

# 学術資料のメタデータ生成とリポジトリ構築における 現実と課題

## —学術資源リポジトリ協議会の活動を通じて—

堀井 洋<sup>1)</sup> 林 正治<sup>2)</sup> 堀井美里<sup>1)</sup> 山地 一禎<sup>3)</sup> 高田良宏<sup>4)</sup> 古畑 徹<sup>5)</sup>  
1) 合同会社 AMANE 2) 一橋大学情報基盤センター 3) 国立情報学研究所  
4) 金沢大学総合メディア基盤センター 5) 金沢大学資料館

著者らは、これまで大学博物館や自治体などの公共博物館に所蔵されている文献以外の学術資料（文献資料との対比として，“非文献資料”）を対象にした学術資源リポジトリの構築に取り組んできた。

本報告では、活動の中心である学術資源リポジトリ協議会の役割と実際の資料公開環境における現状と課題について紹介する。

## Current Status and Issues of Repository Construction and Metadata Generation for Academic Resources.

Hiroshi HORII<sup>1)</sup> Masaharu HAYASHI<sup>2)</sup> Misato HORII<sup>1)</sup> Kazutsuna YAMAJI<sup>3)</sup>  
Yoshihiro TAKATA<sup>4)</sup> Toru FURUHATA<sup>5)</sup>

1)AMANE.LLC 2) Center for Information and Communication Technology,  
Hitotsubashi University 3)National Institute of Informatics  
4)Information Media Center, Kanazawa University 5)Kanazawa University Museum

In recent year, social concern about exchanging and sharing of academic resources included non-bibliographic between universities is increasing. However, in actual situation, there are some problems about preservation and digitalization of academic resources. In this paper, a case study of scientific apparatuses since the Meiji era, we discuss about construction of repository for non-bibliographic.

### 1. はじめに

著者らは、これまで大学博物館や自治体などの公共博物館に所蔵されている文献以外の学術資料（文献資料との対比として，“非文献資料”）を対象にした学術資源リポジトリの構築に取り組んできた。本報告では、活動の中心である学術資源リポジトリ協議会の役割と実際の学術資料公開環境における現状と課題について紹介する。

大学や自治体博物館には、膨大かつ多種多様な学術資料が所蔵されている。これら学術資料の学術分野や所蔵管理組織の垣根を越えた分野・組織横断的な情報共有について、学術成果の公開・共有や学術資料の保存などの観点から、近年、関心が高まっている。その一方で、博物資料などの非文献の学術資料に関しては、文献横断検索環境に

相当する分野・組織を横断したりポジトリの構築は実現されておらず、取り扱いについても学術機関間での統一的な仕組みは未定義または普及していないのが現状である[1]。このような状況を鑑みた場合に、学術資料全般を対象にした横断的な学術資源リポジトリを実現させるためには、資料の所在や由来などの情報（メタデータ）を共通のフォーマットに基づいて収集・整理し、ネットワーク上で共有・公開するための情報環境であるリポジトリの構築が不可欠である。さらに、それを核として学術機関・企業・自治体などの関係者が運営や研究利用について議論するためのコミュニティの構築も重要であり、技術・人的コミュニティ双方を構築・整備することが求められている。

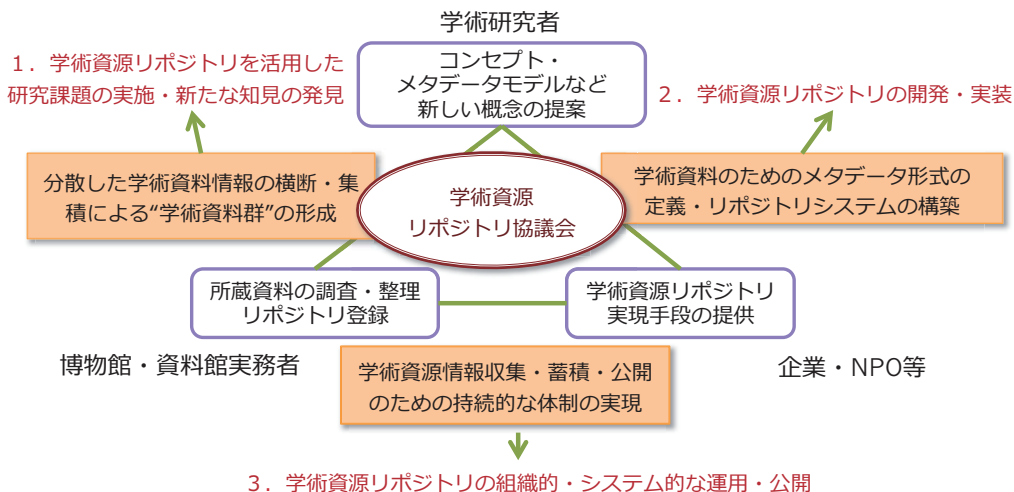


図1： 学術資源リポジトリ協議会の概要

著者らはじんもんこん2012において、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、学術資源リポジトリの概要や現状を発表した[2]。本発表ではそれをさらに拡張し、学術資源リポジトリの普及や継続的運営を想定した、より実践的なメタデータ生成およびリポジトリ公開における横断的な連携に関して論じる。

## 2. 従来研究

博物館資料のデータベース化・アーカイブ化に関しては、これまで概念モデルの提案やそれを反映させた属性・記述タグの定義が行われてきた。概念参照モデル（Conceptual Reference Model）である CIDOC-CRM は、国際博物館会議（International Council of Museums, ICOM）において定義された国際標準である[3][4]。このモデルでは、具体的な個々のデータベースにおける記述項目に共通する「概念」のレベルにおいて関係性を構成・定義し、それらを用いてデータの交換や統合を目指している。学術資料に関する情報を表現するための属性や記述タグは、所蔵機関や学術分野により大きく異なるため、それらを直接的に交換・統合することは現実には困難である。このことから CIDOC-CRM における概念参照モデルの考え方は、非常に有効かつ興味深い。さらに、国内でも博物館所蔵資料を対象にしたミュージアム資料情報構造化モデルが提案されている[5]。このモデルは、博物館および美術館の資料情報システムにおいて、資料情報を共有するためのデータ形式を定義する基盤となることを目指している。「識

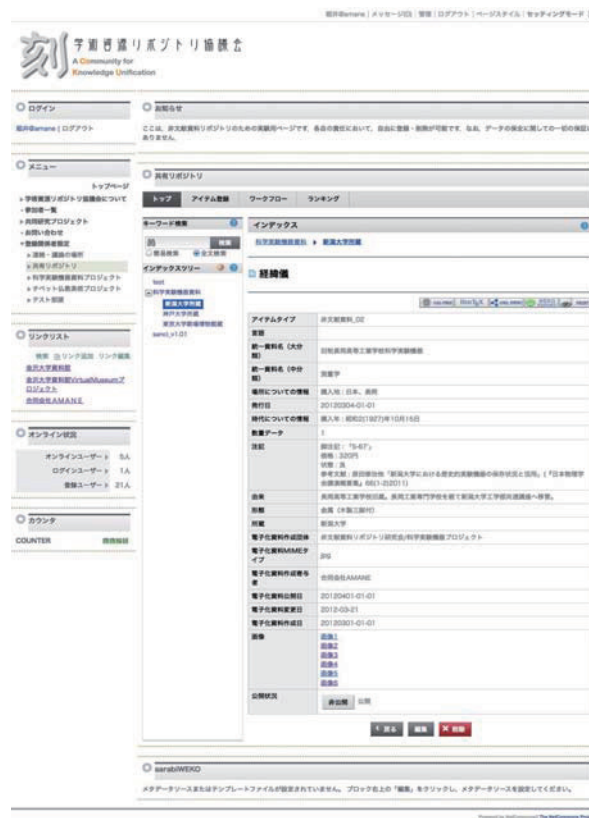


図2：学術資源リポジトリ Web 画面外観

別・特定」「物理的特性」「履歴」「関連・参照」などの対象とする資料の属性を具体的に提示しており、実際の管理業務を強く意識している点が特徴である。

本研究においても所蔵機関や学術分野を横断した異種な資料情報の共有を目指しており、これ

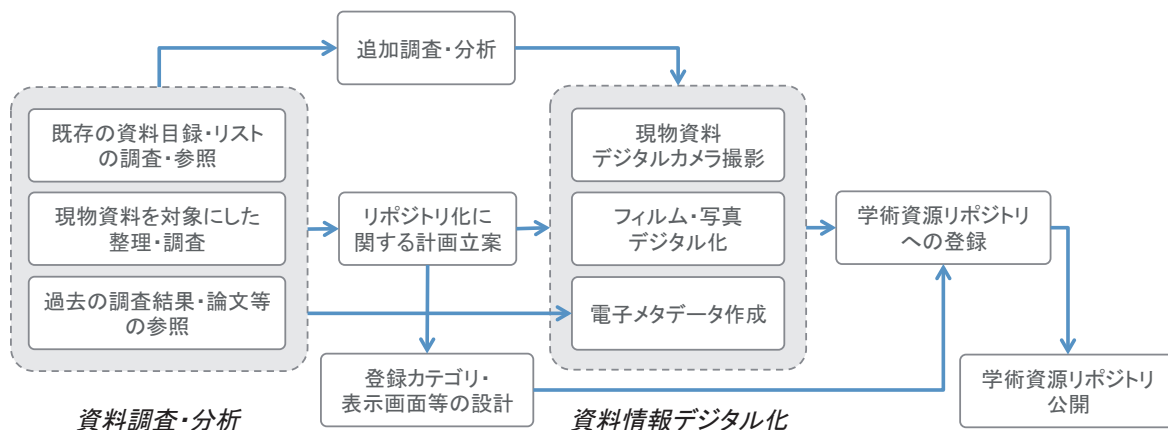


図3： 学術資源リポジトリ構築のプロセスの例

らのモデルを参考にリポジトリの設計・構築を進めている。

石川県立自然史資料館蔵：約110点  
奈良女子大学附属図書館蔵：21点

### 3. 学術資源リポジトリ協議会

合同会社 AMANE・金沢大学・京都大学・国立情報学研究所などの研究者が中心となり進めている学術資源リポジトリ協議会(以下、本協議会)の概要について紹介する。

本協議会は、図1に示すように大学等の学術機関や企業・自治体などの学術資料の保存・活用に関わる研究者が中心となり、2011年に結成された[6]。本協議会は、①学術資料、あるいはリポジトリを研究対象とする研究者、②非文献資料の所蔵管理に携わる博物館学芸員、③リポジトリの構築や博物資料の活用を行う企業・団体職員の三者が協力・連携して、学術資料に関する情報の組織・分野を越えた共有や研究・議論の場の創出を目指している。

現在、図2に外観を示す本協議会 Web サイト (<http://amane-project.jp/hibunken/>)にて公開している学術資料情報を以下に記す。

[明治～昭和期科学実験機器資料]

新潟大学所蔵：20点  
神戸大学所蔵：21点  
東京大学駒場博物館蔵：22点  
石川県立自然史資料館蔵：747点  
大阪教育大学附属図書館所蔵：3点

[明治～昭和期教育掛図資料]

(2013年8月時点・公開準備中)

本協議会の活動は、学術資料情報に関連した研究の実施やリポジトリ化の支援などの研究支援活動、関連する組織や取り扱う学術資料の分野・種類など、様々な面において展開している。

### 4. 学術資源リポジトリ構築に関する現状と課題

#### 4.1 学術資源リポジトリの構築

学術資源リポジトリ協議会では、これまで複数の機関と連携・協力して、学術資料の調査・分析および学術資源リポジトリの構築を実施してきた。ここで、博物館等における学術資料を対象にした学術資源リポジトリ構築のプロセスの例を図3に示す。学術資源リポジトリの構築では、対象となる学術資料の調査・分析と、それを基に計画された学術資料情報のデジタル化に大別される。学術資料の調査・分析では、所蔵機関や関連する分野の研究者が過去に作成した資料目録や調査報告書を参照するか、それらが存在しない場合は現物資料を対象にした調査を実施し、現時点における資料所蔵状況の把握を行う。そして、得られた資料所蔵状況に関する情報から、以後実施する学術資料のデジタル化および学術資源リポジトリ上での公開に関しての具体的な計画立案をおこなう。計画立案に際しては、対象となる学術資料の種類・総量・状態などから画像撮影や資料分析・電子メタデータ作成に要する時間およびコストを算出する。また、対象が特定の個人や団

体に関連した学術資料の場合には、学術資源リポジトリ上の公開に関して、事前に関係者から了解を得ることが必要である。次に、立案した計画に基づき、画像撮影等のデジタル化を実施する。デジタル化では、デジタルカメラやスキャナーを利用して、対象となる学術資料のデジタルを行うことが一般的であるが、その際に撮影者や撮影条件などについても、学術資料に関する学術情報と併せてメタデータに記録する。そして、対象となる学術資料の登録カテゴリや表示画面の構築、資料コレクションの全体の解説等を記述し、生成した電子データを学術資源リポジトリに登録し、インターネット上で公開する。

#### 4. 2 資料情報の整理状況とメタデータ

以上、大まかな学術資源リポジトリの構築プロセスを示したが、実際の資料所蔵の現場では、学術資料情報の把握およびメタデータの作成に関しては、必ずしも理想的な状況にあるとは限らない。

本研究では、所蔵組織における学術資料およびそれらのメタデータに関する整備状態を、以下の5段階に分類・定義する。

**状態 1**：現物資料自体の所在・状態が不明。

**状態 2**：資料のみが存在し、関連する情報（目録・文献・論文等）が一部または全部存在しない、または不明である。

**状態 3**：資料と関連する情報（目録・文献・論文等）が一部または全部整理され存在する。しかし、それらの電子データは存在しない。

**状態 4**：資料と関連する情報・メタデータが可読な形式で一部または全部電子化され存在する。

**状態 5**：既に何らかの形でメタデータが電子的に検索可能な状態で公開されている。

本研究において、所蔵資料に関するメタデータの整理状況に着目した理由は、学術資源リポジトリでの学術資料情報の蓄積と公開を試みる際に、対象となる学術資料とそのメタデータの整理状



図 4： サクラシメジムラージュ標本  
(金沢大学資料館蔵)

況が実現を左右する非常に重要な要素となるためである。上記の定義における状態 1～2（関連する資料情報・メタデータが存在しない）に相当する、所謂“未整理資料”の場合には、対象となる学術資料の調査や整理・メタデータの生成を最初期段階から実施する必要性が生じ、学術資料情報公開に要する時間的・人的なコストは非常に高くなる。状態 3 および 4 についても、現実の所蔵現場においては、目録作成などの資料調査実施時点から長い年月が経過している状況もしばしば存在し、その場合は所蔵資料の实在確認および目録掲載情報との突き合わせ確認作業を行う必要がある。さらに、MS-DOS 以前の計算機環境で作成された古い電子データの場合には、復元の必要性が生じる状況も発生する。

このような実際の所蔵機関における資料整理の現状に対して、本研究では、メタデータの階層的な記述と、その作成作業における役割（ロール）の定義を試みた。

#### 5. 階層的なメタデータ記述と分担者の役割

前述のとおり、博物資料等の学術資料情報のリポジトリ化・公開の実現性は、対象となる学術資料それ自体の特徴や分類の他に、所蔵状況や整理状況、想定される利用目的等に大きく依存する。この問題に対して、本研究では、学術資料情報の階層的な構造を定義し、それぞれを記述的する役割分担者を明確に意識した学術資源リポジトリ構築を実施している。学術資料情報記述における分担者の役割一例として、図 4 に示すサクラシメジムラージュ標本（金沢大学資料館蔵）のメタデ

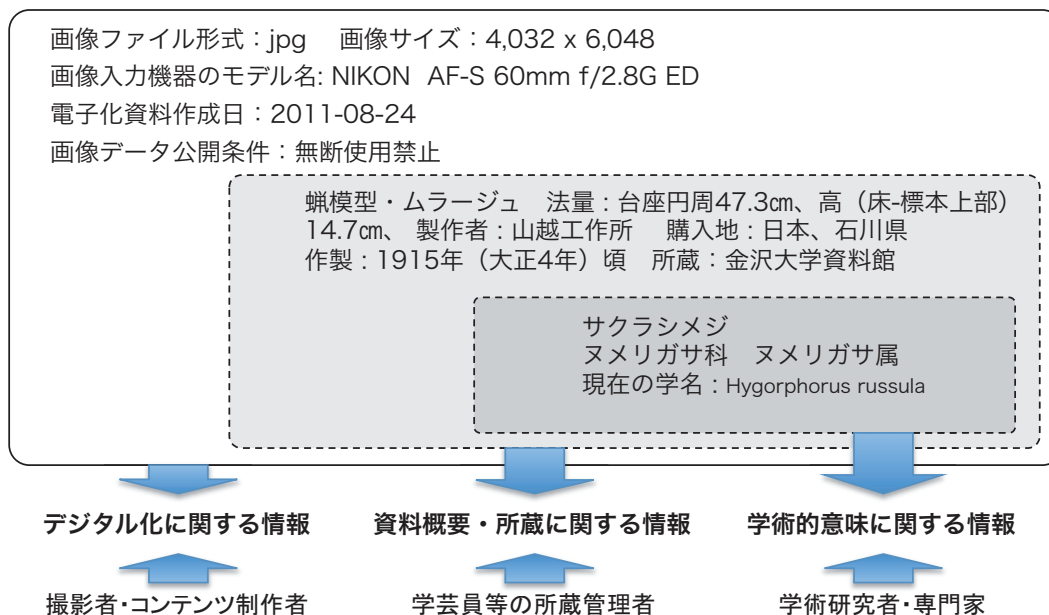


図5： サクラシメジムラージュ標本（金沢大学資料館蔵）のメタデータの例

ータの例を図5に示す[7]。この学術資料情報を記述するメタデータは、資料の寸法や概要に関する情報（ムラージュ標本）、その資料が表現している内容に関する情報（菌類サクラシメジ）、さらにデジタル画像等の資料を表現している媒体に関する情報（jpeg形式等）から構成される。それぞれのメタデータを生成する分担者も、資料の所蔵や資料の外観・概要に関しては学芸員等の所蔵管理者、資料が有する学術的意味・研究的位置付け等に関しては学術研究者、資料を撮影したデジタル画像やデジタルコンテンツについては撮影者・コンテンツ制作者から構成される。

さらに、情報の更新・修正頻度についても、各メタデータによって大きく異なる。所蔵機関などの資料管理情報や資料の寸法・外観に関する情報は、記入ミスや所蔵機関の変更が発生しない限り、生成時点の情報を修正する必要はない。しかしながら、学術資料が有する学術的意味に関する情報については、調査研究の進展や別の学術分野におけるアプローチなどにより、記述内容の修正や新たな知見に関する記述の追加が必要となる事態が発生する。このように、学術資源リポジトリに登録された学術情報の修正や更新という観点からも、メタデータ生成・管理における役割の明確化と記述の階層化は有効であると考えられる。

## 6. まとめと今後について

本研究では、学術資料のリポジトリにおけるメタデータの整理状況に基づいた、各段階の定義を行った。さらに、学芸員等の所蔵管理者、専門知識を有した学術研究者、撮影者・コンテンツ制作者から構成される分担者を定義し、それらを明確に意識した学術資源リポジトリ構築を提案した。現在、NetCommons + WEKO(Ver.2.0.1)上に構築した学術資源リポジトリにおいて、メタデータ定義と学術資料情報の公開を行っている。

本研究における今後の課題は、組織・分野を横断する学術資料情報の共有についてである。本研究では、階層的なメタデータを定義したが、それらは組織・分野を横断する学術資料情報の共有においても有効であると考えられる。横断的な学術情報の共有を実現する際には、共有する情報の構成（項目）とそれらをまとめる単位（粒度）について考慮することが重要である。本稿で述べたメタデータ階層を用いることにより、所蔵機関や電子データに関する情報については、異なる分野や機関間においても統一的なメタデータ記述をすることは容易となる。また、資料概要の記述についても資料形態の違いにより、同一の項目でその概要を表現することには限界があるが、資料を記述する粒度について、一定の統一的な基準を設け

ることは可能であると考え。しかしながら、学術的意味の表現に関しては、専門とする学術分野において標準となる表現に、学術分野毎に独自の記述形式が存在している。さらに、担当する学術研究者により学術的な関心や定義が大きく異なる（前述のムラージュ標本の例では、菌類標本としての観点と、ムラージュの工芸技法としての観点など）。このように現状において、統一的な形式で記述可能な情報と、学術的な情報それぞれについて、分離してメタデータを記述・管理する本提案手法が1つの現実的な解となり得ると考える。

これらのことを踏まえ、今後さらに、分野・組織横断的な学術資源リポジトリの構築を進めていく所存である。

## 謝辞

本研究の一部は、科研費挑戦的萌芽研究（『貴重技術史資料のフォークソノミーによる集合知と連動したデジタル展示システムの開発』、課題番号 23650561）、および科研費基盤研究（『非文献資料リポジトリによる機関横断的学術資源群形成に関する研究』、課題番号 24300310）の支援を受けて行われた。本研究の実施に際して協力頂いた大学・自治体博物館をはじめとする関係機関、ならびに関係者の方々に厚く感謝致します。

## 参考文献

- [1] 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 19(3), 251-263, 2009-10-28.
- [2] 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, Vol.2012.No.7 pp. 17-22, 2012
- [3] <http://www.cidoc-crm.org/>,  
The CIDOC Conceptual Reference Model (CRM)
- [4] <http://icom.museum/>, International Council of Museums

[5] 村田良二, ミュージアム資料情報構造化モデルによる博物館業務支援と情報共有, 情報処理学会研究報告. DD, 2006(11), 9-16, 2006

[6] <http://amane-project.jp/hibunken>, 学術資源リポジトリ協議会

[7] <http://kuvm.kanazawa-u.ac.jp>, 金沢大学ヴァーチャルミュージアムプロジェクト