

研究室に埋蔵されている研究・教育資料公開の試み 非文献コンテンツを対象とした 学術情報リポジトリの構築

高田 良宏, 笠原 禎也
金沢大学総合メディア基盤センター

(Feb 17, 2010 附属図書館シンポジウム/DRF-KanNihonkai)

Information Media Center, Kanazawa University

1

実験開発の背景

大学には...

- 世界的に見ても価値の高い情報
- 日々確実に増加

書誌系の情報



学術論文
紀要
研究報告書

学術情報リポジトリ

集中的に管理

- 世界に向けて発信
- 有効に利用



画像, 動画,
音声などの
コレクション
実験・観測
データ

学術的に非常に貴重
学内外から参照要請

- 公開が遅れている
- 有効に利用されていない

Information Media Ce.

2

非文献コンテンツ

文献コンテンツ

⇒学術論文, 紀要, 研究報告書など, 書誌情報を元に管理できるコンテンツ

非文献コンテンツ

⇒文献コンテンツ以外, 写真, 動画, 音声, 教材などのコレクション, 実験観測データ

大学内に蓄積している学術情報の内, 公開が十分に進んでいない**非文献コンテンツ**に焦点をあてる

Information Media Center, Kanazawa University

3

非文献コンテンツに 対応した学術情報リポジトリの開発

非文献コンテンツ

⇒技術的に学術情報リポジトリの対象外とされている場合が多い
どの様に公開していくか?
流通性, コスト, 先行する文献コンテンツでの実績

既存リポジトリプラットフォームの問題点を改善

非文献コンテンツに対応
共通プラットフォーム
として提供

非文献コンテンツを公開する際の問題点

- 書誌情報の表現に適したメタデータ語彙を非文献コンテンツにどのように対応させるか
- 膨大なコンテンツの分類や登録をどのように行うか
- 異種コンテンツをどのように共存させるか
- 文字情報を含まないコンテンツをどのように検索しやすくするか
- 相互参照をどのように実現するか

4

蓄積している非文献コンテンツ(抜粋)

| コンテンツの内容 | 分野 | 種類 | 全文検索 | 件数等 | 地名等の情報 |
|----------------|------|----------|------|---------|---------|
| インドの仏像・壁画・遺跡 | 人文科学 | 写真 | 不可 | 2万件以上 | 発掘地/所蔵地 |
| 中国語の方言 | 人文科学 | 音声/地図 | 不可 | 約千件 | 調査地 |
| 中国の伝統芸能 | 人文科学 | 動画/写真 | 不可 | 800GB以上 | 撮影地 |
| 岩石標本 | 自然科学 | 写真 | 不可 | 数万件 | 採取地 |
| 「あけぼの衛星」の観測データ | 自然科学 | バイナリ/画像 | 不可 | 約7千件 | 観測座標 |
| 資料館所蔵品 | 共通 | 写真 | 不可 | 数千件 | 作成地/他 |
| e-Learning素材 | 共通 | 作成ソフトに依存 | 一部可 | 約1万件 | 作成地 |

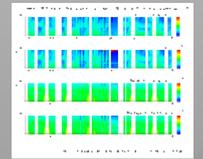
Information Media Center, Kanazawa University

5

蓄積している非文献コンテンツ その2



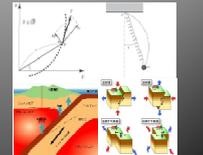
インドの宗教画像(アジア画像集成)



あけぼの衛星の観測データ (サマリープロット)



四高物理機器図録



e-Learning素材

Information Media Center, Kanazawa University

6

実装例:その1 アジア画像集成

このアイテムの引継ぎに以下の識別子を使用してください: <http://hdl.handle.net/10119/10119>

メタデータ簡易表示

タイトル: 永住権
 名称: 高橋
 発行日: 12-Feb-2008
 出版者: アジア図像集成
 ASIAN ICONOGRAPHIC RESOUR
 識別子: 文住権
 URL: <http://hdl.handle.net/10119/10119>
 Google Earthで見ると
 Masaka East

サムネイル

Information Media Center, Kanazawa University

実装例 その2

あけぼの衛星観測データ サマリープロット

e-Learning用素材

四高物理機器図録 (資料館収蔵品)

Information Media Center, Kanazawa University

開発方針

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

- リポジトリプラットフォームのDSpaceをベースにする**
 (KURAをはじめ、国内の機関で最も多く使用されている)
 ⇒ 機能を改良、追加するという形で開発を進める
- メタデータの互換性の確保**
 ⇒ 当該リポジトリでの詳細な定義と他との互換性の両立
- 保守性の確保**
 ⇒ 情報技術に詳しくない管理者でも、容易に管理できる仕組みの導入
- Google Earthによる情報の可視化**
 ⇒ 地理的な位置情報を用いた視覚的な検索機能の導入
- 他リポジトリとの連携**
 ⇒ 文献コンテンツと非文献コンテンツを統一的に検索

メタデータの互換性の確保

メタデータを記述するための語彙: Dublin Core

- 文献コンテンツ ⇒ メタデータの記述法が確立
- 非文献コンテンツ ⇒ 多様で専門的な情報を的確に表現できない
 ⇒ **Dublin Coreを拡張する必要あり!**

無暗に拡張するとリポジトリ間の整合性が取れない
Dumb-Down
 原則

メタデータを拡張する場合の原則
 ⇒ 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

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

メタデータの互換性を確保

当該リポジトリ内での詳細な定義と他リポジトリとの互換性の両立が可能

メタデータの互換性の確保

アジア図像集成用の拡張メタデータ語彙

アジア図像集成用の拡張メタデータ語彙 (抜粋)

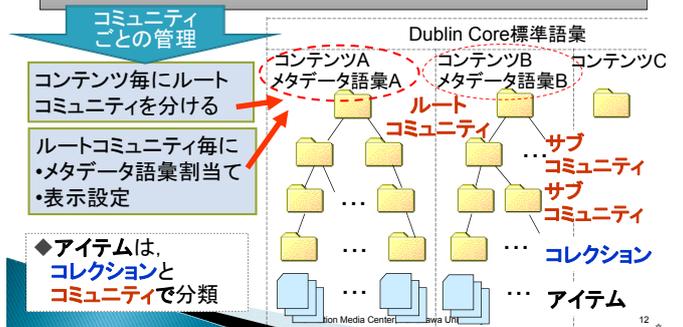
| 基本要素 | 限定子 | 項目名 |
|-------------|------------------|-------|
| Coverage | currentLocation | 所蔵・所在 |
| | originalLocation | 出土地 |
| Description | material | 材質 |
| | dimensions | サイズ |
| Date | datePhotographed | 撮影日 |
| Identifier | original | 画像コード |
| Appendix | nameOfFile | ファイル名 |
| | nameOfFolder | フォルダ名 |

保守性の確保

異種コンテンツの共存

複数の異種コンテンツの管理を容易に行いたい!

⇒ 同一リポジトリに性質の異なる非文献コンテンツを複数登録



保守性の確保

分類とアイテムの登録・管理

管理者にとって最も負担となる部分である！

コンテンツ管理者は、コンテンツに関する情報をExcelなどの表計算ソフトで管理している場合が多い

- 分類の管理
 - ⇒ コミュニティ、コレクションの登録など
- アイテムの管理
 - ⇒ アイテムの一括登録など

エクセル形式のファイルやエクセルの出力ファイル(タブ, CSV形式)を用いて管理を可能とした

保守性の確保

コミュニティとコレクションの記述

エクセル上の記述

| A | B | C | D |
|---|---------|-------|-------------------------|
| 1 | アジア画像集成 | INDIA | Madhya Pradesh |
| 2 | | | Mandu Fort |
| 3 | | | Sanci |
| 4 | | | Sanci, Eastern Torana |
| 5 | | | Sanci, Northern Torana |
| 6 | | | Sanci, Southern Torana |
| 7 | | | Sanci, The Second stupa |
| 8 | | | Sanci, Western Torana |

リポジトリ上の表示

- 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
- 工学
 - 電気学
 - 電気工学
 - 物質化学
- 情報教育
 - プログラミング
 - リテラシー
- 人文学
 - 芸術
 - 心理学
 - 法学
 - 政治学

既存リポジトリの階層をエクセルに保存することも可能

保守性の確保

一括登録

メタデータをExcel等の出力であるタブ区切りテキスト等から読み込む

- 様々なコンテンツ登録に使用できるように、汎用的に作成
 - 1行目: ヘッダ行: メタデータ要素並び
 - 2行目以降: 1行1アイテムの情報(1行目に対応したメタデータ並び、登録先のコミュニティとコレクション、アイテムが保存されているパス)
- 既存リポジトリのメタデータを保存することも可能

表計算ソフトのワークシート(準備するデータの形式)

| 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 | | | | | |

異種コンテンツ共存 実証運用1

Excelでの管理例

| A | B | C | D |
|----|----------|-------|-------------------------|
| 1 | 四高物理機器目録 | 電子 | |
| 2 | | 物理学 | |
| 3 | | 工学 | |
| 4 | | 数学 | |
| 5 | | 電気工学 | |
| 6 | | | |
| 7 | Akebono | MCA | 1989 |
| 8 | | | 1990 |
| 9 | | | 1991 |
| 10 | | | 1992 |
| 11 | | | 1993 |
| 12 | | | 1994 |
| 13 | | | 1995 |
| 14 | | | 1996 |
| 15 | | | 1997 |
| 16 | | | 1998 |
| 17 | | | 1999 |
| 18 | | | 2000 |
| 19 | | | 2001 |
| 20 | | | 2002 |
| 21 | | | 2003 |
| 22 | | | 2004 |
| 23 | | | 2005 |
| 24 | | | 2006 |
| 25 | | | 2007 |
| 26 | | | |
| 27 | アジア画像集成 | INDIA | Madhya Pradesh |
| 28 | | | Mandu Fort |
| 29 | | | Sanci |
| 30 | | | Sanci, Eastern Torana |
| 31 | | | Sanci, Northern Torana |
| 32 | | | Sanci, Southern Torana |
| 33 | | | Sanci, The Second stupa |
| 34 | | | Sanci, Western Torana |

異種コンテンツ共存 実証運用2

コミュニティ・コレクション

IMC Data Repository
Information Media Center of Kanazawa University

コミュニティ・コレクション

- Akebono
 - MCA
 - 1989
 - 1990
 - 1991
 - 1992
 - 1993
 - 1994
 - 1995
 - 1996
 - 1997
 - 1998
 - 1999
 - 2000
 - 2001
 - 2002
 - 2003
 - 2004
 - 2005
 - 2006
 - 2007
- 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
- 四高物理機器目録
 - 電子
 - 物理学
 - 工学
 - 数学
 - 電気工学

異種コンテンツ共存 実証運用3

一覧画面(ブラウザから表示)の表示: コンテンツごとに変更可能

あけぼの衛星の観測データ (サマリープロット)

| 発表発行日/年代 | 登録日 | タイトル | 著者/作成者 | 管理番号 |
|------------|------------|--|--------|----------|
| 1989-03-06 | 2009-07-03 | Akebono MCA Data: ak_11_mca_19890306_v01.cdf | 森澤秀 | IMPS0039 |
| 1989-03-07 | 2009-07-03 | Akebono MCA Data: ak_11_mca_19890307_v01.cdf | 森澤秀 | IMPS0038 |

インドの宗教画像 (アジア画像集成)

| 登録日 | 発表発行日/年代 | タイトル | 著者/作成者 | 管理番号 |
|------------|------------|------------|--------|----------|
| 2009-07-03 | 1989-03-06 | アショカ王柱と菩提樹 | 森澤秀 | IMPS0039 |
| 2009-07-03 | 1989-03-07 | アショカ王柱と菩提樹 | 森澤秀 | IMPS0038 |

四高物理機器図録

| 発行日 | タイトル | そのほかのタイトル | キーワード | 管理番号 |
|------------|------------|--|----------------|-------------|
| 2009-07-03 | ワイヤマン式電流計 | Large Mirror Galvanometer after Weidenmann | 富山県立大学物理情報センター | 123456-0067 |
| 2009-07-03 | ワイヤマン式正誤差計 | Tangent-Busssole nach Weber | 富山県立大学物理情報センター | 123456-0063 |

異種コンテンツ共存 実証運用4

一覧画面(検索した場合)の表示

あけぼの衛星の観測データ
(サマリープロット)

| 登録日 | タイトル | 著者/作成者 | 出版者 |
|------------|--|--|---|
| 7-Jul-2009 | Akebono MCA Data ak_11_mca_20090410_v01.pdf | 高瀬 健治, 渡辺 康志 Kasehara, Yoshiyuki Takata, Yoshihiro | Akebono VLF Team of Kanazawa Univ. in Japan |
| 7-Jul-2009 | Akebono MCA Data ak_11_mca_20050509_v01.pdf | 高瀬 健治, 渡辺 康志 Kasehara, Yoshiyuki Takata, Yoshihiro | Akebono VLF Team of Kanazawa Univ. in Japan |

インドの宗教画像
(アジア画像集成)

| 登録日 | タイトル | 著者/作成者 | 出版者 |
|------------|------|----------|------|
| 7-Jul-2009 | 遺蹟 | 金沢大学 森屋秀 | 金沢大学 |
| 7-Jul-2009 | 生 | 金沢大学 森屋秀 | 金沢大学 |

四高物理機器図録

| 登録日 | タイトル | 著者/作成者 | 出版者 |
|------------|---------|----------|------|
| 7-Jul-2009 | 物理学機器図録 | 金沢大学 森屋秀 | 金沢大学 |
| 7-Jul-2009 | 電気図録 | 金沢大学 森屋秀 | 金沢大学 |

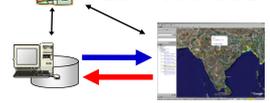
可視性の向上

Google Earthによる情報の可視化

- ▶ 非文献コンテンツの検索性を改善!
⇒ 全文検索を利用できない(メタデータに対する検索のみ)
- ▶ 非文献コンテンツの持つ位置情報に着目

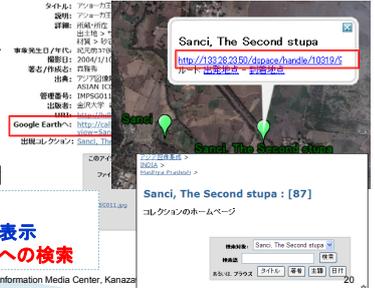
発掘地, 所蔵地, 作成地など

地図上に可視化する仕組み



地図上に情報をプロット

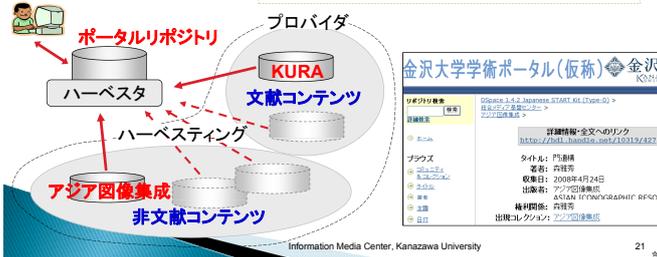
- ◆ DSpaceの情報を地図上へ表示
- ◆ Google EarthからDSpaceへの検索



他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
⇒ ポータルリポジトリ(ハーベスタ)を構築(図書館に協力)

文献コンテンツと非文献コンテンツを統一的に検索できることを実証



まとめ

- ▶ 既存リポジトリプラットフォームの問題点を改善し、非文献コンテンツに適した学術情報リポジトリを構築
 - 今回の改良は多くの非文献コンテンツのリポジトリ化に対して有効である
 - 実装したシステムは最適化を終え一部一般公開中
- ▶ 今後、多様な非文献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行く
 - 実際の運用を見据えた改良を行う

デモ

Google Earthによる情報の可視化

