第 22 回情報知識学フォーラム予稿

金沢大学における

研究データ公開用リポジトリの構築の試み

An attempt to construct a public data repository for research data at Kanazawa University

宮本健弘^{1*}, 笠原禎也¹, 高田良宏¹, 松平拓也¹, 林正治², 松本篤¹, 上田望¹

Takehiro MIYAMOTO^{1*}, Yoshiya KASAHARA¹, Yoshihiro TAKATA¹,
Takuya MATSUHIRA¹, Masaharu HAYASHI², Atsushi MATSUKI¹,
Nozomu UEDA¹

1 金沢大学

Kanazawa University

〒 920-1192 金沢市角間町

E-Mail: miyamoto@cie.is.t.kanazawa-u.ac.jp

2 国立情報学研究所

National Institute of Informatics

〒 101-8430 東京都千代田区一ツ橋 2-1-2

*連絡先著者 Corresponding Author

近年、世界では「オープンサイエンス」の動きが盛んになっている。これに伴い、データ公開システムの重要性は年々増加しており、機械可読性を持ち、人間と機械双方にとって便利なりポジトリシステムの需要が高まっている。しかし、研究機関独自に開発されたデータ公開システムの多くは、汎用性や利便性、システム間連携の面で問題を抱えているのが現状である。これらの背景から本研究では、「JAIRO Cloud」などで実績がある国立情報学研究所開発の WEKO を用いて、金沢大学内の環日本海域環境研究センター及び国際文化資源学研究センターのデータリポジトリの構築を行っている。構築に当たっては、研究データを広く社会に公開する目的に加え、研究データを保有する研究者らがリポジトリの管理に精通していなくとも、自ら登録、修正などを行える環境の整備を進めている。

In recent years, the movement of "open science" has become active in the world. The importance of the data disclosure system has been increasing year by year, and the demand for a repository system for research data that is machine readable and convenient for both humans and machines is increasing. Because most of the data repositories were developed by individual institutes independently, they have problems in terms of versatility, convenience, and coordination between systems. We plan to construct public data repositories for Institute of Nature and Environmental Technology and Center for Cultural Resource Studies in Kanazawa University. In the study, we utilize WEKO, which was developed by the National Institute of Informatics and has already been applied to JAIRO cloud. We aim to design a public data repository for research data not only to open the data widely to society but also to be convenient for researchers to register and correct the contents even if they are not familiar with the internal structure of the system.

キーワード: WEKO, リポジトリ, オープンサイエンス, 研究データ
WEKO, repository, open science, research data

1 はじめに

近年, 世界では論文などの研究成果やその根拠となる研究データを, 広く社会に公開し, 共有化を推進する「オープンサイエンス」と呼ばれる動きが盛んになっている。オープンサイエンスは, 研究の再現性を高めるだけでなく, 既存の研究成果をベースにした新しい発見や学問分野の創出を促進することによる学術研究の効率的な発展が期待されている。これに伴い, データ公開・共有システムの重要性は年々増加しており, コンピュータが自動的に読み込んで処理できるデータ形式, いわゆる「機械可読性」を持ち, 人間と機械双方にとって便利な公開システムの需要が高まっている。しかし, これまで金沢大学を含め各研究機関で独自開発されてきたデータ公開システムの多くは, 汎用性や利便性, システム間連携の面で問題を抱えているのが現状である。これらの背景から, 先行研究^[1, 2, 3]で衛星データを対象にしたリポジトリシステムの開発を試みた。本研究では, 衛星データ用リポジトリの構築・運用の成果を基に, より汎用的なデータリポジトリを構築・運用する場合の指針となるよう, 金沢大学内の環日本海域環境研究センター及び国際文化資源学研究センターが保有するデータを利用してリポジトリの構築を行っている。

2 データ公開システム

リポジトリを構築する際は, 一からシステムを開発するのではなく, 既存のソフトウェアを用いるのが一般的である。代表例としてDSpaceやEPrintsが挙げられる。これらは多くの利用実績があり, リポジトリとしての機能性に関しては十分であるが, 開発元が国外機関であるため, 問題が発生した際に開発元とのやり取りが容易ではない点, 標準で日本語に対応していない点が問題点として挙げられる。これに対しWEKO^[4, 5]は国立情報学研究所(以下, NII)が開発したNetCommons2(以下, NC2)上で動作するソフトウェアであるため, 開発元とのやり取りが容易である点や日本語に標準で対応している点, CMSとしての機能を利用可能な点が

利点として挙げられる。WEKOのシステム構成を図1に示す。

図 1: WEKO のシステム構成

3 環日本海域環境研究センター

3.1 概要

金沢大学環日本海域環境研究センター(以下, 環日センター)は, 環日本海域が直面する自然と人間活動に起因する環境問題の解決と, 持続可能な世界の将来環境の創生を目的とした研究機関である。環日センターでは, 自らの研究活動により取得した様々な観測データや標本データを蓄積している。研究分野は大気, 海洋, 陸域, 統合の4つに分類される。現時点で, 各分野の数値データに加えて, 海洋と陸域に関する標本データを含む11種類のデータの登録を予定している。

3.2 WordPress 版

現在, 環日センターではWordPressを用いて, 同センターが取得・蓄積したデータを公開するシステムを運用している。公開中のWordPress版のページを図2に示す。WordPressはCMSとしても使えるが, もともとはブログソフトウェアであるため, 便利なプラグインが多数存在し, ページの見た目も綺麗に仕上げる事ができる。しかし, データリポジトリとしての運用を考え

た場合、データを1ファイルずつ手作業で登録する必要があり、ダウンロードの際にも1ファイルずつダウンロードすることになるため、登録するデータの数が増えると管理負荷の増加、ユーザーの利便性低下などの問題がある。また、機械可読性も悪いためリポジトリシステムに適したソフトウェアとは言えない。例えば、図2において、各ファイルのリンクをクリックすれば実データをダウンロードすることが可能であるが、そのデータがどのようなデータであるかはファイルを開くまで確認することができない。

3.3 WordPress と WEKO の比較

前節で指摘した問題の解決のため、本研究では、WEKO を用いて新たなリポジトリ構築を試みた。WordPress 版と比較すると、リポジトリソフトウェアであるため、データ管理の点で優れている。専用のソフトウェアを用いることでデータを一括登録できることに加え、WEKO の標準の機能として一括エクスポートがあるため、特に観測データ等の大量のデータを取り扱うことに長けていると言える。また、WEKO に登録したデータはアイテムと呼ばれ、それぞれのデータにメタデータを登録するため、ユーザーがファイルをダウンロードし中身を見なくとも、アイテム詳細ページからメタデータを確認することで、ファイルの情報をある程度知ることができ、検索の際にも必要なデータのみを容易に検索することができる。また、xml 形式でメタデータを保有するため、機械可読性が良く、他のシステムやプログラムから利用することが可能である。また、WEKO にはデータの添付ファイルを差し替えた後も差し替え前のファイルを保有することができる履歴管理機能があり、研究に用いたデータが更新された場合も、必要に応じて元のデータを用いて再現できるなど、研究の再現性を保証することができる。CMS という観点から見ると、見た目に関してはシンプルなデザインになるが、NC2 上で動作しており、アカウント管理機能や、ランキング機能などの、機能性も様々な工夫がなされている。これらを踏まえると、WEKO はリポジトリとしてだけではなく、CMS としての機能も十分に備えており、観測データの公開システムを構築

する際には、より適したソフトウェアであると言える。

4 構築したリポジトリ

4.1 メタデータ

現在、登録済みのデータは以下の通りである。

- ブラックカーボン濃度 (大気)
- 海水温・塩分・気温・湿度データ (海洋)
- 岩石・化石標本 (陸域)

数値データ2つについてはWEKOで利用可能なメタデータの中で、観測データの記述に適したSPASE^[6]を採用した。SPASEはNASAを始めとした国外機関によって2002年より開発が行われている自己記述型のメタデータ形式であり、現在も国内外の機関によって改良が行われながら、主に宇宙科学分野で用いられている。国内では、超高層大気観測データ向けのメタデータ・データベースであるIUGONET^[7]でもメタデータ形式として採用されている。

大気分野のデータであるブラックカーボン濃度は、IUGONETの前例があるためSPASEとの親和性が高いが、海洋分野のデータである海水温・塩分・気温・湿度データは、必須項目でも記述が困難な項目が複数あるため、別のメタデータを検討する必要があると考えている。ブラックカーボン濃度のアイテム詳細ページを図3に示す。

岩石・化石標本についてはメタデータ項目を独自に設定し、外部にメタデータを提供する場合は、世界的に利用されている汎用メタデータフォーマットであるDublinCoreで出力できるように項目の関連付けを行う。この際、WEKO上のメタデータ項目とDublinCoreの要素との関連付けが難しい項目はDescriptionの要素で出力するようにしている。岩石・化石標本のアイテム詳細ページを図4に示す。

図 2: WordPress 版データベース

4.2 管理者自身による管理に向けて

WEKO にアイテムを登録する際、SWORD Client for WEKO(SCfW)^[8] を用いることで一括登録が可能である。SCfW は NII が提供しているソフトウェアであり、データをリポジトリに登録する際の障壁を低くすることを目的として開発された SWORD と呼ばれる軽量プロトコルを使用している。しかし、SCfW は Windows 専用のソフトウェアであるため、Mac や Linux 等の他の OS を用いている場合にデータの一括登録が困難である。現在、この問題を解決するために、ブラウザからメタデータファイルと実データを含んだ圧縮ファイルをアップロードすることで、一括登録可能な機構の開発を行っている。

5 国際文化資源学研究センターへの適用

環日センターと並行して金沢大学国際文化資源学研究センターのリポジトリ構築に着手した。同リポジトリには数値データではなく書籍や写真、映像資料などが主体のリポジトリを構築する予定である。映像ファイルについては、これまで画像や数値データを扱ってきたのに対し、同リポジトリではサイト内での映像・音声の再生について検証が必要である。また、開発中の一括登録・更新機構が人文系のデータでも運用可能であるか検証が必要であり、開発段階で様々なデータに対応可能な汎用性を、意識して開発を計画中である。この機構が、数値・論文・画像・映像等のデータで運用できれば、様々なフォーマットのデータリポジトリの構築手法及びデータの一括登録手法が確立でき、データ登録者の負担軽減が期待できる。

WEEKO 編集

トップ アイテム登録 ワークフロー ランキング

検索 検索

▶ 詳細検索 ◎ 全文検索 ◎ キーワード検索

インデックスツリー

- ☐ 大気環境領域
 - ☐ ブラックカーボン濃度
 - ☐ Granule
 - 2010
 - 2011
 - 2012
 - 2013
 - NumericalData
 - Observatory
 - Instrument
- ☒ 海洋環境領域
 - Person
 - Repository

インデックス

大気環境領域 ▶ ブラックカーボン濃度 ▶ Granule ▶ 2010

NT_abs_201007.txt 利用統計は来月からご利用いただけます

File / Name	License
NT_abs_201007.txt	
📄 NT_abs_201007.txt (60.58KB) [0 downloads]	

[更新履歴](#)

PDF HTML XML JSON WEEKO REPORT PRINT

アイテムタイプ	SPASE.Granule
言語	日本語
キーワード	absorption coefficient Black Carbon concentration
Spase.Granule.ResourceID	spase://Kannichi/Granule/KanazawaU/NOTOGRO/MAAP5012/NT_abs_201
Spase.Granule.ReleaseDate	2017-02-08T15:54:53
Spase.Granule.ParentID	spase://Kannichi/NumericalData/KanazawaU/NOTOGRO/MAAP5012
Spase.Granule.StartDate	2010-07-01T00:00
Spase.Granule.StopDate	2010-07-31T23:59
Spase.Granule.SourceType	Data
Spase.Granule.Source.URL	http://133.28.23.47/permalink.php?keyword=NT_abs_201007.txt
公開状況	公開 非公開に変更

戻る
編集
削除

図 3: アイテム詳細ページ:ブラックカーボン濃度

カガミガイ 利用統計は来月からご利用いただけます

File / Name	License
LGSK060502_mollusca.JPG	
📄 LGSK060502_mollusca.JPG (163.91KB) [0 downloads]	

[更新履歴](#)

PDF HTML XML JSON WEEKO REPORT PRINT

タイトル(英)	Dosinia (Phacosoma) japonica (Reeve)
アイテムタイプ	陸域標本 / Terrestrial Specimen
言語	日本語
タイプ / Item Type	化石 / Fossil
学名 / Scientific Name	<i>Dosinia (Phacosoma) japonica (Reeve)</i>
和名 / Japanese Name	カガミガイ
地質時代 / Geological Age	前期更新世 / Early Pleistocene
採集者 / Collector	塚田真二 / Shinji Tsukawaki
採集日 / Date	2006-05-02
数量 / Number	1
注記 / Remarks	Ogasawara, K., 1977, Paleontological analysis of Omma fauna from Toyama-Ishikawa area, Hokuriku Province, Japan. Sci. Rept., Tohoku Univ., 2nd ser., vol. 47, pp. 43-156.
所蔵 / Storage Location	金沢大学総合教育2号館地学学生実習室 / Laboratory of Geology, General Education Hall, Kanazawa University
用途 / Present Use	教材 / Teaching Material
画像 / Image	LGSK060502_mollusca.JPG

図 4: アイテム詳細ページ:岩石・化石標本

6 まとめ

本研究では、汎用的なデータリポジトリを構築・運用する場合の指針となるリポジトリシステムの構築を目指した汎用データリポジトリの構築を行っている。保有するデータには、大きく分けて数値データと標本データがあるが、数値データとしてブラックカーボン濃度、標本データとして岩石・化石標本のデータを登録したことにより、これらを参考にさまざまなデータ種に対してメタデータの作成を進める予定である。また、使用している OS に依存せずにアイテムの一括登録・更新が行える機構を開発しており、これを用いることで、管理者自身による管理が容易になると考えている。今後の全学展開を視野に、国際文化資源学研究センターのリポジトリ開発を行うことで、WEKO を用いたリポジトリが様々な分野・データフォーマットに対応できることを示し、本学のオープンサイエンスの活発化に貢献していきたいと考えている。

参考文献

- [1] 馬淵嵩大; 笠原禎也; 高田良宏; 松平拓也; 後藤由貴; 松田昇也; 「WEKO を利用したあけぼの VLF 観測データ公開システムの開発」, 平成 26 年度名古屋大学太陽地球環境研究所研究集会及び第 263 回生存圏シンポジウム (京都大学), 2015.2.19-20.
- [2] 馬淵嵩大; 笠原禎也; 高田良宏; 松平拓也; 後藤由貴; 山地一禎; 林正治; 「WEKO を利用した超高層・宇宙科学観測データ公開システムの開発」, 日本地球惑星科学連合 2015 年大会 (幕張メッセ国際会議場), 2015.5.24-28.
- [3] 馬淵嵩大; 笠原禎也; 高田良宏; 松平拓也; 山地一禎; 林正治; 「科学データ解析・公開に最適なデータ公開システムの開発」, 平成 27 年度電気関係学会北陸支部連合大会 (金沢工業大学), 2015.9.12-13.
- [4] WEKO
<http://weko.at.nii.ac.jp/>
- [5] 山地一禎; 青山俊弘; 武田英明; 「学術資源共有基盤 WEKO の開発」, デジタル図書館, Vol.36, pp.51-61, 2009.03.
<http://ci.nii.ac.jp/naid/120001507523> (2017 年 1 月 24 日参照)
- [6] Space Physics Archive Search and Extract <http://www.spase-group.org/> (2016 年 8 月 14 日参照)
- [7] IUGONET
<http://www.iugonet.org> (2016 年 8 月 20 日参照)
- [8] SWORD Client for WEKO(SCfW)
<https://community.repo.nii.ac.jp/news/release-note/SWORD-Client-for-WEKO/> (2016 年 7 月 20 日参照)