

R011-11

C会場：11/7 AM2 (10:45-12:30)

11:45~12:00

分野横断型データ公開プラットフォーム AMIDER の開発

#小財 正義¹⁾, 田中 良昌^{1,2)}, 阿部 修司³⁾, 南山 泰之⁴⁾, 新堀 淳樹⁵⁾

(¹ PEDSC/ROIS-DS, (² 国立極地研究所/総研大, (³ 九大・i-SPES, (⁴ 国立情報学研究所, (⁵ 名古屋大学宇宙地球環境研究所

Development of the cross-disciplinary data publication platform AMIDER

#Masayoshi Kozai¹⁾, Yoshimasa Tanaka^{1,2)}, Shuji Abe³⁾, Yasuyuki Minamiyama⁴⁾, Atsuki Shinbori⁵⁾

(¹ PEDSC/ROIS-DS, (² NIPR/SOKENDAI, (³ i-SPES, Kyushu Univ., (⁴ NII, (⁵ ISEE, Nagoya Univ.

In order to promote cross-disciplinary research and data-driven research, it is essential to arrange an environment to handle scientific data from various fields in an integrated manner. The Polar Environment Data Science Center of the Joint Support-Center for Data Science Research (DS), the Research Organization of Information and Systems (ROIS) aims at promoting the publication and use of scientific data obtained from research activities of the polar science community. One of the key pillars of the PEDSC is development of the AMIDER system, which will enable a cross-disciplinary overview of scientific datasets in various fields. Application of the AMIDER to a wider scientific fields beyond polar science in the future is also considered. Our design concepts are:

- Present datasets in various formats and fields, such as space science, geoscience, and bioscience, in a uniform design and interface.

- Enhance accessibility and usability of scientific datasets by adopting a general web design used in e-commerce, etc.

In this presentation, we will report on the system design, development, and data preparation status of the AMIDER.

The uniform and user-friendly design appears in the display of dataset list consisting of thumbnail images and snippets. The snippets are made as simple as possible so that even non-specialized users can imagine the contents at a glance. The top page consists of this data list display, category filters, and search box. Page transition to each dataset page is designed to be realized with a minimum operation of users.

Each dataset page consists of the main image, datafile download interface, data plots, list of related datasets, and metadata table. These UI parameters (datafile format, etc.) can be customized in a configuration file for each dataset, enabling a uniform UI design to cover various datasets.

The design and operation system are developed so that the accessibility to the datafiles of each dataset is also enhanced. Each datafile is placed in a repository of each research institute, and the AMIDER system acquires the URI. By our careful data curation and support, the datafiles are prepared in a uniform format as much as possible. Common Data Format (CDF) and Network Common Data Form (NetCDF) are the primary data formats. This strategy makes it possible to provide the format conversion function with AMIDER.

This design and operation system has also achieved the presentation of related datasets which have relatively high correlation scores with the respective dataset. The AMIDER system reads the datafiles of each dataset and calculates the correlation score between the datasets. Each dataset page presents related datasets based on this score and induces “walk-around” between datasets. The correlation coefficient and Earth mover’s distance are used as the correlation score. Further enhancements of the functions are possible by implementing batch programs in the future.

Scientific datasets of geospace, meteorites, earthquakes, meteorology, oceans, and biology will be registered, and test operations will be conducted this year. Further optimization will be made based on the results, and we will prepare for the public release in 2023.

The AMIDER project will provide a platform for wide-field researchers or non-specialized educators to discover and use scientific data based on their interests. It will also enable a “walk-around” experience between datasets and create a cross-disciplinary research opportunity.

異分野融合研究やデータ駆動型研究の推進には、多様な分野の科学データを統合的に扱うための環境整備が必須である。ROIS-DS 極域環境データサイエンスセンターは、極域科学コミュニティの観測・研究活動により取得された科学データの公開と利用を促進する活動を展開している。その柱の一つが、多様なデータを横断的に俯瞰できるデータ公開プラットフォーム AMIDER である。将来的には極域科学に留まらないより幅広い分野への適用も目指し、以下の設計コンセプトのもとに開発している。

- 時系列や成分データ、標本情報など、質・分野の異なる科学データであっても統一的なインターフェースで提示する。
- 汎用的な web デザイン (e コマースなどを参考としたもの) を採用し、専門的な科学データのアクセス性と利便性を高める。

本発表では、AMIDER のシステム設計や開発状況、公開データの整備状況について報告する。

統一的かつユーザーフレンドリーな web デザインとして構築したのが、サムネイル画像とスニペットで構成されるデータ一覧表示である。スニペットには可能な限り平易な表現を使用し、専門外のユーザーであっても、多様な科学データのコンテンツを一目でイメージできるよう配慮されている。トップページはこのデータ一覧表示とカテゴリフィルターや検索ボックスで構成され、最小限の動作で各データセットのページへ遷移できるように設計されている。

各データセットのページは、メインイメージ画像、データファイルのダウンロード、データプロット画像、関連するデータの一覧、メタデータテーブルで構成される。これらの表示パラメータ（データファイルの形式など）はデータセットごとに設定ファイルで調整でき、多様なデータセットを包括しつつ統一的なユーザーインターフェースを可能にしている。

各データセットのデータファイルへのアクセス性も高められるよう、設計と運用体制を工夫している。データファイルの実体は各研究機関の外部リポジトリに置き、AMIDER システムではその URI を取得する。きめ細かなデータキュレーションと支援を提供することで、データファイルも可能な限り統一的な形式とする態勢を取っている。Common Data Format (CDF) と Network Common Data Form (NetCDF) をデータフォーマットの核としており、それにより、AMIDER でフォーマット変換機能を提供することも可能とした。

さらに、この設計と運用体制により実現したのが、関連するデータ（相関スコアの高いデータ）の提示機能である。AMIDER システムで各データセットのデータファイルを読み取り、データセット間の相関スコアを計算する。それに基づき、各データセットページで関連するデータを提示し、データセット間の「渡り歩き」へ誘導する。相関スコアとしては、現時点では相関係数と Earth mover's distance を採用している。バッチプログラムの追加により、将来的にさらなる機能拡張も可能である。今年度中に宙空圏や隕石、地震、気象、海洋、生物などの各分野研究者により整備されたデータをシステムへ入力し、関係者内で試験運用を実施する。その結果に基づいてさらにブラッシュアップし、来年度の公開へ備える予定である。

これらの活動を通して、専門外の研究者や教育関係者であってもそれぞれの興味に基づいて科学データを発見し、使用できる環境を整備する。また、異分野のデータを渡り歩く体験を可能とし、異分野融合研究のきっかけを創り出す。