

いまどきの文献画像データベース

平成21年11月13日

図書系職員のためのアプリケーション
開発講習会



もし、文献画像データベースを 作るように言われたら...



Dou-Shi-
You!

業者に外注すれば業者のほうである程度考えてくれるところはあるんですが、なんでもお任せともいかないですね。
一緒にいまどきの状況がどのようなものかみていきましょう！

この機会に、個々の要素について
イチから考えてみましょう！

メタデータ形式はどう決める？

検索システムはどうする？

画像のmisekataはいかに？

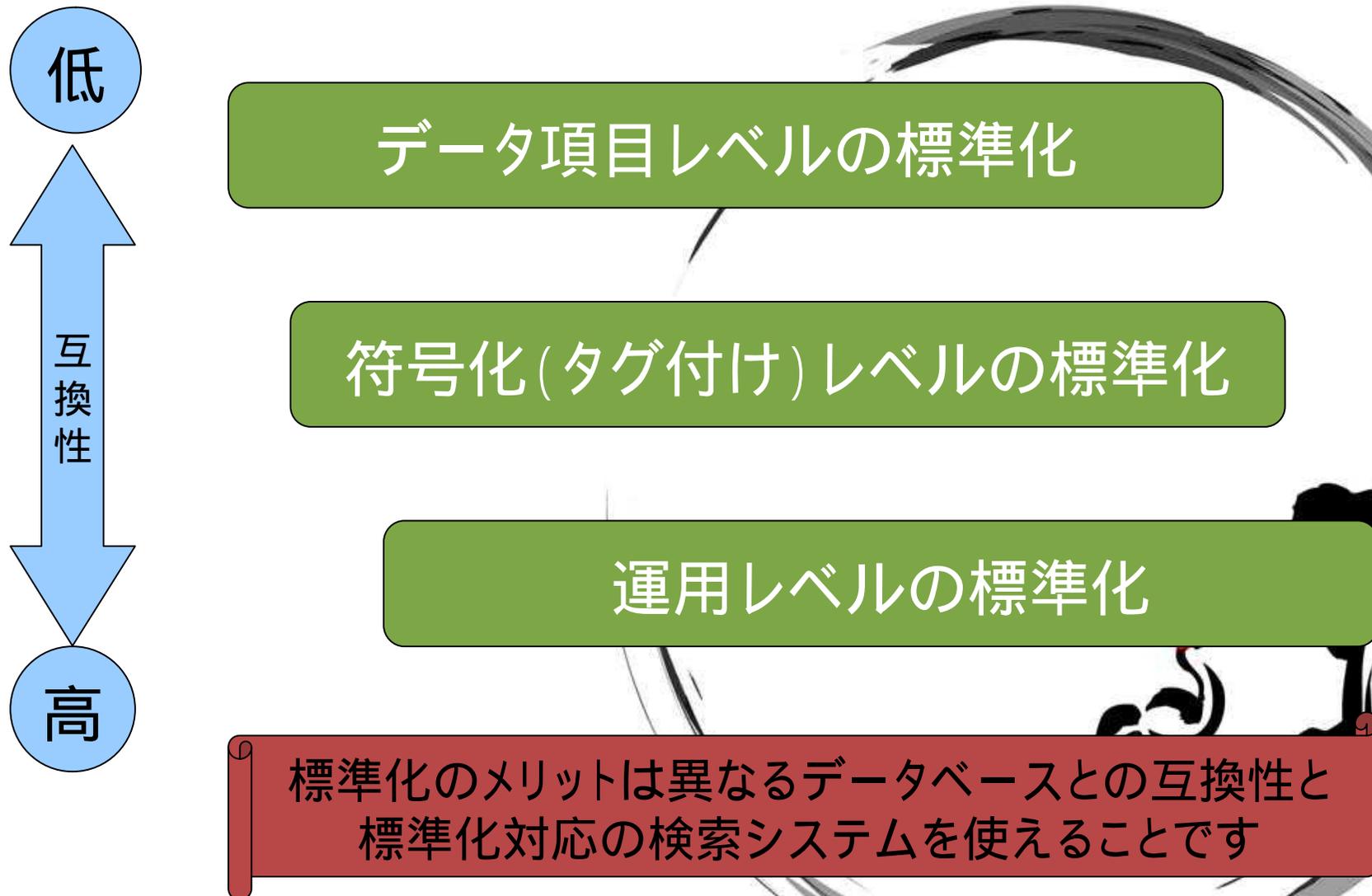


メタデータ形式はどう決める？

- メタデータとは「データについてのデータ」のことです
- 図書目録もその一種ということになります
- グローバルスタンダードというか、標準化の話がいくつかあります



標準化についてモデル的に 考えてみる

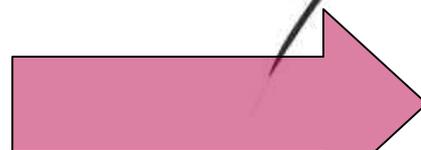


具体的に互換性をみてみましょう

～ダブリンコア同士の昔話2002～

NII (JuNII)

ダブリンコア
(OAI準拠)で
書かれた
メタデータ
A



データ変換して
メタデータBに
統合

Engel (東大経済)

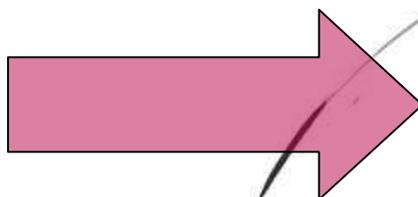
ダブリンコア
(XML)で
書かれた
メタデータ
B

同じダブリンコア間でも微妙に違うデータ形式を
解析してデータ変換プログラムを書くはめに

具体的に互換性をみてみましょう

～ 符号化レベルまで同一の場合～

EADで
書かれた
メタデータ
A

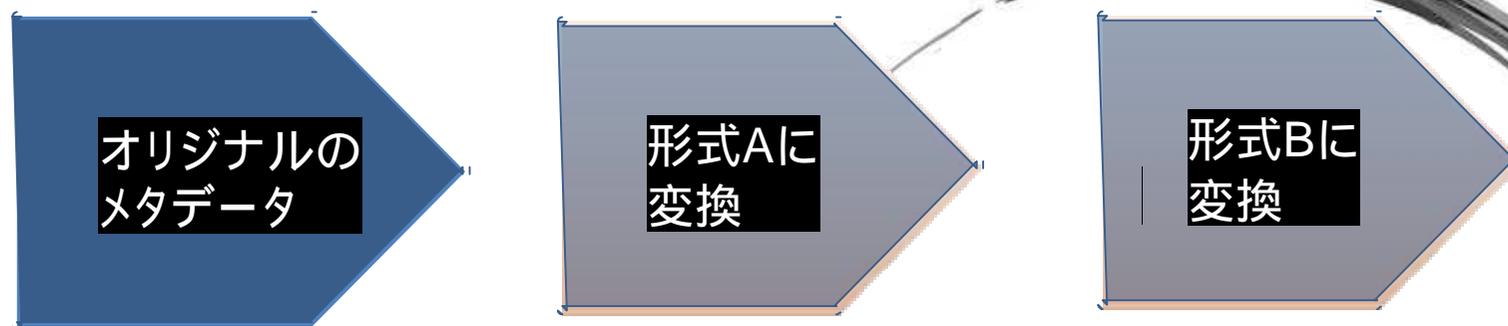


(たぶん)データ変換不要で
メタデータBに統合可能

EADで
書かれた
メタデータ
B

(たぶん)データ変換プログラムを書く必要なし、
そのままEAD対応の検索システムで利用可能。
ただし、運用ガイドラインが異なっていると、
個々のコレクションごとにタグの用途に差異が
できるかも??? (こころへん自信なし)

符号化(タグ付け方式)の変換と情報の劣化



機械的にデータ変換を行うと、**基本的に情報量は変わらないか減る**と考えています

データ形式の変換の際に、データの間違いを直す、増えた項目のデータを人手で調べなおす、などにより、逆に情報を増やすことも考えられます。

規格の列挙

- ISBD (国際標準書誌記述)
- ISAD(G) (国際標準記録史料記述)
- FRBR (書誌レコードの機能要件)
- ISO 11506 (オフィス電子文書のメタデータ規格)
- 各種MARC (NCMARC, JPMARC, MARC21など)
- ダブリンコア
- JuNII2 (NIIによる機関リポジトリ用の規格)
- EAD (*Encoded Archival Description* アーカイブの国際標準)
- MODS (NDLデジタルアーカイブシステムで使用予定)
- XMP (Tiff画像やPDFに埋め込むためのメタデータ規約)
- METS (XML形式)
- 各種運用ガイドライン

整理して示そうと思ったのですが、どうもまえだには、まだ理解しきれていないので列挙どまりまで。

ダブリンコアの動向

- 規制の弱い15エレメント
- 既存の特殊目録をダブリンコアに変換することも容易
- 実は、RSSなど気づきにくいところでも使われています。
- 国会図書館による運用ガイドラインもあります

最近では、基本の15のエレメント(データ項目)に加え、さらにエレメントが追加されているようです

EADの動向

- アーカイブスで使われるメタデータの国際表標準
- SGMLかXML上で使えるみたいですが、XMLでの利用がメインのようです
- EAD対応の検索システムもすでに存在しています
- コレクションを多階層で管理可能です。ただし、1コレクション1ファイルなのでファイルがかなり大きくなりますが...

http://archives.nijl.ac.jp/DAS/projects/eadfa/20051020_gotoh.htm

国際標準として注目はしているのですが、
利用にあたっては対応の学習が必要かと。

junii2の動向

- NIIが機関リポジトリのデータ交換用に作成したメタデータ規格
- もとが、juniiなので、ダブリンコアの影響を受けているかんじ
- 雑誌論文向けのデータ項目を用意

<http://www.nii.ac.jp/irp/archive/system/junii2.html>

MODSの動向

- MARCXMLを簡略化した(サブセットの)メタデータ規格
- もとが、MARCのため図書館屋的には親しみやすいかも
- MARCXML形式の書誌データをMODSに変換するのは簡単みたい
- 将来的にNIIからMARCXMLでデータ入手できるようになれば、まえた作のOPACダウンローダーみたいな苦し紛れな仕組みは不要かも
- 国会図書館でも採用の予定あり

http://www2.aasa.ac.jp/org/lib/j/netresource_j/guenther0306/3.1guenther_j.pdf

検索システムの話

- いまどきの検索システムというと、日本の商用では、InfoCom社が進んでいそうです。
 - データ項目を問わないところで、InfoLib-DBRというシステムがあります。
 - また、EAD対応のシステムもあるようです。
 - もっとも「予算があれば」、の話ではありますが...
- InfoLib-DBRほどの機能はありませんが、データ項目フリーのシステムとして、東京大学経済学部図書館の「古貨幣DB」のシステムをみてみます。



「古貨幣」検索システムとは

- 経緯
 - 東京大学経済学部が古貨幣の画像を公開するにあたり用意した検索システムです
- 特色
 - 表形式のメタデータを簡単に公開できます。
 - データ項目は自由です。
- 問題
 - 複数検索項目をまとめて検索できない(経済学部カスタマイズ希望案件)
 - 新規DBを作るのは簡単なのですが、細かい設定ができないため、機能に不足を感じる部分があります
 - ブラウジング機能をもっていません。
 - 詳細結果画面をもっていません
 - 検索可能な項目がもれなく表示されます
 - 繰り返しに項目に対応していません



「古貨幣検索」システムの使い方

- まず、東京大学経済学部資料室に相談してソースコードをもらいます。
- ソースコードをサーバーにインストールしましょう。必要なものはPHPとMySQLが使える環境です。
- 必要に応じてカスタマイズしましょう！

実際に古貨幣検索システムを流用して東京大学社会科学研究所の
特殊目録データベースを作成しています

https://mbc.dl.itc.u-tokyo.ac.jp/ISS_catalog/

まえだによる古貨幣システムのカスタマイズ

- 文献ごとに簡単な紹介を表示するようにした
- カンマ区切りではなくタブ区切りのデータをロードできるようにした
- DB選択メニューの並び順を変更できるようにした

たぶん、「改善」になっているとは思いますが...

実際の活用例

「東京大学社会科学研究所特殊コレクション一覧」

複数の文献を登録可能で、それぞれのデータ項目を指定することができます

データ項目に応じて、検索指定も変わります

文献に応じてデータ項目を変えることができます。

文献画像があれば、そのリンクを張ることもシステム上、可能です。

社会科 料 - Microsoft Internet Explorer

yo.ac.jp/ISS_catalog/index.php?kind=sogo

研究所所蔵特殊コレクション一覧 (試行)

十河本
元国鉄総裁十河信二氏(1884-1984)

十河本
タイトル関連情報
タイトル関連情報

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 次へ (検索結果 1726件中1~100件目)

タイトル関連情報	版	巻次	年次	出版者
世界現勢報告 / 日本評論社編				東京 : 日本評論社
昭和十三年の国際情勢 / 赤松祐之著				東京 : 日本国際協會
昭和十五年の国際情勢 / 日本外政協會編				東京 : 日本外政協會
新日本同盟會報 / [新日本同盟編]		昭和8年6月		東京 : 新日本同盟
新日本同盟會報 / [新日本同盟編]		昭和8年7月		東京 : 新日本同盟
新日本同盟會報 / [新日本同盟編]		昭和8年8月		東京 : 新日本同盟
新日本同盟會報 / [新日本同盟編]		昭和8年10月		東京 : 新日本同盟
国際経済の諸問題 / ヴァルガ編纂 ; 世界経済研究会[訳]				東京 : 世界経済研究会
アジア民族興亡史観 / 東洋協会調査部 [編]	上			東京 : 東洋協会

いまどきの文献画像表示の話

- 文献画像表示といえば、Webと相性のよいところで、Jpegあたりが使われてきました。あと、PDFとか。
- このところの流行(?)をちょっと追ってみましょう！



Jpeg2000

- 知る人は知っているJpegの後継規格
- 国会図書館の「近代デジタルライブラリ」でも、Jpegと併行して、Jpeg2000を使い始めたようで、要注目です。
- Jpeg2000を使うとカラーの画像をJpeg以上に高圧縮することができます。
- 最近、保存用画像フォーマットとしてTiffとともに使われてきているようです。
- 難点は、ブラウザにプラグインをいれないと見られないことですが...

東京大学経済学部でも、Jpeg2000を使っていますが、同じJpeg2000でもこちらは、Part6 (JPM)という文献むけの規格です。ただし、プラグインの対応っぷりは、無印のJpeg2000のほうがよいみたい

Flashを使った画像表示

- 最近では、文献画像をFlashで表示することも多くなってきています。この講習会では次の3種類について評価を行ないましたので、それをまとめてみます。
 - Zoomify Express
 - iPallet/Lime
 - KmView-f



Zoomify Express

- フリーで使えるFlashビューワー
- 軽さの秘密は、Google MAPと同じく大きな画像をタイルみたいにファイル分割して、その都度必要なファイル(タイル)のみ読み込んでいるから、みたいですね。
- 画質的には、おそらくJpegの品質で75くらいかと。最高画質に比べれば落ちますが、そう悪くはないです。
- 画像の回転などの加工処理まではできないので、画像によっては事前に加工をしておくのがよさそうな

講習会の試験ページ

http://mbc.dl.itc.u-tokyo.ac.jp/murakami/zoomy/226_012/

Zoomify Expressの 画像のロードについて

- 各ディレクトリごとに、HTMLとFlash本体を置く必要がある？
 - いい方法があるかもしれませんが、こころへんはプログラムで一括処理可能
- 画像の一括変換
 - Windowsアプリ(GUI)を使って画像変換を行ないます。が、これでは大量のデータロードは難しい、昔ながらのバッチファイルでなんとか処理できそうですが...

iPallet/Lime

- 大学図書館業界ではメジャー
- 堀内カラーに電子化を依頼すると、もれなくついてくるみたい。
- 堀内カラー講習会内部での利用許諾を得て、ソフトウェアを入手。
- 画像の持ち方は東大が産学連携で開発した方式、Zoomify Expressと同様に画像をタイルよろしくJpeg分割するので軽いかも。

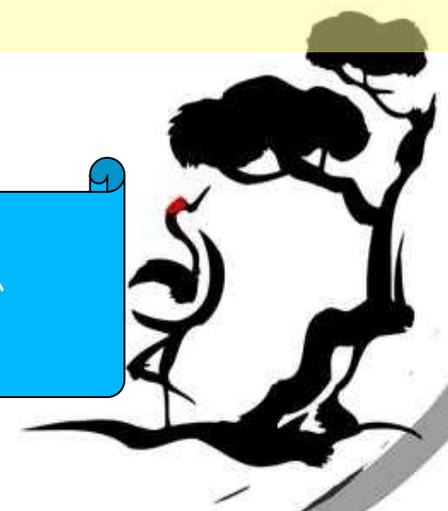
講習会の試験ページ

<https://mbc.dl.itc.u-tokyo.ac.jp/mbc/maeda/iPallet/pages/>

iPallet/Limeの 画像データ一括ロード

- Windowsマシン上で動作
- 最大で6000ピクセル(芝居番付DBの画像では、90%くらいに縮小が必要だった...)
- 1000画像までの一括処理(フォルダごと)が可能

画像データのピクセルに上限があることが少し面倒ですが、これがクリアされていればあとは楽です。



KmView-f

- 国際マイクロ社のFlash画像ビューワー (無償版)
- iPallet/Limeと同じく、東大を含め日本の大学図書館での利用が多いです
- この無償版の利用許諾を講習会用に得ました。
- マルチページに対応！
 - ファイル名が限定されるので、ちょっと注意
- サムネイル表示！
 - サムネイル画像は自分で用意です
- jpegをそのまま利用可!
 - zoomyやiPalletのように、画像を分割するわけではないので、表示が重いといえは重いです。

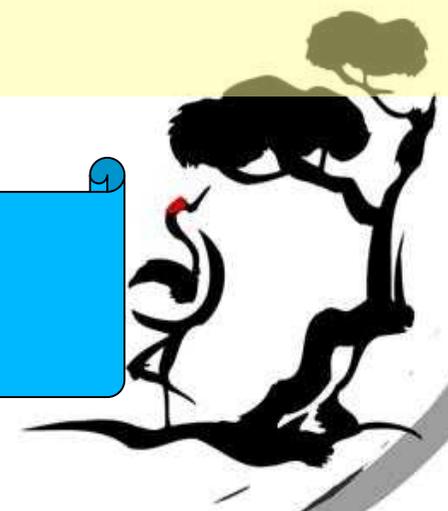
講習会の試験ページ

<https://mbc.dl.itc.u-tokyo.ac.jp/mbc/maeda/kmView-f/kmview.html>

KmView-fの 画像データ一括ロード

- Jpegファイルを命名ルールに従ってディレクトリに配置すればOK！ これでマルチページ表示ができます
- あと、各ディレクトリに設定ファイル(最大ページ数を記述)も必要です。ここらへんも一括処理可能です。

Jpeg形式で文献画像を用意できれば、
使うのはとても簡単です。



ZOOMA Viewer

- ZOOMA社の画像ビューワー
- iPallet/LimeやZoomify Expressと同様に画像をタイル状に分割して使用
- Active X版とFlash版があり



PDFとPDF/A

- PDFはPDFでも、PDF/A という規格のものを使うのがよいみたい
- 長期にわたり利用を可能にするために、フォントの埋め込みを義務づけたり(マシン依存をなくしたいり)するなど通常のPDFと違い制約があります。

