

APHRODITE データの DB 化による品質管理 および web 上でのデータ公開について

谷田貝 亜紀代[†] 今井 雅[†] ○前田 未央[†]
石田 祐宣[†]

A2-3 要旨

APHRODITE-2 では、プロダクト作成の際の品質管理機能を充実化させる目的で、解析フローにデータベース化処理を組み込んだ。また、アルゴリズム改良版プロダクトを公開するにあたり、web 上でデータ公開する仕組みを構築した。

Quality Control by making a database (DB) of Station Precipitation Data for APHRODITE and Product Release via the Web-Site

Akiyo Yatagai[†] Masashi Imai[†]
Mio Maeda[†] and Sachinobu Ishida[†]

We incorporate making a database (DB) in the analysis flow of making APHRODITE daily grid precipitation database for the fulfillment of the quality control of both original and gridded data. We also developed a data release system on the website that makes registration of the user possible.

1. はじめに

APHRODITE プロジェクトでは、APHRODITE-1 (2006 年 5 月～2011 年 3 月) において、アジアの水資源への温暖化影響評価のため、アジア各国から、長期間にわたる降水・気温の観測データを収集し、それらを、解析に便利な 0.25 度/0.5 度の日降水・日平均気温データにグリッド化し、公開した。現在進行中の APHRODITE-2 (2016 年 6 月～) では、極端降水評価と気象解析のため、グリッド化の解析アルゴリズムを改良し、データセットを更新している。

これまでに収集した観測データは、30 カ国、67 年 (1951 年～2017 年) 分に及ぶ。収集データは、図 1 の解析フロー (Task1 : Convert Data, Task2 : Quality Control, Task3 : Climatology, Task4 : Interpolation, Task5 : Re-gridding) を経て、0.25 度/0.5 度のデータにグリッド化され、プロダクトとして公開される。

この度、APHRODITE-2 では、プロダクト作成の際の品質管理機能を充実化させる目的で、Task1 と Task2 の間にデータベース化という処理を組み込んだ。また、アルゴリズム改良版プロダクトを公開するにあたり、web 上でデータ公開する仕組みを構築したので、報告する。

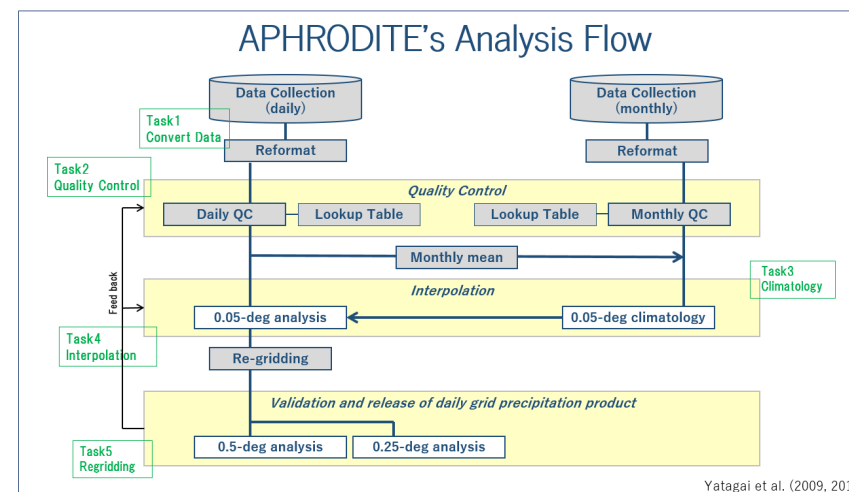


図 1 APHRODITE's Analysis Flow

[†] 弘前大学大学院理工学研究科
Graduate School of Science and Technology, Hirosaki University

2. APHRODITE データの DB 化による品質管理

2.1 解析処理

解析フローの第 1 ステップである Task1 では、様々なフォーマットで収集された観測データを、統一フォーマット（APHRO フォーマットもしくは川本フォーマットと呼ぶ）に変換する。変換されたデータはバイナリ形式である。続いて、フォーマットが統一されたデータに対し Task2 にて品質管理（QC）を施す。QC にてエラーが除外されたデータに対し、後続処理にてグリッド化処理を施す。

2.2 データベース化

Task1 と Task2 の間にデータベース化という処理を組み込んだ(図 2 参照)。目的は、バイナリデータの可視化と、品質管理機能の充実化である。開発は LAMP 環境（OS：Linux，Web サーバ：Apache，データベースソフト：MySQL，プログラミング言語：PHP）で行った。データベースに格納したデータのレコード数は、58,554,876 件であった。

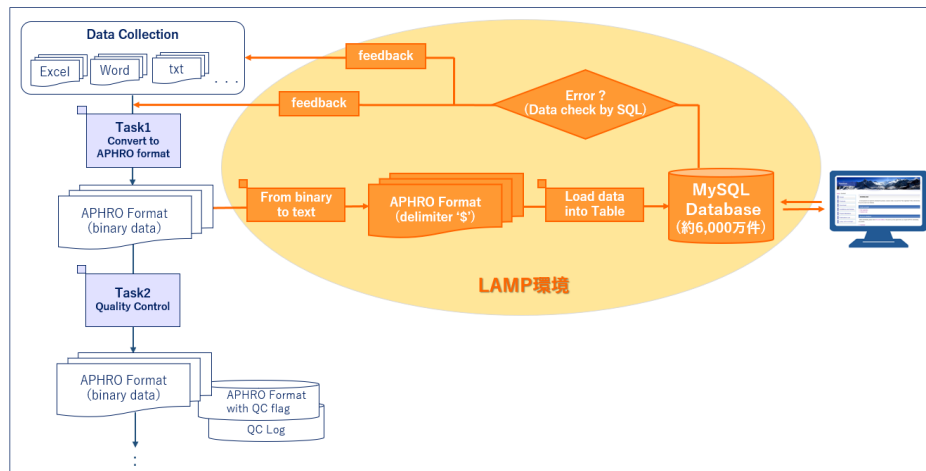


図 2 APHRO フォーマットデータのデータベース化

2.3 データの可視化と品質管理機能の充実化

データベース化することで、バイナリ形式の APHRO フォーマットデータを web 上で参照できるようになり、これまで困難であった中間データへのアクセスが容易となった。

また、データベースでは、膨大な量のデータの中から、目的のデータを SQL で簡単に見つける（検索する）ことができるため、データチェック機能としての役割を持たせることができた。つまり、観測データに含まれる明らかなエラー値、Task1 でのフォーマット変換ミス、後続の QC (Task2) には組み込まれていないため検出されないエラーなどを早期段階で検出できるようになり、それらをフィードバック・解決することにより、プロダクトの品質確保につなげることができた。

3. web 上でのデータ公開

3.1 システム構築の背景と要件

作成したプロダクトについては、これまでも web 上で公開してきたが、今後、データ量が増えていくこと、利用者の増加とともにユーザ管理が大変になっていくこと等が懸念されたため、これらの課題をクリアした新たな仕組みを構築することとした。

システム要件は、50GB～200GB 程度のデータを置くことができること、世界中の scientist がデータをダウンロードできること、ダウンロード時の認証 ID/PW をユーザに自動通知できること、ユーザの email address を把握できること、である。これらの要件を満たすシステムを、APHRODITE ホームページ上に構築した。

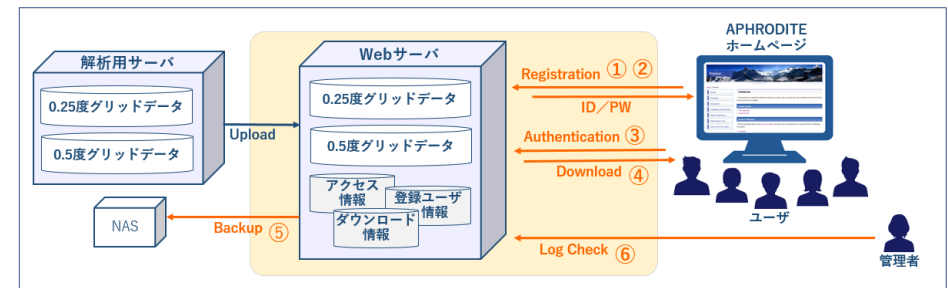


図 3 システム概要

3.2 実装した機能

本システムに実装した機能は、①アカウント登録機能、②パスワードリセット機能、③認証機能、④ダウンロード機能、⑤バックアップ機能、⑥ログ管理機能である（図 3 参照）。

ユーザがプロダクトをダウンロードする際、ユーザ個別の ID・パスワードにて認証することとした。機能①では、ユーザがメールアドレスを登録すると、認証時に必要

な ID とパスワードが記載されたメールがユーザに自動送信される仕組みになっている。パスワードを忘れた際には、機能②にて、メールアドレスを入力すると、新しいパスワードが記載されたメールがユーザに自動送信される。

機能③では、認証方式として、Apache の Basic 認証を採用した。ID・パスワードは、ユーザ毎に個別に設定するため、LDAP (Lightweight Directory Access Protocol) にて、システムが生成したパスワード、ID (①でユーザにメールで自動送信されるもの) により認証する。

機能④では、Apache のインデックス機能を使用しており、プロダクトを置いているディレクトリ階層を順に辿って、目的のファイルにアクセスする。

システム障害等によるデータ損失に備え、バックアップを取得する。機能⑤では、cron にてバックアップスケジュールを設定し、コマンドには、リモートマシン間で高速なファイル同期処理を行う rsync を用いた。

機能⑥では、各種ログから、登録ユーザ情報 (email address 一覧)、ダウンロードファイルリスト、当サイトへのアクセス状況 (アクセスログ解析ソフト「Visitors」を使用) を確認できる。

3.3 システムリリースと利用状況

本システムを、プロダクトの公開に合わせ、2018年9月27日にリリースした。リリースから約5ヶ月経過した2019年2月現在、約500名の方に登録いただいている。登録ユーザの所属機関は、わかっているだけで、22カ国、108機関に及ぶ。これまでの最大アクセス数は、プロダクト公開日 (2018年9月27日) の504件で、その後も一日平均約120件で推移しており、世界中の多くの方々に APHRODITE プロダクトが認知され、活用いただいている。

4. 今後の取り組み

今後は、アジア・世界への応用・貢献として、複数プロダクトの web 上での検索機能の構築、データ提供者へのアルゴリズム改良版プロダクトのフィードバック、アルゴリズム改良版プロダクトの条件指定 (緯度・経度、日付) によるデータの可視化、テキストデータのダウンロードツールの開発、ユーザからのフィードバックに基づく定常的な運用方法の検討に取り組む予定である。

5. 謝辞

本研究は、(独) 環境再生保全機構「環境研究総合推進費」(2-1602) の支援を受け

ている。

参考文献

- 1) Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi and A. Kitoh (2012): APHRODITE: Constructing a Long-term Daily Gridded Precipitation Dataset for Asia based on a Dense Network of Rain Gauges, *Bulletin of American Meteorological Society*, **93**,1401-1415, doi:http://dx.doi.org/10.1175/BAMS-D-11-00122.1.
- 2) Yatagai, A., O. Arakawa and K. Kamiguchi, H. Kawamoto, M. I. Nodzu, A. Hamada (2009): A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges, *SOLA*, **5**, 137-140, doi:10.2151/sola.2009-035.