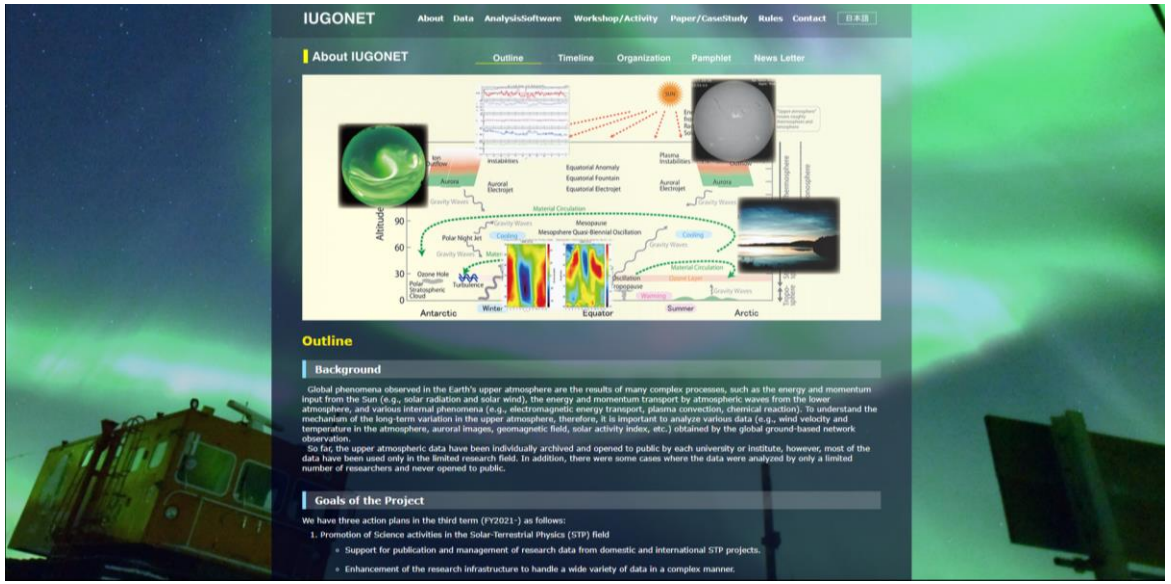
- カタログ表示
- 検索（プリセットキーワード、管理者名など）
- 各データセットページ
  - ✓ 時系列データ
    - データ有無カレンダー
    - データダウンロード
    - フォーマット変換
    - 可視化画像
  - ✓ 標本データ
  - ✓ その他数値データ
- IUGONET・生物標本・隕石標本・ADS
- 英語ページ

System overview



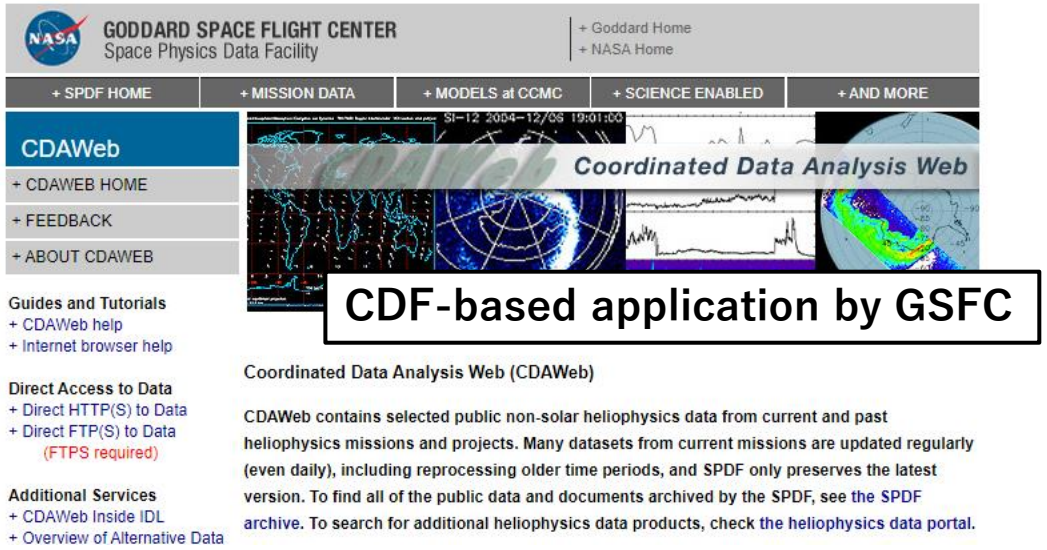
Metadata format

- SPASE (Space Physics Archive Search and Extract)
  - ✓ Managed by SPASE consortium including NASA and JAXA
  - ✓ Standard format in space physics
  - ✓ AMIDER's space science data
- Geographic information/Geomatics schema in ISO standard
  - ✓ Managed by ISO Technical Committee 211
  - ✓ One of the standard formats used by NIPR (National Institute of Polar Research) in Japan



CDF and NetCDF research data formats

- Self-describing and machine-independent binary formats
  - ✓ Interoperability between platforms is ensured
- CDF (Common Data Format)
  - ✓ Managed by GSFC/NASA
  - ✓ Standard format in space science
  - ✓ AMIDER's space science contents (aurora-image, geomagnetic, or upper-atmosphere data)
- NetCDF (Network Common Data Form)
  - ✓ Managed by NSF Unidata community
  - ✓ Standard format in meteorological science, etc.
  - ✓ AMIDER's meteorological and meteorite-sample data

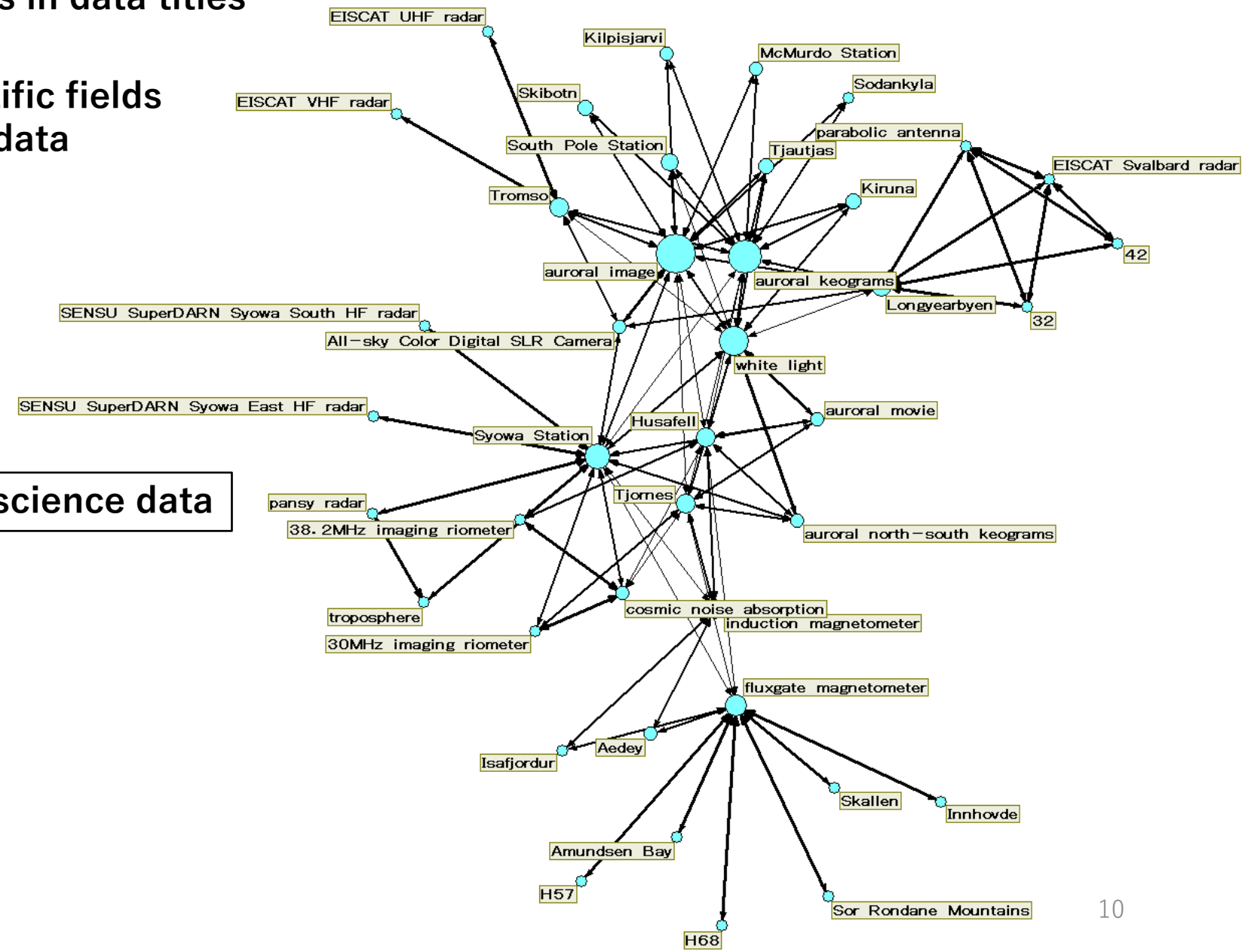


CDF-based application by GSFC

NetCDF-based data center in NOAA

今後の展望

- 公開運用の実績を積みつつシステム・サービス開発
- データキュレーションのDX（digital transformation）によるデータ提供者のサポート
  - ✓ オープンサイエンスの促進
  - ✓ データベースの品質向上
  - ✓ e.g., メタデータの自動作成
- 研究データのプロモーションに関する研究開発
  - ✓ 研究データ、研究活動、研究分野の可視化
- 自然言語処理, Text mining, 画像処理, etc.

Attempt of text mining for the research data promotion

- Co-occurrence network of terms in data titles for space science data
- Visualize the overview of scientific fields from the viewpoint of research data