

AMIDER: 極域科学を中心とした 分野横断型データカタログ

小財正義

情報・システム研究機構

データサイエンス共同利用基盤施設 (ROIS-DS)

極域環境データサイエンスセンター (PEDSC)

特任研究員

AMIDERプロジェクトチーム



大学共同利用機関法人 情報・システム研究機構

データサイエンス共同利用基盤施設

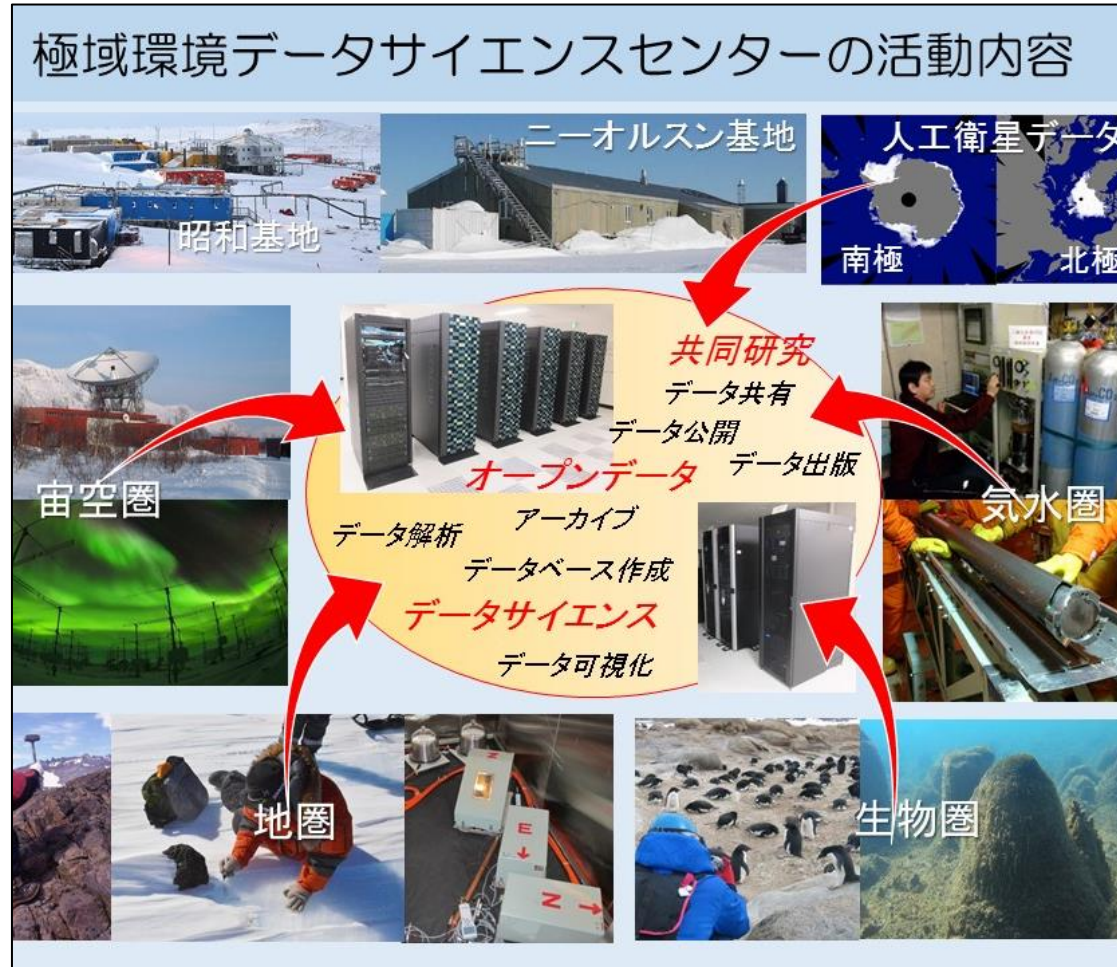
Joint Support-Center for Data Science Research (ROIS-DS)



AMIDER

本プロジェクトは文部科学省「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」の支援を受けています。

AMIDERプロジェクトの背景 - 極域環境データサイエンスセンター



pedsc.rois.ac.jp

- 極域科学を中心とするデータの管理や利活用を支援・促進
- 極域科学がカバーする多様なサイエンス（生命科学、地球科学、宇宙科学、etc）
- 分野横断型データ公開システムの開発（AMIDER）

AMIDERプロジェクト

- 分野横断型データ公開システムの開発（AMIDER←あみだくじ）
- データ公開に関する次世代型モデルケースを創出
 - ✓ 多様な科学分野のデータを包括的にデータベース・カタログ化
 - ✓ 専門外のユーザーによる科学データ利活用の障壁を無くす
 - ✓ システム管理者・データ提供者（研究者）の負担軽減
 - ✓ 外部機関との連携によるさらなる高度化
- 数人体制での実験的プロジェクト
 - ✓ 機動性を活かし、積極的に新しいコンセプトを導入・試行
- 融合研究の促進・データ利活用の促進
- 極域科学をステップとしてあらゆる分野へ展開

- 多様な科学分野へのOne-stop gateway
- 各データセットのビジビリティ・アクセシビリティ向上
- 分野横断的な「渡り歩き」へ誘導・新たな異分野連携の着想



➤ ターゲット

- ✓ 専門分野外のユーザー
- ✓ 多様な科学分野との連携に興味がある研究者
- ✓ 一般市民, 教育関係者, 学生, etc

生命科学

地球科学

宇宙科学

-
-
-

□ ユーザーインターフェースの設計

- トップページ・検索結果のカタログ表示
- ✓ ECサイト（Amazonなど）を参考にしたユーザーフレンドリーなデザイン
- ✓ 多様なデータセットを統一的デザインで表示
- ✓ コンテンツをイメージできるサムネイル画像
- ✓ 専門用語の使用を抑えたスニペット
- ✓ データの「ブラウジング」自体を楽しめるように

The screenshot shows the AMIDER website interface. At the top, there is a logo for AMIDER and navigation links for English, Help, Terms of Use, and Contact. Below the logo is a search bar with the text "調べたいキーワードを入力してください" and buttons for "AND", "OR", "検索", and "条件をクリア". A horizontal menu below the search bar contains categories: 生物学, 気候 気象 大気, 地球科学, 海洋, 宇宙 超高層大気, 隕石, 生物標本, オーロラ, 地磁気.

The main content area features a large banner image of a penguin with the text "人文知による情報と知の体系化" and "データから情報へ". Below the banner, there is a section titled "データ" with two sorting options: "アクセス数順" and "ランダム順".

The data items are displayed in a grid:

- Item 1:** Image of a monochrome camera. Caption: "スウェーデン・チャウチャスのモノクロ全天イメージャ (Watec社製、波長427.8nm) で撮影されたオーロラ画像デジタルデータ".
- Item 2:** Image of a fluxgate magnetometer. Caption: "南極昭和基地のフラックスゲート磁力計で得られた地磁気1秒値データ".
- Item 3:** Image of a monochrome camera. Caption: "スウェーデン・キルナのモノクロ全天イメージャ (Watec社製、波長427.8nm) で撮影されたオーロラ画像デジタルデータ".
- Item 4:** Placeholder image with "NoImage" text. Caption: "1973年に採集された巻貝の一種の標本(登録番号: A00995-0001)".
- Item 5:** Image of a field site. Caption: "南極昭和基地の30Mングリオメータで得られた雑音吸収 (CNA) データ".
- Item 6:** Image of a fluxgate magnetometer. Caption: "南極・アムンゼン湾 (AMB) のフラックスゲート磁力計で得られた地磁気1秒値データ".
- Item 7:** Image of a squid. Caption: "タコの一種の標本(登録番号: A02471-0001)".
- Item 8:** Image of a fluxgate magnetometer. Caption: "南極・セルロンダーネ山地 (SRM) のフラックスゲート磁力計で得られた地磁気1秒値データ".
- Item 9:** Image of a penguin. Caption: "東南極の昭和基地周辺で採集されたアデリーペンギン (成体) の標本(登録番号: A00025-0001)".
- Item 10:** Image of a sponge. Caption: "1984年に東南極のフクロノボリで採集されたカイメンの標本(登録番号: A0166-0001)".

➤ 各データセットページ

- ✓ 専門用語の使用を抑えたタイトル
- ✓ データファイルのダウンロード
 - 実験・観測データ本体
 - 科学データ専用フォーマット（バイナリ）からプレーンテキストへの変換機能
 - 現在はCDF・NetCDF（宇宙科学分野）へ対応
- ✓ 可視化画像の表示
 - 日時を指定可能
- ✓ 各表示設定・パラメータは各データセットのconfigファイル（YAML形式）で定義
 - 統一的デザインを守りつつ、多様なデータに対応

The screenshot displays the AMIDER data portal interface. At the top, the AMIDER logo is on the left, and navigation links for English, Help, Terms of Use, and Contact are on the right. The main content area features a photograph of a white fluxgate magnetometer mounted on a black base. Below the photo, the title reads "南極昭和基地のフラックスゲート磁力計で得られた地磁気1秒値データ". A red heading "データダウンロード" is followed by a form with three sections: "開始日時" (Start Date) with a date picker set to 2022-08-22, "期間" (Period) with a dropdown set to "1 day", and "フォーマット" (Format) with a dropdown set to "CDF" and a "ダウンロード" button. A "ファイル有無の表示" button is also present. Below this, a red heading "可視化画像" is followed by a "開始日時" date picker set to 2022-05-27 and a "plot" button. At the bottom, a series of five thumbnail plots are shown for dates 2022/05/27, 2022/05/28, 2022/05/29, 2022/05/30, and 2022/05/31. Each thumbnail contains multiple stacked line and heatmap plots.

➤ 関連データセットの提示 (recommend機能)

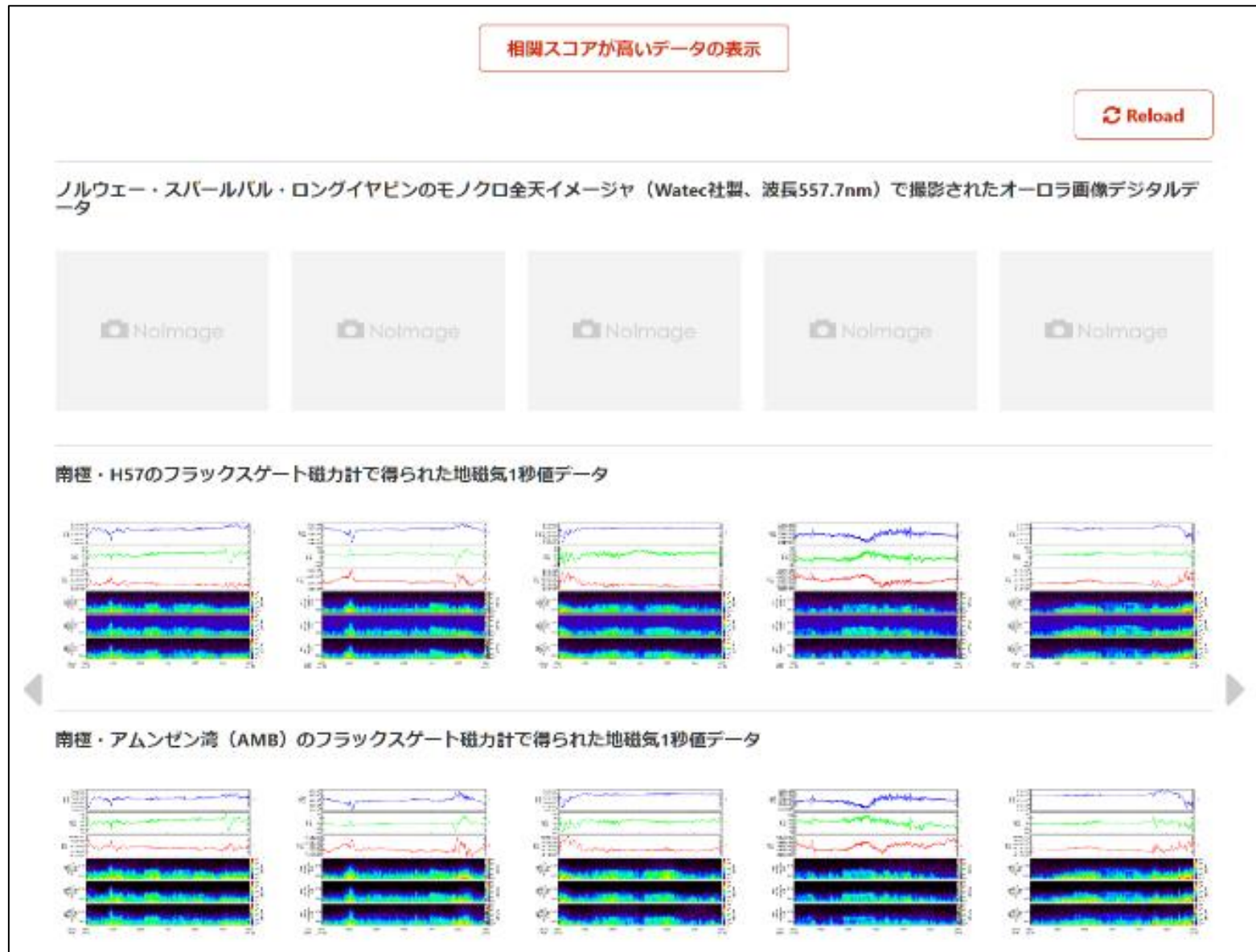
✓ 相関スコアの高いデータセットを提示

- 同時刻データの存在
- データの相関計算

✓ データセット間の「渡り歩き」へ

✓ 今後の発展

- テキストマイニング
 - タイトル、説明文、関連論文
 - トピック分析など
- ネットワーク図など全体像概観
- 外部リポジトリとの連携



➤ メタデータテーブル

✓ 専門用語の使用を抑えた表現

観測データ情報	
説明	南極昭和基地のフラックスゲート磁力計で観測された地磁気1秒値データです。 ❖ 2003年以降のデータが公開されています。オリジナルデータは0.05秒値で取得されています。
キーワード	❖ 地球科学 太陽地球相互作用 電離圏/磁気圏ダイナミクス 磁場/磁流
所有者	❖
観測開始日時	❖ 1998-01-02T00:00:00
観測終了日時	❖
観測場所	緯度（北端）：-69.00 緯度（南端）：-69.00 ❖ 経度（東端）：39.58 経度（西端）：39.58
観測領域	❖ 地球磁気圏 地球表面近傍電離圏E領域
謝辞	論文や講演等でデータを使用する前に、国立極地研究所のこのプロジェクトの担当者（uapdata[at]nipr.ac.jp）にご連絡ください。我々は共同研究を歓迎しており、また、このデータがあなたの研究に適するかどうか確認するために早期にご連絡していただくことを強くお勧めします。データを出版物及び講演で使用する場合、国立極地研究所に対する謝辞を加えていただくと共に、担当者宛てにそのPDFファイルをお送りいただけますようお願いいたします。謝辞の例は、以下の通りです。「地磁気データは国立極地研究所により提供されました。昭和基地の宙空モニタリング観測は、主に文部科学省の日本南極地域観測隊（JARE）の支援を受けています。地磁気データの公開・流通は、IUGONETプロジェクト（ http://www.iugonet.org/ ）の支援を受けています。」観測についての詳細は、文献「JARE data report, Upper atmosphere physics 26, 1-59, 2008-02」を参考にしてください。
データ処理状態	❖ null
観測機器種別	❖ 磁場

データキュレーション支援ツール（今後の課題）

- 標準形式メタデータ作成ツール
- 標準フォーマットへのデータセット変換ツール
- 自然言語処理（テキストマイニング）の活用
 - ✓ 表記揺れの抑制（例：西暦 vs. 元号）
 - ✓ メタデータから一般向け説明文の自動生成
 - ✓ その逆？
- スキーマ等を設定可能に⇒汎用ツールとして提供

データセットタイトルのテキストマイニング試行

➤ 共起表現に基づくネットワーク図

✓ 登録データセットの概観

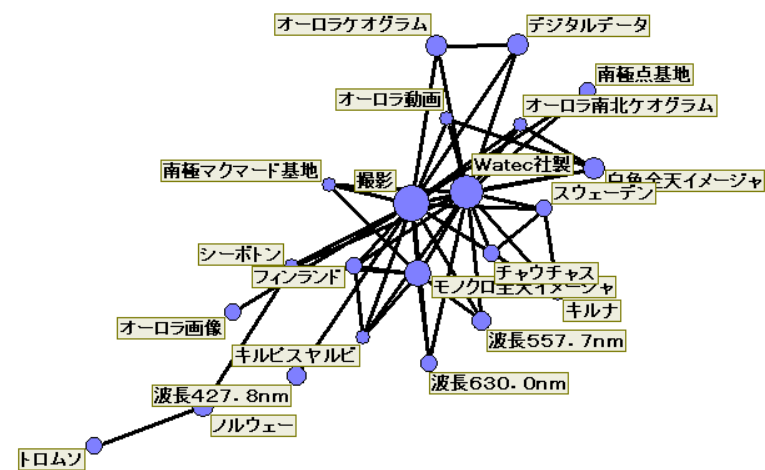
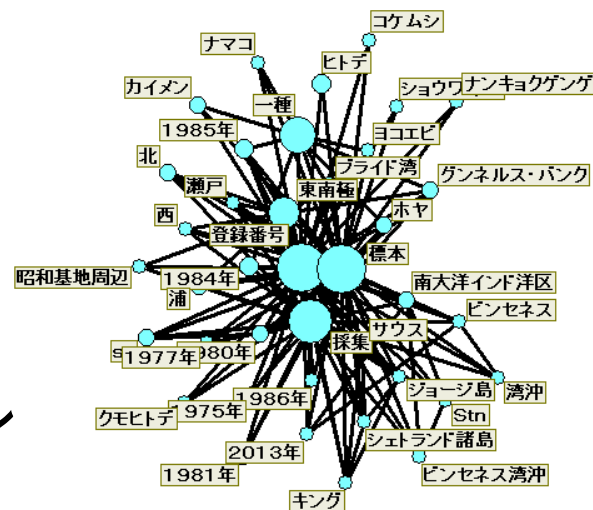
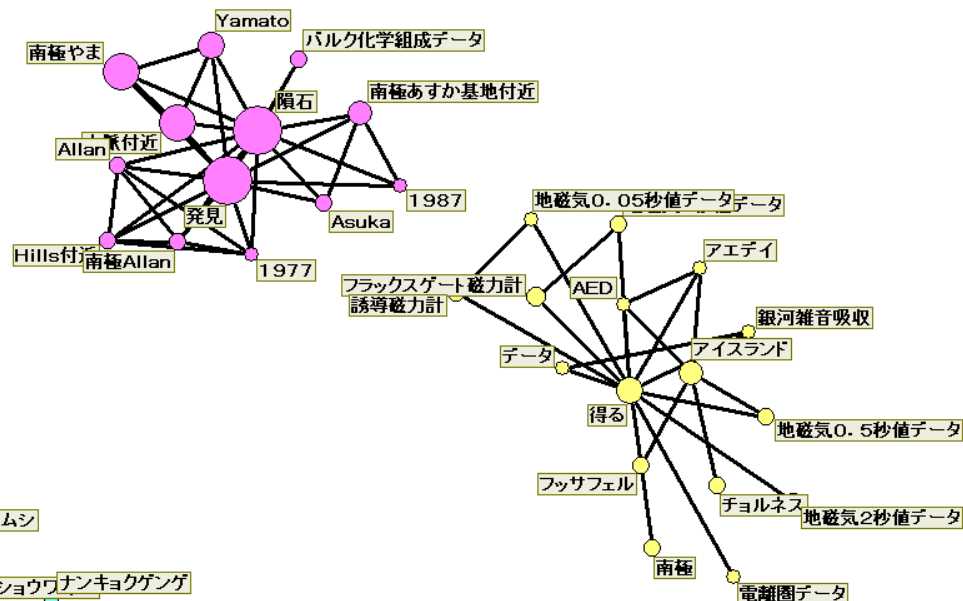
✓ テキストに基づく分野識別

- 隕石試料
- 生物標本
- 高層物理学観測（画像・その他）

✓ タイトル入力時の表記揺れ

- 「得る」 ⇔ 「撮影」
- メタデータ改善点の発見

⇒ データキュレーション支援ツール



スケジュール

➤ 進捗状況

- ✓ Webアプリケーションの初版完成
- ✓ サーバー構築作業中（ほぼ完了）

➤ 今後

- ✓ 間もなく限定的範囲での運用開始予定（同センター内など）
- ✓ 運用体制整備
- ✓ 一般公開へ向けて段階的に公開範囲拡大予定
- ✓ 関係各所での調整・宣伝
 - データ提供者・研究者のレビュー
 - 学会などでの宣伝
- ✓ より発展的な開発
 - テキストマイニング
 - キュレーション支援ツール
 - 外部専門家との連携