

Development of Common Platforms for non-Bibliographic Contents Disclosure in University : Construction of an Academic Resource Repository Excellent in Visibility and Maintainability for non-Bibliographic Contents

メタデータ	言語: jpn 出版者: 公開日: 2017-10-05 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.24517/00028559

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



2010年度論文賞記念講演

大学における非文献コンテンツ公開のための共通プラットフォーム
の開発 --- 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 ---

**Development of Common Platforms for non-Bibliographic
Contents Disclosure in University
--- Construction of an Academic Resource Repository Excellent in
Visibility and Maintainability for non-Bibliographic Contents ---**

高田良宏^{1*}, 笠原禎也¹, 西澤滋人², 森雅秀³, 内島秀樹⁴
Yoshihiro TAKATA, Yoshiya KASAHARA, Shigeto NISHIZAWA,
Masahide MORI and Hideki UCHIJIMA

1 金沢大学総合メディア基盤センター

Information Media Center, Kanazawa University

E-mail: yoshihiro@kenroku.kanazawa-u.ac.jp, kasahara@is.t.kanazawa-u.ac.jp

2 金沢大学自然科学研究科(現在, ソラン 株式会社)

Graduate School of Natural Science and Technology, Kanazawa University

(Currently, SORUN CORPORATION)

3 金沢大学人間社会学域人文学類

School of Humanities, College of Human and Social Sciences, Kanazawa University

4 金沢大学情報部情報企画課

Information Infrastructure Service Division, Kanazawa University

*連絡先著者 Corresponding Author

我々は、大学内に蓄積している学術情報の内、その公開が進んでいない非文献コンテンツに焦点をあて、それらを学内外に公開するための指針となるべく、既存情報インフラ上で運用可能な公開手法を考案し、共通プラットフォームとして提供することを目指している。その一環として、本研究では、非文献コンテンツをリポジトリ化する場合の問題点を整理し、その解決のため、メタデータの互換性、異種コンテンツの共存、大規模コンテンツの管理、コンテンツに関する位置情報を用いた可視化などを考案した。そして、既存リポジトリプラットフォームのDSpaceを拡張し、可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った。本講演では、研究の背景と構築したシステムの概要を述べる。

1 はじめに

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[1]は、我々が、大学における非文献コンテンツ公開のための共通プラットフォームの開発の一環として行ってきた研究で、その成果をまとめ情報知識学会誌に投稿させて頂いたものである。この度、当論文が情報知識学会論文賞に選ばれたことは、誠に光栄なことであり、ご指導ご協力頂いた関係者および推薦委員の先生方に深く感謝申し上げる次第である。

情報化が進む今日、大学は貴重な学術情報を蓄積するだけでなく、世界に向けて発信することを求められている。大学が公開している学術情報は、学内に蓄積している膨大な学術情報の一部であり、多くの学術情報は公開に至らず、学内に死蔵しているケースも少なくない。それに対して、昨今、各大学では学術論文、紀要、研究報告などの学術情報を登録し、機関リポジトリとして公開している。一方、写真、動画、音声などのコレクションや実験観測データをはじめとした文献以外の学術情報は、教育・研究上有用なものでありながら対象外とされている場合が多い。

このような背景のもと、本研究では、学内に蓄積している学術情報の内、公開が進んでいない写真、動画、音声などのコレクションや実験観測データに焦点をあて、それらを学内外に公開するための指針となるべく、既存情報技術を駆使した、既存情報インフラ上で運用可能な公開手法を考案し、共通プラットフォームとして提供することを目指した。

本講演では、まず、研究全体の目的にあたる、大学における非文献コンテンツ公開のための共通プラットフォームの開発について述べる。続いて、今回受賞対象となった、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について、その概要を述べる。

2 非文献コンテンツ公開のための共通プラットフォームの開発

本章では、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を行うに至った背景であり、研究全体の目標にあたる非文献コンテンツ公開のための共通プラットフォームの開発について概説する。

2.1 非文献コンテンツ

大学内に蓄積している学術情報の内、学術論文や紀要、研究報告書などの文献系のコンテンツを文献または文献コンテンツと呼ぶのに対し、文献系コンテンツ以外の写真、動画、音声、教材などのコレクションや実験観測データを非文献コンテンツと呼ぶ。

2.2 学術情報公開モデル

本研究では、大学内に蓄積している学術情報の内、その公開が進んでいない非文献コンテンツに焦点をあて、それらを学内外に公開するための指針となるべく、既存情報インフラ上で運用可能な公開手法を考案し、共通プラットフォームとして提供することを目指した。共通プラットフォームとは、公開手法を汎用的に実装したシステムで、そのままの形で、あるいは、簡単な設定や簡単な追加実装で再利用できる、いわゆる「使い回し」が可能なシステムである。

まず、非文献コンテンツを学内外に公開する際の問題点を洗い出し、それらを解決するために、大学内のすべての種類の非文献コン

コンテンツを網羅する学術情報公開モデルを作成した。このモデルでは、既存情報技術で構築可能であり、既存情報インフラ上で実運用可能な「学術情報リポジトリ」、「Web-DB管理システム」、「データ配信システム」の3つの公開手法で、大学内のすべての非文献コンテンツの公開をカバーするものとした。図1は、作成した学術情報公開モデルを大学の現状に即した形とし、共通プラットフォームを配置したものである。

2.3 各共通プラットフォームの概要

作成したモデルに基づき、それぞれの公開手法の問題点を改善し、「非文献コンテンツに対応した学術情報リポジトリ」、「多様なアクセス制限に対応したWeb-DB管理システム」、および、「汎用データフォーマットを用いたデータ配信システム」として設計を行った。実証運用では、当大学の研究室で実際に蓄積している複数種の非文献コンテンツを用いて、公開システムとして構築・運用し、

共通プラットフォームとしての有効性と汎用性を確認した。以下に、開発した公開システムの概要を述べる。

【非文献コンテンツに対応した学術情報リポジトリ】：非文献コンテンツの中でも、写真、動画、音声などのコレクションを公開するための共通プラットフォームとなるべく、既存プラットフォームの問題点を改善し、可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った[1]。この部分に関しては、3章で詳細を述べる。

【多様なアクセス制限に対応したWeb-DB管理システム】：公開用Web-DBシステムの一元的な管理・公開を可能とする共通プラットフォームの提供を目的とし、ユーザおよび公開用Web-DBシステム配下のデータベースごとに、きめ細かなアクセス制限を行うことができるWeb-DB管理システムの開発を行なった[2,3]。実証として、地球環境観測データに適用し、地球環境データベースシステムを構築した。詳細は参考文献2,3を参照されたい。

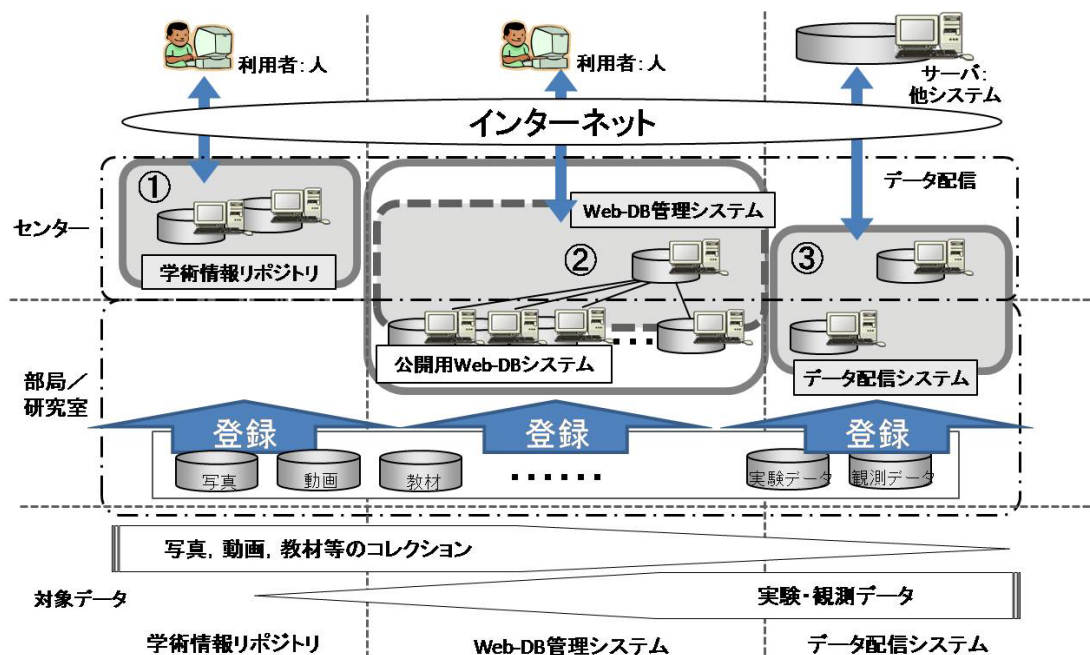


図1 学術情報公開モデル

【汎用データフォーマットを用いたデータ配信システム】:自然科学系の実験観測データを容易に利活用できる形で公開することを目的として,相互配信環境モデルを提案,さらに,その実現のために地球環境観測データに適用し,自然科学データアーカイブシステムの基本構成の開発を行った[4]. 詳細は参考文献4を参照されたい.

3 可視性と保守性に優れた学術情報リポジトリの構築

3.1 設計方針

課題1:先行する文献リポジトリに関しては, Dublin Core[5] (以下DCとする) の考え方に沿った記述法が確立されつつあるが,非文献コンテンツに対しては,それらがもつ多様で専門的な情報をどのようにDCで定義するかが不明確である.

課題2:保守・管理面から考えた場合,管理者が保有する複数の性質の異なったコンテンツ(異種コンテンツ)をどのように管理するか,数千件,数万件以上におよぶコンテンツをどのように分類,登録・保守するか,などの問題がある.さらに,コンテンツ管理者が必ずしも情報技術の専門家であると限らないという問題もある.このため,情報技術に関して専門外である管理者でも,リポジトリの保守管理ができる仕組みの導入を検討する.

課題3:非文献コンテンツは,一般に文字情報を持っていないので,文献コンテンツに比べ利用者に対する検索性が低いという問題がある.一方,非文献コンテンツには,発掘地,所蔵地をはじめとした地名などの情報を持つものが多い.そこで,利用者に対する検索性を高めるため,地名などの地理的位置情

報を得て,より視覚的に検索できる仕組みの導入を検討する.

課題4:各リポジトリに存在する異種コンテンツ同士を横断的に検索,利用することを想定し,リポジトリ間の連携を行い,これらの情報をどのように相互参照するかという検討が必要である.

以上,課題1~課題4を解決し,非文献コンテンツに対応した汎用性の高い学術情報リポジトリを構築するため,表1に示す開発条件を設定した.また,表1には,課題には挙げていないが,⑤プラットフォームの保守性が高いこと,という項目を加え,既存リポジトリプラットフォームをベースにするという条件を設定した.文献コンテンツでの利用実績があり,保守体制が確立しつつある既存

表1 開発条件表

<p>汎用性の確保</p> <p>①データの互換性が確保できること ⇒ 当該リポジトリ上での詳細な定義と,他リポジトリとの互換性を両立(課題1)</p> <p>②他リポジトリとの連携を行えること ⇒ OAI-PMH[6]プロトコルを活用した他リポジトリとの連携を実現(課題4)</p> <p>保守性の確保</p> <p>③複数の異種コンテンツの管理を容易に行えること ⇒ 同一リポジトリ(同一システム)に異種コンテンツを容易に共存させる環境を導入(課題2)</p> <p>④多様かつ膨大な数のコンテンツの管理を容易に行えること ⇒ 分類の登録・管理機能,一括登録機能を導入(課題2)</p> <p>⑤プラットフォームの保守性が高いこと ⇒ 既存リポジトリプラットフォームをベースとする</p> <p>可視性の向上</p> <p>⑥コンテンツが持つ発掘地,所蔵地などの地理的位置に関する情報を可視化し,視覚的な検索機能を提供すること ⇒ 地理的位置情報とGoogle Earthを連携させた検索機能を導入(課題3)</p>

リポジトリプラットフォームを利用することが、保守・管理コストの面で有利と判断したためである。そして、リポジトリプラットフォームのDSpace[7]をベースに、機能を改良、追加する形で開発を進めることとした。

3.2 使用データ

今回は、共著者の森がアジア画像集成として蓄積しているインドの仏像・壁画・遺跡に関する画像を対象にして開発および検討を行った。図2(a)に、蓄積されているコンテンツの例を示す。アジア画像集成は、インドの各地で撮影された画像で、総数は2万件以上に及ぶ。従来、画像のメタデータは、データ管理者の森が、地域、タイトル、所蔵、特徴、サイズ、材質、制作年代などをExcelの表形式で管理していた。また、本システム開発後の検証の目的で、あけぼの衛星の観測データ（図2(b)）、e-Learning素材（図2(c)）、および、第四高等学校物理機器図録（図2(d)）に適用することとした。

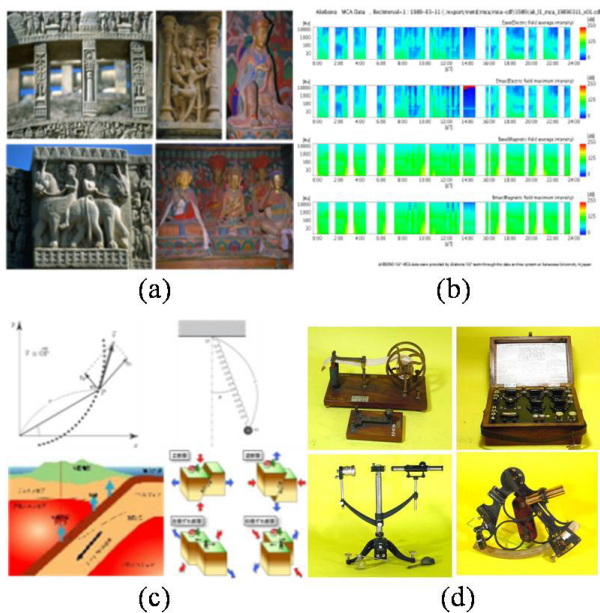


図2 使用したデータ

【(a)：アジア画像集成，(b)：あけぼの衛星の観測データ，(c)：e-Learning 素材，(d)：第四高等学校物理機器図録】

3.3 システム概要

前述の通り、本システムは、リポジトリプラットフォームのDSpaceをベースにして、機能を改良、追加する形で開発を進めた。開発にあたっては、DSpaceの既存クラスを極力書き換えず、クラスをオーバーライドする形で行い、システムの移植やDSpaceのバージョンアップといった場合にも極力影響が出ないように配慮した。図3は、実装したアジア画像集成の一面面（情報表示画面）である[8]。

以下では、メタデータの互換性の確保、保守性の確保、可視性の向上および他リポジトリとの連携について述べる。

3.3.1 メタデータの互換性の確保

DCの標準メタデータ語彙では、非文献コンテンツの情報を的確に表現することは困難なため、各々のコンテンツに合わせ、メタデータ項目を追加した拡張メタデータ語彙を定義する必要がある。今回、各非文献コンテンツ特有のメタデータ項目の定義にDumb-Down原則[9]を導入した。通常、Dumb-Down原則は、組織間の運用ポリシーの違いなどによるメタデータ項目の差異を吸収するために用いられるが、ここでは、各非文献コンテンツの特性の違いによるメタデータ項目の差異を吸収するために用いた。

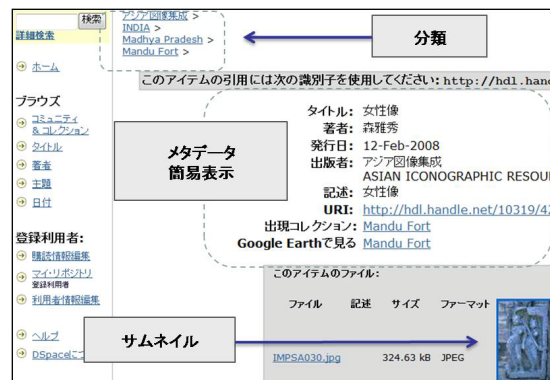


図3 アジア画像集成

3.3.2 保守性の確保

(1) 異種コンテンツの共存

異種コンテンツを同一リポジトリ上に共存させるには、コンテンツごとに適したメタデータ語彙を設定できることと、コンテンツごとに適した一覧表示を行える必要がある。これに対して、メタデータ語彙の設定や検索結果の一覧表示設定をコンテンツ種別に適切な設定に切り換えて利用できるようにコンテンツ別メタデータ語彙の設定法について考案した。図4に異種コンテンツを共存させた場合のイメージを示す。コンテンツの管理単位の最上位がルートコミュニティとなる木構造であることを利用して、種類の異なるコンテンツごとにルートコミュニティを分け、それぞれのコンテンツのルートコミュニティごとにメタデータ語彙の設定と一覧表示の設定を行えるようにした。

(2) コミュニティとコレクションの管理

図4からも分かるように、通常のリポジトリシステムでは、コミュニティとコレクションでアイテムを分類しコンテンツの管理を行っている。この管理のためにWebインターフェイスが準備されるのが一般的であるが、非文献コンテンツの場合、コンテンツの分類が複雑なものが多く、Webインターフェイスを介した方法では体系的な管理が煩雑になりがちである。これに対して、コンテンツ管理者が情報技術の専門家でなくても比較的なじみやすい、Excelファイルを用いたコミュニティとコレクション構造の管理法を考案した。図5は、アジア画像集成で用いたExcelファイルの記述とリポジトリ上での表示例である。

(3) 一括登録

アイテムの登録においても、その数が数千件、数万件以上におよぶ非文献コンテンツに

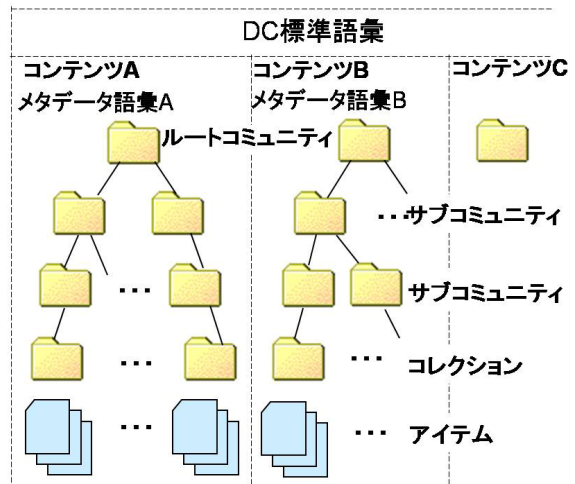


図4 異種コンテンツを共存させた場合のイメージ

	A	B	C	D
1	アジア画像集成	INDIA	Madhya Pradesh	Mandu Fort
2				Sanci
3				Sanci, Eastern Torana
4				Sanci, Northern Torana
5				Sanci, Southern Torana
6				
7				

コミュニティ	コレクション
--------	--------

- アジア画像集成
 - INDIA
 - Madhya Pradesh
 - Mandu Fort
 - Sanci
 - Sanci, Eastern Torana
 - Sanci, Northern Torana
 - Sanci, Southern Torana
 - Sanci, The Second stupa
 - Sanci, Western Torana

図5 コミュニティとコレクションの記述例

対して、Webインターフェイスで逐一登録する方法は、現実的とはいえない。これに対して、前項のコミュニティとコレクションの管理と同様に、コンテンツ管理者が所有のコンテンツを管理するために、情報技術の専門家でなくても比較的なじみやすいExcelなどの表計算ソフトを利用した方法を考案した。表計算ソフト上で管理されるメタデータの出力(TAB区切りまたはCSV形式)を読み込み一括登録が行える仕組みである。具体的には、ファイルの内容を読み込み、メタデータなどの解析を行い、システムが理解できる形式に変更しリポジトリに登録を行うスクリプトを作成した。図6にExcelなどの表計算ソフト上でのメタデータの記述形式を示す。

3.3.3 Google Earthによる情報の可視化

非文献コンテンツでは、全文検索が利用できないため、コンテンツに付与されたメタデータのみによる検索が行なわれる。そのため、利用者は、文献コンテンツの検索と比べると、より適切な検索語句を指定する必要がある。一方、今回利用したアジア図像集成は、図像の出土地や所蔵地といった地理的な位置(地名)を示す情報を持っている。また、アジア図像集成に限らず、非文献コンテンツの中には、地理的な位置情報を持つコンテンツが少なくない。これらコンテンツが持つ地理的な位置情報を利用し、Google Earthと連携して、位置情報を地図上に表示することで可視化し、コンテンツの検索性の改善を図った。

具体的には、アジア図像集成はコレクション名が地名に相当するので、コレクションから座標を取得することとして、2つのスクリプトを作成した。一つ目は、登録されている地名から座標を取得し、Google Earth上に情報を表示させるKMLを生成するスクリプトを作成した。Google Earth上に表示させる情報には、地名の他、アジア図像集成へのリンクを含めることで、Google Earth上からアジア図像集成への検索も可能とした。もう一つは、与えられた地名から地図を表示するための視点を設定するKMLを生成するスクリプトを作成した。図7にGoogle Earthと連携した可視化システムの概要を示す。

3.3.4 他のリポジトリとの連携

他リポジトリとの連携が可能であることを実証するために、学内の学術情報を統一的に公開するポータルリポジトリを構築した。これは、学内に立ち上げられたリポジトリからメタデータをハーベスティングするハーベスタとしてリポジトリを構築し、個別のリポジトリにアクセスすることなく学内の学

表計算ソフトのワークシート

1行目 ヘッダ行	メタデータ要素 1	メタデータ要素 2	...	メタデータ要素 n	コミュニティ・ コレクション	アイテム へのパス
2行目	1個目のアイテムの情報					
...	...					
m+1行目	m個目のアイテムの情報					

図6 メタデータの記述形式

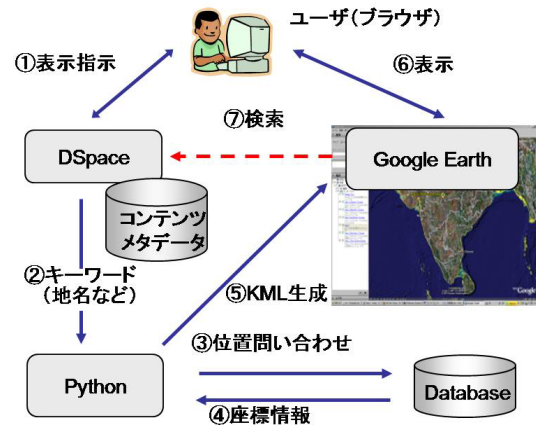


図7 Google Earthと連携した可視化の概要

術情報を横断的に検索できるものである。

今回は、非文献コンテンツに最適化した本システムと文献コンテンツに適した当大学の図書館のKURA[10]で、非文献コンテンツと文献コンテンツを統一的に検索することができることを実証することができた。これにより、学術論文とそれに関連したデータ、図、写真などの非文献コンテンツの一括検索や各リポジトリに存在する異なる分野のコンテンツの横断的な検索が可能となる。

3.4 応用

本研究では、アジア図像集成の他に应用実装を行った。なお、应用実装として、前述のあけぼの衛星の観測データ、第四高等学校物理機器図録、およびe-Learning素材という特性の異なるコンテンツに対しても同一の手法を適用した。図8は、あけぼの衛星の観測データ、第四高等学校物理機器図録、およびe-Learning素材のそれぞれのアイテムの検索結果一覧画面である。

1989-03-06	2-Dec-2008	 Akebono MCA Data ak ii mca 19890	登録日	タイトル	その他のタイトル	11-Dec-2007	 金沢付近の地質鳥瞰図	高原利幸	-	image/jpeg
1989-03-07	2-Dec-2008	 Akebono MCA Data ak ii mca 19890	17-Jul-2009	サーモバイル、熱電管、熱電対列	Melloni's Thermomultiplier	11-Dec-2007	 マグマ発生の様式図	平松良浩	-	image/jpeg
1989-03-08	2-Dec-2008	 Akebono MCA Data ak ii mca 19890	17-Jul-2009	ブルグ氏金属系電計	Breguet's Metallic Thermoelectricity Measurement	11-Dec-2007	 世界の火山	平松良浩	-	image/jpeg
1989-03-09	2-Dec-2008	 Akebono MCA Data ak ii mca 19890	17-Jul-2009	ボロメーター	Bolometer (Linear Bolometer)	11-Dec-2007	 日本列島付近の火山分布	平松良浩	-	image/jpeg
			17-Jul-2009	液体の比熱測定器	Pfaundler's Current Calorimetric calorimeter					
			17-Jul-2009	水銀温度計	Wasser Thermometer					

(a)

(b)

(c)

図 8 応用実装

【(a) : あけぼの衛星の観測データ, (b) : 第四高等学校物理機器図, (c) : e-Learning 素材】

4 おわりに

非文献コンテンツをより広く公開し、教育・研究に有効利用されることを目的に、既存プラットフォームの問題点を改善し、可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った。

今回構築したシステムは、特定分野のコンテンツの特性に依存しない汎用的なものであり、全体目標として掲げた、大学における非文献コンテンツ公開のための共通プラットフォームとして広く応用可能である。また、状況に応じた柔軟な運用が可能で、コンテンツ管理者に掛かる負担も十分抑えることが可能と考えられ、蓄積している様々なコンテンツの公開が促進されると期待できる。

参考文献:

- [1] 高田 良宏, 笠原 禎也, 他3名: 「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.
- [2] 高田 良宏, 笠原 禎也, 他2名: 「多様なアクセス制限に対応した自然科学データベースシステムの開発」, 学術情報処理研究, No.11, pp.50-59, 2007.9.

- [3] Y. Takata, Y. Kasahara, and T. Matsuhira: "Development of a Science Database System Applicable to Various Access Restrictions", *Data Science Journal*, Vol.8, IGY32-IGY43, 2010.2.
- [4] 高田 良宏, 笠原 禎也, 尾崎 友紀: 「汎用データフォーマットを利用した自然科学データアーカイブシステムの開発」, 学術情報処理研究, No.10, pp.5-14, 2006.9.
- [5] Dublin Core Metadata Initiative, <http://www.dublincore.org/> (2009.4.20参照)
- [6] Open Archives Initiative: "Open Archives Initiative Protocol for Metadata Harvesting", <http://www.openarchives.org/pmh/> (2009.4.20 参照)
- [7] DSpace Foundation: "DSpace.org", <http://www.dspace.org/> (2009.4.20 参照)
- [8] 金沢大学: 「アジア図像集成 (Asian Iconographic Resources)」, <http://www.db02.kanazawa-u.ac.jp/dspace>, (2009.4.20 参照)
- [9] 杉本 重雄: 「メタデータについて : Dublin Core を中心として」, 情報知識学会誌, Vol. 10, No. 3, pp. 53-58(2000).
- [10] 金沢大学: 「金沢大学学術情報リポジトリ」, <http://dspace.lib.kanazawa-u.ac.jp/dspace/> (2009.4.20 参照)