

\*本稿は情報科学技術協会発行の『情報の科学と技術』vol.54,no.9(2004) pp. 461-465 に掲載された論文の原稿で、著作権は情報科学技術協会が有しています。発行元の了承を得てセルフアーカイブしているものです。

## OAIS 参照モデルと保存メタデータ

栗山 正光(くりやま まさみつ)

常磐大学人間科学部

〒310-8585 茨城県水戸市見和 1-430-1

**抄録：** OAIS 参照モデルはデジタル情報の長期保存システム構築に関する有力な指針であり国際標準規格ともなっている。本稿では保存のためのメタデータに焦点を絞り、OAIS 参照モデルに示された情報パッケージの概念と、それに基づいて行われているメタデータの枠組み規定の実際について論じる。OAIS 参照モデルはデジタル情報の保存活動を行っている諸機関で広く認知されているものの、それぞれが規定する実際の保存メタデータの枠組みは OAIS の情報パッケージの構成とはかなり異なった形でなされているのが実状である。個々のニーズと相互運用性・再利用性とのバランスが今後の課題となる。

**キーワード：** OAIS 参照モデル、保存メタデータ、デジタル保存、情報パッケージ、メタデータの枠組み

## The OAIS Reference Model and Preservation Metadata

Kuriyama, Masamitsu

Tokiwa University. College of Human Science

1-430-1 Miwa, Mito, Ibaraki, 310-8585 Japan

**Abstract:** The Reference Model for an Open Archival Information System(OAIS) is an important guide to constructing a system to preserve digital information over the long term. It is also an ISO standard. This paper discusses concepts concerning the Information Package in the OAIS model and actual activities to create preservation metadata frameworks based on it. The OAIS reference model is widely recognized among organizations which are active or interested in digital preservation, but actual metadata schemas or frameworks they are developing often have a different structure from the Information Package in the OAIS model. To balance needs of each organization and interoperability/reusability is an emerging challenge.

**Keywords:** OAIS reference model, preservation metadata, digital preservation, information package, metadata framework

### 1. はじめに

OAIS 参照モデル(Reference Model for an Open Archival Information System(OAIS))は、デジタル情報の長期保存システム構築に関する有力な指針である。NASA など世界各国の宇宙開発機関で組織する宇宙データシステム諮問委員会 CCSDS(Consultative Committee for Space Data Systems)が策定したもののだが、宇宙観測のデータに限らず広く情報一般を視野に入れ、OAIS の定義から始まって、情報を保存するアーカイブの責任、機能の詳細、保存のための戦略、さらにはアーカイブ間の連携まで扱った総合的な内容となっている。2002 年、ISO 14721:2002 として国際標準に認定された。

本稿ではデジタル情報の保存のためのメタデータ（保存メタデータ）に焦点を絞って、OAIS参照モデルによって示された諸概念と、実際に諸機関で行われているメタデータ規定の関わりについて論じる。先に筆者は同様のテーマで動向レビュー<sup>1)</sup>を発表したことがあるが、本稿では、その後の展開や別の事例もふまえて異なった角度から考察を行う。

OAIS参照モデルの規格書はインターネットで公開されている。<sup>2)</sup> 本稿のOAISの諸概念の説明はこの規格書に基づいているが、あくまで筆者の解釈によるものであり、理解不足による誤りが含まれている可能性があることをご承知置きいただきたい。なお、OAIS参照モデル全般に関しては、OCLCのLavoieによるレポートが制定の背景なども含めてわかりやすく説明している。<sup>3)</sup> 日本語では、不十分ながら筆者が概略を紹介したものがあるので参考にさせていただきたい。<sup>4)</sup>

## 2. 情報パッケージとメタデータ

OAIS 参照モデルでは保存すべき情報と関連するメタデータを合わせて情報パッケージ(Information Package)というまとまりで扱う。情報パッケージの構成を簡略な形で示すと図1のようになる。このうち内容データオブジェクト(Content Data Object)が保存対象となるオリジナルのデータで、後はすべてそれに関するデータ、すなわちメタデータである。注目すべきは、保存対象データを探索・発見するための情報、いわば目録情報である記述情報は情報パッケージには含まれないということである。

情報パッケージは処理の段階に応じて三つの形態があるとされる。すなわち、情報生産者がアーカイブに提出する SIP(Submission Information Package)、アーカイブ中で保管する AIP(Archival Information Package)、利用者に配布される DIP(Dissemination Information Package)であり、それぞれに格納されるメタデータにも、当然、ずれが生じることになる。しかしながら、話が複雑になるのを避けるため、本稿ではこの三つの区別は考慮しないこととする。

こうした情報パッケージの構成に忠実に従って、具体的なメタデータ要素の枠組み規定を行ったのがOCLC/RLGの保存メタデータに関するワーキンググループ（以下WGと記す）である。このWGが2002年6月に発表した報告書は、先行する四つのプロジェクトのメタデータを比較検討して取り込み、さらに独自の要素も加えて作り上げた、保存メタデータの集大成といった内容となっている。<sup>5)</sup> 以下、このWGの枠組み規定を中心に情報パッケージ中のそれぞれの情報について検討していくこととする。

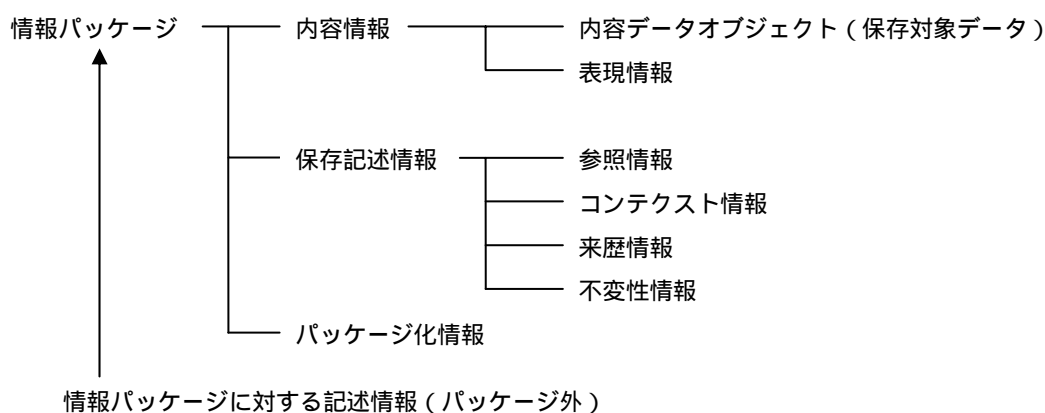


図1 OAIS参照モデルにおける情報パッケージの構成

### 3．表現情報

表現情報(Representation Information)は、内容データオブジェクトを解釈して人間が理解できる形に提示するための情報である。具体的にはデータオブジェクトのフォーマットや使用文字コードなどに関する情報がこれに当たる。ただし単にフォーマットや文字コードの名前を記録しただけでは不十分で、そのフォーマットの仕様書やコード表といった情報も必要となる。さらに、仕様書が外国語で書かれているとすると、今度はその言語の辞書や文法書が表現情報となる。つまり、ある表現情報を理解するためにはそのための別の表現情報が必要であり、芋づる式に表現情報が膨れ上がる。これを表現ネットワークと称している。どの程度の表現情報が必要かは、アーカイブがサービス対象とするコミュニティに、どの程度の知識ベース(Knowledge Base)があることを前提とするかによって変わってくる。もちろんこれは理論上の話であって、実際に個々のアーカイブが外国語の辞書や文法書まで表現情報として付加することは考えられない。

WG では表現情報を内容データオブジェクト記述と環境記述に分けている。前者は保存対象データそのものに関する情報であり、ファイルの構造や大きさなどに関する要素が含まれ、後者はさらにソフトウェア環境とハードウェア環境に分けられる。ソフトウェア環境ではオペレーティング・システムと実行プログラムに関する情報、ハードウェア環境では CPU やメモリ、周辺機器などの必要条件が要素となっている。ひらたく言えば、市販ソフトの箱に示されている必要システム構成の記述と同じようなものであるが、はるかに詳細な記述ができるようになっている。

時代遅れのフォーマットで保存されたデジタル情報を後世に伝えていくための方法として、一般にマイグレーションとエミュレーションの二つが考えられている。前者はフォーマット変換などにより新しい環境に移行していく方式で、後者はデータそのものは変換せず、新しい環境の中に古い環境を再現して利用しようというものである。

WG では内容データオブジェクト記述の中に重要属性(Significant properties)という要素を設け、マイグレーションへの対応を図っている。たとえばテキストの内容だけが重要で、ページ・レイアウトやハイパーリンクなどは二次的ということであれば、マイグレーションの過程でそうした重要でない部分は切り捨てるということも有り得るわけである。関連して保存の過程で失われた機能を記録するための Quirks という要素も用意している。一方、エミュレーションを行うためには詳細な技術情報が必要で、ここで規定されたメタデータ要素だけでは不十分であることをこの報告書自体が認めている。

マイグレーションにせよエミュレーションにせよ、技術情報を含めた表現情報が重要なのは論を待たないが、関連する表現情報を個々のアーカイブが全部保持するのは非現実的であり、どこか一箇所保持して共同利用しようという考え方が当然出てくる。たとえば、デジタル・ロゼッタ・ストーンと名付けられた保存システムのモデルでは、「メタ知識アーカイブ」と称する、メディアへの記録技術とファイル・フォーマットに関する知識を蓄積したデータベースを構築することを提案している。<sup>6)</sup> こうしたメタ知識の集中管理システムができれば、情報パッケージ中の表現情報は外部のメタ知識データに対するポインタを記録しておくだけでいいことになる。

### 4．保存記述情報

保存記述情報(Preservation Description Information)とは保存に関して最も重要な情報のような印象を受ける名前だが、これらの情報がなくても、内容データオブジェクトと表現情報さえ残っていればとりあえず元のデータの利用は可能であり、単なる保存という観点からは、実は二義的な情報である。保存対象データの十分な理解と円滑な管理に資する情報と考えられる。参照情報(Reference Information)、コンテキスト情報(Context Information)、来歴情報(Provenance Information)、不変性情報(Fixity Information)の四種類

に分けられている。

#### 4.1 参照情報

参照情報は内容情報を同定識別するためのもので、ID 番号のほか、必要であれば対象を記述する情報を含めてもいいとのことである。電子図書館コレクションの参照情報の例として「書誌記述」があげられている。ここですぐに疑問となるのが、情報パッケージの外側に置かれている記述メタデータとの違いである。

WGでは、参照情報をアーカイブ内部でのID、グローバルID(アーカイブの外で規定されたID。例えばISBNなど)、資源記述(Resource Description)の三種類に分け、資源記述の目的は「OAISおよび連合した外部システムの内部における資源発見をサポートするのに十分なAIPの記述でIDを補完する」<sup>7)</sup>としている。どの程度の記述にするかはアーカイブの目的などによっても異なるのだろうが、やはり記述メタデータとの違いは曖昧である。別の箇所では、「記述情報は資源発見のためのメタデータで、保存メタデータの範囲外である」<sup>8)</sup>としており、資源記述を参照情報の中に規定するのは矛盾しているようにも思われる。

実際には、資源発見のための記述メタデータと参照情報としてのメタデータとを、別物としてそれぞれ作成・維持するようなシステム設計をすることは考えにくい。現に WG でも独自の記述要素を規定することはせず、MARC やダブリン・コアなど既存のメタデータ要素セットによる書誌レコードをここに埋め込む仕様になっている。概念上の区別と理解しておくべきだろう。

#### 4.2 コンテキスト情報と来歴情報

コンテキスト情報は内容情報と周囲との関係を記録するもので、なぜ内容情報が作成され、どのように他の内容情報オブジェクトと関係するかといったことを含む。一方、来歴情報は内容情報の歴史を記録するもので、内容情報の起源や出所、またその後加えられた変更の記録などである。特別なタイプのコンテキスト情報とみなすことができる。

WGでは、コンテキスト情報を作成理由と関係の二つに分けている。作成理由は、執筆の動機のような知的内容に関わるものものではなく、内容情報がたとえばある実験データのマスター情報源であるといった情報を記す。関係は、他のフォーマットや異版、同一著作を形成する他の内容データオブジェクトとの関係などを示す。来歴情報は起源、受入前、受入、保管、権利管理の五つに分けられ、それぞれに関して起こった出来事をイベントとして記録する。

コンテキストと来歴の違いは必ずしも明確ではない。一応、来歴情報は時間軸に沿った変化を扱うと考えられるが、OAIS参照モデルではコンテキストの例としてキャリアレーション履歴、資金調達履歴などをあげているし、WGが参考に行っている先行プロジェクトの一つCEDARSでは作成理由を来歴情報に入れている。<sup>9)</sup> もっとも、こうした境界線上の要素がどちらのカテゴリーに含まれるかを議論することは、実用上はあまり意味がない。

#### 4.3 不変性情報

不変性情報は、オリジナルのデータに改変が加えられていないか、あるいはデータの消失がないかといったことを確認するための情報で、具体的にはデジタル署名やチェックサムといったものである。WGでは、認証のタイプ、手続き、日付、結果といった要素を規定している。

これは保存記述情報の中でも最もわかりやすい概念である。

### 5. パッケージ化情報

パッケージ化情報(Packaging Information)は、情報パッケージの構成要素を結びつける情報で、たとえば CD-ROM に内容情報と PDI が納められていれば、その CD-ROM のファイル構造などの情報がパッケージ化情報となる。WG では、パッケージ情報は単にデジタルオブジェクトとメタデータを結びつけるだけで、オブジェクト自体の保存には直接関係がないとして検討から除外している。

ここで問題となりそうなのは、パッケージ化情報が内容データオブジェクトを構成するファイル間の関係まで示すのかどうかということである。内容データオブジェクトのファイル構成は表現情報に含めてパッケージ化情報では立ち入らない、と理論上は解すべきかもしれないが、現実に複数の内容データオブジェクト・ファイルとメタデータ・ファイルがあった場合、パッケージ化情報と表現情報としてのファイル構成記述を分離して考えるのは難しく、混乱が生じる恐れがある。

## 6. メタデータ規定の実際

WGの枠組み規定をもとに自館のメタデータ要素を規定しようとしている事例としてコーネル大学図書館があげられる。ここでは研究部門にDPO(Digital Preservation Officer)というデジタル保存専門職員を置き、2003年には保存メタデータのワーキンググループを作って、どの要素を必須項目にするかなど具体的な検討を行っている。<sup>10),11)</sup>

OCLCとRLGでは、WGが定めたメタデータの枠組みを出発点として、実用化へのガイドラインを策定する新たなワーキンググループPREMIS(PREservation Metadata: Implementation Strategies)を立ち上げた。<sup>12),13)</sup>これは二つのサブグループから成る。保存メタデータのコア要素を定めるグループと、実践に向けての戦略を検討・評価するグループである。

このうち実践戦略サブグループでは、2003年11月から電子情報の長期保存に取り組んでいる機関を対象にアンケート調査を行っており、2004年4月現在の回答状況に基づいた報告書の草案を公表している。<sup>14)</sup>それによると、OAIS参照モデルに少なくとも部分的に従っていると回答したのは43機関中24機関で、そのうち4機関は全面的に従っているということである。一方、3つの機関がOAIS参照モデルについて十分に検討していないため答えられないとし、OAISを知らないと答えたところも1つある。欧米においても、OAIS参照モデルに基づいて保存システムを構築しているところばかりではないことがうかがえる。

実際、WGが多くの要素を取り入れている先行プロジェクトの一つ、オーストラリア国立図書館(NLA)のメタデータ要素セットは、解説の中でOAIS参照モデルに言及してはいるものの、表現情報とか保存記述情報とかの区別をすることなく、25の要素を規定している。<sup>15)</sup>WGはここから採用した各要素を表現情報あるいは保存記述情報の中のいずれかに振り分けているのだが、NLA自体はそうした体系を取っていない。

ニュージーランド国立図書館でも2002年にメタデータ・スキーマを発表し、2003年には改訂版を出している。<sup>16)</sup>「OAIS参照モデルで示された諸原則と実用性とのバランスを取るべくデザインされた」<sup>17)</sup>とのことだが、OAISの情報パッケージとは異なり、オブジェクト、プロセス、ファイル、メタデータ修正という四つのエンティティを設け、その下に保存用メタデータの要素を規定している。オブジェクトの下の要素は、ID、ハードウェア環境、ソフトウェア環境などで、OAISでいう参照情報と表現情報に属するものと考えられる。プロセスの下の要素はプロセスのタイプ、目的、行為者、日時など、コンテキスト/来歴情報に属するものである。ファイルの下の要素はオブジェクトを構成する各ファイルの名前、タイプ、フォーマットなど技術情報で、これも表現情報である。メタデータ修正の下には、修正者、修正日付、修正フィールド、修正データなど、メタデータの変更を記録するための要素が規定されているが、これはいわばメタメタデータでありOAIS参照モデルでは該当するものがない。

電子図書館連盟(Digital Library Federation)はXMLによるメタデータ記録方式の標準METS (Metadata

Encoding & Transmission Standard)を策定しており、METS形式の文書はSIP、AIPあるいはDIPの役割を果たすことができるとしている。<sup>18)</sup> METS文書は、METSヘッダー、記述メタデータ、管理メタデータ、ファイル・セクション、構造マップ、構造リンク、動作といった七つの主要セクションから成り、一見したところ、OAIS参照モデルの情報パッケージの構成とのつながりが見出しにくい。このうち管理メタデータは、技術メタデータ、知的財産権メタデータ、ソース・メタデータ、デジタル来歴メタデータにわけられており、OAISの表現情報やコンテキスト/来歴情報に相当するメタデータを記録できる。記述メタデータと管理メタデータは、METS独自の要素を規定しておらず、他に存在するメタデータへのポインタを記録するか、他のメタデータの規定にのっとって記述したものを埋め込むといった仕様になっている。すなわち、記述メタデータ・セクションにはMARCやダブリン・コアなど、管理メタデータではたとえばNISOのZ39.87<sup>19)</sup> などによる記述をすることを想定している。ただし、知的財産権に関しては、METS独自の権利関係メタデータ・スキーマ案を発表している。<sup>20)</sup>

このように、OAIS 参照モデルはデジタル情報の保存活動を行っている機関におおむね認知されているものの、それぞれが規定する実際のメタデータの枠組みはOAISの情報パッケージ構成とはかなり異なった形でなされているのが実状である。

## 7. まとめ

OAIS 参照モデルの情報パッケージの概念と、それに基づく保存メタデータの規定に関して、事例を交えて検討を加えた。また、現実のメタデータ規定は、OAIS 参照モデルに言及しているものでさえ、必ずしも忠実にOAISの情報パッケージの構成を反映したものではないことを示した。OAIS参照モデルはあくまで実用的なシステムを構築する上での思考の手がかりであり、各機関がそれぞれのニーズに合ったメタデータを規定すべきであることは当然と言えば当然である。

しかしながら一方で、現在、保存メタデータだけに限っても、数多くの組織で独自の規定がなされており、しかもそれぞれが改訂を重ねていくつも版が生まれることが予想され、メタデータ規定の過剰が問題になりつつある。こうした事態に対して、メタデータに関する情報を集めた一種の登録簿(メタデータ・レジストリ)を作成しようという提案もなされている。<sup>21)</sup>

保存メタデータに関しても、他の種類のメタデータ同様、個々のニーズと相互運用性・再利用性のバランスをどう取るかが今後の課題と言えよう。

## 参考文献

- 1) 栗山正光. デジタル情報保存のためのメタデータに関する動向. カレントアウェアネス. No.275, p. 13-16 (2003) <http://www.ndl.go.jp/jp/library/current/no275/doc0007.htm>
- 2) Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), Blue Book, January 2002, 147p. <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>
- 3) Lavoie, Brian F. The Open Archival information System Reference Model: Introductory Guide. Dublin, OCLC, 2004, 19p. [http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)
- 4) 栗山正光. 長期保存型電子図書館とOAIS参照モデル. 筑波大学・図書館情報大学統合記念公開シンポジウム「電子図書館の軌跡と未来:ますます広がる電子図書館サービス」報文集. つくば, 筑波大学附属図書館, 2003, p. 21-31. <http://www.tulips.tsukuba.ac.jp/pub/dlsympo/hobunshu/kuriyama.pdf>
- 5) OCLC/RLG Working Group on Preservation Metadata. Preservation Metadata and the OAIS

- Information Model, A Metadata Framework to Support the Preservation of Digital Objects. Dublin, OCLC, 2002. 51p. [http://www.oclc.org/research/projects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/projects/pmwg/pm_framework.pdf)
- 6) Heminger Alan R. and Robertson, Steven B. Digital Rosetta Stone: A Conceptual Model for Maintaining Long-term Access to Digital Documents. Sixth DELOS Workshop on Preservation of Digital Information, Tomar, Portugal 17-19 June 1998. <http://www.ercim.org/publication/ws-proceedings/DELOS6/rosetta.pdf>
- 7) OCLC/RLG Working Group on Preservation Metadata. Op. Cit., p. 31.
- 8) Ibid., p. 9.
- 9) The Cedars Project Team and UKOLN. Metadata For Digital Preservation: The Cedars Outline Specification, 2000, 33p. <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>
- 10) Digital Preservation Officer. <http://www.library.cornell.edu/iris/dpo/index.html>
- 11) Working Meeting on Preservation Metadata. <http://www.library.cornell.edu/iris/dpo/metadata.html>
- 12) Lavoie, Brian F. Implementing Metadata in Digital Preservation Systems: The PREMIS Activity. D-Lib Magazine, Vol. 10, No. 4 (2004) <http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>
- 13) PREMIS (PREservation Metadata: Implementation Strategies) <http://www.oclc.org/research/projects/pmwg/default.htm>
- 14) PREMIS Implementation Survey: Preliminary Summary of Results <http://www.oclc.org/research/projects/pmwg/surveysummary.pdf>
- 15) National Library of Australia. Preservation Metadata for Digital Collections. 15 October 1999. <http://www.nla.gov.au/preserve/pmeta.html>
- 16) National Library of New Zealand. Metadata Standards Framework – Metadata Implementation Schema. 2003, 33p. [http://www.natlib.govt.nz/files/nlnz\\_data\\_model.pdf](http://www.natlib.govt.nz/files/nlnz_data_model.pdf)
- 17) Searle, Sam and Thompson, Dave. Preservation Metadata : Pragmatic First Steps at the National Library of New Zealand. D-Lib Magazine Vol. 9, No. 4 (2003) <http://www.dlib.org/dlib/april03/thompson/04thompson.html>
- 18) METS: An Overview & Tutorial <http://www.loc.gov/standards/mets/METSOverview.v2.html>
- 19) National Information Standards Organization and AIIM International. Data Dictionary—Technical Metadata for Digital Still Images. Released as a Draft Standard for Trial Use, June 1, 2002 – December 31, 2003. [http://www.niso.org/standards/resources/Z39\\_87\\_trial\\_use.pdf](http://www.niso.org/standards/resources/Z39_87_trial_use.pdf)
- 20) Draft Rights Declaration Schema is Ready for Review. 2003. <http://www.loc.gov/standards/mets/news080503.html>
- 21) Day, Michael. Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: a Proposal. DC2003, 28 September-2 October 2003, Seattle, 2003. [http://www.siderean.com/dc2003/101\\_paper38.pdf](http://www.siderean.com/dc2003/101_paper38.pdf)