

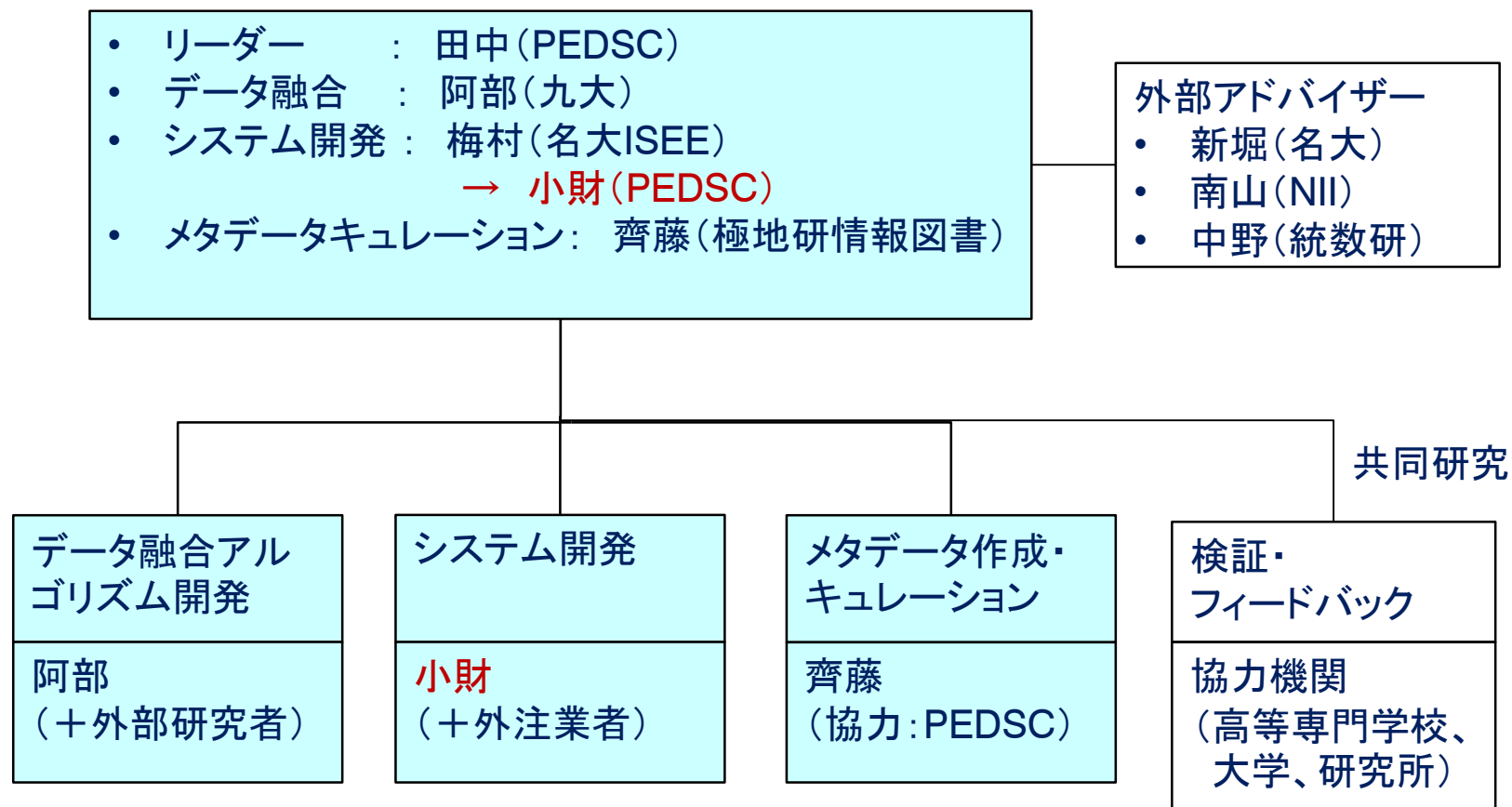
統合データサイエンスプラットフォームの現状

田中良昌^{1,2,3}, 阿部修司⁴, 小財正義¹, 齊藤泰雄²,
南山泰之⁵, 新堀淳樹⁶, 中野慎也⁷

1. 極域環境データサイエンスセンター, 2. 極地研, 3. 総研大
4. 九州大ICSWSE, 5. 情報研, 6. 名大ISEE, 7. 統計数理研究所

1. AMIDERプロジェクトの目的
2. プロジェクトの体制
3. 統合データサイエンスプラットフォームの概要
4. 統合データサイエンスプラットフォームの現状
5. まとめ

- 極地研が所有する多種多様な極域データ(宙空圏、気水圏、地圏、生物圏)のメタデータ、及び、実データを統合的に扱うことができ統合プラットフォームを開発する。
- 将来的にROISや全国の研究機関や大学の所有する幅広い分野のデータへ展開するための手法・ノウハウを構築し、オープンデータ、オープンサイエンスの促進に貢献する。



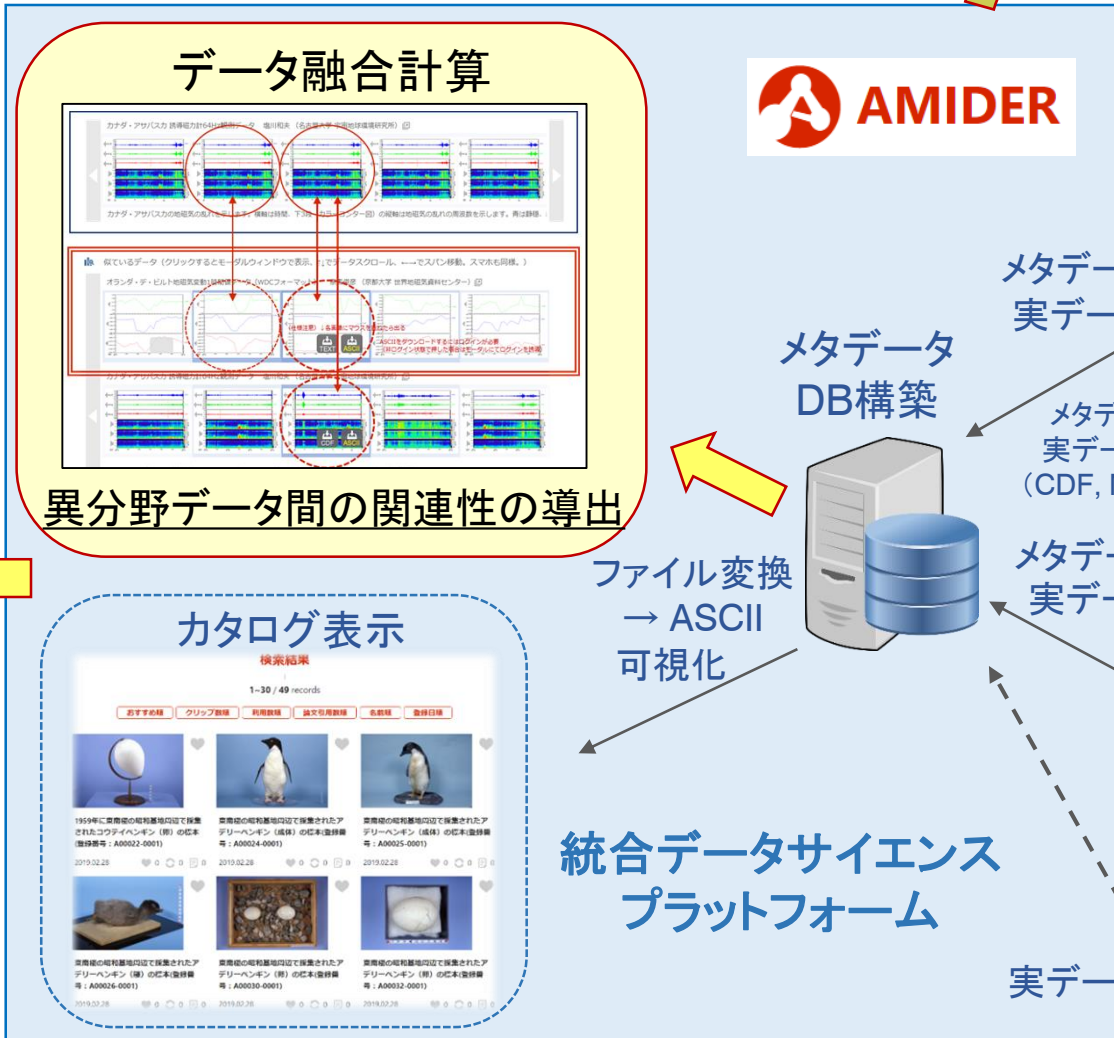
3. 統合データサイエンスプラットフォームの概要

未知の現象の発見
新しい学術分野の
創出へ

ユーザ
(研究者、
一般市民)

データの
相関情報

- 実データ
 - ASCII
 - CDF
 - NetCDF
- カタログ
 - 可視化画像
 - メタデータ



新たな知見
(相関情報)

データ提供者
(NIPRデータPI)

メタデータ
実データ

ドメインDB
• IUGONET
• ADS
•

実データ

リポジトリ
(NII)

4. 統合データサイエンスプラットフォームの現状

AMIDERトップページ(日本語版)

<https://amider.jp>

AMIDER

日本語 ヘルプ サインイン

カテゴリ 何についてしらべたいですか? 検索 条件をクリア

人文知による情報と知の体系化
データから情報へ
情報から知へ
知の触発によって見えるものは何か

データ
Data

2011-09-30 23:30:00
南極昭和基地の白色全天イメージャ (Watec社製) で撮影されたオーロラ動画
2014.03.12

アイスランド・フッサフェル (HUS) の誘導磁力計で得られた地磁気0.5秒値データ
2013.07.23

ノルウェー・スバルバルの EISCATスバルバルレーダー (42mパラボラアンテナ) で得られた電離圏データ
2011.09.22

南極昭和基地のSuperDARN Syowa Eastレーダーで得られた電離圏データ
2011.04.01

アイスランド・フッサフェル (HUS) のフラックスゲート磁力計で得られた地磁気0.5秒値データ
2011.04.01

異分野データを同一システムで管理。

日・英両言語に対応。

4. 統合データサイエンスプラットフォームの現状

検索機能(日本語版)

AMIDER

日本語 ヘルプ サインイン

カテゴリ ペンギン 検索 条件をクリア

日本語、英語による
フリーワード検索

検索結果
1~30 / 49 records

おすすめ順 クリップ数順 利用数順 論文引用数順 名前順 登録日順

検索結果の画像表示

1959年に東南極の昭和基地周辺で採集されたコウテイペンギン(卵)の標本(登録番号: A00022-0001)
2019.02.28

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: A00024-0001)
2019.02.28

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)
2019.02.28

東南極の昭和基地周辺で採集されたアデリーペンギン(雛)の標本(登録番号: A00026-0001)
2019.02.28

東南極の昭和基地周辺で採集されたアデリーペンギン(卵)の標本(登録番号: A00030-0001)
2019.02.28

- 異分野データをまとめて検索、表示することができる統合プラットフォームを構築。
- キーワードやカテゴリ検索で異種データを抽出することで、隣接分野データへの接続を促す。
- データの登録数、利用数を統計・可視化し、データPIIに提供(インセンティブの提供)。

4. 統合データサイエンスプラットフォームの現状

メタデータ表示機能(日本語版)

南極昭和基地の誘導磁力計で得られた地磁気0.05秒値データ

登録日 2019.07.24

クリップ追加

利用できる範囲
Scope of Application

【自然科学】 天文学、宇宙科学

概要
Outline

南極昭和基地の誘導磁力計で観測された地磁気0.05秒値データです。2003年以降のデータが公開されています。

用途
Data Usage

地磁気データから、宇宙起源の約0.1秒~数百秒周期の地磁気の変動を知ることができます。主に共同研究に利用することが可能です。

データの用途など、AMIDER独自のメタデータの追加。

- メタデータフォーマットとしてISO19139、SPASEを採用し、ドメインのサイエンス情報を保持。
- AMIDER独自のメタデータを追加可能。

観測データ情報	
Observation Data Information	
説明	南極昭和基地の誘導磁力計で観測された地磁気0.05秒値データです。2003年以降のデータが公開されています。
所有者	情報・システム研究機構 (ROIS) 国立極地研究所 (極地研) 門倉 昭 *研究代表者 +81-42-512-9105 kadokura [at] nipr.ac.jp
観測開始日時	1998-02-01T00:00:00
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

日本語、英語による分野特有メタデータの表示



4. 統合データサイエンスプラットフォームの現状

解析の強化(データ融合 I)

データダウンロード
Data Download

Date	URL	Download
2003/02/04	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030204_v01.cdf	CDF ASCII
2003/02/05	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030205_v01.cdf	CDF ASCII
2003/02/06	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030206_v01.cdf	CDF ASCII
2003/02/07	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030207_v01.cdf	CDF ASCII
2003/02/08	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030208_v01.cdf	CDF ASCII
2003/02/09	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030209_v01.cdf	CDF ASCII
2003/02/10	http://iugonet0.nipr.ac.jp/data/imag/syo/20hz/2003/nipr_20hz_imag_syo_20030210_v01.cdf	CDF ASCII

可視化データ
Visualized Data

実データを、CDF、NetCDF、GCMD準拠ASCII等の機械可読のフォーマットで公開。

ASCII形式への変換・ダウンロード機能。

QLプロットの表示機能。

- 実データを、分野標準フォーマット and/or 機械可読フォーマット (CDF、NetCDF、GCMD準拠ASCII)で公開し、解析へ誘導。
- 上記実データをASCIIフォーマット変換・ダウンロードする機能を実装。
- QLプロットを自動生成、並べて表示し、現象の発見を支援。

4. 統合データサイエンスプラットフォームの現状

解析の強化、データ駆動型科学の創出(データ融合) II

データ融合計算結果の例:

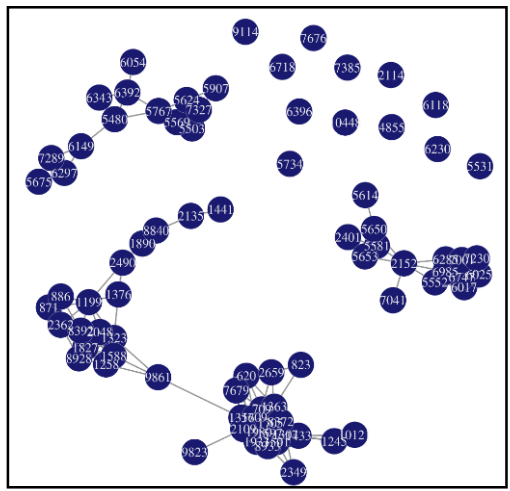
時間連続 vs 時間連続

	A	B	C	D	E	F
1	Base Data					
2	DATASETID: 13453, DATASET_NAME: 10-minute averaged meteorological data (Solar Radiation) obtained at Kizahashi Hama, Skarvsnes on Soya Coast, East Antarctica					
3	PI: Kudoh Sakae (NIPR), FIELD: Natural sciences, FIELD_DETAIL: Earth sciences. Geology, LAT/LON (N, S, E, W): -69.473611, -69.473611, 39.611944, 39.611944					
4						
5	Correlations * The same field data is displayed at the top rows. If you want to revert to the default sort, use column 'A'.					
6	SE	DATA	DATASET_NAME	CORR_20141219	CORR_20141220	CORR_20141221
7	1	13456	10-minute averaged meteorological data (UV) obtained at Skallen on Soya Coast, East Antarctica	0.979567	0.962917	0.940498
8	2	13453	10-minute averaged meteorological data (UV) obtained at Kizahashi Hama, Skarvsnes on Soya Coast, East Antarctica	0.959774	0.965356	0.949691
9	3	13456	10-minute averaged meteorological data (PAR) obtained at Skallen on Soya Coast, East Antarctica	0.973125	0.963216	0.914490
10	4	13456	10-minute averaged meteorological data (Solar Radiation) obtained at Skallen on Soya Coast, East Antarctica	0.970996	0.972709	0.914850
11	5	13450	10-minute averaged meteorological data (UV) obtained at Yukidori zawa, Langhovde on Soya Coast, East Antarctica	0.970750	0.945310	0.943949

	A	B	C	D	E
1	Base Data				
2	DATASETID: 620, DATASET_NAME: Meteorite found around the Asuka Station, Antarctica in the 1986-1987 season (Asuka-8602)				
3	PI: Yamaguchi Akira (NIPR), FIELD: Natural sciences, FIELD_DETAIL: Earth sciences. Geology, LAT/LON (N, S, E, W): -71.7200, -73.00, 28.1700, -21.6700				
4					
5	Correlations * The same field data is displayed at the top rows. If you want to revert to the default sort, use column 'A'.				
6	SE	DATA	DATASET_NAME	CORR	URL
7	1	7679	Meteorite found around the Yamato Mountains, Antarctica in the 1986-1987 season (Yamato-86036)	0.999359	https://amider.jp/data/7679
8	2	2655	Meteorite found around the Meteorite Hills, Antarctica in the 1978-1979 season (Meteorite Hills-78028)	0.999134	https://amider.jp/data/2655
9	3	732	Meteorite found around the Asuka Station, Antarctica in the 1987-1988 season (Asuka-87196)	0.998208	https://amider.jp/data/732
10	4	757	Meteorite found around the Asuka Station, Antarctica in the 1987-1988 season (Asuka-87222)	0.997874	https://amider.jp/data/757
11	5	633	Meteorite found around the Asuka Station, Antarctica in the 1987-1988 season (Asuka-87029)	0.997783	https://amider.jp/data/633

異なる分野のデータを比較し、分野を跨いだ新しい発見を支援する。

成分vs成分



南極で取得された各隕石のバルク成分データの関連性の表示

- 互いのデータの関連性の分かり易く表示する方法を検討中。
- 異種データに相関計算を適用するためのデータの**前処理の一般化・標準化**が課題。

公開に向けての改良・開発

ユーザインターフェース(UI)の改良

- 実験的に実装したユーザアカウント登録等の機能は、公開後の管理コストが大きいため、長期安定運用に適したUIに改良を行う。
- 一般市民を主な対象としてデータの説明等を大幅に簡略化したため、研究者のニーズに寄せた表示に修正を行う。

実装が遅れている機能の開発

- データの一括ダウンロード
- 可視化データの作成・表示
- 関連の高いデータの表示
 - 今年度中の完成に向けて、開発中。

4. 統合データサイエンスプラットフォームの現状

プロモーション、データ掘り起し(PEDSCとの連携)

実データのファイルフォーマットの標準化は高コスト

実データ、メタデータ作成数

分野	データ種	実データ数	フォーマット	メタデータ数	フォーマット
宙空	オーロラ	33	CDF, JPEG	33	SPASE
	地磁気	24	CDF	24	SPASE
	レーダー等	8	CDF	8	SPASE
気水	昭和基地気水モニタリングデータ	4	NetCDF	4	ISO19139
	スバルバル気水モニタリングデータ	4(*)	NetCDF	4	ISO19139
	氷床コア	18(*)	未定	18	ISO19139
	その他	9(*)	未定	9	ISO19139
地圏	地震データ	6(変換中)	CDF	6	SPASE
	インフラサウンド	16(変換中)	CDF	16	SPASE
	磁場・重力	19(*)	CDF	19	ISO19139
	隕石バルク成分	1168	ASCII	1168	ISO19139
	隕石標本			10563	ISO19139
生物	宗谷海岸露岩域気象データ	27	NetCDF	27	ISO19139
	海洋観測データ	(*)	NetCDF	4	ISO19139
	スバルバル環境データ	(*)	NetCDF	8	ISO19139
	生物標本			2678	ISO19139
合計		1,309		14,575	

(*): 2021年度後半に作成予定。

- PIIに個別にインタビューを行い、適切なフォーマット、公開方法を検討。
- 実データは、分野標準のCDF、NetCDF、及び、GCMD準拠ASCII等の機械可読フォーマットで公開。
- メタデータは、キュレーションチームがフォーマットを提案。

4. 統合データサイエンスプラットフォームの現状

NII-RCOSとの連携

- 2020年度から「次期JAIRO Cloud実証実験」に参加。
- 極地研のデータを次期JAIRO Cloudで公開する試みを開始している。

NII RCOS
Research Center for Open Science and Data Platform

test tree
test

隕石成分データの例

データのプレビュー画像

Preview

In [1]:
In [2]:
In [3]:
In [4]:
Out[4]:

<matplotlib.collections.PathCollection at 0x1255ee0d0>



Name / File	License	Actions
<input type="checkbox"/> Kimura 2018 Bulk-Suppl example-11.csv (142.8 kB) sha256: 5e02d2d3a8b505e039fa15eb50b28504c33a6c7f5cfa923656b881f7825c3762		<input type="button" value="Download"/> <input type="button" value="Information"/>
<input type="checkbox"/> nipr-meteorite.ipynb (19.6 kB) sha256: af7dd69024b918e14c013dbf5afaf3436bf89f7cdc872f543bef41011d936c14	<input type="button" value="License"/>	<input type="button" value="Download"/> <input type="button" value="Information"/> <input type="button" value="Preview"/>

ファイルのダウンロード

<https://w3-pedsc.ir.rcos.nii.ac.jp/records/2>

17

views

total

[See details](#)

Versions

Ver.8	2020-09-11 23:23:51.382060
Ver.7	2020-09-11 23:22:13.467164
Ver.6	2020-09-11 23:20:01.834892
Ver.5	2020-09-11 23:19:08.387333

[Show All versions](#)

Share

Cite as
test, 2020.

ランディングページURL (DOIも可)

利用統計

バージョン管理

- 極域データに代表される多様な異種データを統合的に扱えるプラットフォームのベースが完成しつつある。
- UIの改良や、実装が遅れている機能(データ一括ダウンロード、可視化データ表示、関連データ表示)の実装を行い、今年度中のシステムの完成を目指す。
- 極地研の多様なデータの実データ公開とメタデータの付与を継続する。

今後の展望

- 2022年度中にROIS内公開を経て、できるだけ早い時期に一般に公開する。
- ROIS-DS共同研究を通じて、本研究で構築したシステム、手法、ノウハウをROIS-DSの各センターや他大学・研究機関が所有する多様なデータに適用する。