

特集 「デジタルアーカイブ」

デジタルコンテンツのアーカイブとメタデータ

Archive and Metadata for Digital Contents

杉本 重雄
Shigeo Sugimoto

筑波大学図書館情報学系
Institute of Library and Information Science, University of Tsukuba.
sugimoto@slis.tsukuba.ac.jp

Maria Luisa Calanag

図書館情報大学情報メディア研究科博士後期課程
Graduate School of Information and Media Studies, University of Library and Information Science.
calanag@ulis.ac.jp

Keywords: digital libraries, preservation, digital contents, archive, metadata.

1. はじめに

はじめに、アーカイブ (archive) ということばの意味を考えたい。アーカイブは、文書館、文書の保管所であり、また保管所に蓄積された文書のことを意味する。図書館は、文書館とは異なる役割をもっているが、資料を長期にわたって保存するという視点からは同様な役割をもっている。コンピュータ用語では、保管のためにファイルをひとまとめにすること、あるいはひとまとめにしたファイルを保管しているものを意味して用いられている。いずれの場合も、資料を長期に保管することを目的としている。

1990年代のデジタル情報技術を利用した出版と流通環境の発展によって、デジタルコンテンツの種類と量が飛躍的に増えた。それに伴い、デジタルライブラリー (Digital Library) の開発が学術図書館を中心に盛んに進められてきた。その結果、現在の大学図書館ではデジタルライブラリー機能は基幹機能の一つとなっている。図書館や文書館などにとって、電子出版物の保存は従来の資料の保存と同様に基本的な役割である。その中で、デジタルコンテンツの保存は、デジタルライブラリーにおける重要な問題として認められている^{[1][2]}。例えば、アメリカ議会図書館はデジタルコンテンツ、特にはじめから電子的につくられた (born-digital) コンテンツの保存に関する研究開発のために大きな予算を受け、保存のための基盤づくりのプロジェクト National Information Infrastructure and Preservation Program (NDIIPP) を進めている^[3]。デジタルコンテンツは CD-ROM などのパッケージ型の媒体に格納して出版されるもの (パッケージ系コンテンツ) と、ネットワークを介して提供されるもの (ネットワーク系コンテンツ) の2種類に大別される。本稿では、主としてネ

ットワーク系コンテンツの保存について述べる。

デジタルコンテンツのアーカイブを実現するには、保存すべき資料を収集し、コンテンツの利用性を保ちながら蓄積・組織化し、利用者に提供することが求められる。インターネットで提供されるコンテンツの保存の場合は、玉石混交の状態で大規模に提供される多様なコンテンツの中で、保存価値があるコンテンツを低コストでできる限り多く保存することが求められる。収集に関しては Google などの検索エンジンが行う網羅的な自動収集がよく知られている。しかしながら、デジタルアーカイブのためのコンテンツ収集にはアーカイブごとの役割と方針に関わるいろいろな問題がある。また、変化の激しい情報技術環境の中でコンテンツそのものを使える状態に保つための技術が求められる。コンテンツの保存のためには、コンテンツそのものだけではなく、コンテンツの内容やコンテンツの構成方式、保存の履歴などコンテンツに関する情報、すなわちメタデータが必要とされる。このように、デジタルコンテンツのアーカイブに関する中心的話題には、デジタルコンテンツの収集、コンテンツそのものの保存の技術、コンテンツの保存のためのメタデータの3点がある。現時点において、デジタルコンテンツのアーカイブの問題が解決されているとはいいがたい。そのため、本稿ではこの3点に関する課題を中心に述べる。

以下、2章では、ネットワークコンテンツの収集について述べる。3章では、デジタルコンテンツの保存のためのシステムのモデルや仕組みなどについて述べる。4章ではメタデータに関して解説する。利用者に提供するために、コンテンツを収集、固定、蓄積するには知的財産権に関する問題をクリアすることが前提となるが、本稿では触れない。

2. デジタルコンテンツの収集

図書館や文書館では、館ごとに決められた資料の収集方針に基づき、収集対象範囲を決め、資料を収集する。そして、資料の保存方針に従って、収集資料の保存期間や保存形態を決める。したがって、コンテンツの収集はアーカイブの最初の段階である。パッケージ系コンテンツの場合は、コンテンツを納めた「もの」を収集すればよいが、ネットワーク系コンテンツの場合には、自動収集の技術、収集範囲と対象の指定、収集対象の内容評価といった問題がある。

(1) 収集方針と自動収集の技術

収集方針を大別すると、対象範囲のコンテンツを網羅的に収集しようというものと対象範囲の中から保存すべきものを選択的に収集しようというものがある。網羅的収集の場合には、ソフトウェアによる自動収集を行うのが一般的である。選択的収集の場合には、人的に収集対象コンテンツを選定し収集することが行われる。このほか、出版者から保存すべきコンテンツを収集者に提供する場合や出版者と収集者の合意に基づき限定的に自動収集する場合もある。選択的に収集する場合は、コストが高くつき、またカバーできる範囲が限られるといった問題がある。一方、自動収集する場合、収集ソフトウェアのアクセス可能範囲の問題に加えて、ハイパーテキスト構造や動的構造に起因する問題がある。例えば、ハイパーテキスト構造をもつコンテンツの場合、1件のコンテンツが複数のファイルによって構成されることがあるため、コンテンツのまとまりを同定し、かつ構成要素間のつながりに矛盾を生じることなく蓄積することが求められる。また、リンクで結ばれた先の別のコンテンツとの関係に矛盾を来さないようにする必要がある。動的コンテンツの場合、同一のURIでアクセスしても、アクセス方法やタイミングに応じて提供形式が変化するものがあるので、提供元での蓄積形式と収集した形式の違いが生じるという問題がある。また、ホームページのように内容が書き換えられていくコンテンツの場合、版の概念が明確にはないため、保存のために固定される内容が収集のタイミングのみに依存してしまうといった問題がある。

(2) 収集範囲と対象の指定

自動収集の場合、ドメイン名やIPアドレスといったアドレス情報、テキストやイメージといったコンテンツタイプの識別情報を用いて収集対象範囲を限定することが行われている。しかしながら、保存を目的とした収集の場合、コンテンツの内容によって収集範囲と対象を指定することが求められる。例えば、納本図書館としての国立図書館の場合、図書などの従来型の出版物と同様に考えると、その国で出版された、あるいはその国に関連する出版物が収集対象範囲と考えられるが、アドレス情報だけでは対象範囲を指示するのには不十分である。ま

た、データベースや Plug-in ソフトウェアを必要とする動的コンテンツの場合、コンテンツの利用に必要なソフトウェアを何らかの形で提供することが求められるが、識別情報だけからでは必要な情報は得られず、内容を分析する必要が生じる。

(3) 収集対象の内容評価

自動収集の場合、コンテンツの内容が収集方針に合致したものであるかの判断を自動化することが求められる。また、収集したコンテンツを公開するには、非合法コンテンツやプライバシーに関わるコンテンツなど公開に適しないものを公開対象から除外する必要が生じる。また、コンテンツの内容に関する信頼性のように判定の自動化の困難な問題もある。

オーストラリア国立図書館が進めるネットワークコンテンツの保存プロジェクトである PANDORA の場合は選択的収集を行っている¹⁴⁾。合衆国議会図書館と協調して進められている Internet Archive¹⁵⁾やスウェーデン国立図書館¹⁶⁾では、インターネット上で提供されるコンテンツを網羅的に自動収集している。Internet Archive が提供する Wayback Machine は自動収集ソフトウェアによって収集したページを収集時期ごとにまとめて蓄積しているもので、リンク先も Wayback Machine が蓄積している同時期のページに結ばれている。

3. デジタルコンテンツの保存

3.1 デジタルコンテンツの保存に関する基本概念

コンテンツの保存は、基本的には、「もとのままの状態でも利用できること」を意味する。しかしながら、図書館では印刷物をマイクロフィルム化して保存するなど、もとのままの状態でも保存することにこだわらず、保存方針に合わせてコンテンツの内容を利用可能な状態で保存することも行われてきている。デジタルコンテンツに関しても同様のことがいえる。例えば、特定のブラウザに合わせてデザインされたウェブページの場合、保存対象となっているページのソーステキストだけを保存するか、あるいはページの Look and Feel までも保存するためにそれを閲覧するためのブラウザおよびブラウザを動作させるためのハードウェアとソフトウェア（これをコンテンツの利用環境もしくは単に環境と呼ぶ）も合わせて保存するかといった点がある。こうした点は、保存方針によって決められるべきことである。また、「長期保存」が何年くらいを意味するのかについては明確な定義はない。コンテンツの種類や保存の目的によっても異なる。後述する OAIS 参照モデルの文書の中でも保存期間に関しては「それなりの長さ」と定義されている。

保存の単位は、物理的単位と論理的単位の両方で捉える必要がある。物理的単位としてはファイル単位あるいはデータベース単位がある。論理的単位には、1点の電子文書あるいはデータ、資料の集まりとして構成される

ひとまとまりのコレクションがある。前者の例にはワープロで作成された電子文書、Web 文書、物理資料を電子化することでつくられた資料などがある。後者の例には、いくつもの論文や記事を含む 1 タイトルの雑誌、いくつもの Web ページを集めてつくるコレクションなどがある。コレクションの場合は、コレクションに含まれる資料をも含めて保存すると考えるのが自然である。

3.2 デジタルコンテンツ保存のシステムモデル
— OAIS 参照モデル

デジタルコンテンツの保存のモデルとして Consultative Committee for Space Data Systems (CCSDS) の開発した OAIS 参照モデル (Reference Model for an Open Archival Information System)^[7] の国際標準化が進められている。オーストラリア国立図書館の PANDORA プロジェクトをはじめとして、このモデルに基づくデジタルコンテンツの保存のための枠組みとメタデータの開発が進められている。

OAIS 参照モデルでは、図 1 に示すように、情報オブジェクト (Information Object) がデータオブジェクト (Data Object) とその表現情報 (Representation Information) からなるとしている。データオブジェクトは、具体物 (物理オブジェクト, Physical Object) あるいはビットの集まりによって生成されるデジタルオブジェクト (Digital Object) である。表現情報も何らかのデジタル表現された実体である。OAIS 参照モデルでは、情報オブジェクトを保存するために情報パッケージの概念を導入している。情報パッケージは図 2 に示すように内容情報 (Content Information)、保存記述情報 (Preservation Description Information, PDI)、およびパッケージ化情報 (Packaging Information) からできている。また、情報パッケージに関する記述 (Descriptive Information about Package) も必要とされる。この 4 要素のうち内容情報以外はメタデータといえる。PDI は保存対象コンテンツに関する情報をもつのにに対し、他はパッケージに関する情報をもつ。PDI の内容は次の 4 種類の情報に規定されている。

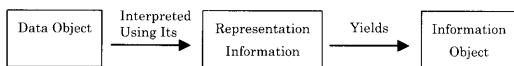


図 1 Data Object, Representation Information, Information Object 間の関係

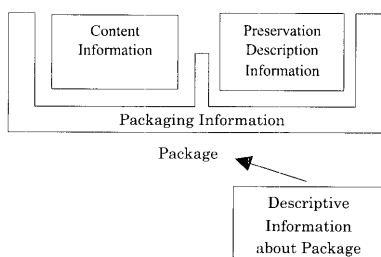


図 2 Information Package の概念構造

- 来歴 (Provenance) : コンテンツの出所やコンテンツの保存に関わる処理の履歴。
- コンテキスト (Context) : 情報パッケージ外のオブジェクトとの関係。例えば、コンテンツがつけられた理由、他のコンテンツとの関係。
- 参照 (Reference) : コンテンツを一意に識別するための識別子、もしくは識別のためのシステム。例えば、ISBN やコンテンツを一意に識別するための属性の集合。
- 不変性 (Fixity) : コンテンツを保護するための情報。例えば、チェックサム。

図 3 は、OAIS 参照モデルの機能実体の関係を表す。SIP, DIP, AIP はそれぞれ寄託情報パッケージ (Submission Information Package), 配布情報パッケージ (Dissemination Information Package) およびアーカイブ情報パッケージ (Archival Information Package) を意味し、コンテンツの蓄積のためのアーカイブへの寄託、利用者からの要求に応じたコンテンツの配布、アーカイブ内での保存のための情報パッケージを意味する。データ管理機能 (Data Management) は、システムに寄託された情報パッケージの登録、利用などの管理を行う。保存用の情報パッケージはアーカイブ記憶 (Archival Storage) に蓄積保存される。利用者からコンテンツのアクセス要求があった場合には、管理機能を用いて検索、利用条件判定などを行い、配布パッケージとして利用者に提供する。管理機能 (Administration) と保存プランニング機能 (Preservation Planning) は、それぞれ保存システム全体の統括管理、状況のモニタリングと蓄積されたコンテンツ保存のためのレポートを行う。

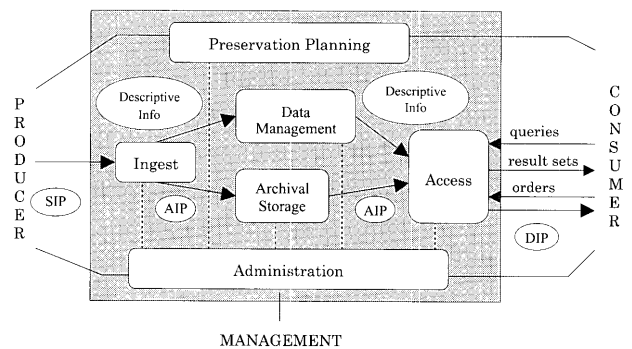


図 3 OAIS 参照モデルの機能要素

3.3 デジタルコンテンツ保存のための方法と取組みの例

デジタルコンテンツの保存が従来の保存と異なる点は、保存対象が記憶装置の中に納められた「内容」であること、コンテンツの利用環境を必要とすることである。「内容」の保存の観点からは、「内容」を表すビット列 (もしくはその集まり) のコピーをとっていけばよい。標準規格が確立し、かつ広く流通しているもの場合は、これでかなり対応できると考えられる。例えば、プレイ

ンテキストの場合、テキストを構成する際に用いられた文字セットがわかっておれば、外字などの例外を除くと安定して保存できると考えられる。イメージデータの場合、データを作成する際の精度や、information loss型と loss-less 型のいずれの形式を使うべきかといった判断は別に必要とされるが、TIFF や JPEG など、標準が確立し、かつ広く用いられるイメージ形式を用いれば安定した保存が可能であると考えられる。

デジタルコンテンツの保存の一般的方法として、移植 (migration) による方法とエミュレーション (emulation) による方法が提案されている。移植は、新しい環境で利用できるようにコンテンツごとに移植する方法である。エミュレーションは、コンテンツを利用するためのコンピュータ環境を実現するエミュレータをつくり、エミュレータを新しい環境に移植していくことで、以前の環境と同様にコンテンツを利用できるようにしようというものである。エミュレーションによる保存の取組みに関して、NEDLIB プロジェクトにおける Rothenberg の報告^[8] や IBM Almaden 研究所の Lorie による報告^[9] がある。ミシガンおよびリーズ大学による CAMiLEON プロジェクトからは移植とエミュレーションに関する報告が出されている^[10]。移植の場合にはコンテンツごとに新しい環境に合わせていく必要があるのに対し、エミュレーションの場合はコンテンツの調整を必要としないので、エミュレーションのほうが少ないコストで実現できるという議論がなされているが、現時点では完全な解決策は得られているとはいえない。

前述の NDIIPP や PANDORA プロジェクトのほかにも、ヨーロッパの国立図書館の共同による NEDLIB プロジェクト (オランダ国立図書館を中心とする)^[11] など国立図書館によるデジタルコンテンツ保存プロジェクトが進められている。国立国会図書館の場合、インターネット上のコンテンツを収集し、保存・提供するサービス WARP の試験的運用を行っている^[12]。イギリスの CURL (Consortium of University Research Libraries) が eLib プログラムの助成のもとで進めたデジタルアーカイブに関する研究プロジェクト Cedars から、メタデータ、知的財産権、保存戦略、コレクション管理など、デジタルコンテンツの保存の重要課題に関するレポートが出されている^[13]。NSF ほかによる Digital Library Initiative^[14] では、CAMiLEON プロジェクトの他、コーネル大学の PRISM プロジェクト^[15] といったデジタルコンテンツの保存を指向したプロジェクトを進めている。

学術雑誌の電子ジャーナル化が進む一方、雑誌の保存が出版社だけに任されるのは不安であるという議論がある。スタンフォード大学図書館が中心になって進める LOCKSS (Lots of Copies Keeps Stuff Safe) プロジェクトでは、多数のクライアントを協調的に利用し、クライアント側に残された雑誌記事を残すことで保存を図る

うという取組みを進めている^[16]。

4. デジタルコンテンツの保存のためのメタデータ

4.1 メタデータについて

メタデータは「データに関するデータ」と定義され、記述対象に関する記述はすべてメタデータであるといっ構わない。例えば、Dublin Core^[17] はいろいろな情報資源 (Resource)、すなわちコンテンツの発見を目的とするもので、タイトルや作成者などコンテンツがもつ種々の属性を表すためのものである。コンテンツのデジタルアーカイブを実現するには、発見のために有用な属性以外に、コンテンツの構造や管理情報、保存の履歴、利用のための条件といったさまざまな情報が必要であり、これらはすべてメタデータである。メタデータを記述するためには、記述対象のどのような属性をどのように記述するかを決める規則が必要である。メタデータ記述のための属性の名前と意味、属性値の表現の形式などを決める規則をメタデータスキーマ (metadata schema) と呼ぶ。

保存のためのメタデータスキーマとして、OAIS 参照モデルに基づくメタデータスキーマの検討が NEDLIB 他で進められた。また、アメリカの Digital Library Federation (DLF) 主導で開発されたデジタル資料のためのメタデータスキーマ METS の開発が進められている。これらについては 4.2 節、4.3 節で紹介する。メタデータの記述対象をいかにとらえるかもメタデータスキーマにとっては重要である。デジタルライブラリーでは、個々のコンテンツの管理だけではなく、コンテンツを集まり (コレクション) として管理することも求められる。そのため、4.4 節ではコレクションの視点からの保存のためのメタデータに関して述べる。従来の書誌データ記述の場合は、冊子体の図書や雑誌、論文などを対象としてきたので、記述対象が固定され、「もの」としての同定も行いやすかった。これは、配架や貸出しといったサービスの単位であるということもできる。ところが、デジタルコンテンツの場合、ファイルなどの物理的な単位にあまり意味がないため記述対象の単位の把握が必ずしも容易ではない。また、デジタルコンテンツの加工のしやすさと多様性の故に記述対象がどのような実体であるかも明確ではない。そのため、4.5 節では IFLA の Functional Requirements for Bibliographic Records を紹介し、デジタルコンテンツの書誌データの視点からメタデータに関する基本的な問題を紹介する。

4.2 OAIS 参照モデルに基づくメタデータスキーマ

デジタルコンテンツの保存のためのメタデータスキーマに関する検討が進められ、Cedars や NEDLIB、OCLC (Online Computer Library Center) と RLG

表 1 表現情報のための推奨メタデータエレメント
(太字はエレメントのカテゴリー)

エレメント名	エレメントの意味
Content Data Object Description	
Underlying abstract form description	ファイルシステムやデータベースの概念要素など内容データオブジェクト (CDO) の基礎的情報に関する人間向け記述
Structural type	静止画や音声、テキストといったCDOが表すデジタルオブジェクトのクラス
File description	ファイル形式や解像度など、CDOを構成するファイルの技術的仕様
Installation requirements	オブジェクトをインストールするために必要な手続き情報
Size	オブジェクトのサイズ
Access inhibitors	暗号化や透かしなどアクセス制限のためのCDOの特性記述
Significant properties	将来の保存プロセスにおいて保存しない維持が必須であるCDOの利用のために必要な特性
Functionality	現時点での表示形式に基づいて、CDOの機能あるいは "Look and Feel" の記述
Description of rendered content	CDOの内容がどのように表示されるべきか、解釈されるべきかの記述
Quirks	保存プロセスや手続きに起因して起こったCDOの機能や "Look and Feel" の損失の記述
Documentation*	CDOの表示もしくは解釈に必要ないし有用な文書。(以下の各カテゴリーにも同様にDocumentationエレメントがある。)
Software Environment · Rendering Programs	
Transformation process	ファイルの解凍やコンパイルなどCDOのインストールのための変換の記述
Display/Access Application	CDOを表示しないしアクセスすることのできるプログラムの記述
Software Environment · Operating System	
OS name	CDOを利用するためのプログラムを実行するプラットフォームの名前ないしは呼称
OS version	OS nameで指示されたオペレーティングシステム(OS)の版
Location	OS nameおよびOS versionで指示されたOSの実稼働システム (working copy) を入手可能な場所
Hardware Environment · Computational Resources	
Microprocessor requirements	CDOのソフトウェア環境の動作に必要なマイクロプロセッサ仕様の記述
Memory requirements	CDOのソフトウェア環境の動作に必要な記憶装置資源の記述
Hardware Environment · Storage	
Storage information	CDOのソフトウェア環境あるいは利用プログラムの動作に必要な永続的記憶装置資源の記述
Hardware Environment · Peripherals	
Peripheral requirements	CDOの利用ないし表示のために必要な周辺装置の記述
Hardware Environment · as a whole	
Location	CDOの利用に必要な物理的装置を利用しないし入手できる場所

(Research Library Group) の共同ワーキンググループでは OAIS 参照モデルに基づくメタデータスキーマスキーマの報告を出している^{[18]~[20]}。ここでは、OCLC と RLG の共同ワーキンググループのレポートを参考にメタデータについて紹介する。

図 2 に示したように、情報パッケージは内容情報と PDI からなっている。内容情報は、ビットストリームないしビットストリームの集合であるコンテンツデータオブジェクト (Content Data Object) と、このビットストリーム (あるいはその集合) をアクセス可能でかつ意味のある知識の表現に変換するための表現情報 (Representation Information) からなる。表現情報は、コンテンツデータオブジェクト記述 (Content Data Object Description) と環境記述 (Environment Description) からなる。さらに、環境記述をソフトウェア環境とハードウェア環境に分け、ソフトウェア環境は再生用プログラム (Rendering Programs) とオペレーティングシステム、ハードウェア環境は計算リソース (Computational Resources)、記憶装置 (Storage) および周辺装置 (Peripherals) に分けて、メタデータの記述要素 (メタデータエレメント) を定義している。PDI

表 2 保存記述情報 (PDI) のメタデータエレメント

エレメント名	エレメントの意味
Reference Information	
Archival system identification	アーカイブシステムの中で内容データオブジェクト (CDO) およびそれに関連付けられたメタデータ (保存情報パッケージ, AIP) を一意に識別するための識別子
Global identification	アーカイブの外部のシステムからCDOおよびそのメタデータ (AIP) を一意に識別するための識別子
Resource description	リソース発見のための情報であり、(利用可能な場合には) もととなるメタデータから抽出、あるいはリソースへのアクセス機能のためにアーカイブ自身によって作られる。
Context Information	
Reason for creation	CDOが作られた理由に関する情報の記録
Relationships	当該CDOと他のCDOとの間の主要な関係の記録
Manifestation	当該CDOと同一のオブジェクトであって、かつ異なる実体との関係。例えば、CDOを移植した場合、以前のオブジェクトとの関係
Intellectual content	当該CDOの知的内容と他のCDOの知的内容との間の関係の記録
Provenance Information	
Origin	当該CDOが作成されたプロセスの記述
Pre-ingest	メンテナンス、内容の変更、保管など、当該CDOの作成からアーカイブへの寄託までの履歴に関する記述
Ingest	当該CDOがアーカイブに収録されたプロセスの記述
Archival retention	アーカイブ内での保存中における当該CDOのメンテナンス、内容の変更、管理などに関する記述
Rights management	当該CDOに関わる法的事項の記述
Event*	内容や形式、所蔵など当該CDOに関わる何らかのイベントの記述。 (*Provenance Information のエレメントに共通)
Fixity Information	
Object authentication	当該CDOとその内容の認証のためにアーカイブの最小要求要件に合致するための十分な情報。例えば、デジタル署名、透かし、チェックサムなど

については 3.2 節に示した四つのカテゴリーに関してそれぞれメタデータエレメントを決めている。表 1 と表 2 に表現情報および PDI の推奨メタデータエレメントを示す。なお、推奨エレメントを実際に利用する際の規則については応用ごとに決まることになる。なお、この表では一部を除き各カテゴリーのトップレベルのエレメントのみを示し、サブエレメントは省略した。

4.3 Metadata Exchange and Transmission Standard (METS)

METS^[21] は Encoded Archival Description (EAD)^[22] などでの経験をもとに、Digital Library Federation を中心につくられたデジタルコンテンツのためのメタデータ記述の標準である。METS は OAIS 参照モデルにおける情報パッケージ (SIP, AIP あるいは DIP) として用いることができる。

METS の記述には下に示す要素が含まれており、また XML によるメタデータ記述の形式を決めている。

- **Descriptive Metadata** : コンテンツの検索などのために利用する記述的メタデータ。外部の記述的メタデータ (例えば、MARC レコードなど) を参照する場合もある。
- **Administrative Metadata** : ファイルの生成と管理の記録、知的財産権情報、デジタルオブジェクトの作成の対象となった、もとの非デジタルオブジェクト、デジタルオブジェクトの来歴情報など管理のための記述。
- **File Groups** : デジタルオブジェクトを構成する一つもしくは複数のファイルをグループ化したも

の、同一の内容を異なる方式によって作成したファイルのリスト。

- **Structural Map** : デジタルオブジェクトの構造を表す。例えば、セクションに分かれた録音データの場合、全体とセクションとの関係を記述する。
- **Behavior** : オブジェクトがもつ動的性質の記述。

4.4 コレクションレベルのメタデータ

デジタルライブラリーでは資料の集合としてのコレクションを管理の対象とすることがあるため、コレクションの記述のためのメタデータが提案されている^[24]。コレクションも保存対象を定義するうえで重要であるため、オブジェクト単位の保存のためのメタデータ記述だけではなく、コレクションを対象とした記述も必要とされる。例えば、カリフォルニア大学バークレー校の場合、印刷物などの従来のコレクションに関して、

- **Comprehensive** : できる限り網羅的
- **Research** : 研究および大学院博士で必要とするレベル
- **Study** : 学部および大学院での教育で必要とするレベル
- **Basic** : 主題領域を知るうえで基本的な資料のレベル
- **Minimal** : ごく限られた資料だけを収集しているレベル

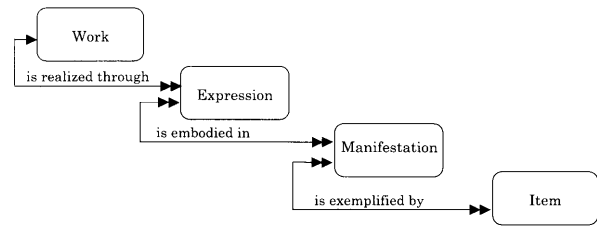
というコレクションのレベルを定義している。デジタルライブラリーの場合、これとは別に、

- **Archived** : 長期に保存している
- **Served** : 図書館のサーバに蓄積してサービスをしている
- **Mirrored** : ミラーサイトを運営して、サービスをしている
- **Linked** : 外部のサーバへリンクしている

といったサービスの違いを明示している^[24]。著者らは、従来のコレクションレベルの記述とデジタルコレクションのレベルの記述を組み合わせ、コレクションレベルでのメタデータの記述の深さを決めるためのモデルを提案した^[25]。

4.5 書誌記述のための基本モデル— Functional Requirements for Bibliographic Records

図4はIFLAの報告書^[26]の中で示されているリソースに関する4階層のモデルである。Workは表現方法を問わない知的内容、Expressionは何らかの表現方法によってWorkを表したもの、ManifestationはExpressionを何らかの媒体上に実現したもの、ItemはManifestationの個々のものを意味する。例えば、「蜘蛛の糸」という小説を仮定すると、作者の考えたストーリーとしての小説はWorkであり、それを具体的に表現したもの（小説の場合はテキスト中心に表現されたもの）



単純矢印は対応する実体が一つ、二重矢印は対応する実体が複数あることを意味する。両端が二重矢印の場合多対多、一方が二重矢印の場合は1対多の関係を表す。

図4 IFLA FRBRにおける知的内容を表す実体 (Group 1 Entity) 間の関係

がExpressionである。同じWorkであっても、例えば大人向けと子供向けといったように異なるExpressionをもつ場合がある。例えば、オンライン書店で「蜘蛛の糸」で検索し、「蜘蛛の糸、ポプラ社文庫—日本の名作文庫 (ISBN 4-591-00906-8)」という本を見つけ、それを購入したとすると、検索結果として示されたものは一つのManifestationであり、購入した1冊が一つのItemである。Manifestationはそれからつくり出されたItemの総称ととらえても構わない。

ネットワークコンテンツの場合、コンテンツとしてのビット列がExpressionからItemのいずれになるのか明確ではない。例えば、コンテンツを利用環境ごとにインストールして利用する場合、Item、すなわちコンテンツの実体（ファイル、あるいはビット列）は置かれている場所に依存する個別の実体であると考えられるので、一つのManifestationからいくつもの（異なる）Itemがつけられるととらえることができる。一般のHTML文書や単一のイメージデータの場合、ManifestationとItemの間の区別はつけづらい。また、スタイルシートによって表現形式を変化させることのできるXML文書の場合、XML文書をExpression、スタイルシートを適用してつくられる表示用の文書、例えば、HTML文書をManifestationととらえることができる。時とともに内容が変わる組織や個人のホームページの場合、WorkとExpressionが変化し、それに伴ってManifestationも変化するがManifestationに与えられる識別子(URL)が変化しないという場合もある。

5. おわりに

資料を保存することは図書館の中心的な機能の一つであり、従来から酸性紙問題対策やマイクロフィルム化など保存のための努力がなされてきた。コンテンツとその利用環境のデジタル化が進んだことによって新たな課題が生じてきたといえる。保存技術、効率的な収集と選択の技術とともに現時点では十分とはいえず、知的財産権管理などの問題と併せてこれからの取組みが求められる問題である。

デジタルライブラリーの分野ではデジタルライブ

ラー間の相互運用性 (Interoperability) が重要な課題と認められている。その中で、保存は時を超えた相互運用性 (Temporal Interoperability) の問題であるとも言われている。コンテンツをもとのままの形で利用可能な状態に保つこと、いわば厳密な相互運用性を将来のシステムに求めることは現実的であるとは思えない。また、大量のコンテンツを扱うためのスケーラビリティも重要である。そのため、コンテンツそのものにとらわれず、コンテンツの知的内容を保存するには十分な方法、いわば緩やかな相互運用性を実現する方法を含めて保存の課題に取り組まねばならないと考えられる。

◇ 参 考 文 献 ◇

- [1] Hedstrom, M.: Digital Preservation: Problems and Prospects, *Digital Libraries*, No. 20, pp. 3-15 (2001): http://www.dl.ulis.ac.jp/DLjournal/No_20/1-hedstrom/1-hedstrom.html
- [2] Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, National Research Council: LC21: A Digital Strategy for the Library of Congress, 288p., National Academic Press (2001): <http://books.nap.edu/html/lc21/>
- [3] NDIIPP: National Digital Information Infrastructure and Preservation Program: <http://www.digitalpreservation.gov/ndiipp/>
- [4] National Library of Australia: Pandora Archive - Preserving and Accessing Networked Documentary Resources of Australia: <http://pandora.nla.gov.au/index.html>
- [5] Internet Archive: <http://www.archive.org/>
- [6] National Library of Sweden: Kulturarw3, Long time preservation of electronic documents: <http://www.kb.se/kw3/ENG/>
- [7] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1 Blue Book (2002.1): <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>
- [8] Rothenberg, J.: An Experiment in Using Emulation to preserve Digital Publication, NEDLIB Report series, No. 1 (2000): <http://www.kb.nl/coop/nedlib/results/NEDLIBemulation.pdf>
- [9] Lorie, R. A.: Long Term Preservation of Digital Information, Proceedings of JCDL'02, pp. 346-352, ACM Press (2002)
- [10] CAMiLEON project: CAMiLEON Project: Creative Archiving at Michigan and Leeds Emulating the Old on the New: <http://www.si.umich.edu/CAMILEON/>
- [11] NEDLIB homepage: <http://www.kb.nl/coop/nedlib/>
- [12] 国立国会図書館: インターネット資源選択的蓄積実験事業: <http://warp.ndl.go.jp/>
- [13] Consortium of University Research Libraries: CURL Exemplars in Digital Archives: <http://www.leeds.ac.uk/cedars/index.html>
- [14] Digital Library Initiative Phase 2: <http://www.dli2.nsf.gov/>
- [15] Prism project: Prism Home Page — Information Integrity in Distributed Digital Libraries: <http://www.prism.cornell.edu/>
- [16] LOCKSS Web Site: <http://lockss.stanford.edu/index.html>
- [17] Dublin Core Metadata Initiative: <http://dublincore.org/>
- [18] Cedars Project: Cedars Guide to Preservation Metadata (2002): <http://www.leeds.ac.uk/cedars/guideto/metadata/>
- [19] Lupovici, C. and Masanes, J.: Metadata for Long Term Preservation, NEDLIB Report series, No. 2 (2000): <http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf>
- [20] The OCLC/RLG Working Group on Preservation Metadata: Preservation Metadata and the OAIS Information Model — A Metadata Framework to Support the Preservation of Digital Objects (2002): http://www.oclc.org/research/pmwg/pm_framework.pdf
- [21] METS: Metadata Encoding & Transmission Standard: <http://www.loc.gov/standards/mets/>
- [22] Encoded Archival Description (EAD): <http://lcweb.loc.gov/ead/>
- [23] Powell, A., Heaney, M. and Dempsey, L.: RSLP Collection Description, *D-Lib Magazine*, Vol. 6, No. 9 (2000.9): <http://www.dlib.org/dlib/september00/powell/09powell.html>
- [24] Berkeley Digital Library SunSITE: Collection and Preservation Policy: <http://sunsite.berkeley.edu/Admin/collection.html>
- [25] Calanag, M. L., Tabata, K. and Sugimoto, S.: Linking Collection Management Policy to Metadata for Preservation, *Proc. DC-2002*, pp. 35-43 (2002): <http://www.bncf.net/dc2002/program/ft/paper4.pdf>
- [26] International Federation of Library Associations and Organizations (IFLA): Functional Requirements for Bibliographic Records (1998): <http://www.ifla.org/VII/s13/frbr/frbr.pdf>

2003年4月3日 受理

著 者 紹 介



杉本 重雄 (正会員)

京都大学工学部情報工学科、同大学院工学研究科情報工学専攻修了。京都大学工学博士。京都大学工学部助手を経て、1983年図書館情報学部に転任。2002年10月筑波大学との大学統合により、筑波大学図書館情報学系教授。知的コミュニティ基盤研究センター勤務。デジタルライブラリー全般に関心をもち、研究の中心はメタデータとそれに関するソフトウェア技術を中心に研究科都度を進めている。情報処理学会、ACM、IEEE-CSなどの各会員。Dublin Core Metadata Initiativeでは評議委員会、諮問委員会のメンバー。



Maria Luisa Calanag

フィリピン東マニラ大学卒業。図書館情報学大学院図書館情報学専攻修士課程修了。修士(図書館情報学)。現在、図書館情報学大学院情報メディア研究科博士後期課程に在学中。フィリピンでは、フィリピン工科大学大学院図書館、European Commissionほかで図書館員、ドキュメンタリストなどの経験をもつ。デジタルコンテンツの保存とコレクション形成に関心をもつ。IFLA会員。