

非文献資料のための学術資源群による サブジェクトリポジトリの構築（構想と進捗状況）

高田良宏*1, 林正治*2, 堀井洋*3, 堀井美里*3, 山地一禎*4, 山下俊介*5, 古畑徹*1

(*1 金沢大学, *2 一橋大学, *3 合同会社 AMANE, *4 国立情報学研究所, *5 北海道大学)

yosihiro@kenroku.kanazawa-u.ac.jp

概要: 近年, 研究資料に関するキーワードとして, オープン化, ビックデータおよびオープンサイエンスが挙げられ, 資料を永続的に蓄積し再利用可能な情報基盤の整備が急がれている. 我々は, 整備が進んでいない非文献資料において, 非文献資料版ビッグデータ・オープンデータ化に対応した情報基盤の開発を進めている. 本稿では, 全体構想と現在までの成果である学術資源群を基にしたサブジェクトリポジトリ構築の進捗状況等について述べる.

1 はじめに

近年, オープン化, ビックデータ, および, オープンサイエンスが研究資料に関するキーワードとなっている. 特にオープンサイエンスに注目が集まっており, 論文等成果に加え, その根拠となる研究資料のリポジトリ化・オープン化が急がれる. 国内でもオープンサイエンスの推進に向け, 内閣府の「国際的動向を踏まえたオープンサイエンスに関する検討会」が, 我国におけるオープンサイエンス推進のあり方を取りまとめている[1].

一般にオープンサイエンスの推進が議論される場合, リポジトリ化・オープン化の対象になるのは, 実験観測データなどの数値データを始めた自然科学分野の研究資料となりがちである. しかし, 文化財・博物館学や人文・社会科学分野で取り扱われることが多い歴史的・文化的資料(以降, 非文献資料)もリポジトリ化・オープン化の対象とすべきである.

我々は, 2013年に非文献資料のリポジトリ化推進のため, 学術資源リポジトリ協議会[2]を設立し, 非文献資料の再利用化を目的とし, デジタル化, リポジトリ化に関する議論を進めてきた[3,4]. その一環として, オープンサイエンス時代に向けた非文献資料版ビッグデータ・オープンデータ化に対応した情報基盤の確立を目的とするプロジェクトを発足させ, 研究開発を進めている.

本稿では, プロジェクトの全体構想と現在までの成果である学術資源群を基にしたサブジェクトリポジトリの構築の進捗状況等について述べる.

2 非文献資料

非文献資料(あるいは非文献コンテンツ)とは, 広義においては, リポジトリ化が定着している論文を始めとした文献資料に対して, それ以外の研究資料をさしたものである[5]. その中でも, 今回対象とするのは, 文化財・博物館学や人文・社会科学系の分野での利用が中心の文献以外の歴史的・文化的資料である.

既に, 自然科学系の分野での利用が中心の実験データを始めた数値データに関しては, いくつかの分野でリポジトリ化・オープン化の試みが実施されている. 一方, 今回対象とする非文献資料に関しては, 文献資料や数値データとは違い, 物理的, 意味的, あるいは履歴的にも多種多様な特性を持つ資料が多いことなどから進んでいないのが現状である. 非文献資料にも有効なリポジトリ基盤の確立が課題である.

多様な非文献資料に対して, 文献資料では一般的な機関リポジトリを対象に行われているメタデータ収集による横断検索を適用したイメージが図1である. 資料群固有の情報が埋もれてしまい(メタデータの欠落と元情報との乖離が発生), 学術的精度を担保することは困難であ

る。

我々は、リポジトリ化する単位として学術資源群を考案した[6]。学術資源群とは非文献資料を学術的な視点から俯瞰的、意味的に分類した組織に非依存な資料情報の集合である。資料が関係する分野、利用法、形状、起源など研究者の視点から分類を行う。様々な分類が考えられるが、後述のサブジェクトリポジトリ構築の基本単とするため、資料の当時の社会での分類(主に使われ方)を基本とした群とした。

3 学術資源群によるサブジェクトリポジトリ

本章では、プロジェクトの全体構想とその実現手法を概説する。

3.1 全体構想

図2に示すように、信頼性を評価する格付けモデルを適用した学術資源群によるサブジェクトリポジトリにより、非文献資料の全登録化・全公開化・再利用化に耐え得るリポジトリ基盤の実現を図る。これにより、次のような検索法を提供でき、研究目的や手法に応じたアクセス

が可能となる。

- 全体検索(従来の横断検索)(図2の①)
- 適切な学術資源群への検索(ピンポイント検索)(図2の②)
- 適切な格付けの資料への検索(格付けによるフィルタリング)(図2の③)

3.2 資料情報格付けモデル

一方、全登録・全公開を行うことで、様々な品質(精度)の資料情報が混在し、リポジトリ基盤全体の資料情報の質の低下が心配される。全登録と学術的な精度の担保という、相反する要求を解決する必要がある。この問題に対して

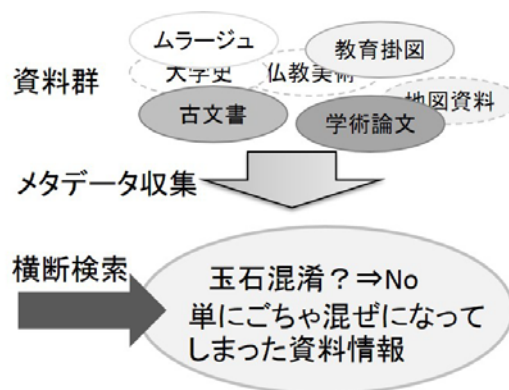


図1 非文献資料に対する従来の横断検索

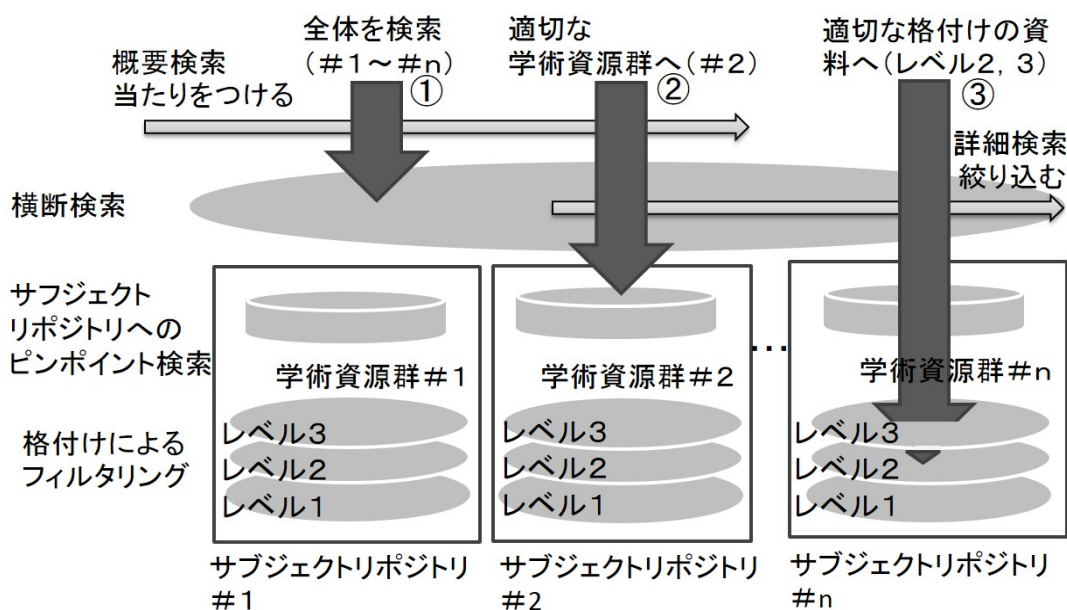


図2 学術資源群によるサブジェクトリポジトリ

- ①全体検索(横断検索)
- ②適切な学術資源群へ(ピンポイント検索)
- ③適切な格付けの資料へ(格付けによるフィルタリング)

は、資料情報を登録から公開までの過程において、データの体裁、信頼性、有用性等で評価し格付けを行う非文献リポジトリ用の資料情報格付けモデルと、その運用ワークフローを導入する。図3に登録～公開までのプロセスにおける資料情報格付けモデルの概要を示す。

3.3 高付加価値メタデータ

付加価値メタデータ（高品位な資料情報）を導入し、前節の格付けで、上のレベルへの移行を許可する（いわゆるデータ査読）の指標の一つとなる。資料情報の学術的精度を担保し、資料情報の再利用に資する。加えて、仮登録・登録用の「最小限のメタデータ」、および、公開するために推奨される「公開用のメタデータ」も定義する（図3）。

4 進捗状況

4.1 サブジェクトリポジトリの構築

NII が提供するリポジトリプラットフォーム WEKO[7]を用い、学術資源群を基にしたサブジェクトリポジトリを構築した。現在、表1に示すように、明治期から昭和初期までの旧制高等学校由来の科学実験機器資料と教育掛図資料の

2つのサブジェクトリポジトリを構築し試験公開している。また、天文学機器資料のサブジェクトリポジトリの準備を進めている。図4に科学実験機器資料に関するサブジェクトリポジトリの公開例を示す。

4.2 今後の展開

現在は、図2の①の状態であるが、資料情報に対するハーベスティングと DOI 付与を進めており、それらが完了すれば、図2の②の状態に移行する。今後、高付加価値メタデータの検討、資料情報の格付けモデルとワークフローを開発実装することで、図4の③の実現を目指す。また、サブジェクトリポジトリの整備も進める。

5 おわりに

本稿では、学術資源リポジトリ協議会が開発中の非文献資料版ビッグデータ・オープンデータ化に対応した情報基盤の開発プロジェクト全体構想と進捗状況を報告した。構築したサブジェクトリポジトリの評価や、現在開発中の諸機能については次の機会に報告する。当リポジトリ基盤の整備により、非文献資料の原則全登録化・全公開化・再利用化が進み、社会の様々な

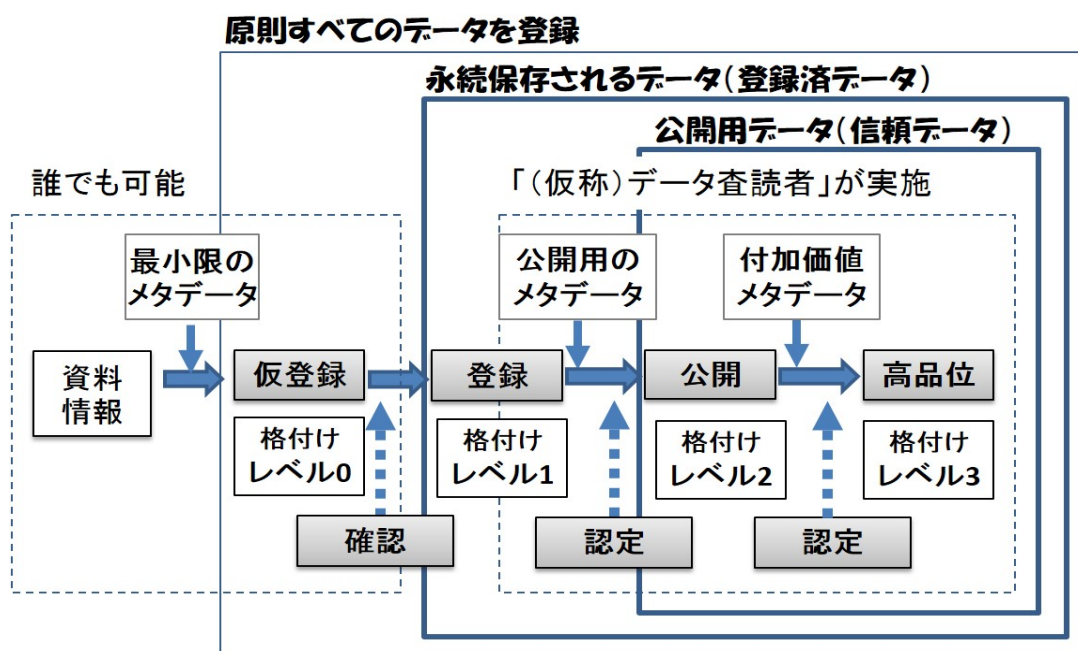


図3 資料の登録～公開までのプロセス (資料情報格付けモデルの概要)

分野で利用が始まったビッグデータの実現（情報系で取り扱うビッグデータとは規模が違うが、資料の持つ意味、および、研究スタイルが違う）も当該分野では十分にビッグデータと言える）も可能であり、データ駆動型の研究課題での利用など、今までにない新たな学術的価値が発見され、当該・隣接分野を始めとする多分野で活用が進むと期待できる。

謝辞

本研究の一部は、科研費（基盤研究(B): 24300310, 挑戦的萌芽研究: 25560140, 基盤研究(C): 15K00446), および、課題設定による先導的人文・社会科学研究推進

表 1 サブジェクトリポジトリの公開状況

資源群	状況	所蔵機関
科学実験機器	公開	<ul style="list-style-type: none"> ・新潟大学 ・神戸大学 ・東京大学駒場博物館 ・石川県立自然史資料館 ・大阪教育大学附属図書館
教育掛図	公開	<ul style="list-style-type: none"> ・石川県立自然史資料館 ・大阪教育大学附属図書館 ・奈良女子大学附属図書館
天文学機器	準備中	案：石川県立自然史資料館 案：国立天文台

事業「地域に現存する学術資料を活用した地域学術観光創出に関する研究」によるものである。

参考文献

- [1] 内閣府：国際的動向を踏まえたオープンサイエンスに関する検討会，<http://www8.cao.go.jp/cstp/sonota/openscience/>, (2015年9月29日参照).
- [2] 一般社団法人学術資源リポジトリ協議会 (Re*poN), <http://www.repon.org/>, (2015年9月29日参照).
- [3] 高田 良宏, 林 正治, 堀井 洋, 堀井 美里, 山地 一禎, 上田 啓未, 古畑 徹, 学術資源情報の共有と「場」の創出 ～学術資源リポジトリ協議会の活動の展開～, 大学 ICT 推進協議会 2013 年度年次大会 (AXIES2013) 論文集, T1A-1, 2013.12.
- [4] 高田 良宏, 林 正治, 堀井 洋, 堀井 美里, 山地 一禎, 上田 啓未, 古畑 徹, 学術資源リポジトリ協議会の活動の展開 (続報) ～試行から事業への展望～, 大学 ICT 推進協議会 2014 年度年次大会 (AXIES2014) 論文集, T2A-20, 2014.12.
- [5] 高田 良宏, 笠原 禎也, 西澤 滋人, 森雅 秀, 内島 秀樹, 非文献コンテンツのための可視性と保守性に優れた学術情報リポジトリの構築, 情報知識学会誌, Vol.19, No.3, pp.251-263, 2009.9.
- [6] 堀井 洋, 林 正治, 堀井 美里, 山地 一禎, 高田 良宏, 塩瀬 隆之, 古畑 徹, 学術資源リポジトリ構築に向けた現状と課題 - 明治期以降の科学実験機器資料のリポジトリ化を事例として -, 人文科学とコンピュータシンポジウム論文集, IPSJ Symposium Series 12(7), 17-22, 2012.12.
- [7] 国立情報学研究所 : WEKO, <https://weko.at.nii.ac.jp/>, (2015年9月29日参照).



図 4 学術資源群 (科学実験機器) によるサブジェクトリポジトリの実装例