

# ADSにおける可視化アプリケーションの 構築とデータ共有について

杉村 剛

極地研・北極域データアーカイブシステム (ADS)

# はじめに

- 我々は、北極域に関わる様々なデータセットの収集・保管・公開するための基盤システムとして、北極域データアーカイブシステム (ADS) の構築を行っている。
- ADSでは、「誰でも容易に望んだ情報が得られるシステム」を目指しています。
- データベースは、データの検索・取得を手軽に行えることが魅力

だがしかし、そのデータベースは

本当にユーザが望む情報を与えられていますか？

ユーザに過剰な時間と労力を強いていませんか？



# データベース

- 一般的に、データベースは、
  - データの取得・作成・登録に関する情報をメタデータとして公開
  - メタデータの情報に基づき検索システムを構築
  - 検索されたメタデータに付随する実データ等をダウンロード可能にしている。
- ユーザ側は必ずしも取得に関する情報からデータを探す訳ではない
  - 2018年に南極昭和基地で測った大気温度 ⇒ OK
  - 2010年代に立川に大雨をもたらした時の大気温度 ⇒ NG
    - ✓ 2010年代に立川の大気温度を持っているデータセット：検索可能
    - ✓ 立川での大雨という情報：**検索不可能**
- データが内包する情報はデータ取得とは直接関係ないので、従来のデータベースでは検索できない



# データからの情報抽出



- 「2010年代に立川に大雨をもたらした時の大気温度」を得るためには
  1. データベースで「2010年代の立川」の降水量データを検索
  2. 実データをダウンロードし、可視化ソフトで降水量を可視化  
データの種類に合った可視化ソフトが必要
  3. 上記日時のうち、大雨をもたらした日時を抽出  
時間を変えながら可視化を繰り返す必要がある
  4. データベースで「抽出した日時の立川」の気温データを検索
  5. 実データをダウンロードし、可視化ソフトで気温を可視化  
降水量と違うフォーマットの可能性も

⇒ステップ数は少ないが、実はけっこうな労力と時間が・・・

現状では、ユーザの努力に頼っている部分が多いが、  
データを提供する側は、これらの**コストを軽減する努力**をすべき

# 解決案

- ビッグデータ解析
  - 登録されているデータセット全てに対してデータ解析を実施し、内包する情報自体もデータベース化
  - ✓ 開発側に資金と労力が必要
  - ✓ ユーザ毎の異なる要求全てに応えることは困難
- 知的可視化
  - データから特異点抽出、類型化、構造把握などを自動的に行い、可視化・画像化したものをデータベース化する
  - ✓ 開発に専門的な人材が必要
  - ✓ それぞれの分野の専門的な知識が必要
  - ✓ データベースへの登録方法、および検索方法が不明
  - ✓ 発展途上の分野
- 工程の簡略化
  - 検索・可視化部分を簡略化し、それぞれの工程の時間短縮・労力軽減を図る

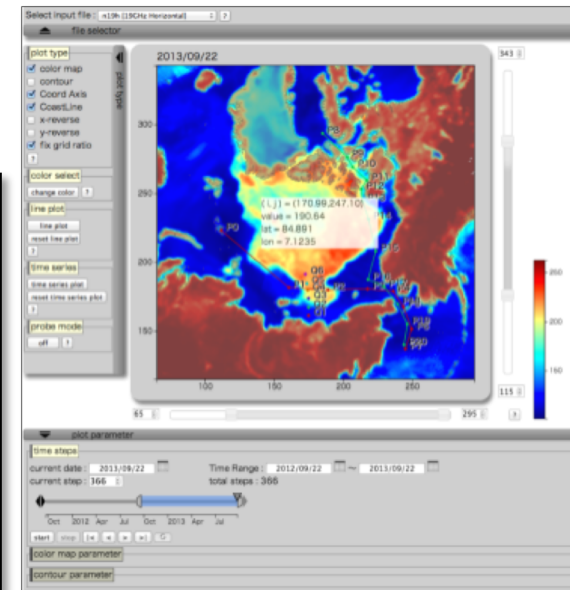
⇒ 現実的な解決策

# 開発の方向性

- 検索システム
  - ✓ DBの高速化
  - ✓ web画面上での操作性の向上
  - ✓ 検索結果の見やすさ
    - ⇒ バージョンアップ作業中
- **可視化システム**
  - ✓ webアプリケーション化
  - ✓ データ自動読み込み
  - ✓ 容易な可視化作業

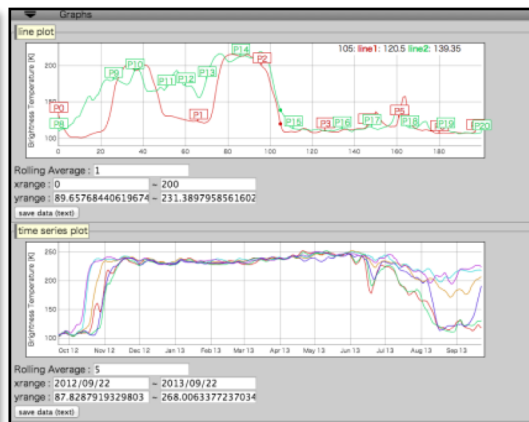
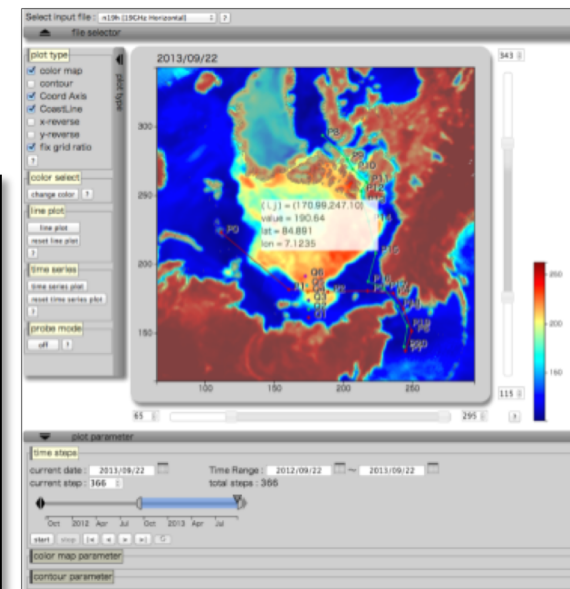
ADSに公開中、および次期バージョンの  
可視化システムについて紹介する

- 現在ADSで公開バージョンの可視化システム
- webアプリケーションとして開発することにより、特別なソフトウェアのインストールが不要
- DB上の実データを自動的に読み込める
- 画面上のGUIだけでデータの可視化・解析ができる





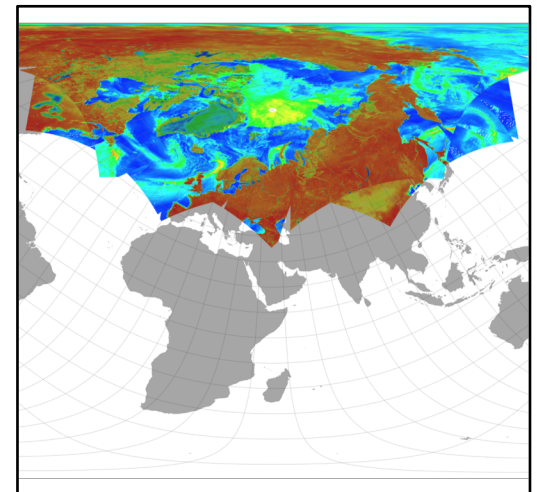
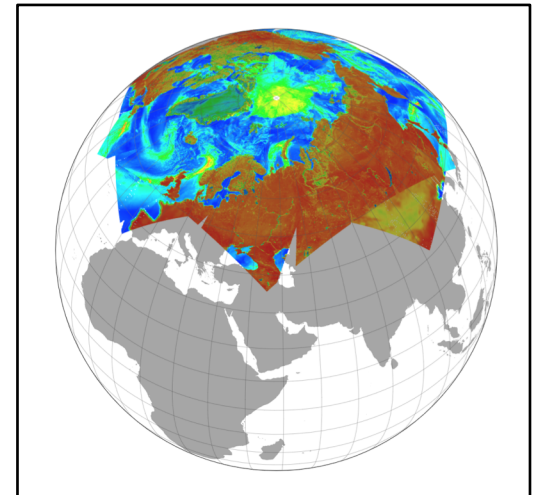
- 実装済みデータ
  - AMSR-E/AMSR-2衛星プロダクト（極域/全球）
  - SSM/I衛星プロダクト（極域）
  - 再解析データ、モデルアウトプット等
- 機能
  - カラーマップ、コンター図、時系列アニメーション
  - 断面グラフ、時系列グラフ、データ出力
- 欠点
  - 登録データセットが少ない
  - 一度に1データしか可視化できない
  - 可視化速度がクライアントPCのCPUに依存





# 次期バージョン可視化システム

- 改良点
  - ✓ netCDFに対応
    - ⇒ 読み込み可能データの増加
  - ✓ 描画データ毎のレイヤ化
    - ⇒ 複数データの同時プロット & 比較が可能
  - ✓ 様々な投影法での描画
  - ✓ WebGL使用
    - ⇒ GPUを用いることで描画の高速化を図る
- 課題
  - ✓ 読み込みデータ数増加によるメモリオバーフローへの対処
  - ✓ 複数レイヤの解析方法
  - ✓ GPU非搭載マシンでのCPUへのfallback機能



# まとめ

- DBからのデータ検索は「取得・作成・登録に関する情報」を基に構築されている
- 現状、データが内包する情報を基にした検索は難しい
- ユーザ自身が可視化しながら確認する作業が必要
- データの検索&可視化作業は時間と労力を要する
- このようなコスト軽減のためにオンライン可視化アプリケーション「VISION」の開発を行った