リポジトリ用データ管理システムの構築

メタデータ	言語: Japanese 出版者: 公開日: 2019-04-11 キーワード (Ja): キーワード (En): 作成者: 宮本, 健弘, 笠原, 禎也, 高田, 良宏, 松平, 拓也, 林, 正治, 松木, 篤, 上田, 望, Miyamoto, Takehiro, Kasahara, Yoshiya, Takata, Yoshihiro, Yoshihiro, Matsuhira, Takuya, Takuya, Hayashi, Masaharu, Matsuki, Atsushi, Ueda, Nozomu メールアドレス: 所属:
URL	https://doi.org/10.24517/00053841

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

第 23 回情報知識学フォーラム予稿

リポジトリ用データ管理システムの構築
Construction of data management system for repository

宮本健弘[1]*, 笠原禎也[1], 高田良宏[1], 松平拓也[1], 林正治[2], 松木篤[1], 上田望[1]

Takehiro MIYAMOTO[1]*, Yoshiya KASAHARA[1], Yoshihiro TAKATA[1], Takuya MATSUHIRA[1], Masaharu HAYASHI[2], Atsushi MATSUKI[1], Nozomu UEDA[1]

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

近年のオープンサイエンスの活発化に伴い, 人間と機械の双方に可読性を持つリポジトリシステムの需要が高まっている. このような背景から, 本研究では, 国立情報学研究所が開発した WEKO を用いて, 金沢大学環日本海域環境研究センター及び同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. 従来, WEKO へのコンテンツの一括登録には, Windows でのみ動作するソフトウェアが必要であった. 我々は, ICT に精通しているとは限らないデータ所有者が, リポジトリ上のデータ管理を行えるように, OS に非依存なブラウザからファイルをアップロードするだけでコンテンツの一括登録に加え, メタデータやコンテンツの追加登録・更新等が可能なデータ管理システムを構築した. 特に更新機能では, 「更新フラグ」属性を用いて, コンテンツの世代管理とその公開方法の制御を可能にした. 本稿では両センターのデータリポジトリと構築したデータ管理システムの概要を述べる.

Along with the recent activation of open science, the demand for a repository system that is readable to both humans and machines is increasing. In this context, we constructed data repositories for Institute of Nature and Environmental Technology and Center for Cultural Resource Studies in Kanazawa University using WEKO, which is a repository platform developed by the National Institute of Informatics. In the original design of the WEKO, it is necessary to utilize software for Windows for batch registration of data contents. In the present study, we developed a new user interface for data management for data owners who are not familiar with ICT. By using this system, it is possible for data owners to operate a batch processing for additional registration, update of the registered contents and modification of the metadata simply by uploading a file via browser without depending on the OS. Particularly, it is innovative that out software supports revision control and data access control of the contents by setting "update flag" attribute. In this paper, we introduce the user interface developed for the data repositories in Kanazawa University.

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

近年のオープンサイエンスの活発化に伴い，機械可読性を持ち，人間と機械双方にとって便利なリポジトリシステムの需要が高まっている．本研究では，国立情報学研究所開発の WEKO[1] を用いて，金沢大学環日本海域環境研究センター及び同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．WEKO，両センターの概要および構築したデータリポジトリの設計については参考文献 [2] をご参照されたい．

本稿では，ICT に精通しているとは限らないデータ所有者が，リポジトリ上のデータ管理が容易に行えるように開発を行ったデータ管理システムについて述べる．また，WEKO におけるアクセス履歴の確認方法をまとめ，今回導入した Piwik[3] についても説明する．

2 データ管理システム

WEKO にデータを登録する際，SWORD Client for WEKO(SCfW)[4] を用いることで，アイテムの一括登録が可能である．しかし，SCfW は Windows 専用のソフトウェアであるため，Mac や Linux 等，他の OS を用いている場合に利用できない．本研究では，この問題を解決するため，OS に非依存なブラウザからファイルをアップロードするだけで，WEKO に対してアイテムの一括登録やデータの追加・更新を行うことが可能なデータ管理システムを構築した．

ICT に精通していない研究者でも容易に利用できるための条件として，分かりやすさと敷居の低さが求められる．分かりやすさを実現するために，各メタデータ属性ごとの違いの説明やメタデータファイルの編集が複雑になってしまう部分を半自動で行うこととした．また，敷居の低さを実現するために，ソフトウェアをダウンロードして使うという形式ではなく，ブラウザからアップロードするという形式をとった．

WEKO 上のデータ管理を行う際の手順を図 1 に示す．ここからは，各手順において，データ管理システムが関与する機能について説明する．

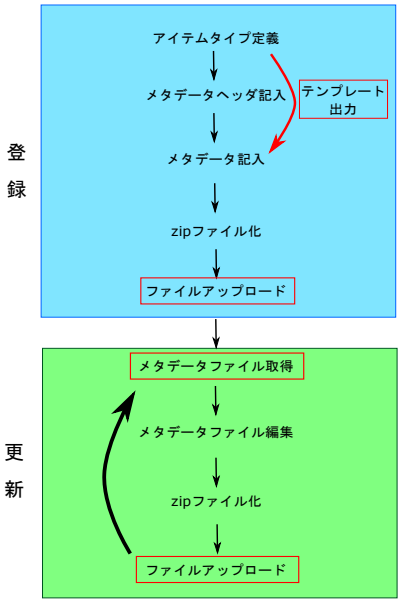

図 1: データ管理の流れ

2.1 データ登録

WEKO では，アイテムタイプというメタデータ属性の集合によってアイテムが分類されている．アイテムタイプは WEKO の GUI で自由に新規作成や編集を行うことが可能である．メタデータファイルの例を図 2 に示す．

アイテムタイプ"sample"のメタデータ属性

アイテムタイプ	タイトル	言語	著者	ページ数
sample	サンプル1	日本語	宮本	10
sample	sample2	英語	Miyamoto	10
sample	サンプル3	日本語	宮本	15

各行が1つのアイテムとして登録される

図 2: メタデータファイル例

メタデータファイルを作成する際は，WEKO で定義したアイテムタイプ通りにファイルの 1 行目にあたるヘッダにメタデータ属性を書き込み，2 行目以降に実際のメタデータを書き込む必要がある．今回構築したデータ管理システムには，アイテムタイプごとにメタデータテンプレートファイルを出力する機能を実装した．この機能を用いることで，ヘッダに応じたメタ

データを2行目以降に書き込むだけで, メタデータファイルの作成が可能になるだけでなく, 全角半角の違いや文字の抜けなどによって起こるデータの抜けを防止している. データ登録を行う場合, 作成したメタデータファイルと実データファイルを zip ファイル化する. その後, 登録先インデックスを選択し, アップロードすることでアイテムの登録が完了する.

2.2 データ更新

この節では更新機能について説明する. WEKO 上のデータを更新する際は, WEKO 上に存在するアイテムと, 更新用のデータの紐づけを行う必要がある. この際用いるのが, WEKO_URL という属性である. この属性はアイテム ID の情報を持っているため, メタデータが WEKO 上のどのアイテムと関連しているかを判別することが可能である.

また, 研究の再現性を保証するといった観点から, 一度公開した実データは残しておく必要がある. WEKO には実データファイルの更新を行った際, 過去のファイルを履歴として残すことができる更新履歴機能が備わっている. この履歴は公開・非公開を決めることができるため, この機能を活用するために「更新フラグ」という属性を新たに定義した. 更新フラグに以下のいずれかを記入することで対応した更新を可能とした. この機能によって, データ更新時にコンテンツの世代管理が容易に行えるようになった.

- m:メタデータ更新
- rh:実データファイル更新 (履歴公開)
- ph:実データファイル更新 (履歴非公開)
- replace:実データファイル置き換え

”m”は”m&rh”のように”&”でつなぐことで, 他のモードと併用できる. また, 更新フラグ属性が空欄だった場合はそのアイテムをスキップし, 更新は行わない. また, replace モードはアイテムが未公開の場合のみ更新が可能である. これは, 一度公開したアイテムを削除することが, 研究の再現性を損なう可能性があるためである.

更新を行う際は, ヘッダにこれらの属性が追加され, WEKO_URL の情報が記入されているファイルが必要である. WEKO_URL の情報が記入されているファイルを入手するには, WEKO エクスポート機能を用いるか, SCfW によってデータを登録した際のファイルを所持している必要がある. しかし, 前者は 1 度に 100 件の制限があり, 後者には, OS の制限がある. この問題の解決のために, メタデータ出力機能を開発した.

今回開発した機能では, WEKO に登録されているアイテムのメタデータをアイテムタイプごとに出力することができる. このとき出力されるメタデータファイルには, WEKO_URL と更新フラグの 2 属性がヘッダに追加されている. ユーザは, このファイル内の更新したい項目を追加・修正し, 更新フラグを記入することで, 更新ファイルを作成する. その後, 作成したファイルを登録時と同様に zip ファイル化し, アップロードすることで, 更新が完了する.

3 アクセス管理

WEKO におけるアクセス管理機能としては, 以下の 3 つが挙げられる.

- カウンタ (NetCommons)
- ログレポート
- 利用統計

カウンタは, NetCommons のモジュールの 1 つである. サイトへのアクセス数を管理するものであるが, bot のアクセスもカウントしてしまうため, 正確なアクセス数を把握することはできない. ログレポートは, 接続元 IP などを確認することができるが, 接続元の国籍を把握することができない. 利用統計は, 国籍を確認することができるが, アイテムごとの情報であるため, サイト全体としてのアクセスの把握には適していない.

ログレポートと利用統計はそれぞれ, サイト全体のアクセス数の把握, 国籍情報の把握, というメリットを持っている. これらの両方の機能を合わせ持つアクセス解析ツールが必要と考え, Piwik を導入した. Piwik のページを図 3 に示す.

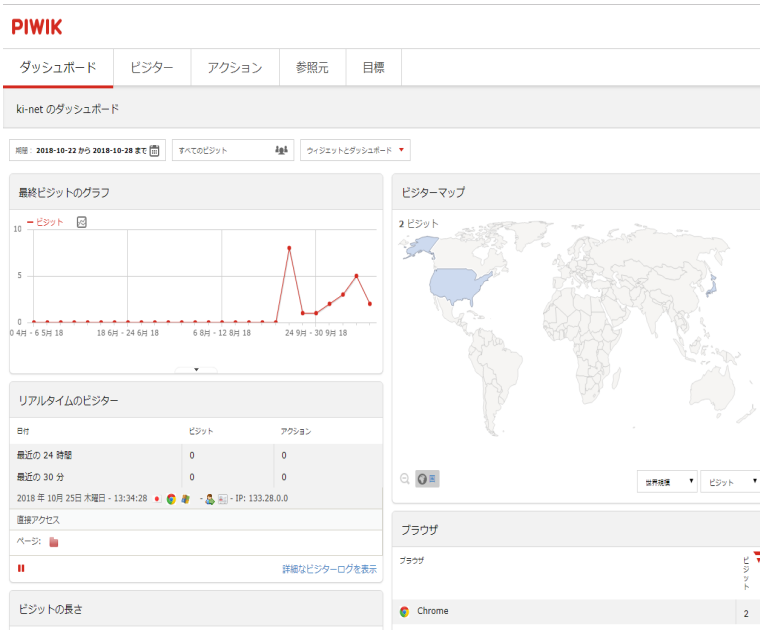

図 3: Piwik の動作画面

Piwik を用いることで，日別や週別のアクセス数をグラフで表示することができ，アクセス元の国籍を世界地図で確認することが可能となった．また，サイト訪問者の遷移を表示することができるため，データの関連性などの点からその後のデータの登録インデックスなどを決める際の参考になることが期待できる．

4 まとめ

本研究では，ICT に精通していないデータ所有者でも，OS に非依存なブラウザ経由で，WEKO 上のデータ管理を容易に行うためのデータ管理システムを構築した．テンプレートファイル出力機能やメタデータ出力機能により，データ登録・更新時の負担やミスが減ることが期待できる．

また，アクセス解析ツール Piwik を導入し，アクセス数や国籍などを直感的に確認できるようになった．

今後は，今回構築したデータ管理システムを用いて多くのデータが登録され，リポジトリ上のデータがさらに充実していくことが期待される．そのためにも，利用者がシステムの使い方を容易に理解できるような解説ページを作成する予定である．

謝辞

本研究の一部は，科研費 (18K18525）および平成 30 年度 NII 公募型共同研究 (会合 4) の支援を得て実施している．

参考文献

[1] WEKO
http://weko.at.nii.ac.jp/

[2] 宮本健弘; 笠原禎也; 高田良宏; 松平拓也; 林正治; 松木篤; 上田望; 「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」, 情報知識学会誌, vol.27, No.4, pp.337-342, 2017. https://doi.org/10.2964/jsik_2017_037

[3] Matomo(Piwik) https://matomo.org/

[4] SWORD Client for WEKO(SCfW)
https://community.repo.nii.ac.jp/news/release-note/SWORD-Client-for-WEKO/