

地球観測データに対するメタデータ登録システムの構築 —データセットドキュメントメタファを利用したメタデータ登録—

絹谷 弘子[†] 清水 敏之^{††} 吉川 正俊^{††} 喜連川 優[‡]

[†] 東京大学地球観測データ統合連携研究機構 ^{††} 京都大学大学院情報学研究所 [‡] 東京大学生産技術研究所

1 はじめに

我々の「データ統合・解析システム (DIAS)」は地球規模観測や各地域での観測で得られたデータを収集し、永続的な蓄積を行い、それらのデータの統合利用により環境問題や危機管理に有益な情報を提供することを目指している。我々が収集するデータは多様であり、現状ではデータセットを説明するメタデータを作成する手段がない¹⁾。多様なデータを統一的な枠組みで管理し、利活用するためには各データセットの持つ特性を取得することが重要である。データセットに関する情報はデータ提供者から取得する必要がある。そのためメタデータ登録システムを構築してきた。しかしメタデータになじみのないデータ提供者にはハードルが高く入力方法にはメタデータを意識させない工夫が課題であった。一方データ提供者の一部にはデータセットに関するドキュメントを書いた経験がある。この経験に基づいてドキュメントを作成する際の内容をメタデータ作成に反映させることにし、我々はドキュメントメタファを利用したメタデータ登録システムを構築した。このシステムによりデータ提供者には統一された形式のドキュメントを、そして我々はメタデータを取得することが可能となった。

本稿ではこのデータセットドキュメントメタファを利用したメタデータ登録システムについて紹介する。2 節で背景と目的、3 節では構築したメタデータ登録システムについて述べ、最後に 4 節で関連研究と今後の課題について述べる。

2 背景と目的

衛星観測、海洋観測、陸上観測、気候モデルシミュレーションなどの様々なデータを科学的・社会的に有用な情報に変換し、その結果を社会に提供するために、我々は地球観測データ統合・解析システムを開発している。我々は先行研究において地球観測データに対するメタデータ処理システムを設計し、地上観測データに対するメタデータ登録システムのプロトタイプを構築した。またメタデータ管理の枠組みの中でのメタデー

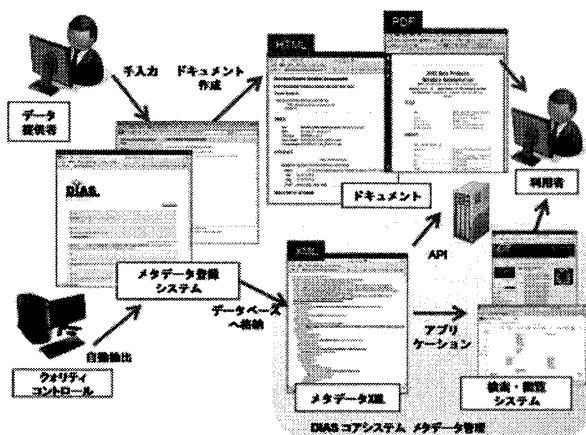


図 1: メタデータレジストリフレームワーク

タレジストリフレームワークについての検討を行った [1, 2, 3].

DIAS では多種多様なデータに対してメタデータを記述する必要があり、メタデータ形式とメタデータ入出力インタフェースをデータセットごとに設計する方法ではデータ種類分の時間を要することになり現状に対応できないことが明らかになった。そのため多様なデータに適用可能なメタデータ形式を作成することをめざすことにした。まずデータ提供者が作成しているデータセットを紹介するドキュメントを分析することでドキュメントに含まれているメタデータを集め、先行研究で利用している地理空間情報に対するメタデータ標準である ISO19115[4] の項目との対応付けを行った。さらにメタデータの中で検索を目的として利用されるキーワードをあらかじめ選定しやすくするため、GCMD サイエンスキーワードをはじめ公開されているキーワードリストを収集した。また入力インタフェースにおいて必須項目、オプション項目としてどのような入力形式がよいのかを検討した。まず 6 つのデータセットを選定し、データ提供者と協同で試行錯誤を繰り返しながら対象データの範囲を拡大させることにした。

3 メタデータ登録システム

図 1 は、我々のメタデータレジストリフレームワークである。メタデータの長期的な維持管理は DIAS コアシステムが行う。メタデータ登録システムはデータ提供者の手入力データとデータに対して行ったクオリティコントロールなどからの自動抽出データを入力し、データセットドキュメントの生成とメタデータ XML

Development of a Metadata Registration System for Earth Observation Data
—A Metadata Registration using Dataset Document Metaphor—

[†] Hiroko KINUTANI (kinutani@tkl.iis.u-tokyo.ac.jp)

^{††} Toshiyuki SHIMIZU (tshimizu@i.kyoto-u.ac.jp)

^{††} Masatoshi YOSHIKAWA (yoshikawa@i.kyoto-u.ac.jp)

[‡] Masaru KITSUREGAWA (kitsure@tkl.iis.u-tokyo.ac.jp)

Earth Observation Data Integration and Fusion Research Initiative (EDITORIA), The University of Tokyo ([†])

Graduate School of Informatics, Kyoto University (^{††})

Institute of Industrial Science, The University of Tokyo ([‡])

1) メタデータの知識を前提としたメタデータ作成システムは存在する。

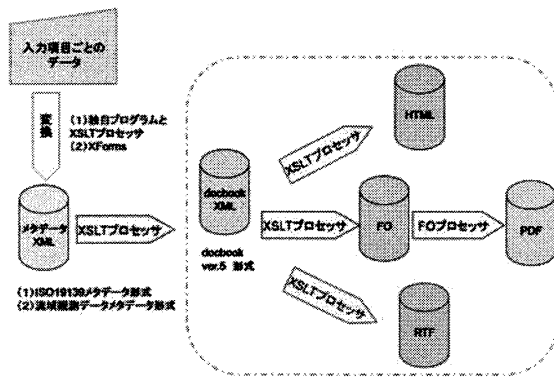


図 2: docbook XML ファイルを介したドキュメント生成

ファイル生成を同時に行う。それらは DIAS コアシステムに投入され検索に利用される。

我々は現在 2 種類のメタデータ登録システムを開発しているところである。(1) DIAS ドキュメント・メタデータ作成システムと (2) アジア水循環機構 (AWCI) の地上観測データに対するメタデータ登録システム [2] である。(1) は DIAS が保有しているデータセットすべてを対象とし、データセット全体を記述するドキュメントとメタデータを作るシステムである。(2) はアジア 18 カ国の河川流域に関係した観測データを対象とし、観測ステーションごとにデータセットを記述するドキュメントとメタデータを作るシステムである。データセットを対象としたメタデータは一番粒度の大きなメタデータであるのに対し、各観測ステーションの観測データ (1 ファイル) を対象としたメタデータは一番粒度の細かいメタデータである。例えば地球温暖化で話題となる IPCC の気候変動予測モデルの出力結果に対するドキュメント・メタデータを (1) を利用して作成している。(1) のインタフェースではメタデータの必須項目とオプション項目で入力欄を分類し、必須項目を入力すればドキュメント・メタデータが作成できる。英語と日本語に対応し、各データセットごとに両言語でドキュメント・メタデータを作成する。(2) のインタフェースでは作成するドキュメント形式の章建てで入力欄を構成し、必須項目に加えデータのクォリティチェック結果を自動的に加えてドキュメント・メタデータを作成できる。

図 2 は入力された項目からドキュメントを作成する過程である。どちらも docbook [5] 形式の XML ファイルを作成し、docbook の変換ツールを利用して HTML, PDF を作成する。データ提供者にとってメタデータを入力する労力と負担の増加をデータセットドキュメントを別途作成する手間の削減で補う効果を期待している。

図 3 は (1) のシステムを利用して作成されたメタデータ数の推移である。対象とするデータによっては記述しにくいものもあるため試作段階で中断しているものも多いが、2010 年 1 月現在、およそ 70 のドキュメント、35 種類のデータセットを対象としてこのシス

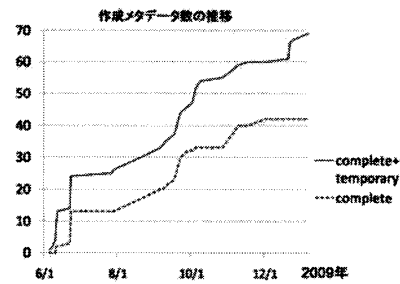


図 3: ドキュメント・メタデータ作成システムを利用して作成されたメタデータ数の推移

テムを試用しているところである。ドキュメントは分野を超えてデータの提供をする際の概要を示す文書として既に利用されている。さらに、これらのメタデータに基づいてデータセットをマトリクスに分類表示する検索インタフェースを開発しているところである。

4 関連研究と今後の課題

本稿では、ドキュメントメタファを利用することによってメタデータになじみのないデータ提供者にエディタに近い感覚でメタデータの入力作業を行うシステムを紹介した。メタデータ標準の XML を作成するための汎用的なメタデータ作成ツール、エディタは多数存在する。しかし、いずれもメタデータに関する知識が必要である。特定分野を対象としたメタデータ登録システムは多数存在するが、多様なデータセットを対象とする試みは、我々の知る限りは存在しない。

利用者の意見を反映させ、開発したシステムのユーザビリティを向上させること。そしてデータセットを対象としたメタデータとファイル単位のメタデータの間の中間メタデータへの対応が今後の課題である。

謝辞

本研究は、文部科学省による委託研究費「データ統合・解析システム」からの支援を受けて行われた。またメタデータ登録システム開発にあたり京都高度技術研究所 山田篤氏、海洋研究開発機構 福田和代氏、市野美夏氏、川本温子氏、東京大学地球観測データ統融合連携研究機構 小野雅史氏を始め DIAS の関係者の皆様にご協力いただいた。ここに記して謝意を表します。

参考文献

- [1] 絹谷弘子, 生駒栄司, 高橋慧, 吉川正俊, 喜連川優. 地球観測データに対するメタデータ処理システムの設計. 電子情報通信学会第 19 回データ工学ワークショップ, 第 6 回日本データベース学会年次大会 (DEWS2008) 論文集, March 2008. C9-6.
- [2] 絹谷弘子, 生駒栄司, 吉川正俊, 喜連川優, 小池俊雄. 地球データに対するメタデータ登録システムの構築. 電子情報通信学会 第 13 回 Web インテリジェンスとインタラクション研究会, 2008.
- [3] 立床雅司, 高橋慧, 絹谷弘子, 吉川正俊. メタデータレジストリフレームワークの構築. 電子情報通信学会 第 13 回 Web インテリジェンスとインタラクション研究会, 2008.
- [4] International Organization for Standardization. *ISO 19115:2003 Geographic information - Metadata*, 2003.
- [5] Norman Walsh. *Docbook 5.0: The definitive guide*. O'Reilly Media, Aug. 2008.