

O-012

デジタルコンテンツに適した学術情報リポジトリの構築 Construction of the Academic Information Repository Suitable for Digital Contents

高田 良宏[†] 笠原 禎也[†] 西澤 滋人[†] 森 雅秀[†] 内島 秀樹[†]
Yoshihiro Takata Yoshiya Kasahara Shigeto Nishizawa Masahide Mori Hideki Uchijima

1. はじめに

大学には世界的にみても価値の高い学術情報が数多く蓄積されている。これらの情報を蓄積・管理し、教育・研究に利活用するためのシステムが部局、学科、研究室など、さまざまな単位で個別に構築・運用されることは、それに費やされる資金や労力の面からみても非効率である。また、仕様が異なるさまざまなシステムが運用されても、情報の流通を阻害し、教育・研究に利活用することはできない。このような問題に対し、学術情報を低コストで流通性を維持しつつ、学内外に公開することを目的として、電子的な形態で集中的に蓄積・管理し、機関リポジトリとして公開する動きがある。本学でも附属図書館が金沢大学学術情報リポジトリ (KURA) を立ち上げ運用を行なっている[1]。しかし、現在、運用されている機関リポジトリは、主に学術論文、紀要、研究報告などの書誌情報 (以降、書誌コンテンツとする) を対象としている。一方、書誌コンテンツ以外の写真、動画、実験・観測データをはじめとした多くの学術情報 (以降、デジタルコンテンツとする) は、教育・研究上非常に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。この理由として、コンテンツの種類が多様多岐なため、書誌情報を元に情報管理を行うリポジトリへの登録が厄介である、コンテンツを検索する場合に利用者への検索性が確保できないなどが挙げられる。このような問題を解決し、デジタルコンテンツをリポジトリ化できれば、学内の多様な学術情報が教育・研究に有効利用されると期待できる。本研究では、書誌情報以外のデジタルコンテンツをリポジトリ化する場合の指針となるべく、KURA と連携可能な機関リポジトリプラットフォームの一つである DSpace[2] を改良し、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目指した。

2. 開発諸元

本研究で目標とした開発諸元を次に示す。

- (1) 汎用性の確保: プラットフォームおよび、メタデータの互換性が確保できること
- (2) 保守性の確保: 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。また、同一リポジトリに異種コンテンツを容易に共存させることができること
- (3) 可視性の向上: 所在情報などのコンテンツの地理的位置に関する情報を用いた視覚的な検索機能を導入すること
- (4) 他リポジトリとの連携: OAI-PMH[3]プロトコルを活用した他リポジトリとの連携が可能なこと

今回は、金沢大学内で蓄積されている様々なデジタルコンテンツの内、筆者らがアジア画像集成[4]として蓄積している学術資料 (画像) と、金沢大学総合メディア基盤セン

ターで e-Learning 素材データベース[5]として蓄積してきた e-learning 素材を対象にして開発および検討を行った。図 1 に、蓄積されているコンテンツの例を示す。

3. システムの概要

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。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。

3.1 メタデータの互換性の確保

機関リポジトリでは Dublin Core[6] (以降、DC とする) と呼ばれるメタデータ語彙を使用している。DC は 15 の基本要素と、より詳細な情報を表すための限定子から構成されている。本来 DC は書誌情報を記述するためのメタデータであるため、そのままではデジタルコンテンツに適さない。そのため、本研究では、前述の 2 種類のコンテンツに対して、互換性を保ちつつ必要な項目を追加して拡張メタデータ語彙を定義した。すなわち、限定子を取り除いた場合でも基本要素と値との間に矛盾がないという Dumb-Down 原則に従って定義することであり、これにより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させることができる。

3.2 保守性の確保

3.2.1 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。メタデータ語彙の設定に関しては、登録するそれぞれのコンテンツに適したメタデータ語彙がその種類分設定可能で、さらに、登録の際に容易に切り替えて利用できるようにした。一覧表示に関しても、それぞれのコンテンツに適したメタデータ項目の設定がコンテンツの種類分設定可能とし、表示の際に利用できるようにした。

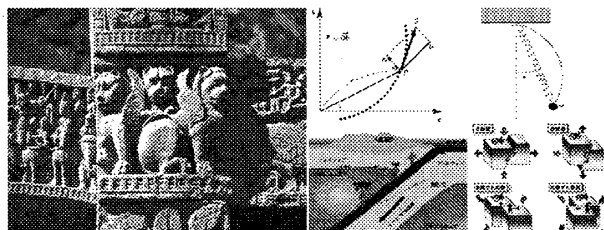


図1 蓄積されているコンテンツの例 (左: アジア画像集成, 右: e-Learning 素材データベース)

[†] 金沢大学 Kanazawa Univ.

3.2.2 コミュニティとコレクションの管理

DSpaceはコミュニティやコレクションでコンテンツの実態(アイテム)を分類しコンテンツの管理を行っている。これらの作成はWebインターフェイスに頼っており、コンテンツの登録とともに負担となる部分である。ここでは、Excelファイルを用いてコミュニティとコレクションの構造の管理を行えるようにした。この仕組みを導入することで、DSpaceの管理者のコミュニティとコレクション管理に掛かる負担が軽減されるとともに、リポジトリの構造の再現を容易に行うことが可能となった。

3.2.3 一括登録

デジタルコンテンツの場合、アイテムが数千件、数万件以上におよぶ場合が多く、これらの登録を簡単な準備で一括登録できるスクリプトを開発した。Excelなどの表計算ソフトの出力ファイルを読み込み、一括登録が行える仕組みを導入した。具体的には、メタデータ、アイテムおよび登録先のコミュニティとコレクションなどの情報を記述したTAB区切りまたはCSV形式の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、DSpaceが理解できる形式に変更し登録する。この方法を導入することにより、登録の際の負担が軽減されるとともに、コンテンツごとの再現を容易に行うことが可能となった。

3.3 Google Earthによる情報の可視化

デジタルコンテンツでは、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、書誌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画像集成は、画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、アジア画像集成に限らず、学内に存在するデジタル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。今回、Google Earthとの連携にはKML(Keyhole Markup Language)を用いた。

具体的には、アジア画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、説明文の他、アジア画像集成へのリンクを含めることで、Google Earth上からアジア画像集成への検索も可能となった。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2にGoogle Earthと連携した可視化システムの概要を示す。

3.4 他のリポジトリとの連携

筆者らが中心となり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の試作を行なった。図3に試作した金沢大学学術ポータル上での検索結果の例を示す。これは、OAI-PMHプロトコルに従い、学内のリポジトリのメタデータをハーベスタにより収集するものである。今回はデジタルコンテンツに最適化した当システムと書誌コンテンツ

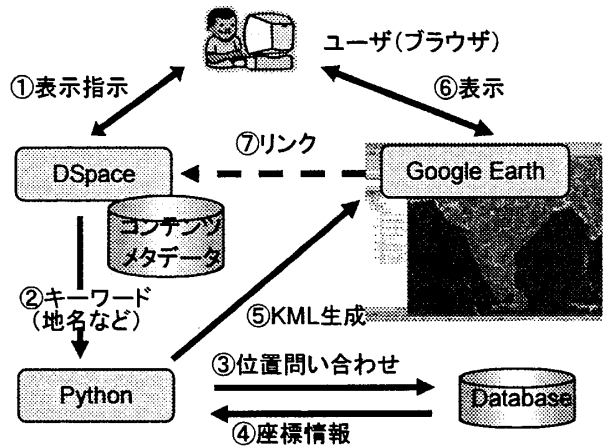


図2 Google Earthと連携した可視化の概要

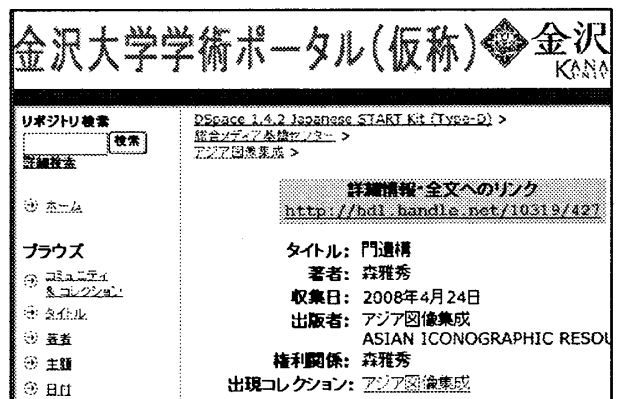


図3 金沢大学学術ポータル

に適した図書館のKURAで実証し、書誌コンテンツとデジタルコンテンツを統一的に検索することができた。これにより、当システムと現況のリポジトリ(KURAなど)が連携可能であることを示すことができた。

4. まとめ

本研究では、デジタル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、既存のプラットフォームの問題点の改善を行った。特に保守性・可視性の改善は多くのデジタルコンテンツのリポジトリ化に対して有効である。現在、当システムは最適化を終え一部一般公開中である。学術情報を統一的に公開するための金沢大学学術情報リポジトリも実証運用を開始した。今後は、学内の多様なデジタルコンテンツに多数適用し、問題点の洗い出しと改良を加えていく。

参考文献

[1]金沢大学学術情報リポジトリ(KURA), <http://dspace.lib.kanazawa-u.ac.jp/dspace/>
 [2]DSpace.org, <http://www.dspace.org/>
 [3]Open Archives Initiative Protocol for Metadata Harvesting, <http://www.openarchives.org/pmh/>
 [4]アジア画像集成, <http://air.w3.kanazawa-u.ac.jp/>
 [5]高田 良宏, 笠原 禎也, 佐藤 正英, 鈴木 恒雄, 松本 豊司, 森 祥寛, e-Learning 素材管理・再利用システムの開発, コンピュータ&エデュケーション, Vol.20, (2006).
 [6]Dublin Core Metadata Initiative, <http://www.dublincore.org/>