

統合データサイエンスプラットフォームの現状と今後の計画

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

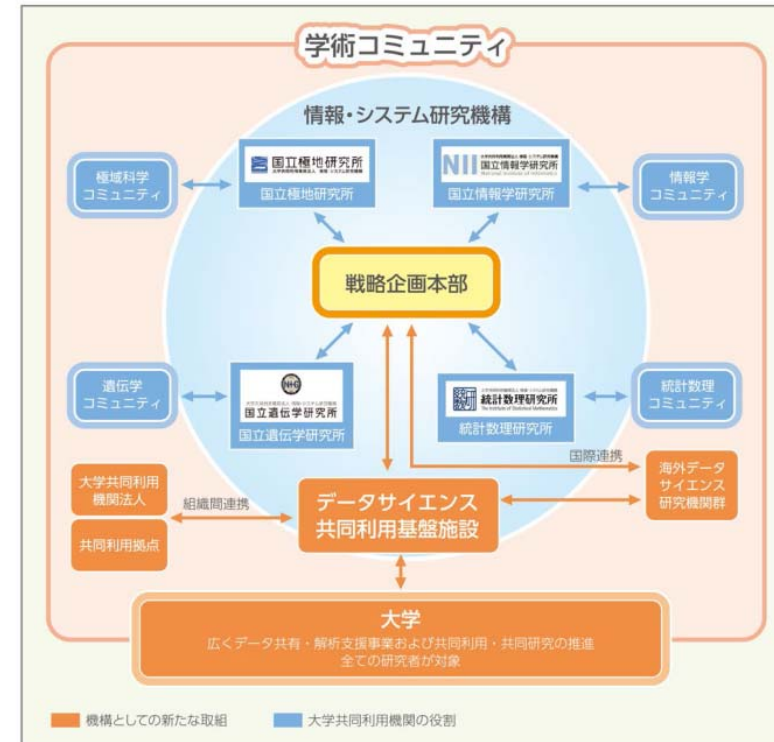
1. 極域環境データサイエンスセンター, 2. 極地研,
3. 名大ISEE, 4. 九州大ICSWSE, 5. 東京財団政策研究所,
6. 統計数理研究所

背景



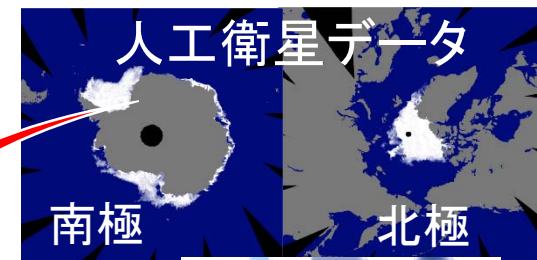
三位一体の活動により、大学のデータ駆動型研究を積極支援

- 1 データ共有支援 生命、地球環境、社会、人文学における大規模データの共有支援
- 2 データ解析支援 最先端の解析支援、分野を横断する高度な数理的手法の開発と支援
- 3 データサイエンスのための人材育成 大学等におけるデータ駆動型の研究と異分野交流・新分野創成の促進



- ライフサイエンス統合データベースセンター
..... 生命科学分野のデータベース統合化によるオープンサイエンスの推進
- 地球環境データサイエンスセンター (平成29年度 設置予定)
..... 極域の環境データに関するデータベースの整備及び共同利用の推進
- 社会データ構造化センター
..... 社会調査データ、公的統計ミクロデータ、ソーシャルビッグデータのデータ共有の推進
- 人文学オープンデータ共同利用センター (平成29年度 設置予定)
..... データサイエンスに基づく人文学(人文情報学)の創生及びデータのオープン化の推進
- ゲノムデータ解析支援センター
..... バイオインフォマティクス技術を駆使したゲノムデータの解析支援
- データ融合計算支援プロジェクト
..... データ融合計算技術を駆使した諸科学・産業界における課題の解決支援

極域データの統合データベース



極地研の多種多様な
データの検索、データ
取得、描画、解析等が
可能なデータベース



ROIS-DS施設の目的

- データサイエンスの観点から、大学等の多様な分野の研究者に対し、大規模データ共有支援事業およびデータ解析支援事業と人材育成を行なうことで、我が国の大学等の研究力強化に貢献するとともにオープンサイエンスの実現を目指す。
- また、研究コーディネータによる学会ブース展示等により、事業紹介や相談受付、解析手法のアドバイスをを行い、大学等のデータ駆動型学術研究を支援する。

AMIDERプロジェクト

- 2年間の機構長裁量プロジェクト(2018～2019年度)

メンバー:

- 田中良昌(PEDSC)、梅村宜生(名大)、阿部修司(九大)、齊藤泰雄(極地研情報図書)、南山泰之(東京財団政策研究所)、新堀淳樹(名大)、中野慎也(統数研)

目的:

- 異種・異分野データを統合的に扱うことができるデータベースを開発し、将来的に、ROISの各研究所、及び、全国の研究機関や大学の所有する多種多様なデータに展開する。
- 異種・異分野のデータ間の関連性を計算機により導出し、新たな知見を得るデータ融合の手法を確立する(データ駆動型科学(Data-Driven-Science)の推進)。

ゴール:

- 統合データサイエンスプラットフォームをPEDSCが公開する多種多様な極域科学データに応用し、データサイエンスを全国に展開するための手法・ノウハウを確立する。

統合データサイエンスプラットフォーム(1)

<https://amider.jp>

ターゲット:

- 研究者、及び、一般市民(主として教員や大学生・高校生)。
- 一般市民により重心を置く。

一般市民の興味を引く写真ベースの表示

日・英両言語に対応

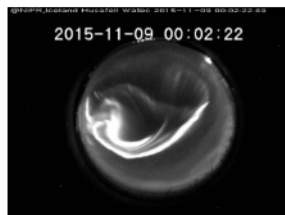
データ
Data



Image	Title	Date	Interactions
	南極昭和基地のSuperDARN Syowa Southレーダーで得られた電離圏データ	2011.04.01	0
	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	0
	南極ドームふじの氷床コアの気温復元 (過去34万年間) の初期結果	2018.06.13	0
	アイスランド・フツサフェル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	0
	南極・H57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	2011.04.01	0
	南極・昭和基地フラックスゲート磁力計で得られた地磁気2秒値データ	2011.04.01	0
	アイスランド・アエディ (AED) のフラックスゲート磁力計で得られた地磁気2秒値データ	2011.04.01	0
	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	2013.07.09	0
	南極・インホブデ (IHD) 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	2011.04.01	0
	アイスランド・チヨルネス (TJO) の誘導磁力計で得られた地磁気2秒値データ	2013.07.23	0

現在、いくつかの極地研究所が所有する研究データ(宙空圏、気水圏、地圏、生物圏等)を登録し、関係者公開中。

統合データサイエンスプラットフォーム(2)



アイスランド・フッサフェル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



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

2013.07.09



アイスランド・イーサフィヨルズウル（ISA）の誘導磁力計で得られた地磁気2秒値データ

2013.07.23



ASCIIデータをダウンロード機能、可視化グラフの表示機能実装

観測領域	<ul style="list-style-type: none"> 地球磁気圏 地球表面近傍電離圏E領域
謝辞	<p>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資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。データが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。</p>
データ処理状態	<ul style="list-style-type: none"> 未校正
観測機器種別	<ul style="list-style-type: none"> 磁場

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

書籍・出版物情報

解析の強化、データ駆動型科学の創出を目指す

可視化グラフ投入：8月初旬

2019/08/27

極域のオープンデータ・オープンサイエンスに関する研究集会

6

統合データサイエンスプラットフォーム(3)

現時点で登録済みのメタデータの種類:

- (a) 地球物理系観測データ: 高層大気44
- (b) 地球物理系採取データ: 気水圏24
- (c) 鉱物等標本系データ: 生物標本2818、隕石標本10563

成分含有率等の試料系
ヒストグラムデータの掘り起し:
9月下旬

2019年度8月現在、実データの公開があまり進んでおらず、登録数が少ない状況。
今後、年末にかけて作業を加速する計画。

検索結果

91~120 / 10563 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



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74063)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74071)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74078)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74085)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74092)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74093)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74094)

2019.02.28



1974-1975年シーズンに南極やまと山脈付近で発見された隕石 (Yamato-74095)

2019.02.28



南極の昭和基地周辺で採集されたアデリーペンギンの標本(登録番号: A00028-0001)

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の標本(登録番号: A00029-0001)

2019.02.28



1980年に東南極のグエーシー基地周辺、シャリー島で採集されたアデリーペンギン(卵)の標本(登録番号: A00031-0001)

2019.02.28



南極の昭和基地周辺で採集されたアデリーペンギンの標本(登録番号: A00022-0001)

2019.02.28



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



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

データPIの方は、ご協力をお願いします！

統合データサイエンスプラットフォーム(4)



AMIDER
XXXX

田中 良昌

変更

お気に入り
Favorite



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



ノルウェー・トロムソのEISCAT VHFレーダー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

AMIDER サインイン

Email

パスワード

サインイン

AMIDER 新規アカウント作成

姓 名

Email

パスワード パスワード確認

アカウントを作成する

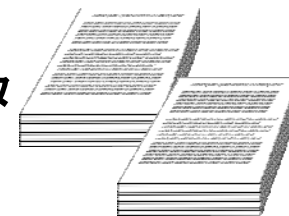
他にも、以下のような機能を有する。

- ユーザアカウントを作成して、お気に入りデータを登録
- データ所有者へのデータ利用申請
- データの参照数、登録数、利用数を統計化・可視化してPIIに提供
- 似ているデータの表示(2019年度開発)

データ提供者(PI)が得るリターンと問題点

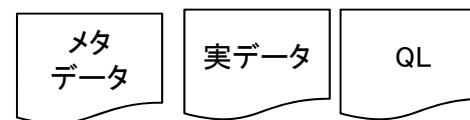
得られるリターンは？

- データ検索、描画、実データダウンロード、類似データ表示サービス
- 新たなデータユーザ(研究者、教育関係者)
- 研究成果のクレジット(共同研究論文の増加)
- ステークホルダーや一般市民への成果公開・アウトリーチ
- 論文投稿時のデータ公開の省略(近年は、データ公開が義務化していることが多い。)



データ提供者が行うべき作業

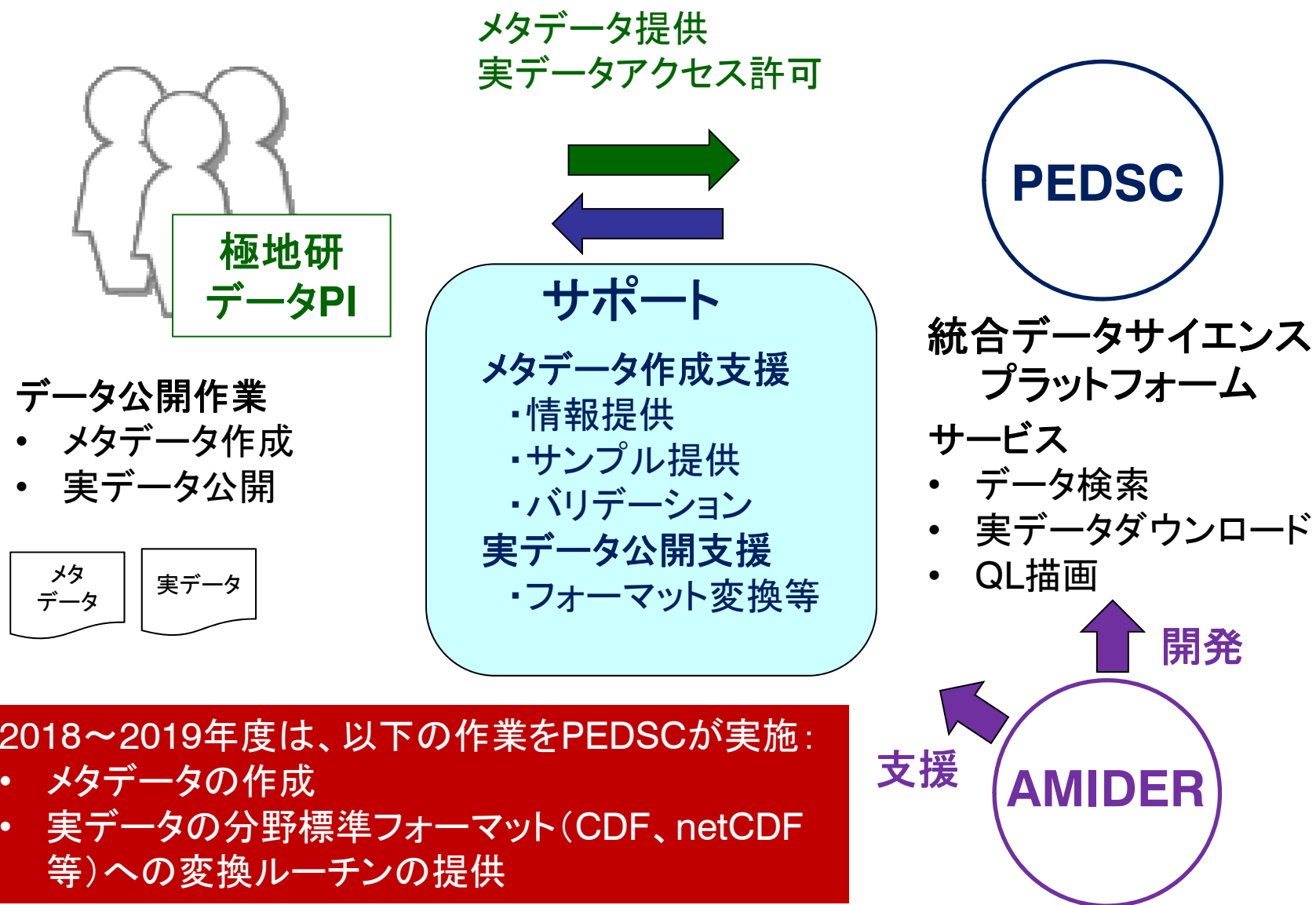
- メタデータの作成
- 実データの公開
- クイックルック(QL)の公開(任意)



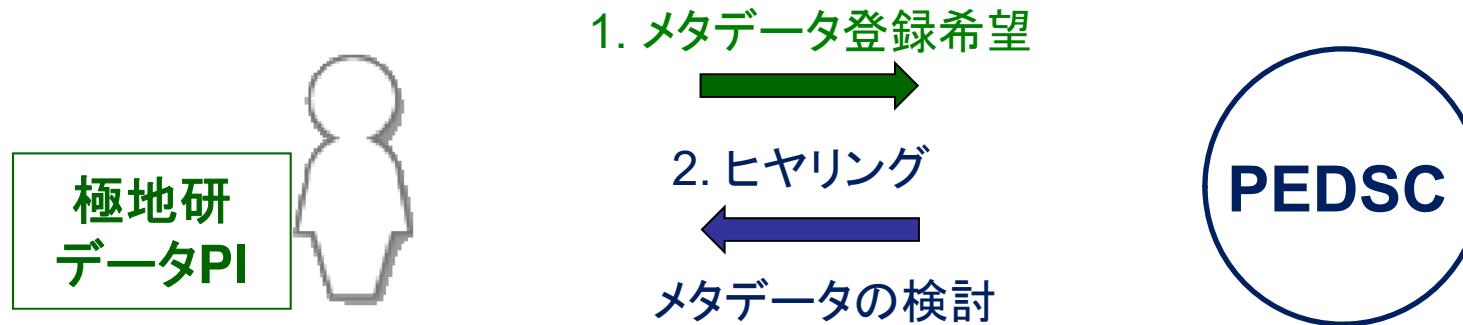
問題点

- マンパワー、資金が足りない。
- やり方が良くわからない。

データ提供者(PI)の作業と得られるサポート



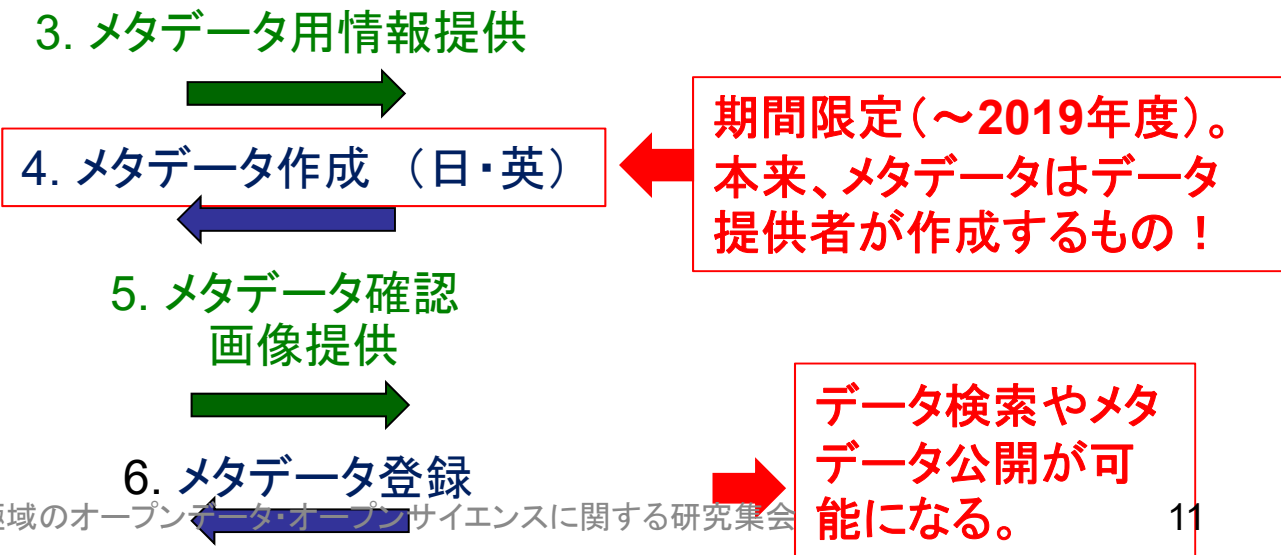
作業フロー(メタデータ登録)の例



既にメタデータがある？

Yes → 基本的に、それをそのまま使う。

No → 分野特有のメタデータフォーマットがあれば、それを利用。
無ければ、既存のフォーマット(DIF, ISO19139, spase)を利用。



まとめ

- 統合データサイエンスプラットフォームは、極地研の多種多様なデータを統合的に扱えるサービスを提供します。ターゲットは、一般市民（特に、教員や大学生・高校生）を含みます。
- 上記プラットフォームにより、共同研究の促進、研究成果の発信等、多くのメリットがあります。
- 2019年度8月現在、いくつかの極域データを登録し、関係者公開を実施中です (<https://amider.jp>)。
- 今なら、（本来、データ提供者が行うべき）**メタデータ作成や実データの公開をPEDSCが支援します**ので、ご協力をよろしくお願いします。（～2019年度まで期間限定！）