

アーカイブズの構造情報とデジタルデータの仕様情報の入力を重視した Digital Cultural Heritage とそのコンテンツマネージメントシステム

研谷紀夫

東京大学大学院 情報学環

文化資源をデジタル化して格納・公開するシステムである Digital Cultural Heritage においては、資料の出所などに関する階層構造化された情報を格納することが望ましい。さらに、格納されたデジタルデータの仕様に関する情報なども、画像評価のために記録されていることが必要である。本論考では、このような背景を踏まえ、Digital Cultural Heritage においてアーカイブズの出所に関する構造的な情報と、デジタルデータの仕様情報を反映・表示させるコンテンツマネージメントシステムの構築例を示す。

Digital Cultural Heritage and Its Content Management System, with a Focus on Archive Structural Information and Digital Data Spec Information Entry

Norio Togiya

Interfaculty Initiative in Information Studies University of Tokyo

With regards to Digital Cultural Heritage, the system for digitally storing and publicizing cultural resources, we feel that structured data such as place of origin of materials, etc. should be stored. Additionally, for the sake of image assessment, information regarding the specifications of stored digital data and such need to be recorded as well.

In this discussion, in accordance with this context, we will present a sample construction of a content management system which reflects/displays data such as specification information of digital data and structural information of the Digital Cultural Heritage archives, such as place of origin.

1. はじめに

文化資源をデジタル化して格納・公開するシステムであるデジタルアーカイブや Digital Cultural Heritage (以下総称して DCH と述べる) などにおいては、図書資料、文書資料、芸術作品資料など多様な資料が格納される。

この中で、特に文書資料に関しては、文書が作成された出所の階層構造を記録し、それらをアーカイブズの目録情報などに反映させる必要がある。これらを含めた、国際的なアーカイブズの目録記述の基準は、ICA(International Council on Archives)などによって策定され ISAD(G)(General International Standard Archival Description) [1]において示されており、これらに適応した、メタデータを入力することが推奨されている。

また、格納されたデジタルデータも、評価のために、デジタルデータの仕様に関する情報などを掲載されることが望ましい。本プロジェクトでは、このような現状を踏まえ、DCH においてアーカイブズの構造情報とデジタルデータの仕様情報を反映・表示

させるコンテンツマネージメントシステム (以下 CMS) の構築例を示す。

2. 構造情報とデジタルデータの仕様情報を反映させるシステム

文化資源をデジタル化して公開する DCH システムにおいては、図書、絵画、文書、映像、造形物、無形文化財など多様な文化資源をデジタル化したコンテンツが格納される。このような資料の中で、図書や絵画といった資源は、一定のジャンルや資料群ごとにグルーピングなどが行われるが、基本的には、一点一点のアイテムが資料情報化の基本単位となる。図書の分野における構造化モデルとしては、FRBRモデルなどがあるが、これは書誌を作成する様々な要素を構造的に捉えモデルである[2]。そのため、資料が成立した場所や組織、保存・保管されていた環境や場所に関する階層構造のモデル化がその主目的ではない。

これに対して、主に公文書や歴史史料などの多くは、文書の出所などにおいて階層構造が生じるため、それらに応じて目録の情報化を行う必要がある。具

ンド、シリーズ、アイテムの全てを使用する必要はなく、必要な要素だけを組み合わせることが可能である。さらに、フォンドの下にサブフォンドなどを追加し、要素を増加させることも可能である。

また、アーカイブズのメタデータの数は極めて多数に及ぶことや、設計はされていてもデータが入力されない項目などもある。そのため、表示するメタデータと表示しないメタデータを設定することも必要となる。これらを反映させる画面として、**図 3** に示されるように、表示情報を管理する機能を設けた。



図 3：メタデータの表示設定画面

本機能によって、設定された、メタデータは、管理画面から入力することが可能であり、実際のデータは、**図 4** のように、利用者が閲覧するフロント画面から参照することができる。

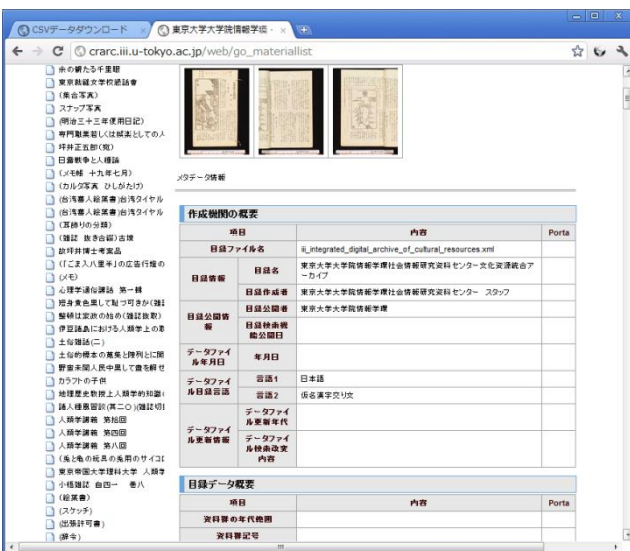


図 4：メタデータの表示

また、本システムでは、資料を 1 点 1 点登録するのではなく、CSV フォーマットでダウンロードし、それらに情報を入力し、**図 5** のように、アップロードすることで、メタデータの登録を可能とする機能を設けた。これによって、あらかじめ設定した構造

に基づいて、システム技術者ではなくてもデータを入力できるシステムを構築した。

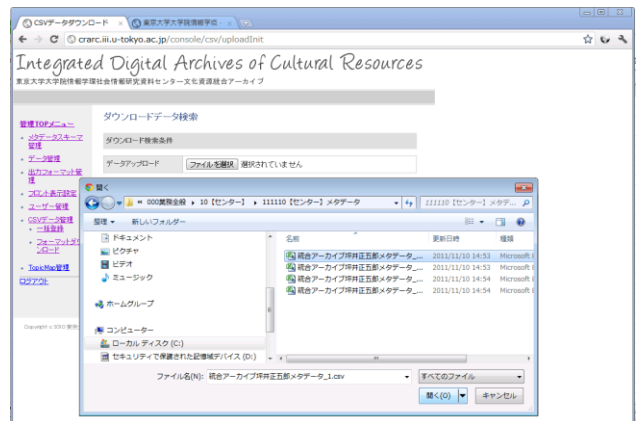


図 5：CSV ファイルのインポート機能

4. デジタルデータの仕様に関する情報

本システムにおいては、格納したデジタルデータに関する仕様情報を格納できる機能を設けている。著者は資料のデジタル化に関する、仕様情報の入力について、拙論にまとめている[3]。本システムはこれらのデジタルデータの資料情報を格納する機能実装した。

これらの情報の取得は主に「(A) 資料内容調査とデジタル化の計画」、「(B) デジタル化」、「(C) 評価」、「(D) データ管理」のフェーズに分けて、情報を取得する。この中で、最初の「(A) 資料内容調査とデジタル化の計画」は文化資源そのものについて調査を行い、それに基づいて資料の電子化の計画をたてるフェーズである。本フェーズでは主に資料の内容に関する情報が調査の上で格納される。また、その上で、デジタル化を行う範囲、予算、時期、担当者などを決め、ドキュメントに記録する必要がある。

次の「(B) デジタル化」ではオリジナルの文化資源からデジタルデータを作成するフローに応じて、デジタル化に関する仕様情報が収集され記録される。

デジタル化に関する情報については、デジタル化の責任者、時期などの入力情報共通のプロファイルのほか、入力機器別のプロファイルと入力環境についてのプロファイルおよびカラーマネジメントプロファイルを記録する。その中で、**表 1** で示された内容はデジタル化全般に関する総合的な内容である。

表 1：データに関する総合的な情報項目

ID	項目名	内容
B-01-01	識別子	
B-01-01-01	レコード ID	ひとつのデジタルデータレコードを一意に同定するための識別子
B-01-01-02	旧ファイル名	入力機器によって自動的に与えられたファイル名
B-01-01-03	ファイル名	任意に名付けたファイル名
B-01-02	ソフト情報	データ作成用のソフト名及び RAW データとアプ

		リケーションの対応
B-01-02-01	ソフト情報 - 製品名	
B-01-02-02	ソフト情報 - パージョン	
B-01-02-03	ソフト情報 - タイプ	
B-01-02-04	ソフト情報 - その他	
B-01-03	フォーマット	データのフォーマットタイプ
B-01-04	ソースタイプ	データのソースが、一次的文化資源か、文化資源のアナログ複製(画像、映像、音声等)かを区別する
B-01-05	ソース情報	
B-01-05-01	ソース情報 - 対象の部分情報	ページ、フォーカス部分など、もとのソース全体のうちデータに記録されている部分を特定するための記述
B-01-05-02	ソース情報 - 対象の範囲	(組物などコンポーネントのうちの範囲、コレクションのうちの範囲)
B-01-05-03	ソース情報	計測対象、計測位置(画像情報などがあればその座標との対応など)
B-01-06	作成(撮影)者名	デジタルデータの作成者の名前
B-01-07	作成(撮影)日	デジタルデータを作成(撮影)した日付

また、表 2 の内容は、デジタルカメラ及びスキャナなどの機材や、種別などの入力機器全般に関する総合的な内容に関する仕様情報である。

表 2 : 入力に関する総合的な記録

ID	項目名	内容
B-02-01	入力機器	入力機器について
B-02-01-01	入力機器 - タイプ	入力機器のタイプ
B-02-01-02	入力機器 - メーカー	入力機器のメーカー
B-02-01-03	入力機器 - 型番	入力機器の型番
B-02-02	入力設定	機器の設定に関する情報
B-02-02-01	入力設定 - シャーブネス	
B-02-02-02	入力設定 - ノイズリダクション	

さらに、表 3 においては、デジタルカメラを使用した場合の、機種名やレンズの種類などの、デジタルカメラに関する入力情報全般のメタデータが構成されている。

表 3 : デジタルカメラの入力に関するメタデータ

ID	項目名	内容
B-03-01	デジタルカメラ-入力付属機器	
B-03-01-01	デジタルカメラ-入力付属機器-レンズ型番	撮影に用いたレンズの型番
B-03-01-02	デジタルカメラ-入力付属機器-フィルター	撮影に用いたフィルターの種類
B-03-02	デジタルカメラ-入力設定	
B-03-02-01	デジタルカメラ-入力設定-ホワイトバランス	デジタルカメラの場合、撮影時に設定したホワイトの色温度(自動的に画像のヘッダー情報に記録される場合もある)
B-03-02-02	デジタルカメラ-入力設定-感度	撮影時に用いたフィルムの ISO 感度、デジカメであれば設定した ISO 感度設定
B-03-02-03	デジタルカメラ-入力設定-焦点位置	焦点位置(解像度を判定するために用いる)(デジタルカメラであれば自動的に画像のヘッダー情報に記録される場合もある)
B-03-03	デジタルカメラ-画像 bit 数	各チャンネル毎に画像のビット数を記述する(繰り返し可)

そして、表 4 においては、スキャナを使用した場合の、機種名やレンズの種類などの、スキャナに関する入力情報全般のメタデータが構成されている。

表 4 : スキャナの入力に関するメタデータ (一部略)

ID	項目名	内容
B-04-01	スキャナー-入力設定	
B-04-02	-解像度	
B-04-02-01	-解像度-倍率	倍率
B-04-02-02	-解像度-線数	線数
B-04-03	-出力設定	出力の種類(ネガ/ポジ)
B-04-04	-その他	

最後に、表 5 においては、三次元のデジタルデータを使用した場合の、仕様情報に関するメタデータが構成されている。主な構成要素としては、解像度や、倍率、出力設定などが該当する。

表 5 : 三次元データに関するメタデータ

ID	項目名	内容
B-06-01	三次元情報-入力環境	入力環境について解説する。
B-06-02	入力方法	入力方法について解説する。
B-06-03	頂点数	
B-06-04	ポリゴン数	
B-06-05	付帯データ	
B-06-05-01	付帯データ-テキストチャデータ	当該の三次元データがテキストチャデータを持っているか
B-06-05-02	付帯データ-質感データ	当該の三次元データが質感データを持っているか
B-06-06	その他	

さらに「(C) 評価」においては作成したデータの評価に関する情報項目である。これまで、デジタル化されたデータは、オリジナルの現物と比較して色彩や形状を正確に反映しているか、あるいはそれらと比較することのできるデータが記録されているかを評価する指標が確立されていない。そのため、本項目では、主に色彩などを中心に、カラーマネージメントの評価や、指定された環境におけるモニターと現物の相互評価などのデジタルデータの評価を行う。また本項目では視覚データのみを扱ひ、聴覚データには対象外とした。評価は、作業に使用する機器の評価を最初に行い、さらにそれらの機器を用いた視覚を中心とした評価を行う。そして、これらの評価結果を、特定のフォーマットにあわせて記録される。

そして、「(D) データ管理」では、一次的文化資源データの管理およびそのデータから作成する二次的文化資源データの管理についての情報が記録され、「(E) メタデータの構成」のフェーズでは、上記までの各フェーズで取得した各プロファイル情報を整理して記述し、デジタルデータについての情報を記録したデータである「メタデータ」として格納される。

この「(E) メタデータの構成」のフェーズにおいては、図 6 に示されるように、CMS 機能を使用して、

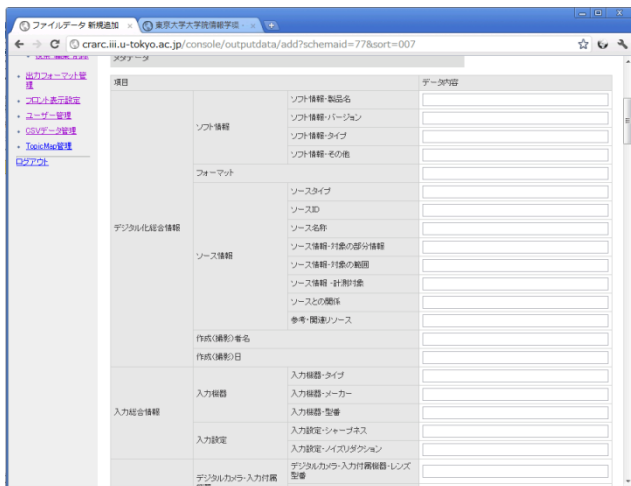


図 6: デジタル仕様情報の入力フォーム

デジタルデータに関する情報を入力できるような機能を設けた。これらは、静止画像だけではなく映像などの情報にもメタデータを付与することが可能である。

そして、これらの情報は図 7 で示されるように、資料画像の右上をクリックすることによってメタデータを表示をすることができる。



図 7: 資料画像の表示とデジタルデータの仕様に関する情報

さらに、本 CMS では、格納したデータを、特定のスキーマを設定した上で、出力する機能を設けている。図 8 に示されるように、設計したスキーマを XML 方式で出力する方式を設定し、それらを設計することによって、入力したデータを指定した XML 形式で出力ができる機能を設けた。これらにより、CMS で GUI 上で構成したメタデータスキーマを XML 形式で出力することが可能となった。

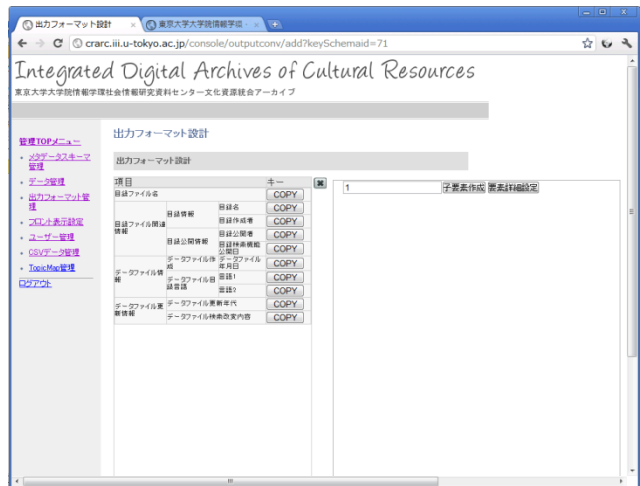


図 8: メタデータの XML 出力設定画面

5. 実証と課題

本機能は実装後、カスタマイズ機能などを使用して、社会情報研究資料センターに収蔵されている、坪井正五郎関係資料、大戦間期プロパガンダポスターコレクション資料と、新聞学の祖である小野秀雄資料に関するメタデータを実験的に入力した。データ数は両コレクションとも約 100 件程度を登録し、各データはフォンド、シリーズ、アイテムの各項目が設計され、それらに応じたメタデータを設計した。両者においては、アイテムレベルのメタデータにおいては数項目において異なったエレメントを設定した。

そして、三つの資料群について、構造情報を反映させながらも、一部は異なったメタデータを設計できる CMS の特徴を利用して設計し、メタデータを入力した。坪井正五郎、小野秀雄、大戦間期プロパガンダポスターコレクションなどは、それぞれ構造やメタデータのエレメントセットが異なるが、カスタマイズ機能などを設けて、それぞれの資料に応じて設計して、データを入力した。

また、多くの資料が格納される小野秀雄資料などは、CSV ファイルを使用して、多くのメタデータの入力を行い、その後、1 点ずつ資料を登録する場合は、フォームを用いて、資料の登録を行った。2 つの入力機能があるため、資料の量や内容によって、メタデータの入力方法を選択することができる。

また、資料の表示においては、図 8 に示されるように、構造化された資料情報はディレクトリの形式で表示され、ディレクトリの構造から資料を選択して表示できるようにしている。そして、デジタルデータに関しても、全ての情報は格納されていないが、デジタル化に関する主要な情報については、メタデータに格納した。

一方で、今後の課題としては、フォームの設計画面などが、やや複雑になり、それを扱うまでに、多少時間を要する課題などをあげることができる。特

にアーカイブズに関する情報項目は、極めて多く、それらを網羅的に扱うと、極めて多くの入力項目を用意する必要がある。そのため、そのエレメント数が多くなるが、それらをより扱いやすく設計するインターフェイスを設けることが必要である。

また、デジタルデータに関しては、項目数が多く、入力に時間を要する場合が多い、そのため、より効率的にこれらの情報を入力できるようなシステムの改善が必要である。

いずれにしても、このようなシステムを導入することによって、専門的なデータベースエンジニアでなくとも、多様な資料の構造を捉えながら入力していくことの可能なコンテンツマネジメントシステムが実現すると考えられる



図 8: 資料群の表示

参考文献

[1]ISAD(G):[http://www.icacds.org.uk/eng/ISAD\(G\).pdf](http://www.icacds.org.uk/eng/ISAD(G).pdf)(Accessed:2011-11-01)

・ISAD (G)の基本概要については以下の論文にまとめられている。

田窪直規,国際標準記録史料記述一般原則 : ISAD (G) : その基本構造・考え方と問題点,レコード・マネジメント : 記録管理学会誌(44),pp1-22,2002

・アーカイブズにおける編成記述とメタデータの関係については以下の論文にまとめられている。

坂口貴弘,アーカイブズの編成・記述とメタデータ,情報の科学と技術 60-9,pp384-389,2010

・また、ISAD (G)の概念を取り入れたデジタルアーカイブに関する事例は以下の論文などにまとめられている。

小川 千代子,ISAD(G)の実装 : アジア歴史資料センターの階層検索システム(イノベーションとしての記録管理) ,レコード・マネジメント : 記録管理学会誌(45), pp10-25, 2002

牟田昌平,本格的デジタルアーカイブを目指して : アジア歴史資料センターの実験,情報知識学会研究報告

会講演論文集(10),pp 65-70, 2002

[2]Patrick Le Buf,Functional Requirements for Bibliographic Records (FRBR) ,Routledge,2005

[3] 研谷紀夫 北岡タマ子 高橋英一,文化資源の電子化における記録情報管理を重視したガイドラインの提案とそれに基づく Digital Cultural Heritage の構築,人文科学とコンピュータシンポジウム予稿集 pp317-324,2009

[4] EAD: <http://www.loc.gov/ead/>(Accessed:2011-11-01)

EAD の紹介として以下のような論文がある。

安澤 秀一,エンコードド アーカイヴァル デスクリプション EAD : SGML-XML の応用形として,人文科学とコンピュータ研究会報告 2001(67),pp17-24,2001

[5]五島敏芳,日本における記録史料記述 EAD/XML 化の実践—記録史料管理と EAD (符号化記録史料記述) 利用をめぐる—,情報処理学会シンポジウム論文集,pp217-224,2002

日本の記録史料記述 EAD/XML 化と記録史料管理 記録史料管理過程における EAD 利用の位置をめぐる,情報知識学会誌,12 巻 4 号,pp3-21,2003

[6]村越一哲,表計算ソフトを利用した史料目録 EAD 化のためのツール,『国文学研究資料館,アーカイブズ情報の共有化に向けて』,岩田書院,2010

[7] 後藤真,京都府行政資料のデジタル・アーカイブ化とその課題-EAD/XML の適用の可能性と歴史学,都市文化研究 (8) , pp2-15, 2006

[8]国立公文書館 EAD 定義 第 1.07 版 :

http://www.digital.archives.go.jp/howto/pdf/naj_ead107.pdf (Accessed:2011-11-01)

[9]Spreadsheets to EAD:

<https://github.com/jronallo/stead> (Accessed:2011-11-01)

[10] EADitor: <http://code.google.com/p/eaditor/> (Accessed:2011-11-01)

[11] California Digital Library Online Forms: <http://www.lib.hitu.ac.jp/about/da/index.html> (Accessed:2011-11-01)

[12]<http://www.archivists.org/saagroups/ead/tools.html> (Accessed:2011-11-01)

[13] 東京大学社会情報研究資料センター Digital Cultural Heritage

<http://crarc.iii.u-tokyo.ac.jp/web/>

(現在は研究メンバー向けのみの公開であるが、2012 年 3 月に一般に公開する予定である)