

# 英国のリポジトリをリンクする

機関リポジトリ等のデジタルリポジトリを対象とする  
リポジトリ横断型のユーザー指向サービスを支援する技術・組織モデル

## 予備的評価研究報告書: 付録

Alma Swan (キー・パースペクティブ有限会社)  
Chris Awre (ハル大学)

プロジェクト参加機関:  
キー・パースペクティブ有限会社  
ハル大学  
ノッティンガム大学 SHERPA  
サウサンプトン大学電子工学・計算機科学科



## 付録: 技術アーキテクチャおよびインフラストラクチャ

### 目次

- [A1 リポジトリ環境](#)
  - [A1.1 環境の不均一性](#)
  - [A1.2 オープンアクセス](#)
- [A2 リポジトリの種類とその概要](#)
  - [A2.1 機関リポジトリ](#)
    - [A2.1.1 JISC FAIR プログラム](#)
    - [A2.1.2 英国における進展](#)
    - [A2.1.3 世界における進展](#)
  - [A2.2 データセット](#)
    - [A2.2.1 eBank UK](#)
    - [A2.2.2 JISC Digital Repository Programmeのプロジェクト](#)
    - [A2.2.3 グリッドとリポジトリ](#)
  - [A2.3 マルチメディア資料](#)
    - [A2.3.1 Community Led Image Collections \(CLIC\) 調査](#)
    - [A2.3.2 ビデオと音源](#)
  - [A2.4 学習オブジェクト](#)
    - [A2.4.1 複合型オブジェクト](#)
- [A3 ユーザー指向サービス](#)
  - [A3.1 既存の活動の概要](#)
  - [A3.2 ユーザー指向サービスの設計](#)
    - [A3.2.1 配信オプション](#)
    - [A3.2.2 ワークフロー](#)
    - [A3.2.3 リポジトリの隠蔽と顕在化](#)
    - [A3.2.4 集約か、分散か](#)
    - [A3.2.5 認証要件](#)
  - [A3.3 ユーザー指向サービスの機能](#)
    - [A3.3.1 全般](#)
    - [A3.3.2 デポジットと受入](#)
    - [A3.3.3 ハーベストとアグリゲーション](#)

- [A3.3.4 検索とブラウズ](#)
    - [A3.3.5 表示、出力、変換](#)
    - [A3.3.6 プッシュと通知](#)
    - [A3.3.7 リクエスト](#)
    - [A3.3.8 入手とリンク付け](#)
    - [A3.3.9 コメント、アノテーション、タグ付けと編集、更新](#)
    - [A3.3.10 テキストマイニング](#)
    - [A3.3.11 保存](#)
    - [A3.3.12 その他](#)
  - [A3.4 ユーザー指向サービス提供のレベル](#)
- [A4 メタデータ](#)
  - [A4.1 メタデータ標準](#)
  - [A4.2 メタデータのリッチさ](#)
  - [A4.3 メタデータの粒度](#)
  - [A4.4 アプリケーションプロファイル](#)
  - [A4.5 メタデータのマッピング](#)
  - [A4.6 メタデータのシンタックス](#)
    - [A4.6.1 Topic Maps](#)
  - [A4.7 メタデータの自動生成](#)
  - [A4.8 メタデータの品質](#)
  - [A4.9 主題分類](#)
    - [A4.9.1 OAIセット](#)
  - [A4.10 メタデータの変更と追加](#)
  - [A4.11 識別子](#)
    - [A4.11.1 識別子の粒度](#)
    - [A4.11.2 著者識別子](#)
    - [A4.11.3 識別子の解決](#)
    - [A4.11.4 識別子の取り決め](#)
  - [A4.12 複合型オブジェクト](#)
    - [A4.12.1 パッケージング](#)
  - [A4.13 相互運用性](#)
  - [A4.14 メタデータとフルコンテンツを同時に扱う](#)
- [A5 インターフェース](#)
  - [A5.1 Webインターフェース](#)
  - [A5.2 マシンインターフェース](#)
  - [A5.3 インターフェースの選択肢](#)
    - [A5.3.1 OAI-PMH](#)
    - [A5.3.2 RSS](#)
    - [A5.3.3 ATOM](#)
    - [A5.3.4 SRW/U](#)
    - [A5.3.5 Z39.50](#)
    - [A5.3.6 OpenURL](#)
    - [A5.3.7 SSI](#)
  - [A5.4 相互運用性に対する軽量なアプローチ](#)
  - [A5.5 リポジトリ仕様](#)
  - [A5.6 Webサービスインターフェース](#)
  - [A5.7 Web 2.0](#)
- [A6 アーキテクチャ](#)
  - [A6.1 アーキテクチャの選択肢](#)
  - [A6.2 中間的共通基盤サービス](#)
  - [A6.3 認証と承認](#)
  - [A6.4 サービス指向アーキテクチャ](#)
  - [A6.5 リポジトリサービス層](#)
  - [A6.6 ローカルインフラストラクチャへのリポジトリの統合](#)
- [参考文献](#)

## A1 リポジトリ環境

近年、様々な部門で独立にリポジトリの開発が行われている。たとえば、機関リポジトリと学習オブジェクトリポジトリがしばしば並行して開発されており、これらの部門における活動から多くのことを学ぶことができる。最初に開発されたオープンアクセスリポジトリは、特定の主題分野のコミュニティのニーズに焦点を絞ったものであった (arXiv<sup>1</sup>など)。そして、これらのニーズがリポジトリの保管物の種類を大部分決めていた。その他多くのリポジトリが、大学をはじめとする組織の要求を満たすために開発されている (SHERPA e-プリントリポジトリ<sup>2</sup>の多く)。多くのリポジトリは単一のアイテム種別あるいはフォーマットを対象としている。たとえば、White Roseリポジトリ<sup>3</sup>はe-プリントを、BioMed Image Archive<sup>4</sup>は画像を対象としている。しかし、機関が必要とするすべてのものをリポジトリに保管することを決めた機関もあった (ケンブリッジ大学<sup>5</sup>やエジンバラ大学<sup>6</sup>)。単一のファイルからなる単純なアイテム (PDF文書やJPEG画像など) を保管するリポジトリもあれば、関連する複数のアイテムで構成される複合型オブジェクト (文書と画像、データベースからなる学位論文など) を保管するリポジトリもある。各組織の具体的なニーズがこれらの決定の基礎となり、エンドユーザサービスを提供する豊かで変化に富むリポジトリ環境をもたらしている。

## A1.1 環境の不均一性

この豊かで変化に富む開発は、同時に、様々な目的で構築され、様々なコンテンツを持つ数多くのリポジトリが存在する極めて不均一な環境を作り出した。このため、不均一性は常に存在するものであり、今後も続くことは間違いないと思われる。取材した人たちは皆、これがエンドユーザサービスを管理する上で主要な問題であることに同意した。この不均一性により、この環境に存在するリポジトリの使用が断片化され、横断検索の完全性に大きな影響を及ぼす可能性がある。

この不均一性を克服し、利用可能な様々なリポジトリをより統一的な形でエンドユーザに提供する様々な方法が提案されている。

- ポータルを使うことにより、提供 (presentation) レベルで様々なリポジトリを統一することができる。
- 融合 (fusion) レベルで統一することができる。たとえば、Googleは、極めて不均一な環境からメタデータとコンテンツを収集して、Webサイトを通じてこれを提供している。
- 共通のメタデータ標準を使ったり、異なるメタデータフォーマットを持つレコードをマッピングしたりすることにより、メタデータレベルで統一することができる。
- 上記3方法の組み合わせ。単独で完全に満足させる方法があるとは思われない。

オープンな標準やプロトコルの使用は、この状況を克服するための必須のツールである。リポジトリとリポジトリの保管物を利用したいと望んでいるサービスでは異なるレベルで取り決めが行われるからである。自動的に2つのリポジトリから同時に情報を持ってくると、それらの間の違いを処理するために使用されるメカニズムにより情報を失う結果になる。最下位のメタデータレベルにおいて統一すれば、問題が生じる前にこれに対応できる可能性がある。

不均一性が問題となる可能性があると認識していたとしても、すべての問題を解決しようとする前に、様々なコミュニティにそれがどの程度影響を及ぼすかを理解することが重要である。不均一性は同時に、リポジトリの進化にとって最大の希望であるとも考えられる。最終的には、様々なリポジトリをその違いを考慮に入れた上で同時に利用することができる方法を探す必要がある。

## A1.2 オープンアクセス

リポジトリを構築する際に検討すべき問題の<sup>1</sup>つは、コンテンツへのアクセスの公開・非公開をどのようにするかであろう。保存を目的とするリポジトリでは非公開のアーカイブとして作成される場合もあり、この場合、短期的なアクセスは重要な機能ではない。しかし、多くのリポジトリにとってその目的は何らかの方法でコンテンツをエンドユーザに提供することである。このアクセスをどのようにコントロールすべきだろうか。言い換えれば、保管しているメタデータとコンテンツをどの程度公開し、どの程度アクセスを制限するかである。広義には、通常、機関リポジトリは、ほとんどがe-プリントや雑誌論文という形の研究成果に対するオープンアクセスを促進するために開発されてきたものである。実際、制限なしのアクセスはその目的の不可欠な部分である。「機関リポジトリ」という言葉は、ほとんどこのオープンアクセスアプローチの同義語になっている。しかし、中には、単なるe-プリントを超えるものを保管しており、公開するものとしたくないものを決定する必要のある機関リポジトリも存在する。たとえば、機密資料を

含む学位論文など、資料へのアクセスに特別な条件や制限のある場合がこれに該当する。

その他の種類のリポジトリにおいて資料をオープンアクセスとして提供する範囲は、多くの場合、エンドユーザの識別が必要であるかどうかに等しい。オープンアクセスとは、その定義により、すべての人が好きな時にコンテンツを自由に利用できるようにするものである。しかし、ある種の資料にとっては通常、誰がそれにアクセスしているかを知ることが重要である。学習用オブジェクトを所有する多くのリポジトリはこの方法で運用している。通常、コンテンツの知的財産権の保護がアクセス制限を設ける理由となっているが、エンドユーザが誰であるかを知ることが、そのようなエンドユーザを対象を絞ったサービスを開発する上での参考になるという利点も持っている。エンドユーザを知らないということは、幅広いユーザーを得た場合に、諸権利がどの程度順守されるかをリポジトリがあまりコントロールできないことを意味するが、知的財産権をオープンアクセス資料に設定することも可能である。

以下の章では、多かれ少なかれオープンアクセス機能が実装に組み込まれているリポジトリの開発を取り上げる。エンドユーザサービスは、ユーザーが利用できるものとできないものを明確にするために、利用したいと考えるリポジトリについてよく理解する必要がある。

## A2 リポジトリの種類とその概要

以下の節では、エンドユーザサービスの開発に影響を与えると思われるリポジトリの領域で行われているいくつかの開発の概要を述べ、エンドユーザサービスの提供に関係する問題を検討する。

### A2.1 機関リポジトリ

2003年、CNIのClifford Lynchは機関リポジトリの定義を次のように提案した ([Lynch, 2003](#))。

「私が考えるに、大学における機関リポジトリとは、大学がその構成員に提供する、大学とその構成員により作成されたデジタル資料を管理し発信するための一連のサービスである。ここで最も重要なことは、これらのデジタル資料を組織し、アクセスを提供し、あるいは配布し、適切な場合には長期保存を行うなど、資料の受託管理に対して組織として責任を持って取り組むことである。… 機関リポジトリは単なるソフトウェアとハードウェアの固定化されたセットではないのである。」

上で述べたように、機関リポジトリは時に、研究成果を置く場所を提供し、それにより研究成果の組織化とアクセスを容易にすることにより、オープンアクセスを支援するリポジトリという概念の同義語になっている。しかし、リポジトリは想定される一連のサービスの方向へ徐々に進化していると思われる。これらのサービスは、様々な種類の共通性のあるコレクションを1つのものとしてオープンアクセスにすることをめざすだろう。本報告書は、事実上大部分がオープンアクセスである開発を取り上げるが、「機関の」リポジトリに対する目的は異なるので、各機関が採用している戦略には相違があり、これがリポジトリの将来のオープンアクセスの程度に影響を与えるかもしれない。

#### A2.1.1 JISC FAIR プログラム

FAIRプログラムの14のプロジェクトのうち、3つのプロジェクトが主にe-プリントを発信するオープンアクセス機関リポジトリの開発を対象とした。サウサンプトン大学のTARDis<sup>7</sup>とグラスゴー大学のDAEDALUS<sup>8</sup>、ノッティンガム大学のSHERPAである。SHERPAプロジェクトは多くのパートナーを持ち、さらに<sup>17</sup>機関と英国図書館（所属機関のない研究者向け）でリポジトリが構築されることになった。さらに、セントアンドリュース大学はHaIRSTプロジェクト<sup>9</sup>で、エジンバラ大学はTheses Alive!プロジェクト<sup>10</sup>で各々リポジトリを開発した。これらのリポジトリの開発は、機関リポジトリが採用できる様々なアプローチを明らかにした。概して、すべてのプロジェクトが研究成果とe-プリントを対象とした。ただし、TARDisは、研究成果の意味するものを比較的広く捉えて、多くの種類の研究成果物を含めることを追求した。対照的に、SHERPAは雑誌論文のプレプリントとポストプリントだけを対象を絞った。DAEDALUSは両者を対象としたが、管理上とサービス上の理由から、両者を<sup>2</sup>つの異なるリポジトリに分配した。

#### A2.1.2 英国における開発

FAIRによる開発とこれに並行した開発により、英国のオープンアクセスリポジトリの数は大きく増加した。サウサンプトン大学のRegistry of Open Access Repositories (ROAR) [11](#)には、2006年5月現在で英国のリポジトリが70、リストアップされている。これらには機関のリポジトリ、複数機関による合同のリポジトリ、部局レベルのリポジトリに加えて雑誌出版社により運用されているリポジトリや、データセットへのアクセスを提供しているリポジトリなど少数の特殊な事例も含まれている。部局リポジトリの多くは、その起源により主題ベースのリポジトリであるとも考えられる。Lynchの引用の末尾の記述とは合わないが、ほとんどすべての機関リポジトリは想定されるある一定の機能やサービスを提供する特定のソフトウェアパッケージを使って構築されている。これは英国においては通常、サウサンプトン大学のEPrintsシステム [12](#)かMITのDSpaceシステム [13](#)の採用を意味している。

### A2.1.3 世界における開発

リポジトリの開発におけるこの進展は英国だけに留まらない。2005年5月のJISC、SURF、CNI共催の会議において、世界の13カ国におけるリポジトリの設置に関する貴重なデータが作成された ([van Westrienen, 2005](#))。少数の国では依然として少ないものの、多くの国ではリポジトリの設置は増加中である。この傾向は機関リポジトリが機関のインフラとして定着することを導いていると、この論文の著者は結論付けた。また、そのようなリポジトリインフラの存在はエンドユーザサービス層を支援するだろうとも述べている。エンドユーザサービスの開発支援がリポジトリの存在に依存していることは注目に値する。

### A2.2 データセット

多くの分野の研究では大量のデータを生み出す。公的助成を受けたデータは自由に利用できるべきであるということは、OECDを通じて国際レベルにおいて広範囲に合意されている [14](#)。英国の研究会議の多くは助成を行った研究で得られたデータを全国データセンターを通じて保管している。研究活動から得られるデータを自発的に、あるいは要求によりデポジットするために利用できる中央サービスが数多く存在する。たとえば、Arts & Humanities Data Service (AHDS) [15](#)やEconomic and Social Data Service (ESDS) [16](#)、British Atmospheric Data Service (BADC) [17](#)などである。データセットはその他の組織や機関にも保管されている。機関レベルでは、データセットを保管する独自の施設を構築する傾向があった。これらは機関全体で調整されたものではなく、データを生成する部局の要求によるものであった。その結果、実験データとそれにより作成された出版物の連携が同じリポジトリで行われることは稀である。実際、データと出版物を一緒に保管することを支援している様子はほとんどなく、ほとんどの活動はこれらを独立に保存することに関心を向けている。しかしながら、追加情報を得たり、提示された考え方がどのように発展してきたかを理解したり、データ出力がどのように記述されたかを理解したりするためにはデータと出版物が連携できることが望ましい。たとえば、Entrezのヌクレオチドデータベースやゲノムデータベース、タンパク質データベースとPubMedの連携は、国立医学図書館の中央リポジトリを中心に構築された定評のあるエンドユーザサービスである [18](#)。

#### A2.2.1 eBank UKプロジェクト

そのような連携は、現在、Digital Repositories Programmeの多くのプロジェクトで研究されており、特に、分散して存在するリポジトリとつなぐ方法についても追求されている。これらはUKOLNとサウサンプトン大学による初期のeBank UKプロジェクトに刺激されたものである ([Lyon, 2003](#))。eBank UKでは、EPrintsを改造したシステムに結晶学のデータセットをデポジットすること、および、(ダブリンコア由来のスキーマを使って) 関連するメタデータを作成することを研究した。このメタデータを収穫して、結晶学の出版物からのメタデータと共にアグリゲートすることができる。これらの相補的な情報と利用できる場合は両者の間で確立されたリンクに対する統合エンドユーザ検索サービスが提供された。プロジェクトの第2期では、よりリッチな情報とそれを使ったエンドユーザサービスを提供するためにデータセットのさらなるメタデータ記述を検討中である。この作業を支えるものは、研究や学習、教育のワークフローの中にデータセットを正しく位置づけたいという願いである。

eBank UKの経験は、European Research Area (欧州研究圏)における将来の知識基盤のためのテストベッドを確立することをめざす、より対象範囲の広いEU DRIVERプロジェクト案に活かされている。このプロジェクトは、データのための既存のインフラであるGEANT2を補足することをめざしており、また、これらのデータと知識基盤が相互作用できる方法についても検討する予定である。

#### A2.2.2 JISC Digital Repositories Programmeのプロジェクト

データセットをリポジトリ環境に持ち込み、それにより、データセットをより広く発信する道を開くさらなる研究が次のように行われている。

- CLADDIER<sup>20</sup>は、データを科学者が発見し補足した場所から、生成された新たなデータを保管するデータリポジトリへシームレスに移動しデポジットできるようにするワークフローを研究している。様々な出版物とデータオブジェクトの識別がこれを利用する鍵である。
- GRADE<sup>21</sup>は、地理空間データセットの交換と他の地理空間資源との連携を容易にするために地理空間データセットを保管するリポジトリの可能性を検討している。
- R4L<sup>22</sup>は、実験データを操作、発信、報告する場合の問題に対応するサービスとツールのプロトタイプを開発している。これには、その後の利用のために効率的なアグリゲーションを可能とするメタデータの研究を含んでいる。
- SPECTRa<sup>23</sup>は、実験データの消失を防ぎ、さらなる研究や教育を支援するためにオープンアクセスで利用できるようにするという観点から、実験データのデポジットを容易にするツールを開発している。
- StORe<sup>24</sup>は、一次研究データリポジトリと出版物リポジトリに対する研究者のニーズの違いと、これら二つの連携を容易にするために必要となるミドルウェアや共通のインフラを研究している。

### A2.2.3 グリッドとリポジトリ

世界中で行われていた一般的なグリッドやコンピュータ基盤の開発は、リポジトリの開発や関連するエンドユーザサービスに利用できる多くのリソースをもたらした。グリッドは、ストレージ機能をネットワークレベルに移すことにより、利用可能な容量を増加させる機会を提供する。以下は、これが役に立つと考えられる2つのケースである。

- デジタルコンテンツを（メタデータと共に、あるいは独立に）自らエンドユーザサービスとして提供する場合、ファイルサイズや迅速なアクセスに必要な帯域幅といった問題が存在する。ネットワークキャッシングサービスは、コンテンツの移動が頻繁に生じる場合に発生する可能性のあるボトルネックを緩和するのに役立つだろう。
- ARC OAI-PMHハーベスタの開発者は、より多くのメタデータをハーベストするにつれ、システムがその負荷に容易に対処できなくなることに気づいた。彼等の解決策は、ハーベスティング処理で行われる負荷、特にアグリゲーション処理自体の負荷をグリッドに分散することであった。すなわち、複数のデータプロバイダがコンテンツを公開している。個々のグリッドノードは必要に応じてこれをハーベストする。そして、横断検索モジュールですべてのハーベスタノードを横断検索するのである。

EU DILIGENT (Digital Library Infrastructure on Grid ENabled Technology) <sup>25</sup>プロジェクトは、仮想的研究グループのためのオンデマンド・デジタルライブラリを開発するためにグリッド技術を全面的に利用している。

### A2.3 マルチメディア資料

インターネット、特にWebの発展は、一般的な用途や研究目的において画像資料や音源資料、ビデオ資料の広範囲な利用を促した。これらのリポジトリは多くの場合オープンアクセスで利用可能であるが、著作権やコレクションを商業的に利用したいと考えるコンテンツ所有者の希望により、利用に大きな制限が設けられているリポジトリもまた同じくらい多く存在する。画像やビデオのコレクションは、高等・成人教育の学術コミュニティではナショナルライセンスにより利用可能であるが、EDINAのEducation Image GalleryやEMOL<sup>26</sup>のような国立データセンターのエンドユーザサービスを利用することによりオープンアクセスで利用することもできる。しかし、このような資料の多くがコミュニティ内に存在し、たとえコンテンツの作成者が望んでいたとしても、現在のところそれが公開されていないことは明らかである。

#### A2.3.1 Community Led Image Collections (CLIC) 調査

多くの画像コレクションは依然として非公開であり、エンドユーザサービスを通じて利用することができない。これらは主に学術コミュニティにおいて自らの利用のために作成されたコレクションである。JISCの助成を受けたCLIC調査<sup>27</sup>は、これらのコレクションを調査し、より広く公開するための最良

の方法を研究するために始められた。調査により、すべてのコレクション所有者はコレクションを公開するという考えを条件付ではあるが受け入れることがわかった。コレクションのうち利用可能なものをリストアップする一覧表や、目録やハーベストしたメタデータに基づく発見サービスは歓迎された（80%以上）が、全国サービスに貢献するという考えはあまり歓迎されなかった（30%程度）。コレクション所有者は自分のコレクションが発見されることは望んでいるが、自分のWebサイトを通じて公開した場合に得られると考えられるブランド化を失うことは望んでいなかった。

調査では、エンドユーザサービスを可能とするインフラが不足していることも明らかになった。多くのコレクションは単なるMicrosoft Accessのファイルであり、メタデータはそれほど広くは作成されていなかった。これらのことから、サービスの実現を容易にするには、画像を公開する手軽なアプローチが必要であるという勧告が出された。

#### A2.3.2 ビデオと音源

2003年に行われた動画と音源のためのポータルを実現するためのユーザー要件調査<sup>28</sup>でも、CLICと同じように、これらの資料の発見を容易にするための解決策の実施が支持された。また、コミュニティが作成した資料を保管・発見される場所についても大きな関心が示された。ただし、この発見への関心は、何らかの中央発見サービスへ資料を提供することに対する関心とは必ずしも一致しなかった。関係者は、絶対に自分の手の内は見せることなく他人のすることを見たいと望んでいた。CLICにより宣伝された手軽なアプローチは、この調査の回答者のニーズにも合致すると思われる。

#### A2.4 学習オブジェクト

先の節で示した資料の多く、実際には、オープンアクセスで公開されているすべての資料は、潜在的には学習オブジェクトであると考えられることができる。しかし、研究成果のためのリポジトリの開発では、ほとんどの場合オープンアクセスによる公開が組み込まれているが、学習オブジェクトのリポジトリでは、これが必ずしも主要な優先事項にはなっていない。学習オブジェクトリポジトリと研究オブジェクトリポジトリの開発は並行して行われてきたが、両者の方向性の違いはおそらく、表<sup>1</sup>で示したような学習オブジェクトと研究オブジェクトに対する一般的な認識の違いに関係しているだろう。

学習オブジェクト	研究オブジェクト
作成者は手の内に置くことを好む	作成者は手から放すことを好む
エンドユーザは利用することを求められる	エンドユーザは読むことを求められる
分割・別目的の利用が支持される	分割・別目的の利用はまれ
リポジトリはまだ便利なものだと考えられていない	リポジトリは便利なツールだと考えられている
リポジトリソフトウェアはオーダーメイド	リポジトリソフトウェアはオープンソース

表 1: 学習オブジェクトと研究オブジェクトの相対的な性格

オブジェクトが何のためのものであるかについての理解が異なっているので、他人に何をすることを許すかについて異なる見解が生じることになる。研究オブジェクトは1つの研究を発信するようデザインされており、これをオープンアクセスとして公開する際には、おそらく暗黙のうちに、これは決して変更されることはないという認識がある。対照的に、学習オブジェクトは単に読まれることではなく、利用されることを意図している。学習オブジェクトを分解して、他のものと組み合わせることで別の目的に利用することは奨励されている。しかし、この変更の自由は、時に不本意なまま放置されている場合もある。JISCのX4Lプログラム<sup>29</sup>の調査から、作成者は多くの場合、学習オブジェクトは、一般に分割することなく利用できるような方法で作成されるべきだと考えていることが示された。また、多くの作成者は開発したオブジェクトを商業的に利用したいと考えている。それゆえ、学習オブジェクトのリポジトリは必ずしもオープンアクセスではなく、ユーザー登録や有料化、指定コミュニティの会員限定など、何らかのアクセス制限を設けている。JORUMリポジトリの利用条件など、英国ではこれが特に当てはまるように思われる。一方、米国では、iLumina<sup>30</sup>やSMETE<sup>31</sup>のように、オープンアクセスの学習オブジェクトリポジトリが数多く存在する（ただし、これらの名称はどちらもデジタルライブラリのものであることに注意せよ）。学習オブジェクトはWeb上で見つけることもできるし、リポジトリ以外で利用できるものもある。作成者は、作成した資源のメタデータを投稿する場としてMERLOT<sup>32</sup>などの発見サービスを利用することができる。これにより、メタデータを公開する一方で、オブジェクト自体は手元に置き、コントロー

ルを保つことができる。

#### A2.4.1 複合型オブジェクト

多くの学習オブジェクトは単一の画像や文書である単純型オブジェクトであるが、多くの要素で構成される複合型オブジェクトであるものも多い。これらの複合型オブジェクトはIMSコンテンツパッケージ仕様を使って組織化することができる。この仕様には、各要素がお互いにどのような関係にあるかを記述したマニフェストファイルが含まれている。複合型オブジェクトは分割や別目的の利用をすることができるので、この型の学習オブジェクトの再利用を可能とする。複合型オブジェクトは、これを管理し、正しい方法で提供できるシステムを必要とする。現行のVLE（仮想的学習環境）システムは必ずしもこれを効率的に処理していない。これが学習オブジェクトへの一般的な理解に影響を与えているかもしれない。

### A3 ユーザー指向サービス

#### A3.1 既存の活動の概要

オープンアクセスのリポジトリ横断型アクセスを提供するイニシアティブは数多く存在する。そのうちで重要な開発をここで手短かに紹介する。オープンアクセスとOAIは密接に関係するので、ここで紹介するサービスの多くはOAIハーベスティングモデルに基づいており、OAI-PMHを使用している。しかし、最近では、オープンアクセスによる公開を進めるために非OAI技術を使った開発も数多くあり、これについても検討が必要である。

Arc – Arcは、OAI-PMHに基づく初めてのエンドユーザ横断検索サービスである。OAI-PMHがリリースされた直後の2000年にオールド・ドミニオン大学で構築された。Arcは概念証明型のサービスとして設計され、既知のデータプロバイダからハーベストする独自のエンドユーザーサービスとして維持されているが、おそらく、サービスを行うために作成したハーベスタや検索用のソフトウェアの方でよく知られている。これらのソフトウェアは、その後、数多くのOAIサービスプロバイダで利用されており、以下で説明するePrints UKの開発にも使用されている。このソフトウェアは、簡略検索、詳細検索、対話式検索、アンテーションサービス、検索結果のブラウズとナビゲーション機能を提供する。

Arcサービス: <http://arc.cs.odu.edu>

Arcソフトウェア: <http://sourceforge.net/projects/oaiarc/>

ePrints UK – JISC FAIRプログラムの助成を受けたプロジェクトで、ePrints UKは英国の機関リポジトリに存在するe-プリントへのアクセスを提供することを追及したが、BioMed Centralのような英国の学術コミュニティが関心を持つリソースへのアクセスも追加されている。ePrints UKはArcソフトウェアを使用している。サービスはプロジェクト終了後も引き続き稼動しているが、活動はそれほど活発ではない。プロジェクト自体は、Webサービスを使ってe-プリントのフルテキストを補足し、e-プリントの主題、引用、著者の各メタデータを追加する可能性を調査した。ダブリンコアを使ってe-プリントの目録をとる方法およびDCレコードからフルテキストにリンクする方法に関するガイドラインを作成することにより、ハーベストされるリポジトリ間である程度のメタデータの一貫性を提供することが追求された。

ePrints UK サービス: <http://eprints-uk.rdn.ac.uk>

ePrints UK プロジェクト: <http://www.rdn.ac.uk/projects/eprints-uk/>、および  
[http://www.jisc.ac.uk/fairsynthesis\\_eprintsuk.html](http://www.jisc.ac.uk/fairsynthesis_eprintsuk.html)

OAIster – OAIsterプロジェクトは、2001年にOAIサービスプロバイダの開発を研究するためにメロン財団が助成を行った7つのプロジェクトのうちの1つである (Waters, 2001)。これらのプロジェクトのうち3つは公開されており、中でもミシガン大学のOAIsterが最も有名である (他の2つは、エモリー大学のAmerican Southプロジェクトとイリノイ大学のDigital Gateway to Cultural Heritage Materialsである)。OAIsterは知られているすべてのリポジトリをハーベストしているが、実際にオブジェクトにリンクしているレコードだけを保管している。すなわち、書誌レコードだけのレコードは保管していない。リポジトリによりメタデータに不整合が見られるので、OAIsterではアクセスを容易にするために数多くの正規化を行っている。たとえば、資料種別に関する制限リストへのレコードのマッピングである。この作業は現在手作業で維持している。OAIsterは、独自のWebインターフェース以外のインターフェースの普及を追及しており、ハーベストしたメタデータをYahoo!に公開したり、アグリゲートしたデータへのSRU検索インターフェースを提供したりしている。

OAIsterサービス: <http://oaister.umdl.umich.edu/>

DAREnet – オランダのSURFによるDAREイニシアティブは、オランダのすべての大学に機関リポジトリを実装した。DAREnetはこれらのリポジトリに対する横断検索サービスを提供するOAIサービスプロバイダである。ここでは明らかにオープンアクセスが強調されており、DAREnetに含まれるのは、コンテンツ全体が自由に利用できるオブジェクトのメタデータだけである。DAREnetの一部であるCream of Scienceサービスは、オランダの上位207名の研究者が生産したすべての出版物を対象としている（そのうち約60%がオープンアクセスとして自由にアクセス可能である）。3つ目のサービスとして開発されているPromise of Scienceは博士論文へのアクセスを提供するものである。DARC (Distributed Africana Repositories Community) <sup>33</sup>などの小規模な主題指向のサービスプロバイダにより、全国サービスは補完されている。オランダは、現在のところ、全国レベルのオープンアクセスエンドユーザサービスを持つ世界中で唯一の国である。

DAREnet サービス: <http://www.darenet.nl>

Cream of Science サービス: <http://www.creamofscience.org/>

ARROW – オーストラリアのARROW (Australian Research Repositories Online to the World) イニシアティブはメルボルンのモナシュ大学が本拠地であるが、パートナーはオーストラリア全土に存在し、データプロバイダとサービスプロバイダの両者を含んでいる。特に、オーストラリア国立図書館 (NLA) はARROW Discovery Serviceを担当し、数多くのリポジトリを横断するOAIサービスプロバイダ機能を提供している。このサービスにおける整合性を保つために、NLAはハーベスティングへのレコード提供方法に関する指針を制定している。NLA自身も様々なOAIコレクションをハーベスティングに提供しており、中でもGoogleに便宜を図っている。画像はARROW内で独立にハーベストされ、PictureAustraliaやMusicAustraliaで使用されている。

ARROW プロジェクト: <http://arrow.edu.au>

ARROW 発見サービス: <http://search.arrow.edu.au/apps/ArrowUI/>

グラスゴー大学 – リポジトリ横断のエンドユーザサービスを提供するのは何も国レベルや機関合同レベルである必要はない。グラスゴー大学は出版された論文用とプレプリント・灰色文献・学位論文用という2つの異なるリポジトリを運用している。これら2つのリポジトリを利用しやすいように共通のアクセスポイントを提供するローカル用のOAIサービスプロバイダをPKPハーベスタソフトウェアを基に開発中である。グラスゴー大学のリポジトリのコンテンツは、もうひとつのアクセスポイントとしてGoogleによるクローラに公開されている。

DAEDALUS プロジェクト: <http://www.lib.gla.ac.uk/daedalus/>、および

[http://www.jisc.ac.uk/fairsynthesis\\_daedalus.html](http://www.jisc.ac.uk/fairsynthesis_daedalus.html)

Citebase – この実験的なOAIサービスプロバイダはサウサンプトン大学のTim Brodyにより開発され、数多くの様々なリポジトリを対象とする汎用的な横断検索を提供している。Citebaseは論文の引用を見ることができ、また、これを基に引用分析をすることができることが注目に値する。

Citebase サービス: <http://www.citebase.org/>

BASE/Omega/Scirus – ドイツのビーレフェルト大学のBASE (Bielefeld Academic Search Engine) サービスは、OAIデータプロバイダとその他の情報源を組み合わせ、これらに対するFASTベースの検索インターフェースを提供している。これは、OAI-PMHがWeb検索エンジン技術と連携して使用されている一例である。検索結果からGoogle Scholarへのリンクが張られており、これをたどることにより、エンドユーザは調査を続けることが可能となる。ユトレヒト大学のOmegaメタデータベースも、横断検索を提供するために、電子ジャーナルを主とする内外の様々なソースからOAIでメタデータをハーベストしてアグリゲートしている。DAREnetを通じて提供するためにこのメタデータはどの後再ハーベストされる。ElsevierのScirus Web検索エンジンも、BASEと同じFAST検索エンジン技術を使い、Webクローラを通じてリソースの横断検索機能を提供しているが、クローラがリポジトリを知っている場合はオープンアクセスリポジトリへのアクセスも提供している。

BASE: <http://www.base-search.net/>

Omega: <http://omega.library.uu.nl/>

Scirus: <http://www.scirus.com/>

その他 – <http://www.openarchives.org/service/listproviders.html>にあるOAIのWebサイトには、その他にも数多くのOAIサービスプロバイダがリストアップされている。このリストの中で注目すべきものは次のとおりである。

- DL-Harvest – Digital Library for Information Science and Technology (DLIST) の一部として開発されたサービスプロバイダである。情報科学に主題を絞っているが、これは適当なりポジトリのみをハー

ベストの対象とすることにより実現している。METALISはこの分野におけるもう一つのサービスであるが、ヨーロッパのリポジトリに重きを置いている。

- MeIND – ドイツの様々なリポジトリを対象とする一般的な検索サービスであるが、同時に、コンテンツをコンテンツ種別と主題で木構造に分けてブラウズできるようにしている。
- NCSTRL – e-プリントと共にテクニカルレポートへのアクセスに焦点を絞ったサービスプロバイダである。

Web検索エンジンの開発 – 既に述べたように、最近、一般的なWeb検索エンジンでオープンアクセス資料へアクセスできるよう開発が進んでいる。しかし、オープンアクセス文献への第一アクセスポイントとしてのWeb検索エンジンの価値をより広げるような多くの動きがあることがますます明らかになっている。2006年3月に行われたサウサンプトン大学における機関リポジトリのアクセス分析により、アクセスの64%はWeb検索エンジン経由であり、リポジトリに直接アクセスしたり、独自の検索インターフェースによるものではないことが明らかになった<sup>34</sup>。ロスアラモス国立研究所で最近行われた調査から、OAI-PMHリポジトリからハーベストした330万件のレコードのうち、65%はYahoo!で、44%はGoogleで、7%はMSNサーチで索引化されており、21%はこれら3者のいずれでも索引化されていないことがわかった (McCown, 2006)。2つのシステムがこの公開に取り組み、対応しようと試みている。

- DP9 – Arcサービスプロバイダのチームは、リポジトリとWeb検索エンジンの間に置き、検索エンジンがリポジトリのメタデータを索引化できるようにするこのツールも開発している。これは、リポジトリレコードに固定のURLがない場合に特に重宝する (ただし、DSpaceやEPrintsなど多くのリポジトリソフトウェアシステムはこの機能を提供している)。  
DP9: <http://arc.cs.odu.edu:8080/dp9/index.jsp>
- Google sitemaps – 索引化のためにサイトを登録するこの機能は、現在ではOAI-PMH 2.0準拠のリポジトリの登録を許可している。これは、上で述べたオーストラリア国立図書館が使用しているルートである。  
Google sitemaps: <http://www.google.com/webmasters/sitemaps/login>

メロン財団が助成を行ったmod.OAIの開発はこの逆を追求したものであり、Apacheウェブサーバ上のWebコンテンツをOAI-PMHを使って公開できるようにするものである。これは、Webサーバに対して、例えば、ある日以後に追加・更新されたすべてのファイルをリクエストしたり、あるMIME型のすべてのファイルをリクエストしたりできるようにする。

mod\_oai: <http://www.modoi.org/>

Peter Suberが2005年11月に行った新たなオープンアクセス検索プロジェクトのレビュー<sup>35</sup>により、OAIとWeb検索エンジンの両アプローチをカバーするこの分野にはさらに多くの開発が行われていることが明らかになった。

- OJAX – ユニバーシティカレッジ・ダブリン校の情報図書館学科で開発されたOAI準拠のリポジトリのためのAJAXを使ったメタ検索エンジンである。  
OJAX: <http://ojax.sourceforge.net/>
- Choogle – 化学のためのオープンアクセスWeb検索エンジン。これは、商用で作成され、機関内で使用するためにはライセンスが必要であるが、無料のサービスである。テキストによる検索だけでなく、化学構造式を作成して検索することもできる。  
Choogle: <http://www.choogle.com/>
- Science.gov – 米国の政府機関が作成したオープンアクセス資料に対するポータルでWeb検索エンジン機能も持つ。  
Science.gov: <http://www.science.gov/>
- ブラウザのツールバー – NCBIは所有するすべてのデータベースを検索できるFirefox用とIE用のNCBI検索ツールバーを提供している。OAsesは、Web検索エンジンだけでなくオープンアクセスサービス (OAster) やリポジトリ (DOAJ) <sup>36</sup>も横断検索するIE用のツールバーである。  
NCBI: [http://www.nlm.nih.gov/pubs/techbull/nd05/nd05\\_toolbar.html](http://www.nlm.nih.gov/pubs/techbull/nd05/nd05_toolbar.html)  
OAses: <http://psyplexus.com/oases/>
- Creative Commons検索エンジン – クリエイティブ・コモンズ・ライセンスで検索することができ、エンドユーザは利用可能な資料を直ちに見つけることができ、さらにその利用条件を知ることができる。この機能はしばらくの間、クリエイティブ・コモンズのWebサイト<sup>37</sup>で提供していたが、現在は、GoogleとYahoo!でも提供している。

Web検索エンジンを知り、可能であれば協定を結ぶことはリポジトリにとって有益である。現在のところ、多くの場合、互いに許可を求めているので、両者の関係は不明確である。検索エンジンがキャッシュコピーを提供している場合、これはアイテムの取り消しという問題を引き起こす可能性がある。

RSS – RSSを使ってリポジトリ間の相互作用を促進するには2通りの方法がある。個々のリポジトリがRSSフィードを提供し、各アグリゲータがこれを収集してエンドユーザに提供する方法と、OAIアグリゲータがアグリゲートしたメタデータに基づいてRSSフィードを作成して、エンドユーザに提供する方法である。そのようなフィードはコンテンツが定期的にデポジットされる際にフィルタリングする必要がある。そうでないとフィードを解釈し利用することが難しくなるからである。ただし、RSSは必要とするアグリゲーションと情報のレベルと幅をエンドユーザが決めることを可能にしている。次のような例が存在する。

- グラスゴー大学は、出版済み論文用のリポジトリに新規に登録されたレコードのRSSフィードを提供している<sup>38</sup>。
- UKOLNのIMeshプロジェクトは、OAI-PMHを使って公開されたレコードをRSSを使って公開するモジュールを開発した (Duke, 2003) 。

ARROW Discovery Serviceは、RSSは使用していないが、選択した主題に基づいて新規コンテンツをメールで知らせるサービスを提供している。これら両技術的アプローチは、情報の選択的配信 (SDI) 機能を提供するものであり、これは検索サービスの付加価値的補助サービスとして長年にわたりその価値が証明されている確立されたアプローチであり、サービスである。

### A3.2 ユーザー指向サービスの設計

カナダのサイモンフレーザー大学のGriff Richardsは、本研究のインタビューでリポジトリ横断型のエンドユーザサービスに必要なものについて尋ねられた際に、次のように強い調子で答えた<sup>39</sup>。

「もしもサービスがあるリポジトリを対象とせずに構築されたとしたら、そのリポジトリは消える運命にある。もしもあるリポジトリのアーキテクチャがサードパーティによるサービスの自発的開発を妨げるとしたら、そのリポジトリは消える運命にある。一連の健全なサービスはコンテンツが生きていることを示している。もしも相互運用性がないとしたら、それは利用可能なコレクションではなく単なる知識の貯蔵庫になるに過ぎない。」

前の節で述べたサービスはいずれもリポジトリが単なる知識の貯蔵庫にならないことを目指したものである。これらのサービスは、単にオープンアクセス資料に基づくサービスだけでなく、そのような多くのサービスを企画開発する際の問題に対処しなければならなかった。この節では、サービスを企画する際に検討を必要とする様々な問題を取り扱う。これらの問題を現実置き換える際には、CNRIのLarry Lannomが示唆したように、スーパーマンモデルを考慮し、物理的パラダイムは信頼しないことが役に立つだろう<sup>40</sup>。

#### A3.2.1 配信オプション

アグリゲータとリポジトリは、自身をエンドユーザに利用させる方法、および、これを支援するために必要なインターフェースを決定する必要がある。独自のWebインターフェースで配信を行うとブランド化することができるが、エンドユーザに自身のWebサイトを発見するよう要求し、さらに、サービスを訪問することを求めることになる。Webインターフェースの開発はさらに、エンドユーザが扱わなければならない無数のWebインターフェースをさらに増加させることになる。そのため、たとえば、情報環境にe-プリントサービスを追加することはエンドユーザに混乱を与え、利用を減らすことになるだけかもしれない。広く利用されているサービスの一部になることやユーザーワークフローに合わせることで、認知度を高めることになる。たとえば、リポジトリをより広いデジタルライブラリの中に置けば、その特徴が際立ち、それを使ったサービスを開発するより大きなインセンティブができることになる。認知度が高まった後も、利用可能となった検索などのオプションがエンドユーザにとって意味のあるものであり、また、サービスが提供するものが明らかであることを保証することが重要である。

代替手段として、マシンインターフェースを開発することは、コンテンツへのさらなるルートを提供することになり、ブランド力は低下するが、コンテンツが発見され利用される機会ははるかに大きくなるだろう。マシンインターフェースはエンドユーザのWeb環境で利用され、サービスをエンドユーザに届けるこ

となる。これは、コンテンツをWebサーチエンジンに公開しているリポジトリで起きていることである。OAIsterが提供している<sup>41</sup>ように、SRW/Uインターフェースの提供は、メタサーチサービスなどの構造化検索のルートにオープンアクセスリポジトリのメタデータを含めることを可能にする。たとえば、d+ツール<sup>42</sup>を使うことによりSRUターゲットを実現でき、機関ポータルやVLEに検索機能を追加することができる。機関ポータルなどの環境への組み込みには、WSRP<sup>43</sup>を使ってそのような組み込みを複数のプラットフォームで可能にするポートレットの開発の恩恵を受けるかもしれない。どちらの配信ルートを採用するかを決断する際には、対象となる開発作業と支援の量により最適な解が存在するだろう。どちらの手順を採るべきかについては確固とした規則は存在しないが、対象とするエンドユーザがどこにいるかを検討し、サービスとコンテンツをそちらに向けて公開するよう作業すれば、最大の利用を得ることができるだろう。

### A3.2.2 ワークフロー

通常、エンドユーザサービスを企画する際には、リポジトリとの相互作用のワークフローに注意する必要がある。オープンアクセスの分野では、これは特に、研究者のワークフローを明らかにし、リポジトリ横断型のエンドユーザサービスをそれに合わせることに該当する。リポジトリへのアクセスを研究者の環境に組み込むことができさえすれば、リポジトリが物理的にどこに存在するかはほとんど無関係である。そのようなワークフローの確立には、ワークフローを構成する様々な要素が実際に処理を行うことを保証するためにワークフローの各段階の品質保証が必要となる（ワークフローがSOAアプローチに基づいている場合は特にそうである）が、エンドユーザとその必要とする情報との相互作用を容易にするので長期的な利益を提供することができる。ワークフローを検討する際に解決しなければならない特別な問題は、ワークフローをどこで終了させるかである。ユーザーのワークフローは情報の利用を含んでいるが、多くの情報プロバイダモデルはエンドユーザへの提供の時点で終了しており、その利用は対象としていない。エンドユーザサービスにより配信される情報の考える利用のシナリオを作成することは、そのような情報を第一にどのように提供するかを考える際に役に立つだろう。

### A3.2.3 リポジトリの隠蔽と顕在化

エンドユーザサービス提供への次のステップは、リポジトリをエンドユーザに見えるようにするか、リポジトリを隠し、利用していることをエンドユーザが意識しないようにするかを決定することである。どちらのアプローチにも利点があり、エンドユーザが情報にたどり着くことを容易にすることができる。技術的には、おそらく中間的なリポジトリサービス層を経由することにより、リポジトリとエンドユーザサービスの関連が見えないように両者の関係を構造化することにメリットがある。ただし、エンドユーザサービスがリポジトリのための単なるシンレイヤのプレゼンテーションエージェントとして機能し、情報連鎖のある時点でエンドユーザをサービスからリポジトリ自体へ渡した方がメリットのある場合もある。これにより、エンドユーザサービス自体では直接提供できないより高度な機能やよりリッチな情報、フルコンテンツへのアクセスを提供できるようにすることができる。

この決断をするに当たって、利用可能なソフトウェアが提供する機能に影響を受けすぎないことが重要である。これは考慮すべきことではあり、作業を進めるために決断を必要としている機関では特にそうであるが、機関レベルで管理できる最高のサービスは何であるか、ネットワークレベル（すなわち、機関横断）で提供できる最高のサービスは何であるかを検討することが役に立つ。多くの情報資源に基づく商用サービスはこのネットワークレベルで運用しており、その対象範囲の広さがもたらす価値により広く使用されている。ネットワークレベルで稼動する最高のサービスは何かを考えることが、それを提供できた場合に得られる最高の価値をもたらす助けとなるだろう。エンドユーザの観点からは、リポジトリの場所とエンドユーザサービスの場所は無関係である。エンドユーザは情報が元々どこにあったとしても1つの場所で利用できるように統合されることを望むだろう。

### A3.2.4 集約か、分散か

上で示したように、利用可能な様々なサービスプロバイダを考えると、アグリゲーションのレベルやアグリゲータの数は増加している。これは、オープンアクセスの分野もそうであるが、特に商用部門に当てはまる。これらの多くは、分散アクセスモデルに従うのではなく、検索に提供したいと考えるあらゆるものをまとめて収集することに集中している。逆に、図書館ポータル製品は、図書館が提供するサービスのポートフォリオに追加するものとして人気があることを証明している。検索対象のスクレーパビリティは検討すべき問題の1つである。この件に関して分散検索には限界のあることが明らかであるが、アグリゲー

ションはこれを克服することができる。どちらのアプローチが最もふさわしいかを決定する際に鍵となる要因は、ユーザー要件を満たすことができる柔軟性である。

### A3.2.5 認証要件

オープンアクセスリポジトリは、すべての人が自由に利用できる資料を主に保管することになるが、たとえ限られた期間であるにしてもアクセスに何らかの制限を持つ資料のために認証と承認を行う必要が常に残されている。特に、学位論文がこれに該当する。学位論文には、商業的理由や知的財産権上、あるいは政治的な理由で取り扱いに注意が必要な資料を含んでいる場合があるからである。したがって、リポジトリソフトウェア自体の機能（FedoraのXACMLポリシー<sup>44</sup>など）を使用したり、ハーベストできる資料を限定（OAIセットを利用するなど）したりして、ある程度のフィルタリングを行う必要がある。利用できるものとできないものを区別することにより、メタデータとコンテンツが分けて考えられるようになるかもしれない。ARROW Discovery Serviceはメタデータを無料で公開しているが、フルコンテンツのアクセスに何らかの制限を設けることは認められている。

### A3.3 エンドユーザーサービスの機能

エンドユーザーサービスで提供できる機能には多くの種類がある。本研究で取材した人々により価値があると考えられている機能を表<sup>2</sup>に示した。各機能についてはこの後、詳細に検討する。

検索	ブラウズ	ハーベスト/アグリゲーション
コメント/アノテーション	編集/更新	出力
プッシュ/通知	保存	デポジット/受入
リンク付け	テキストマイニング	変換
入手	リクエスト	表示

表 2: エンドユーザーサービスに含めることが可能な機能

#### A3.3.1 全般

SCRANの評価により、エンドユーザが望む機能には、検索、ブラウズ、アグリゲーション、コメント、編集、出力が含まれることが明らかになった。この表におけるワークフローの進行は、エンドユーザが自分のワークフローをどのように考えているか、また、ワークフローにあわせるためにどのような情報資源を望んでいるかを明らかにしており、興味深い。エンドユーザは、Webという常に利用できる世界にいたので、一年中休みなくアクセスできることを期待していることは驚くことではない。これを提供するためには、実行不可能なものには手を出さず、提供可能なソリューションに的を絞り、ユーザーの視点にたって設計することに集中することが重要である。

#### A3.3.2 デポジットと受入

2005年11月に開催されたJISC CETIS会議<sup>45</sup>におけるリポジトリ機能のギャップ分析では、コンテンツをリポジトリにデポジットする標準的な方法が欠けていると結論付けた。デポジットは、主に個々のリポジトリに焦点が絞られる機能の分野であるが、様々なリポジトリにデポジットできるようになれば、たとえば、機関リポジトリや主題リポジトリなどの、異なるルートを通じてコンテンツを利用可能にするという問題を解決するかもしれない。JISC Digital Repositories Programme支援チームはこの問題に取り組んでいる<sup>46</sup>。また、2006年4月に開催された「学術リポジトリ間の相互運用性を拡大する」と題したメロン会議においてもこの問題は議論された<sup>47</sup>。

問題の核心には、様々なリポジトリソフトウェアパッケージで使用されているデポジット技術や方法の相違がある。デポジット機能はほとんどの場合ソフトウェアに埋め込まれているが、標準的なAPIがあれば独立したサービスとしてリポジトリから取り出すことができるだろう。SRW Update<sup>48</sup>はこれを提供しようと試みている。また、DSpaceはWebDAVを利用することで、これにある程度成功している<sup>49</sup>。しかし、両者とも今のところ、標準的なデポジットAPIに相当するものは提供していない。デポジットにその他の機能、たとえば、受入時の通知機能などを組み込むとこの処理に付加価値を与えることになるだろう。

### A3.3.3 ハーベストとアグリゲーション

リポジトリのメタデータやコンテンツのアグリゲーションを容易にする既存の様々なサービスについては既に説明した。アグリゲートされた情報は、その後、検索やブラウズで公開したり、あるいは他の機関によるさらなるハーベストに公開したりすることができる。エンドユーザは通常、何らかの別の方法でアグリゲートされた情報にアクセスすることになり、とりあえずアグリゲーションを開始することはないだろう。OAI-PMHの場合、ハーベスティングは水面下で行われ、その結果作成されるアグリゲーションをエンドユーザやその他のサービスに公開する。サーチエンジンはWebをクロールし、エンドユーザは通常検索やブラウズを通じてアグリゲートされた情報を利用する。RSSアグリゲーションは、適当なRSSフィードを選択するという意味でよりエンドユーザのコントロール化にあるといえる。しかし、ここでも、アグリゲーション処理はアクセスポイントの背後に隠れている。つまり、エンドユーザサービスとアグリゲーションは密接に関連しているが、アグリゲーションは必ずしもエンドユーザサービス自体の機能ではないのである。

### A3.3.4 検索とブラウズ

検索はそれ自体エンドユーザサービスが提供できる最も重要な機能である。エンドユーザは検索語を使ってリポジトリと相互作用することができ、情報はWebを通じてアクセスされるという幅広い期待に合致している。しかし、提供される検索の種類は異なり、数多くの選択肢が存在する。主題に基づくコミュニティを対象とする場合は主題検索が最も有用であるが、そのような主題検索を可能にするメタデータを提供することは難しいことが広く知られている。そうではあるが、これは、明確なユーザーの要求に応えるためにさらなる注目をする価値のある目標である。著者検索も重要であるが、これは主題検索の特別な形であると考えることができる。

その他の検索オプションには、電子学位論文のためのEThOSプロジェクト<sup>50</sup>やOAIsterのコンテンツ種別アクセス（ただし、これは高度な正規化を必要とする）のようなコンテンツ種別による検索、マルチメディアオブジェクトにアクセスする際に特に重要になるデータフォーマットによる検索、Citebaseサービスで提供されているような引用検索、以前に説明したクリエイティブ・コモンズ検索エンジンにより提供されているようなライセンス検索がある。既存の検索結果を、異なる検索タイプ、すなわち、異なるフィルターを適用してさらに検索する能力は検索プロセスに付加価値を与える。

検索は、自分が探しているものを知っているエンドユーザが既知の単語を使って行うには便利なものであるが、潜在的にはそれを知らないエンドユーザにとっても便利なものである。Webサーチエンジンはしばしば、ある言葉から始めて「探しているものを知る」という検索の開始点として利用されている。ブラウジングは、エンドユーザが探しているものを事前に知らなくても、案内することにより利用可能なコンテンツへ導くもう一つのルートを提供する。また、ブラウズを使うことで検索の開始を支援することができる。ブラウズの実装は複雑な場合がある。ブラウズを行うためには、リポジトリやアグリゲータにおいてメタデータやコンテンツが構造化されている必要があるからである。しかし、この試みは他の方法では実現できないようなコンテンツへ至る方法を提供することができる。ブラウジングは検索と比較すると、両者とも様々なユーザグループに対応し、そのニーズを満たすことができるにもかかわらず、現在のところ過小評価されている機能である。エンドユーザはその時点のワークフローにあったものが提供されるのであれば、どちらを使用しているかは気にする必要はないのである。

検索とブラウジングは、リポジトリが公開しているものをエンドユーザが探索することを可能にする。この探索機能に代わるものとして、近年、ビジュアル化が多くの関心を集めている。Grokker<sup>51</sup>はGoogleの検索を視覚的に表示する方法の一例である。また、Choogleを使った化学構造式による検索が、化学の分野で特に注目されている。現在のところ、ビジュアル化による代替手段の採用は一般的には限られているので、この対象を絞ったアプローチがビジュアル化にとって最適であるかもしれない。この例外は画像のブラウジングである。これは、その視覚的性格により画像へのアクセスを提供する一般的な方法であることが証明されている。

### A3.3.5 表示、出力、変換

検索とブラウズ機能の提供は、発見機能の半分に過ぎない。リポジトリは、エンドユーザサービスが便利な形に加工できる適当な派生物を返すことができなければならない。リポジトリとアグリゲータは検索やブラウズのリクエストにどのように応えるかを検討する必要があり、一方、エンドユーザサービスはリポジ

トリの出力をどのように扱うかを検討する必要がある。サービスは、メタデータとフルコンテンツをダウンロードするためのリンクを表示するだけでよいのか。あるいは、分析やコメント、再利用ができるようにサービスとしてコンテンツを表示したり変換したりできるようにすべきか。書誌データを様々な引用スタイルで出力することは提供できるサービスの1つであるかもしれない。

米国のロスアラモス国立研究所で開発されたaDOReアーキテクチャでは、リポジトリで利用できる配信物に関する情報を示すPahtways InterDisseminator<sup>52</sup>を提供している。エンドユーザは、利用可能な選択肢から配信物をその場で変換できるモジュールを選択する。複合型オブジェクトでは特別な問題があり、エンドユーザサービスはオブジェクト全体を構成するために様々なサブコンポーネントを適切な順番で受け入れ、加工する必要がある（これを実現するために、ISM CPではマニフェストファイルが設計されている）。そのような状況においては、初めに様々なコンポーネントをまとめてアグリゲートすることがこの処理を助けることになる。これに代わる方法としては、分散して存在するサブコンポーネント間の関連を関係（relationship）サービスが常に把握し、これらに関するその後の作業を容易にする方法が挙げられる。両者とも、これを行う能力と妥当性は、初めにサブコンポーネントを分散させた理由と目的により影響を受けるだろう。

### A3.3.6 プッシュと通知

情報をリポジトリやアグリゲータからエンドユーザにプッシュできると、公開されているものを情報源側で制御したり、構造化の程度を決めたりすることができるようになる。エンドユーザ側もプッシュされた情報を受け取るか否かを決定する余地があり、これにより自ら制御している感じを持つことができる。ただし、そうすることにより、リポジトリがエンドユーザのためにパッケージ化したものを受け取ることになる。したがって、情報のプッシュは、公開するものを決定できるという意味でリポジトリやアグリゲータにとって利益になる。また、既知の情報源から価値ある通知を受け取れるという意味でエンドユーザにとっても利益になる。通知の受信は、情報の配信方法としてまったく好みの問題であるが、DSpaceサイトで検索・ブラウズに次ぐ<sup>2</sup>番目の位置を占めるなど非常に人気のあることが証明されている。エンドユーザが通知を開始するので、技術的に言えば、これはプッシュ技術ではない。しかし、エンドユーザレベルでは通常、情報がプッシュされるものであると考えられている。CORDRAモデル<sup>53</sup>では、リポジトリを公開するために情報を適当なレジストリにプッシュするものであると論じているが、プッシュであれ、プルであれ、ハーベストや同等な機構を通じたこの環境における最適な処理は、依然として確立されていない。

プッシュ技術は実装のための作業をリポジトリに要求するが、この技術を導入することにより付加価値が得られる。たとえば、RSSフィードを部局のWebページに埋め込むことにより、機関にある情報やリポジトリの存在を目立たせることができる。RSSをこの目的で使用することはSDIサービスが図書館の提供する定番サービスであった時代を思い出させる。そのようなフィードは一般的なものでもよいが、関連するメタデータが利用できれば対象となる利用者のニーズに合わせた主題ベースのものであることが望ましい。

### A3.3.7 リクエスト

必要なコンテンツを発見したが表示や配信ができない場合は、別の場所にコンテンツのコピーをリクエストする機会をエンドユーザに提供すると便利である。オープンアクセスの環境では、うまくいけばフルコンテンツが自由に利用できるの、これは必要でないかもしれない。しかし、そうではない場合もあるだろう。Stevan Harnadは、利用制限がある場合にe-プリントのコピーを著者にリクエストするという考えを推進しており、最近、この機能がEPrints<sup>54</sup>とDSpace<sup>55</sup>に実装された。また、エンドユーザが別の機関に所属する場合は、リポジトリの管理者にリクエストを送る必要があるかもしれない。これらのリクエストはリポジトリ環境向けのものではあるが、従来のILLリクエストと同じようなものである。

### A3.3.8 入手とリンク付け

リンク付けはフルコンテンツの場所を容易に位置づけるための機能であり、主に検索やブラウズを行い、その結果を使って行うものである。これは発見されたメタデータからリポジトリにあるフルコンテンツへのリンク付けを意味する場合がある。ePrints UKサービスもARROW Discovery Serviceもこのアプローチをとっている。また、メタデータからどこか別の場所で利用できるフルコンテンツへのリンク付けを意味する場合もある。たとえば、メタデータから、図書館が購読することにより利用できるコンテンツや別Web

のサイトで利用できるフルコンテンツへのリンク付けである。オープンアクセスの環境では、それがどこにあると何が利用できるかがわかることが重要である。

これらのリンクは直接作成することもできるが、OpenURL<sup>56</sup>を使って文脈依存の形で作成することもできる。この場合、他のリソースを通じてさらなる発見を容易にすることができる。サウサンプトン大学で行われたリポジトリコンテンツにOpenURLを作成する研究は、これが必ずしも容易ではないことを示した。必要とされる情報の粒度が常に利用できるわけではないからである。しかし、グラスゴー大学は、自身の図書館目録からリポジトリへのOpenURLリンクを作成しているし、CERNはCDSwareリポジトリから他のリソースへのリンクにOpenURLを使っている。アグリゲータもOpenURLを利用している。OAsterは、完全準拠ではないが、OpenURLからOAsterを検索することを可能にする技術を提供している。また、METALISサービスプロバイダは、アグリゲーションデータに基づいたOpenURLを生成して、他のサービスにリンクしている。OpenURL 1.0 Z39.88-2004標準は、書誌的資料に限定されていた以前の0.1版に比べて大きな柔軟性を提供しており、まだ研究はされていないが、様々なコンテンツ種別をリンクする数多くの可能性を提供している。

#### A3.3.9 コメント、アノテーション、タグ付けと編集、更新

Amazonブックレビューやランク付けなどが存在する現在では、同様な機能が他の場所でも要求されることが多くなっている。同様な機能をリポジトリで提供すれば、コミュニティで情報を交換したり、見解や意見をお互いに提供したりすることができるようになるだろう。そのような機能を個々のリポジトリレベルで提供することは不可能かもしれない。また、アノテーションは、むしろ、ネットワークレベルやアグリゲーションレベルで提供することができるサービスだと思われる。JISC 5/99 Learning & Teaching ProgrammeのRESULTSプロジェクトは、エンドユーザが作成した教材のリンク集に基づくこの種の機能を推進した。自分で選んだキーワードによるタグ付けやフォークソノミーアプローチの開発は、エンドユーザが作成したコンテンツを既存のリソースに追加するもう一つの方法である。

情報をリソースに付加することはコンテンツ作成の一種である。編集や更新もそうであるが、セキュリティ上より厳重な認証と、引き続き行われる編集により作成される様々な版を追跡するための変更追跡記録を必要とする。このレベルの機能は、これらの必要な許可を行うことができる個々のリポジトリレベルで行うことがより妥当である。

#### A3.3.10 テキストマイニング

テキストマイニングは、利用可能なデータから概念と解釈を導くことができる分析技法を適用することにより情報資源を深く探索することを可能にする。利用できる情報の粒度が高く、利用できる情報が多いほど、探索できるレベルは高くなる。したがって、テキストマイニングはアグリゲーションレベルでより良く実行することができる。マンチェスター大学にある国立テキストマイニングセンター<sup>57</sup>は、これまでは入手が容易であったので抄録を使用してきた。オープンアクセス環境では、分析のために利用できるメタデータやコンテンツを開拓する大きな可能性が存在する。リポジトリを横断してテキストマイニングを直接行うことも可能であるかもしれない。ただし、コンテンツのフィルタリングが可能でないと、非常に多くのノイズ要因が存在することになるだろう。

テキストマイニングでは、抽出された情報の著作権保持者を確定することが重要である。これは新しい情報だと考えることができるからである。これはオリジナル文書がオープンアクセスで利用可能な場合にも当てはまるかもしれないが、これについては検証が必要であろう。

#### A3.3.11 保存

保存は単なる<sup>1</sup>機能ではない。該当資料の保存を最終目的とする一連の機能の集まりである。保存に関する個々の機能は保存要件を満たすために必要な作業に関係する。開放型アーカイブ情報システム(OAIS)の参照モデル<sup>58</sup>には、探索と検索は存在しないが、リポジトリはDIP(配信情報パッケージ)へのリクエストを受け付けることができる。これは、保存プロセスの一部としてリポジトリ間でDIPを移動することに関係する。このDIPの交換はプッシュ技術やプル技術で行うことができるだろう。保存機能もまた個々のリポジトリレベルがよりふさわしいものであるが、リポジトリ間でのDIPの移動をサードパーティが調整する余地が存在する。

#### A3.3.12 その他

既に述べた以外にも、リポジトリ環境における様々な機能を考えることができる。

- エンドユーザが研究者や学生などではなく管理者の場合は、リポジトリやアグリゲータの利用に関する統計データを収集できる能力には価値がある。アイテムが受けた引用の数やフルコンテンツのダウンロード回数は個々のリポジトリにとって特に参考になる。一方、アグリゲータは、クリックしてデータ元のリポジトリを訪れた回数やアグリゲートした情報をアクセスするためにエンドユーザが利用している技術的ルートを知りたいと考えている。EPrintsとDSpaceは共に統計機能のアドオンを持っている。
- e-サイエンスの世界では、クエリをリモートのデータセットに対して発行することができる。これは通常、機関内、あるいは機関を横断するバーチャル組織を通じた協力の結果である。クエリが発行されると適当なグリッド資源でジョブが実行される。結果はある時間が経ってから返され、オープンアクセスやその他の配信や分析のために必要に応じて様々な環境や方法で出力することができる。

### A3.4 エンドユーザサービスの提供レベル

エンドユーザサービスの提供には様々なレベルが考えられる。サービスの提供は個人レベルから世界レベルにまで及ぶ。このレベルの違いは必ずしもサービスの違いを必要としないが、ユーザーのニーズを満たすためにリポジトリのビュー (view) に違いが現れるかもしれない。オープンアクセスリポジトリ横断型のエンドユーザサービスを提供するために最もふさわしいレベルは今のところ明確ではないが、各レベルのサービスがどのように稼動しているかについての経験は増加している。

リポジトリを使ったエンドユーザサービスの構築は、リポジトリの設立母体の単位に影響を受ける。リポジトリの多くは機関や主題レベルで構築されている。それゆえ、このレベルにおける初期画面の作成は比較的容易であり、また、様々なユーザグループやRAEなど様々な目的に応じて、幅広いニーズに対応することができる。リポジトリの開始画面から、リポジトリのコンテンツを使ったもっと詳細なビューを作成することができる。たとえば、リポジトリに対する個人や部局の貢献を示す画面の提供などである。MITでは、研究者がまず、リポジトリにコンテンツをデポジットし、次に、そのメタデータを抽出して個人や部局のWebページでこれを提供できるようにした<sup>59</sup>。ロチェスター大学では (Foster, 2005) インターフェースを作成してリポジトリ上に個人用画面を提供したが、このような方法でもこの種のサービスを提供することができる。また、RSSも同様な対象を絞ったビューを提供できる。

アグリゲーションを通じて、リポジトリを横断するより広いビューを作成することができる。たとえば、ePrints UKは英国のe-プリントに対する全国的ビューの提供をめざし、IRIScotland<sup>60</sup>は、そのサブセットとして、スコットランドを対象に同じことをめざした。DAREnetはオランダにおいて同じことを行っている。これら3つのプロジェクトは、アグリゲーションにおいて包括的な方針を採用し、まったくフィルタリングすることなく、利用可能なすべてのリポジトリを対象とするビューを提供している。オランダのリポジトリを対象とする別のビューを提供するためにDAREnetのサブセットが作成されている。オランダの先導的研究者が作成したコンテンツを対象を絞っているCream of Scienceエンドユーザサービスである。国家的イニシアティブはリポジトリコンテンツの収集やアグリゲーションを行うための便利な組織体制を提供している。

この全国的、あるいは地域的ビューは、そこで行われているものを発見する、また、ある国で行われている研究を奨励する——DAREnetやIRIScotlandの背後にある要因である——という意味で利点を持つだろう。しかし、オープンアクセスの世界では、地理的境界には限界があるだろう。世界的規模の研究コミュニティは通常、境界とは関係なしに研究が行われているところを見ており、これを利用できるようにするためには、より幅広いアグリゲーションやエンドユーザサービスが必要とされる。OAIsterやARCは世界規模のビューを採用しており、これがその広い知名度や利用を説明しているかもしれない。メロン財団の助成を受けたその他のOAIサービスプロバイダプロジェクトが、おそらく特定のリポジトリを対象を絞ったビューを提供したために、それほど広く知られていないことは注目に値する。これは、プロジェクトの対象コミュニティに対する有用性や価値を傷つけるものではない。重要なのは世界規模のビューであり、必ずしも世界規模の利用者ではないからである。Googleも世界規模のビューを採用し、非常に評判が高いことを証明した。世界規模のビューから始めて、必要に応じてフィルタリングすることが、エンドユーザサービスに最も高い柔軟性を提供するが、フィルタリングのレベルはメタデータの粒度に依存することになる。

必要に応じてエンドユーザサービスを提供するためには柔軟性が鍵となる。機関リポジトリを動的に結合する機能は、エンドユーザが自分でアグリゲーションを構築するチャンスを提供するが、これはアグリゲーションを行うための技術に影響を与える。RSSはこの方法で容易に結合することができるが、OAI-PMHは多くの設定を必要とするからである。エンドユーザサービスは、エンドユーザがコンテンツと相互作用するところで成功すると思われる。したがって、様々なユーザグループがオープンアクセス資料にアクセスする必要があるのはどこかという質問に答えることは、どんなエンドユーザサービスを開発する必要があるかについて情報を提供することになる。エンドユーザサービスを構築する際にサービスの基準を設定しておけば、その価値と利点を後で評価する際に役に立つだろう。

多くの場合、必要とされる相互作用は主題レベルのものである。このレベルのエンドユーザサービスの提供は本質的に難しいものであるが、主題ベースのエンドユーザサービスへの要望は高い。教育の観点からは、講師は自らの専門分野に沿って考えるので、主題ベースのビューでコンテンツが利用できることを好むだろう。利用を促進するためには、機関や地理上の境界があるリポジトリやアグリゲーションを主題ベースのインターフェースの背後に隠す必要があるかもしれない。しかし、これは、これを可能にするために必要なメタデータのレベルや詳細さに大きく関係する。このアプローチはePrints UKで最初に検討され、e-プリントに主題メタデータを追加することにより、RDNのハブを通じて主題ベースのビューを提供することを試みた。ベータテストまで行われたが、当時はこれをさらに進めるほど技術が成熟していなかった。しかし、アプローチは検証済みであるので、主題分類の問題のソリューションが見つかりさえすれば、今でも行う価値がある方法であると考えられる。

リポジトリ間で何らかの調整や合意がなされており、NEREUSコンソーシアム<sup>61</sup>のように共通の主題分類体系が存在すれば、主題に基づく（NEREUSの場合は経済学分野）ビューを提供する可能性がある。OAIのセットはリポジトリのコンテンツを主題により分類する方法を提供する可能性があるが、それには、セットと合意された主題語彙の両者の幅広い採用と応用が必要になるだろう。あらゆるもののクローラでGoogleが採用しているフリーテキストアプローチは、検索の際ヒットすることもしないこともあるが、開始点として価値があることが証明されている。Google Baseは、この「リポジトリ」にレコードをアップロードする際に構造化された記述を追加する機会を提供している。Google Baseはタギングアプローチ、あるいはフォークソノミーアプローチを採用している。このアプローチは、主題ベースのビューでうまく利用できないかさらに調査する価値のあるものである。

ドイツのビールフェルト大学のNorbert Lossauはリポジトリのサービスレベルの分類を作成した。これはe-サイエンス用に設計されたものであるが、オープンアクセス環境で研究を行う研究者がリポジトリを利用したいと考えるレベルを明らかにしている。これはおそらくアグリゲーションにも適用できるだろう。

- 情報の入手・発見・検索・アクセス・ナビゲーション
- 情報の管理（個人的・組織的、公開・非公開）
- 情報の取り扱い（アノテーション、操作）
- 学術コミュニケーション（学術成果の創造による連続したプロセス）
- 出版（最終地点）

リポジトリコンテンツに文脈依存リンクを適用することにより、あらゆるレベルのサービスの粒度をより詳細にすることが可能である。エンドユーザサービスがユーザーが誰であるかを知っていれば、たとえば、OpenURLのContextObjectsを実装技術として使うことによりフィルタリングを行うことができ、当該ユーザーのコンテンツへのアクセスの可否や、利用できるコンテンツのバージョンや範囲を決定することができる（Blanchi, 2006）。このプロセスは、図書館の参考カウンターで教授と学部1年生が同じ質問をすると違う回答が返ってくることに同じである。

## A4 メタデータ

前章で説明したサービスは程度の大小はあれ、すべてメタデータを使用している。Googleでさえ、提供するサービスのためにハーベストしたフルテキストからメタデータを抽出している。メタデータの影響を調査することなしにサービスの検討を行うことは不可能である。

### A4.1 メタデータ標準

利用できるメタデータ標準の種類は「標準について素晴らしいことは選択肢が非常に多いことだ」という

格言をまさに表している。各標準は独自の目的と起源を持っており、各々を差別化している。採用する標準を決める際には、メタデータを使用する目的が、コンテンツの組織化である場合の要件と、メタデータとそれが記述するコンテンツへのアクセスを可能にすることである場合の要件との間のバランスを考慮することが重要である。メタデータには、記述メタデータ、技術メタデータ、管理メタデータ、権利メタデータ、保存メタデータの区別があり、各メタデータはそれにふさわしい異なる標準を持っている。表<sup>3</sup>は、どれほど多くの標準があるかを感じられるように、現時点で利用可能な多くの（主に記述メタデータ用の）標準の一部をまとめたものである。これほど多くの標準があるにもかかわらず、すべてのコンテンツを完全かつ正確に記述するには既存の標準だけでは不十分である（たとえば、地理空間データにはISO19115があるものの、データセットを記述するための標準は現在のところ限られている）。しかし、現在ではこれらを作成する際に利用できる様々な分野における多くの経験が存在している。

メタデータ標準	本来の目的	備考
Dublin Core <sup>62</sup>	相互運用可能な検索と発見の支援、および、複数のコンテンツ種別の簡単な記述子付き)	最大公約数的なアプローチはリポジトリに相互運用性を与えることができる。ただし、より情報量のあるメタデータからマッピングした場合、情報のロスが生じる可能性がある。
MARC <sup>63</sup>	書誌リソースの記述	物理的アイテムの記述のために設計されたが、デジタルリソースの記述にも何とか利用することが可能である。また、本来、アクセスではなく記録の交換を意図したものである。
MODS <sup>64</sup>	書誌リソースの記述	MARCを改良・拡張して、形式的なMARCタグとXMLに基づくより柔軟な記述を組み合わせることを可能にしたもの。
VRA Core <sup>65</sup>	ビジュアル文化の作品および関連する画像の記述	ビジュアルコレクションの記述を支援するための開始点となることを意図した標準であるが、完全な解ではない。
MIX <sup>66</sup>	静止画像のための技術情報	デジタル画像コレクションの管理を支援するために現在標準化作業中である。
EAD <sup>67</sup>	アーカイブの発見ツールの記述	アーカイブや博物館で作成される所蔵品目録や索引、関連ツールの記述を可能にするために設計された。
LOM <sup>68</sup>	学習オブジェクトの記述と文脈説明	オブジェクトに関する記述情報および管理情報を包含する詳細な標準。英国版であるUK LOM Core <sup>69</sup> はこれから派生したものである。
TEI <sup>70</sup>	電子テキストの表現	研究や教育における電子テキストの使用を支援するために使用される。メタデータだけでなく文書の完全な表現を提供することを目指している。
CDWA <sup>71</sup>	アート・カルチャーやマテリアル・カルチャーの作品のためのコアレコードの記述。	アート作品を記述するためのカテゴリーの外延的セット。相互運用性を目的としたCDWA-Liteも存在する。
ISO 19115 <sup>72</sup>	地理空間データセットの記述	英国における学術的利用のためにEDINAにより、HE/FEプロファイルが策定された <sup>73</sup> 。幅広く利用されていたDGDC標準 <sup>74</sup> のISO 19115へのマッピングが作業中である。
ETD-MS <sup>75</sup>	電子学位論文の記述	NDLTD <sup>76</sup> により作成された国際標準である。電子学位論文のための英国のメタデータセット <sup>77</sup> がElectronic Thesesプロジェクトで策定されている。

表 3: 現在利用できるメタデータ標準の例

メタデータ標準を選択する際には、それが目的にあってることが重要である。これは、対象となるコンテンツを記述する能力が要件——数多く存在するが中でもアクセスや保存、管理など——を満たすことを意味するだろう。また、標準は透過的であり、かつ、相互運用や共有、支援をより簡単にするためにできるだけ広く採用されているべきである。多くのメタデータ標準は、エンドユーザのアクセスや関連するサービス機能の要件以上にコンテンツの管理を助長する傾向にあることは注目に値する。使用するメタデータ標準の選択による直接的な利益と目的はコンテンツ管理に対するもので十分であろうが、アクセス

や保存を将来にわたって支援するというコンテンツを記述するより大きな目的を検討する価値がある。複数のコミュニティにとって価値があるコンテンツについては、特にそうである。コミュニティによってはメタデータの異なる側面を利用してコンテンツにアクセスしたいと考える場合があるからである。目的が違えば必要な標準も違うかもしれない（たとえば、記述メタデータに保存メタデータを追加したり、異なる目的のために2つの記述レコードを組み合わせるなど）。このように、システムは複数の標準に対応し、これらを管理できるようにしておく必要がある。

#### A4.2 メタデータのリッチさ

使用するメタデータ標準を決定する際には、収集するメタデータの種類とその理由を考える必要がある。エンドユーザサービスで使用するようにメタデータを提供する場合、メタデータがどのように使用されるかを正確に予想することは困難である。できるだけリッチなメタデータセットを提供することにより、利用できる選択肢の幅や柔軟性を増すことができる。このリッチなメタデータは単一のメタデータレコードから得られる場合もあるし、各々別の目的で作成された数多くの関連するメタデータレコードとして公開される場合もある。後者のアプローチはCORDRAプロジェクト<sup>78</sup>で研究されている。このプロジェクトはエンドユーザサービスがメタデータを利用する能力や選択肢を最大にするためにできるだけ多くのメタデータを公開することを目的としている。

このようなアプローチは、シンプルダブリンコアメタデータ標準に従うことにより、リポジトリに相互運用性を与えるメタデータを作成したいという願いと矛盾すると思われるかもしれない。この標準は、検索と発見を容易にすることを目的として設計されたものであり、比較的限られた数のフィールドを提供することにより、多くの様々な分野や機関が使用しているメタデータの違いを隠し、これらのリポジトリを横断するアクセスを可能としている。しかし、ダブリンコアは記述目的で使用するには限界があり、ダブリンコアメタデータをもっとリッチなメタデータから抽出する際には、ほとんど常に何らかの情報のロスが発生することになる。検索や発見のためにダブリンコアにマッピングできるリッチなメタデータベースレコードをリポジトリが持つことができれば、エンドユーザに明確な入り口を示すことにより、D2D（配信から発見）の連鎖の一部として、ブラウザや所在特定などのさらなるエンドユーザサービスのためにオリジナルのメタデータを引き続き使用することができる。このリッチなメタデータレコードを使用して、リポジトリレベルでより詳細な情報や機能を提供するためには、発見結果にメタデータを提供したリポジトリへのリンクを表示する必要がある。

#### A4.3 メタデータの粒度

可能な限りリッチなメタデータを保管する事と同時に、できるだけ細かい粒度でメタデータとコンテンツを保管することも重要である。エンドユーザサービスにコンテンツやメタデータを公開すると、サービスは与えられたものを利用することになる。粒度が荒いと、サービスは個々のオブジェクトをそのまま扱うしか方法がない。粒度が細かいと、場合によっては、エンドユーザサービスでオブジェクトを分解して、これをエンドユーザに付加価値を与えられるように再構成することで、提供されたものをすべて利用したり、その一部のみを利用したりすることができる可能性がある。様々なコンテンツ種別とそのメタデータを扱う方法を定義しているコンテンツのモデル化は、リポジトリやそのコンテンツに最もふさわしい粒度の決定に役立つだろう。

#### A4.4 アプリケーションプロファイル

シンプルなメタデータとリッチなメタデータの両レベルにおいて、コミュニティはメタデータを使用、または応用するための独自の要件を持つことになる。リッチなメタデータレコードとそのダブリンコアへのマッピングとの間の関係（および、その他のメタデータ標準との関係）を規定する特定の応用や機能、コミュニティ、環境に対応するアプリケーションプロファイルの使用は、これら特定のニーズを対象とするエンドユーザサービスを支援することになる（[Heery, 2000](#)）。アプリケーションプロファイルはローカルの応用から作成される場合もあれば、正式な手続きの結果として作成される場合もある。どちらの場合も、適用するメタデータ標準を選択するための情報を得るにはこれらのプロファイルについて知っていることが役に立つ。プロファイルの詳細を提供するInformation Environment Metadata Schema Registry (IEMSR)<sup>79</sup>などのレジストリは、その元となるメタデータ標準に関する情報を提供すると同時に、コミュニティが適切なプロファイルを採用したり、標準化したりする際の参考になるだろう。

#### A4.5 メタデータのマッピング

多くのリッチなメタデータ標準は公式化されたダブリンコアへのマッピングを持っている (MARCやMODS<sup>80</sup>など)。これは相互運用性への要求から作成されたものである。シンプルダブリンコアは、最低限の要件としての必須に過ぎないが、OAI-PMHにおいて最も広く使用されているメタデータ標準である。既に触れたように、このようなマッピングは情報のロスをもたらす場合がある。よりリッチなレコードにアクセスすることにより、D2D連鎖のさらなる破れを軽減する可能性がある。マッピングは、メタデータの管理とエンドユーザサービスへの提供における別の側面を助長するのにも役立つ。

- リポジトリが内部で使用するメタデータと外部世界に公開するメタデータは同じである必要はない。Fedoraリポジトリシステムは内部的にはFOXML (Fedora Object XML) <sup>81</sup>を使用しているが、外部への公開やインポート/エクスポートには、ダブリンコアやMETSを使ったパッケージ化されたメタデータなど、別のメタデータ標準にマッピングしている (パッケージ化についての詳細はA4.12.1を参照)。この柔軟性により、リポジトリは目的に最もふさわしいメタデータ標準を間違いなく使うことができ、必要に応じてこれを変えることもできる。
- ベースレコードとしてリッチなレコードを持つことにより、WORM (ライトワンス・リードメニー) パラダイムを適用することができる。すなわち、他のシステムへのメタデータの提供など、必要に応じてベースレコードから派生物を生成し、特定の目的のために使用することが可能となる。

メタデータのマッピングはメタデータクロスウォークとしても知られている。アプリケーションプロファイルに関しては、クロスウォークの詳細を記録し共通に参照できる場を持つことはコミュニティにとっての貴重な資産になると思われる。OCLCはそのような目的でレジストリを構築している (Godby, 2004)。ただし、クロスウォークはメタデータの移行と交換を容易にするとはいえ、間違っただけのメタデータ自体を修正することはできないことに注意すべきである。

マッピングやクロスウォークは柔軟性を提供するが、この点において、もし選択肢が1つだけでない場合は、メタデータ標準とアプリケーションプロファイルの限定セットを標準化して公開するようコミュニティに奨励することが現実的である。これは、共通の応用を通じてメタデータの一貫性を高めることになり、潜在的には、より詳細で一貫性のある情報に基づいたエンドユーザサービスを可能にする力を持っている。

#### A4.6 メタデータのシンタックス

メタデータは記録される情報のシンタックスを提供する。しかしながら、メタデータ標準はメタデータレコード全体におけるメタデータ要素間の関係を定義することはあまり得意ではない。パッケージング標準 (A4.12.1節参照) は、メタデータと共に時にはコンテンツ自体を構造化された形で伝達する付加的なシンタックスを提供することができる。XMLによるパッケージング標準であるMPEG-21 DIDLは、そこに含まれるメタデータシンタックスを定式化する方法を提供している。すなわち、この標準はコンテンツを含めることができ、コンテンツと関連するメタデータを同時に含める構造を提供することができる標準である (Bekaert, 2003)。これに代わるものとして、RDFとWebオントロジ言語 (OWL) <sup>82</sup>は、メタデータとコンテンツについて、これらは何か、どのように使うことができるか、使うべきか、といった意味論的な説明を行う可能性を提供すると同時に、詳細なシンタックスを提供することもできる。両者とも、そのようなシンタックスの導入を可能としているのは抽象化モデルの存在である。既に説明したコンテンツモデルを含む抽象化モデルは収集されるべきメタデータを決定するのに役立つ。また、エンドユーザの開発も可能にする。基礎となるモデルに基づいてサービスを構築することにより、実際の様々なメタデータやコンテンツをより効率的に管理することが可能になるからである。このアプローチはMITのSIMILE RDFプロジェクトで実現されている。このプロジェクトは、様々なメタデータスキームを同等に組み合わせるためにRDFも使用している<sup>83</sup>。4.13節で示したように、これらのクロスウォークはメタデータ標準間の効率的な移行を可能にする。RDFが提供する抽象化は複雑であるが、大きな柔軟性を提供する可能性を持っている。従って、その可能性を追求すべきである。

##### A4.6.1 Topic Maps

Topic Maps<sup>84</sup>はメタデータを構造化する方法としてRDFに代わるものであり、その使用法はRDFより簡単である。あるレベルでは、(オーストラリア国立図書館で使用されているように) 複数の分散したサイトを統一されたWebフロントエンドにまとめるWebポータルインフラとして機能する。別のレベルでは、複数のメタデータレコードを同時に記述することを可能にし、コンテンツ要素間のつながりと関係を示す

共通項によりメタデータ間にリンクを作成することを可能にする。Topic Mapsは効率的に機能するためには高いメタデータ品質を必要とする。また、レコードを正確にリンクするためにユニークな識別子の使用を必要とする（これはTopic Mapsを使用しなくても、関係を成立させる際には重要な機能である）。また、Topic Mapsはその基礎となるオントロジを必要とし、その周囲に結果として生ずるリソースを取り付けるための構造を追加する。Topic Mapsの使用法に関してはまだ理解されていないことが多い。RDFアプローチと対比するために、Topic Mapsの研究とさらなるオントロジの構築が必要である。それには英国以外での経験、特にノルウェイ教育省の経験を手がかりにすることができる。

#### A4.7 メタデータの自動生成

エンドユーザサービスとTopic Mapsなどの構造化標準の基礎となるリッチなメタデータレコードを作成することは決して簡単なことではない。必要なすべてのメタデータの作成をコンテンツ投稿者に期待することは無理であり、そのための人的資源をリポジトリ管理者が持っているとも、特にリポジトリが成功しているとしたら、思えない。（Arts & Humanities Data Service (AHDS) とEconomic and Social Data Service (ESDS)で行ったように）承認されたワークフローに従って異なる関係者、たとえば、投稿者とリポジトリ管理者の間で作業を分けることが、手作業によるメタデータの作成処理を向上させる一つの方法である。しかし、その経験が示しているように、これが必ずしも常に最も効率的な方法というわけではない。

メタデータの自動生成は、これら手作業の不足を解決する方法を追求したものである。これを可能にする技術は対象となるメタデータの種類により異なる。デジタルオブジェクトの技術メタデータは、ハーバード大学のJHOVEサービス<sup>85</sup>や英国国立アーカイブのDROIDサービス<sup>86</sup>などの適切なツールを使うことで抽出することが可能である。この技術メタデータは部分的に保存目的で使用することができるが、ニュージーランド国立図書館が開発したツールもファイルヘッダから保存メタデータを抽出することを目的としている<sup>87</sup>。

コンテンツから記述メタデータを抽出するツールは存在する（Kea<sup>88</sup>など）が、現在のところ、その処理は他の種類のメタデータほど簡単ではない。自然言語処理（NLP）ツールは、まったくの分野限定であるか、抽出すべき語を学習するために一連の文書を必要とするかのいずれかである。このような処理は生物医学分野では定着してきたが、その他の分野では行われていない。NLPツールのトレーニングは現在のところ簡単な作業ではなく、ほとんどの機関では実行することができないだろう。関連キーワードの抽出を学習ではなく統計を利用して行うNLPツールは正確さには劣るが、記述メタデータの自動生成を実現するためのより現実的な方法を提供するだろう。

NLPはテキストマイニングを構成する様々なサービスの一部である。国立テキストマイニングセンター（NaCTeM）は、この機能を補助する様々なツールを提供している<sup>89</sup>。ただし、テキストマイニングも簡単な処理ではなく、現在のところリポジトリのワークフローに簡単に位置づけることはできないだろう。そうではあるが、テキストマイニング技術を個々のソースではなく一連のリポジトリに適用し、大規模なメタデータの生成に役立てることは実現可能であるかもしれない。テキストマイニングには考慮しなければならない独自の問題が存在する。たとえば、テキストマイニングにより生成されたメタデータに関連する権利（権利の所有者は誰か）やテキストマイニングを実行するためには（現在のところ、ほとんどのテキストマイニングはフルテキストではなく抄録を使って実行されているが）リポジトリのフルコンテンツを集める必要があることなどである。

メタデータの自動生成は図書館界で大きな関心を引いている。報告書「カリフォルニア大学における書誌サービスの提供方法を再考する」には、できればメタデータの自動生成に移行したい旨の明確な希望を示す文章が数多く存在する（[University of California Libraries Bibliographic Services Task Force, 2005](#)）。これには、他の場所に存在するメタデータを集めることも含まれている。これは（たとえば、書誌レコードをCURLやRLGから調達するなど）英国では数年前から普通に行われてきたことであるが、格納すべきコンテンツが別の場所にも存在するリポジトリを考えると検討する価値のあるアプローチである。他のプロセスの一部として生成されるようなメタデータは存在するだろうか。ローカルのユニークな資料を保管するリポジトリにとってこれは考えられないことであるが、たとえば、出版済みの雑誌論文に相当するe-プリントの場合、再利用可能なメタデータレコードが二次情報プロバイダにより生成されている可能性がある。

#### A4.8 メタデータの品質

究極的には、リポジトリに入力されるメタデータの内容と品質が高ければ高いほど、エンドユーザサービスがこれを効率的に利用することが簡単になり、公開されるコンテンツをエンドユーザが利用できる確率も高くなる。入力されるべきメタデータを必須化することがこの問題を解決するかどうかを考えることは有益であろう。実際、これは内閣府電子政府部による統合公共部門語彙の適用において採用されたアプローチである。しかしながら、学術の世界ではこれが成功するとは思われない。品質の問題に取り組むことができる<sup>3</sup>つのレベルが存在する。

- リポジトリレベル: メタデータ標準の実装のためのガイドラインは、選択されたメタデータ標準の使用を助長するという意味ですべてのコミュニティにとって役に立つ。ほとんどの標準は汎用的なガイドラインを伴っているが、その応用に対して抽象的に過ぎる場合がある。たとえば、ePrints UKプロジェクトが作成した文書「e-プリントの記述にダブリンコアを使用する」(Powell, 2003)や「e-プリント詳細表示ページにフルテキストリンクを設ける」(Tourte, 2004)のような特定の目的のための具体的なガイドラインが非常に役に立つだろう。このようなガイドラインは特にダブリンコアにおいて役に立つ。なぜなら、この標準は各フィールドに入力される内容に関して比較的緩やかであり、たとえば、広く採用され使用されている制限語彙がなく、そのため、大きな変異をもたらすからである。幅広い報知と採用はガイドラインの作成に伴い発生する問題である。しかし、このようなガイドラインは特定の目的のために作成されるメタデータのある程度一致させる可能性を提供するので奨励する必要がある。A4.4節で述べた合意されたアプリケーションプロファイルの順守はこの道に沿った第一歩である。ガイダンスは、対象となるニーズに最も関連するサービスのレベルにより(ハーベスト目的でARROWプロジェクトの一環としてオーストラリア国立図書館が推進したように(Campbell, 2005))国レベルで提供することもできるし、(メタデータの交換のためにOLACコミュニティが独自のダブリンコアプロファイルを作成した<sup>90</sup>ように)コミュニティレベルで提供することもできる。
- アグリゲータあるいはエンドユーザサービスレベル: アグリゲータやエンドユーザサービスが手元にメタデータを収集することができれば、エンドユーザにメタデータを提供する前に品質の問題を解決する機会を持つことになる。OAIsterサービスは、ハーベストしたメタデータに整合性を持たせるために正規表を広範囲に使用している(たとえば、フリーテキストによるアイテム種別を制限語彙によるアイテム種別にマッピングしている)。この整合性により、メタデータを使ったエンドユーザサービスの構築が可能になっている。OAIsterの例では、アイテム種別検索の提供である。整合性はサービスを提供する際の重要な機能である。Googleは独自の正規化を行い、そのままではまったく整合性のないクロール後のメタデータを解体して、独自のガイドラインに従ったメタデータを作成している。リポジトリが公開しているメタデータに多様性があるのは避けられないが、この多様性を制限するのではなく、アクセスを可能にするために管理する必要がある。
- 中間サービスレベル: メタデータの自動生成を可能にする可能性のあるルートについては4.7節で説明した。これは、どちらがシステムの適用が容易であるかにより、リポジトリレベルあるいはエンドユーザレベルで実装することができるだろう。テキストマイニングをリポジトリ横断的に適用することもできるだろうが、同様に、ハーベストしたメタデータを持つエンドユーザサービスも、同じ目的でフルコンテンツへのリンクにアクセスすることができるかもしれない。後者のアプローチについては、ePrints UKプロジェクトにおいてWebサービスを使って主題メタデータや典拠コントロールされた著者メタデータを付与するテストが行われた。正規化サービスもリポジトリやエンドユーザサービスレベルで適用される中間サービスであると考えることができる。

メタデータ品質の向上はそれ独自の問題を引き起こす可能性がある。さらに詳細なデータを追加してメタデータの品質を高めると、データが詳細すぎることにより生じる可能性のある相互運用性の低下をきたす危険がある。しかし、品質の高いメタデータレコードは相互運用性をもたらす方法に選択肢を提供する。相互運用性を増すためにいくらかレベルを下げる必要があるとしても、リッチなメタデータがあれば情報連鎖のさらに下流で付加価値を提供することになる。

#### A4.9 主題分類

主題ベースのエンドユーザサービスは非常に重要であると考えられている。多くのエンドユーザは主題を通じてコンテンツを利用したいと考えるからである。この事実は主題分類の必要性を高める。主題語は、目録作業の一環として手作業で追加することができる。その一例はグラスゴー大学であり、一貫性を保つためにリポジトリと図書館目録で共に議会図書館件名標目を使用している。形式的な主題分類の代替となるのは、非形式的なタグの使用である。これはコンテンツの投稿者により付与されるもので、エジンバラ大学のERAリポジトリで既に実行されているが、2002年のRESULTSプロジェクトで提案されたものである。コンテンツのタグ付けは、現在、Web上の様々な「Web 2.0」サービス(Flickrなど)で広く受け入れ

られている機能である。形式的・非形式的の両者を持つことは有益であり、メタデータを様々な方法で提供する可能性を開くものである。両形式を組み合わせることも、現在ノルウェイでテストされているように、Topic Mapsを使うことにより可能である。

主題ベースのアクセスの要望は、リポジトリ横断型の検索や発見を容易にするダブリンコアの最大公約数的なアプローチがこれらの要求を満たすには十分でないことを意味しているのかもしれない。本研究で行ったインタビューの意見では、シンプルダブリンコアはリポジトリの横断検索には不十分であり、主題分類のための大きな潜在能力を持つリッチなメタデータフォーマットがより適していることが示された。限定子付のダブリンコアを使用することもできるが、限定子は（アプリケーションプロファイルで規定されてはいるものの）必ずしも標準化されていないので、相互運用性を高めるのではなく減少させるという欠点を持つ可能性がある。

#### A4.9.1 OAIセット

主題分類を広範囲に適用できる<sup>1</sup>つの方法は、OAI-PMHによるハーベスティングで利用できるOAIセット<sup>91</sup>のメタデータへの適用である。2006年2月、SHERPAプロジェクトは、OpenDOAR (Directory of Open Access Repositories) <sup>92</sup>掲載の英国のリポジトリで現在利用できるセットの調査を行った。その結果、利用されてはいるものの予想していたよりも利用は少ないことが明らかになった。ほとんど主題分類をしていないリポジトリがある一方で、詳しくすぎるセットを提供しているリポジトリもあった。エンドユーザサービスにとって便利なものになるためには、何らかのOAIセットの広範囲にわたる採用についての共通の合意が必要となるが、これによりエンドユーザサービスは有用な主題ベースのアクセスを提供することが可能になる。

#### A4.10 メタデータの変更と追加

メタデータを作成する方法を検討する際には、メタデータの作成は必ずしも<sup>1</sup>回限りではなく、時間の経過と共に変更される場合があることを思い出すことが重要である。コンテンツはイベントや時間に関係する場合があり、必要に応じてイベントに関するメタデータをレコードに追加しなければならないことがある。これはアノテーションやメタデータ階層化の一形態であり、オリジナルレコードに付加価値を与え、データの長期的妥当性を高める。フィードバックやユーザーのコメント、ランク付け、サービス対象となる様々なコミュニティに特化した追加メタデータなど、その他のアノテーション機能も追加することができる。この種の機能を提供しているエンドユーザサービスはそのようなアノテーションがオリジナルメタデータレコードに明確に関連付けられていることを保証する必要がある。問題は、適切な識別子を使った明確な関連付けが必要となることである。これらのメタデータの変更は、版付けの問題や、あるレコードが単なる別バージョンではなくそれ自体新しいレコードになるのはいつなのかという問題を引き起こす。メタデータの変更管理は、現在のところ一般に開発が遅れているが、デジタルコンテンツや関連のメタデータの存在は当たり前になっているので開発を行う必要があるだろう。

#### A4.11 識別子

識別子はすべてのメタデータレコードに不可欠の要素であり、実際には関連するコンテンツ自体にも不可欠な要素である。これらの識別子は、長期にわたり価値を持ち利用できるようにするために以下のような鍵となる特徴を持つべきである。

- 識別子は永続的であるべきである。ただし、永続性とは必ずしも未来永劫を意味するのではなく、単に識別されるオブジェクトの存続期間を意味していることに注意するべきである。世界中に配達される小包で配送識別子が必要なのは配達されるまでであり、その後はもはや意味がない。しかし、他の多くの識別子は保存目的で使用するために何年にもわたって永続化される必要があるだろう。
- 識別子はユニークであるべきである。ユニークでない識別子を使うと重複によりエンドユーザを混乱させる危険性がある。よく知られている識別子体系の中には、たとえば、出版中止後ある期間が経ったISBNのように、識別子の再利用を行っているものがあり、そういう状況でも問題なく対処できることに注意するべきである。ただし、一意性の確保により、さらに下流で発生する可能性のある問題を避けることができる。

ユニークで永続的な識別子を持てば、メタデータやコンテンツを処理する際に一つの識別子を使いまわす

ことが可能になる。その識別子は関連する構成要素が存在する場所を常に知っているからである。この利点は、オブジェクトが複数の識別子を持つ（通常オブジェクトがリポジトリに受け入れられるたびに作成される）ことにより危うくなるかもしれないが、物理的なオブジェクトの世界ではこの問題は何年も前から存在し、それなりに対処されてきた。通常、識別子は各々特定の目的を持っているからである（たとえば、雑誌はタイトルにはISSNを、巻号にはSICIを持っている）。原則的に、各識別子は1つのものだけを識別すべきである。

#### A4.11.1 識別子の粒度

4.3節でメタデータの粒度に関して述べたように、メタデータやコンテンツは荒い粒度で保管するより細かい粒度で保管した方がエンドユーザーサービスにとっての利用価値ははるかに高くなる。これには数多くの利点が存在する。識別子がコンポーネントやサブコンポーネントに明確に関連付けられていることが重要である。各々を簡単に参照できるからである。粒度が細かいと、複合型オブジェクトを分解して、必要に応じて再構成することも可能である。各コンポーネントを識別することができれば、それらを統合することができるが、粒度が荒いと分解することがはるかに難しくなる。粒度の細かい識別子は関連するメタデータとコンテンツを明確に結合することを可能にするので、たとえば、エンドユーザはメタデータレコードがフルテキストに関連付けられているか否かをはっきりと知ることができる。細かい粒度により、コンテンツやメタデータと同じ方法でリポジトリに独自の識別子が付与されたと考えられるので、リポジトリを個々のオブジェクトと見なし、必要に応じて具体的に参照することが可能である。

#### A4.11.2 著者識別子

メタデータレコードやコンテンツの構成要素を識別する上で著者を一意に識別することは、まだ十分に理解はされていないが、レコードのあらゆる面を正確に識別するために検討を要するであろう重要な分野である。オランダのあるプロジェクトは、全国のすべての科学者にユニークなデジタル著者識別子を付与しようとしている。大きな国ではこの試みが実現できるとは思われず、著者名の典拠リストを維持することは簡単な作業ではないだろう。しかし、著者名典拠ファイルは完全にユニークな著者識別子を実現するために必要な第一歩であると考えられる。そのようなファイルを中央にまとめて持つこともできるが、各機関の職員録を使って著者を識別することにより、分散して持つこともできるだろう。サウサンプトン大学のリポジトリは、少なくともローカルレベルで著者を正確に識別するために職員IDデータベースへのリンクを持っている。同様な職員録の利用は別の機関でも検討されている。商用レベルでは、Scopusが著者による検索を可能にするためにScopus Author Identifier<sup>93</sup>を導入し、CSAは特定の著者の全研究成果へのアクセスを容易にするために著者にタグ付けすることを開始している。この問題の可能性を評価するためにさらなる研究が必要である。

#### A4.11.3 識別子の解決

識別子を付与する際に鍵となる問題は、識別子にその出自の場所を示すべきか、場所独立にするべきかという問題である。URLはドメイン名によりその出自を示すが、一意性と永続性のどちらも保証することができない（URLにあるコンテンツが変更されるかもしれないし、URLが再利用されるかもしれないからである）。理想的には、識別子は場所独立であるべきであり、（識別子名前空間により判断することにより）透過的な解決サービスにより自身を解決すべきである。解決の実装方法は、それが世界規模で行われ、ローカルで独自の解決機能を実装する必要がないものであれば、どんなものでもエンドユーザーサービスにとって価値がある。

#### A4.11.4 識別子の取り決め

聞き取りを行った複数の人が指摘したように、現在、識別子に関するソリューションは数多く存在し、共通のソリューションというものは存在しない。多くの人々は他人の判断を（まるでゴドーを待つかのように）待っている。実際、どんなスキームであれ、真の永続性を唯一保証するのは実施に当たったの政策的・組織的な支援である。識別子のモデル化は、基礎となるコンテンツとデータでのモデルの使用と同じように、どんな種類のどの程度の識別が必要であるかのヒントを与えることにより、意思決定プロセスに情報を提供することができる。しかし、先の節で推奨した世界規模の解決は、1つの識別子スキームに関して幅広い合意がなされると有益であることを示唆している。ただし、これには非常に幅広い合意と数ある識別子標準に対する相対的メリットの評価を必要とする。また、システムを円滑に運営するには識別子のシンタックスと登録方法に関する合意も必要である。

合意が得られたとしても、既に数多くの標準が存在し使用されているという問題が残るだろう。変換はほとんど有効ではないと思われる。まだテストされていないが、状況に応じて新しい識別子を適用するために識別子を抽象化することが、整合性を持たせるための答えになるかもしれない。識別子に関するどんな研究においても、その使用法とその使われ方の例は幅広く役立つだろう。

## A4.12 複合型オブジェクト

複合型オブジェクトに関しては、学習オブジェクトリポジトリの文脈において既に触れているが、個々のオブジェクトを合成することができるので、より価値のある学習資料を作成することができる。その他の種類の複合型オブジェクトとしては、データセットやマルチメディアオブジェクト、あるいは関連する出版物を含むe-リサーチオブジェクトがあげられるだろう。ここでは、複合型オブジェクト自体を検討し、さらに、（複合型オブジェクトを形成する複数のインスタンスを持つ）メタデータとコンテンツをパッケージ化してシステム間で移動・利用したいという要望から生まれたパッケージング標準に関連させて検討する。

複合型オブジェクトは潜在的に非常にリッチなものであり、また、複雑な構造体である。メタデータの自動生成の貴重なユースケースが見つけられるのは複合型オブジェクトの生成時である。メタデータの自動生成と同じように、複合型オブジェクトをさらに研究し、その役割、構成、利用法を評価する必要がある。複合型オブジェクトのための適切なデータモデルやコンテンツモデルを作成することにより、複合型オブジェクトに含めたいと思うコンテンツとメタデータを理解することは、そのようなオブジェクトを効果的に使用するための重要なステップとなる。そのようなモデル化は、様々な構成要素が互いにどのように関連するか、各構成要素をどのように使用することができるかということに関する情報を提供すると同時に、オブジェクトにおけるメタデータ標準の使用と様々なインターフェース標準によるオブジェクトの使用の両者のための正確で明確なシンタックスを提供する。また、モデル化は、複合型オブジェクトの様々な構成要素を独立に使用したり評価したりすることができるようにするために使用する識別子の正しい粒度を決定することになる。コンテンツとパッケージ自身の識別子を混乱することなく管理するためには、両者を明確に区別する必要がある。

コンテンツのモデル化は、特にボーンデジタルなコンテンツにとって必要である。コンテンツを構成するために現在使用されている物理的パラダイムはもはや適用できないからである。たとえば、電子学位論文は複数の種類のコンテンツを複雑な方法で含むことがある。複合型オブジェクトがボーンデジタルなコンテンツとアナログな参照を同時に含むかもしれない。これらをエンドユーザに効率的に提供するためにはこれらを扱う新しい方法が必要となる。これは、デジタルライブラリのエンドユーザサービスを通じた提示および個人的なりポジトリコレクションの両者について当てはまる。

複合型オブジェクトはそれ自体、それを構成する個々の要素、すなわち、メタデータやコンテンツの品質を向上させるものではない。低品質のメタデータやコンテンツは役に立たない複合型オブジェクトを生み出すことになる。また、複合型オブジェクトは主にシンタックスに焦点を絞っている。したがって、複合型オブジェクトの意味解釈は、さらなる研究や開発を必要とする分野であり、これにはRDFやOWLが主要な役割を演じるかもしれない。

### A4.12.1 パッケージング

複合型オブジェクトにパッケージ化する目的は様々であるが、主な目的は、資料をある場所から別の場所へ移動するための手段である。この移動は、2つのリポジトリ間の輸送のためであったり、エンドユーザサービスを通じた利用のためであったりする。複合型オブジェクトは、個々の構成要素間の関係を確立しているので発見を容易にし、複数の構成要素をまとめて配信できるので配信を容易にする。また、パッケージ化は、リポジトリ間でコンテンツを移動することができるので保存を容易にする。最近、米国議会図書館で行われたArchive Ingest and Handling Test (AIHT) は、正確な移動を保証するためには移動に使用するパッケージングフォーマットを標準化する必要があると勧告した<sup>94</sup>。このような特定の事例のための標準化は明らかに役に立つものである。別の方法でパッケージを標準化するには、標準されるべきものは何か、何故かについて同様に明らかにする必要がある。パッケージの移動は、リポジトリが投稿されるパッケージを扱うことができるかどうかにかかっている。この機能は、リポジトリレベルで提供されるサービスである場合もあるし、パッケージとリポジトリの間に位置する別の層として提供され、複数のリポジトリで共通に利用されるサービスである場合もある。たとえば、DSpaceは任意のパッケージング標準を処理し受入を可能にする独立したプラグインモジュールを持っている。DSpaceは、この相互運用性を容易にするための方法としてRDFを使用している（パッケージング自体をRDFを使って行うことも可能

であろう)。

コンテンツと関連するメタデータ以外にも、パッケージに含めると便利で、すべての複合型オブジェクトの追加要素となる情報が数多く存在する。このような付加的なメタデータを更新・保守するためのベストプラクティスは今のところ明らかではない。

- コレクション記述: パッケージに含まれているものを記述する方法であり、その有用性を決定する際に利用できる。IMS Content Packaging (CP) のマニフェストファイルがその例である。
- アノテーション/ランク付け: 先に述べたように、複合型オブジェクトに含まれるメタデータは1回だけ作成されるのではなく、イベントベースのメタデータやアノテーション、ランク付けのようなフィードバックが時間が経つにつれ追加される場合がある。

パッケージやそれに含まれる複合型オブジェクトを作成すると複雑性が増すので、何故パッケージが必要であるかを理解し、これらを作成する費用対効果を評価することが必要である。パッケージの作成には利点があるだろうが、その利点はより詳細な調査を必要とする。単にその利益があるがために作成することに価値があるとは思われない。特に、エンドユーザサービスによるパッケージの使用はこれに当てはまる。エンドユーザサービスでは、そのようなパッケージをどのように扱ったら良いのだろうか、また、効率的に作成するには何を使ったら良いのだろうか。

表4に示したように、コンテンツをパッケージ化するために利用できる多くの標準が存在する。メタデータ標準と同じように、これらは異なる部門で作成されたものであるが、その目的は同じである。しかし、各標準は独自の特徴を持っており、ある状況にどれが最適であるかを判断する上で役に立つ。そのような判断を支援するために様々な種類のコンテンツにどのパッケージング標準がふさわしいかを評価する研究をさらに行う必要がある。

パッケージング標準	備考
MPEG-21 DIDL <sup>95</sup>	ISO標準。XMLベースであり、商用エンターテインメント産業に由来し、主にマルチメディア向けに設計された。高度に構造化され、柔軟かつ汎用的であるが、複雑である。資産の移動用に設計されている。抽象モデルを持ち、個々の構成要素の識別に優れる。Base64エンコーディングすることによりコンテンツを含めることができる。
METS <sup>96</sup>	コミュニティベースのオープンな標準であり、米国議会図書館で管理されている。コンテンツではなくXMLメタデータフォーマットをパッケージ化するために設計され、必要に応じてコンテンツへのポインタ(識別子)を持つ。MPEG-21 DIDLほど構造化されていないし、柔軟性もないが、おそらく実装が簡単である。抽象モデルは持っていない。
IMS Content Packaging <sup>97</sup>	IMSの標準。コンテンツとメタデータをZIP圧縮したファイルを基本とするが、XMLバイディングも利用できる。主に学習オブジェクトなどの(具体的な機能を持つ)IMS仕様のオフラインの移動のために設計されたが、一般的な目的にも使用することができる。構成要素に関する情報と構成を提供するXML形式のマニフェストファイルを持つ。抽象モデルを持っている。
ATOM <sup>98</sup>	XML形式の配信フォーマットで、現在IETFで標準化作業中である。Base64エンコーディングすることによりメタデータとコンテンツの両者を含めることができ、パッケージングフォーマットの役目も果たすこともできる。抽象モデルは持っていない。
XFDU <sup>99</sup>	OAIS参照モデルの開発を支援したチームがMETS標準を改造して、分岐したものである。現在の開発段階は明らかでない。

表 4: パッケージング標準の例

MPEG-21 DIDLは米国のロスアラモス国立研究所(LANL)のHerbert Van de Sompel等の活動<sup>100</sup>により学術コミュニティやデジタルライブラリコミュニティの注目を集めるようになった。以来、規模は限られるが、オランダのDAREnetにおける多くのプロジェクトや機関で使用されており、この複合型オブジェクトをハーベストする経験が蓄積されている。

MPEG-21 DIDLの使用には特許に関する懸念が存在する。LANLのJeroen Bekaertはこの件が議論された2006年1月のMPEG会議に参加した。この会議で取り上げられた特許問題に関する疑義は、もし特許問題が存在するとしたら、それはデジタル権利宣言を表現するために使用されているXMLフォーマットに関するものだけであり、パッケージングフォーマット自体には関係がないと結論づけられた。特許に関する

る不透明さは、まだ解決されていない、デジタル権利宣言にXMLフォーマットを使用することに関する継続中の包括的な特許問題によるものである。したがって、MPEG-21 DIDLパッケージング標準の使用に関しては特許による障害は存在しない。一般的に、パッケージ内の権利宣言の使用に関する特許問題は将来に対する影響を評価するためにモニタリングする必要がある。

METSは、移動や資源発見後の配信のために同一のオブジェクトに関する様々なメタデータレコードと一緒に運ぶ方法として図書館コミュニティで広く使用されるようになってきている。特に、資料の交換を成功させるには共通の利用法に関する合意が絶対に必要である。必ずしもすべての図書館向けのメタデータ標準が、METSのXMLという性格に合致しないのではないのかという小さな懸念が存在するが、今では、通常使われている標準には対応するXMLスキーマが存在する。抽象モデルを持っていないので、文書のモデル化は難しく、様々なメタデータレコードをまとめて保持するというシンプルな機能を超えてパッケージを構築することも容易ではないかもしれない。PORTICO電子ジャーナルアーカイビングプロジェクト<sup>101</sup>は、その本来の用途にMETSを採用し、これにMPEG-21 DIDLのいくつかの側面を含めることによりこの問題を扱っている。

IMS CPは堅牢で成功したパッケージング標準であり、数年間にわたって主に学習オブジェクトであるリソースの移動に使用されてきた。最新版である1.2ではマニフェストファイルを採用し、これによりパッケージ内のリソースを参照するだけでなくパッケージ外のコンテンツを指定することも可能になった。California Digital Libraryは、マニフェストファイルに相当するものとしてMETSを使用し、ほとんど同じ方法で分散する資源を指定している。これは、コンテンツが明確に識別されており、必要に応じてコンテンツにリンクすることができる場合のパッケージ化の価値と必要性に関するCDLで現在も続いている議論を開始させた。

#### A4.13 相互運用性

メタデータと同じように、リポジトリの内部で使用する複合型オブジェクトの構造は必ずしも、外部に公開するパッケージの構造と同じである必要はない。Fedoraは内部的にはFOXMLを使用しているが、外部にはMETSパッケージを配信することができる。内部構造から外部構造へ変換する機能は有用である。Van de Sompel等は主要な3つのオープンソースのリポジトリシステム、すなわち、DSpace、Fedora、EPrintsのために、MPEG-21 DIDLへの変換システムを開発した。このような他のパッケージング標準への変換システムが利用できればリポジトリの運用に柔軟性を与えるだろう。しかし、これに伴い、クロスワーク作業の調整やクロスワークにより生じるかもしれない情報欠落の詳細な調査が必要となる。これらの変換はXMLベースの標準に適用するものであるが、IMS CPによる公開は、リポジトリがZIP相当のファイルを処理できることも前提にしている。

パッケージによりもたらされる構造は、相互運用性や発見を容易にすることができる。NDLTDは現在公開のためのパッケージを提供していない。この構造の欠如により、Scirus検索エンジンによる索引化は複雑になっており、将来METSを使用することが現在検討されている。

複数のパッケージング標準が存在するということは、不可避免的に、リポジトリやエンドユーザサービス間の完全な相互運用性のためには、複数の標準を相互に変換したり、複数の標準を管理する機能が役に立つことを意味する。RAMLETイニシアティブはこの可能性を調査しており、現在、標準間の変換を容易にするためにOWLの使用を検討している<sup>102</sup>。上で述べたように、このような変換は情報を欠落させる恐れがあるので、これを詳細に監視する必要がある。基礎となるデータモデルが存在しても、これを示すセマンティックマークアップがパッケージ内に無いと、真の相互運用性を阻害するかもしれない。

#### A4.14 メタデータとフルコンテンツを同時に扱う

多くのパッケージング標準が使用されるということは、コンテンツがメタデータと一緒に転送され、リポジトリやエンドユーザサービスで使用されることを意味する。パッケージはコンテンツをユーザーに直接届けるが、コンテンツをパッケージや簡単なメタデータレコードから指定することも可能である。エンドユーザはどちらのアプローチが取られているかについては興味がないだろうが、どちらの場合も、その意図しているのは、エンドユーザが「ダウンロード可能なビット列」の所在を知り、D2D連鎖の主要な目標であるコンテンツへのアクセスを得ることである。資源発見はフルコンテンツへのアクセスが無ければその価値は減少する。これは、オープンアクセスリポジトリを考えた場合、特に当てはまる。フルコンテンツへのアクセスが主な目的の1つであるからである。フルコンテンツへのアクセスが得られるもう1つの利点に、フルテキストの索引化により発見処理を支援することができることが挙げられる。これはGoogleが

採用している方法論である。ePrints UKプロジェクトで研究されたメタデータの品質向上サービスも関連するメタデータレコードの品質を向上させるための分析にはフルコンテンツへのアクセスに依存していた。

しかし、フルコンテンツを扱うとそれ独自の問題をもたらすことになる。

フルコンテンツがメタデータレコードから参照されている場合、メタデータレコードからのリンクが成功し、エンドユーザの期待を裏切らないことを保証するためには、明確かつ永続的でユニークな識別子が必要になる。また、メタデータレコードからのリンク先——オブジェクト自体を指定するか、オブジェクトに関する付加的な周辺情報を提供できるが、エンドユーザの処理をさらに一手間増やすジャンプオフページを指定するか——に関する標準化が必要である。ePrints UKプロジェクトは、Webサービスを使った処理でフルコンテンツを捕捉するために、ハーベストしたメタデータレコードに含まれていたコンテンツ識別子を使用することによりこの問題に対処しようとした。その結果、コンテンツ自体へのリンクではなくジャンプオフページへのリンクを広範囲に使うことになり、この処理は複雑になった。直接的なリンクが適用できない場合は、OpenURLリンクが文脈依存で永続的なリンクをエンドユーザに提供することができ、情報検索が行き止まりになることを防ぐことができる。複合型オブジェクトパッケージの各構成要素を識別する際にも、明確な識別が必要となるだろう。

今日では、大量のコンテンツを格納したり、ネットワーク上で移動したりすることが非常に容易になっているとはいえ、フルコンテンツ、特に3Dやデータセット、ビデオ、オーディオなどの大容量のアイテムを含むコンテンツへのアクセスを考える際には、容量や帯域の問題を思い出す必要がある。アクセスを容易にするためにパッケージにコンテンツを含めて移動させる場合は、特にそうである。Web上のキャッシングがWebページへのアクセス速度を上げるように、オブジェクトをキャッシュするサービス（たとえば、Akamai<sup>103</sup>で提供されているサービスやもともと電子ジャーナルコンテンツ用であったLOCKSS<sup>104</sup>など）は役に立つだろう。キャッシングは、オリジナルコンテンツを所有するリポジトリにアクセスできない場合のバックアップにもなる。機関を離れたネットワークレベルでのサードパーティによるストレージ機能は既に役立っている（たとえば、英国で利用可能なAHDSやESDS、UKデータアーカイブなどのデータサービスにおいて）。コンテンツの量と複雑性が増すにつれ、ネットワークレベルのストレージをさらに提供する必要が増すものと考えられる。

リポジトリとサービスの間で複製を行うことはスケーラビリティがなく、また、データの真正性や来歴、完全性という問題も引き起こす。コンテンツが更新されたり、そのライフサイクルの間に目的が変更されたりした場合、（するべきだとしても）別の場所にある複製版にどうやってこれを適用するのか。そのような変更のためには、双方向性の標準が必要となる。OAI-PMHはハーベスティングに関して単方向性であるが、オリジナルのリポジトリが、変更されたコンテンツを後で再ハーベストする時はハーベスタの役目を果たすことになる。RSSは、双方向の非同期式交換を可能とするためにマイクロソフトのSimple Sharing Extensions<sup>105</sup>などのアドオンを組み合わせて使用することができる。コンテンツが別場所に保管されている場合、コンテンツ所有者のプライバシーや知的所有権に関する懸念が存在する。エンドユーザサービスや他のリポジトリは、アクセスしたり入手したりしたコンテンツを使って何をするのでいいのか。これは必ずしも技術的な疑問ではないが、技術的な回答が必要であり、注意を向ける必要があるだろう。また、保存の問題も存在する。どの版がどこで保存されるのだろうか。しかし、コンテンツをリポジトリ、アグリゲータ、エンドユーザサービスの間で移動する機能はサードパーティによる保存に関して数多くの可能性を提供する。SHERPA-DPプロジェクト<sup>106</sup>やPRESERVプロジェクト<sup>108</sup>は、e-プリントコンテンツの保存におけるまさにこのタイプの役割を調査している。Hybrid Archivesプロジェクト<sup>104</sup>も、このモデルを使ってコンテンツを捕捉しているが、エンドユーザサービスへのコンテンツの提供は依然としてオリジナルホストにより提供されている。

権利の問題は、JORUM学習オブジェクトリポジトリの開発の中心課題であった。このリポジトリはコンテンツとメタデータの両者を保管し、これを使って一次エンドユーザサービスの役割も果たしている。コンテンツは別利用や再利用を容易にするために保管されている。MERLOTが行っているように、リンクのリポジトリとして機能することはおそらくはるかに簡単なモデルだろうが、JORUMで可能な追加機能を可能にすることはできない。

メタデータとコンテンツを分離させた方が望ましい場合も存在する。資源発見を考えた場合、メタデータはアナログアイテムとデジタルアイテムの両者に対するアクセスポイントを等しく提供することができる。発見することができれば、これらいずれへのリンクも作成することができる。メタデータのあまりにも近くにフルデジタルコンテンツを格納すると、同等以上の価値のあるアナログ資源へのアクセスに比べ

て、デジタルコンテンツへのアクセスに偏る可能性が生じる。

エンドユーザサービスがコンテンツを効果的に使用するためには、配信されたコンテンツフォーマットを利用できなければならない。たとえば、適当なプラグインを使ってPDFファイルを読むことができるように、コンテンツを利用するためのプラグインやツールがどこでも利用できるべきである。配信フォーマットにどれほど柔軟性があるかは、リポジトリとアグリゲータあるいはエンドユーザサービスの間で「オンライン」フォーマット変換サービスが利用できるかどうかにかかってくるだろう。

## A5 インターフェース

本研究に指定されている対象はユーザー指向サービスに関するもの、すなわち、エンドユーザが直接利用することができるサービスである。しかし、このサービスを支えるマシンインターフェースにも関心が集まっている。リポジトリが公開するインターフェースは、リポジトリがそのコンテンツへのアクセスをエンドユーザに提供する方法を決定する。すなわち、Webインターフェースを通じて直接提供するか、マシンインターフェースオプションを通じてリポジトリのコンテンツを提供するかである。したがって、ここではその両者を検討する。インターフェースに関するほとんどの議論は、関連する標準やプロトコルの使用に集中しているが、この節では、リポジトリが提供するインターフェースの様々な側面を検討するだけでなく、これらに関連する問題も扱うことになる。

リポジトリが提供するインターフェースは、リポジトリ間で可能となる相互運用性のレベルも決定する。個々のリポジトリへの単純なWebインターフェースは、通常、（裏でメタサーチツールが複数のHTTPリクエストを発行しない限りは）この要求を満たさない。一方、別の場所で使用するためのコンテンツを公開するマシンインターフェースは、柔軟性を提供することができ、標準を順守させることができれば、リポジトリ横断型の相互運用可能なサービスを確立するための共通のプラットフォームを提供することもできる。したがって、Webインターフェースとマシンインターフェースの開発のうちリポジトリがどちらを優先するかは、可能となる相互運用性のレベルを検討する上で重要である。インターフェースとそれを利用するサービスやアプリケーションは必ずしもリポジトリ自体にあるわけではないことを思い出すことが重要である。これらの間には階層が存在し、この階層化はリポジトリがコンテンツを公開する方法に柔軟性を提供する。

本研究でインタビューした者はおおよそ、インターフェース標準が単純で広く採用されるほど、相互運用性にとっては望ましいという意見に合意した。たとえばRSSのように、少しの努力で多くのことができる標準が特に歓迎される。また、リポジトリを横断する相互運用性を可能にするために必要な標準は極めてうまく機能しているという幅広い合意がある。最も優先すべきことは、アプリケーションのサポート、標準の使用法を決定するための方針、確実な使用を保証する必要性の組み合わせである。方針を決定することは、たとえば、リポジトリシステムの実装経験から最近判明した標準的なデポジットAPIに存在するギャップなど、以前には現れなかったギャップを識別するための助けとなる。リポジトリを更新する機能にもギャップが存在する。SRW Update<sup>109</sup>の開発状況は明らかでないが、WebDAV<sup>110</sup>の経験がこの分野での価値を示しているとはいえ、リポジトリがWebサーバではないという認識に基づいた懸念も存在している。

リポジトリは、可能としたい相互運用性のレベルを見極める必要がある。これが最適なインターフェース標準の選択に影響を及ぼすからである。一般的な利用を除けば、様々な利用場面やコンテンツ種別にとってどのインターフェース標準が最適であるかは不明であり、リポジトリもエンドユーザサービスもこれについてより詳細に調べる必要がある。様々なインターフェースを提供することは、これを評価する助けとなる。実際、すべてのリポジトリはアグリゲータやエンドユーザサービスとの柔軟な相互作用を可能にするために複数のインターフェースの提供を検討すべきである。拡張性があり、柔軟な標準に基づいたアプローチは、リポジトリが将来の未知の要求にも適用できることを保証することになる。この柔軟性は、革新を可能にするためにも、また、標準の順守が推進役になるどころか重荷になることを防ぐためにも重要である。

### A5.1 Webインターフェース

相互運用性を容易にするためにはマシンインターフェースが必要であるが、ここではまずWebインターフェースの問題を取り上げる。このリポジトリコンテンツへの入り口は、アグリゲータと同じように、多くのリポジトリが最優先で開発する第一のものである。インタビューした者の間でもこれは広く受け入れられていたが、一般に考えられるように、マシンインターフェースの方が価値があり、それ故、より優先

すべきであることも受け入れられており、何故そうすべきなのかは明らかではなかった。おそらく、ユーザーの要求により導かれるWebインターフェースの開発という簡単に手に入れられる果実をとりたいとの欲望があるからであろう（もっとも、きちんとしたWebインターフェースの作成は時間もリソースもかかるものである）。そのようなインターフェースはWeb上での視認性をもたらし、個々のリポジトリやその所有者をブランド化するからである。あるいは、マシンインターフェースを使用しているもっと特別な検索ツールを利用しないカジュアルユーザの要求を満たす必要があるのかもしれない。コンテンツごとにURLを持つWebインターフェースは、Googleなどのサーチエンジンのクローラにとっても理想的なものであり、この目的でリポジトリコンテンツを公開する場合には大きな価値がある。同様に、Googleインターフェースデザインを真似たWebインターフェースをリポジトリに設けることをユーザーが期待しているという別の「Google効果」にも注目することができる。

Webインターフェースは結局、直接アクセスに適しており、そのように使用されている。一方、マシンインターフェースは今のところ、本来あるべきほどには使用されておらず、研究もされていない。マシンインターフェースははるかに大きな可能性を提供すると思われるが、未だ手が付けられていない。その一方で、WebインターフェースとそのWebクローラによるアグリゲーションは、リポジトリに対する第一の入り口やインターフェースを提供している。

## A5.2 マシンインターフェース

リポジトリ上にマシンインターフェースを開発する主な理由は、リポジトリコンテンツのその後の利用方法における柔軟性を高めることである。Googleのアグリゲーションを基礎にサービスを作成することは可能であるが、リポジトリメタデータは通常、その他のWebコンテンツと混合されるので、対象を絞ったサービスを提供することが難しくなる（この点でGoogle Scholarには価値があるかもしれないことに注意する必要があるが）。マシンインターフェースは、より明確で対象を絞った公開やリポジトリのメタデータやコンテンツの組み換えを許すので、たとえば、主題指向のアクセスをサポートするなど、特定のニーズに合ったエンドユーザサービスやコンテンツに対する様々なビューを構築することを可能にする。したがって、おそらく自分でレコードをコントロールできないようなどこか別の場所で使用されても構わないとリポジトリが思えるような大きな付加価値を与えることができる。マシンインターフェースとWebインターフェースの両インターフェースを提供すれば両方のアプローチをカバーすることになる。リポジトリは両者を検討すべきである。機関リポジトリに関しては、広く使用されているリポジトリソフトウェアシステムに内蔵されているOAI-PMH機能により、容易にこれを行うことができる。しかし、リポジトリは、単に与えられたものを受け入れるだけでなく、マシンインターフェースに対する要望や要件を積極的に取り上げ、解決すべきである。両者を提供するために必要な努力の量が問題になるかもしれない。しかし、マシンインターフェースがリモートサービスを提供するだけでなく、ローカルのWebインターフェースの基礎としても使用されるとすれば、重複を省くことができる。マシンインターフェースは許容量が小さいので設定が難しい。また、多くのコンテンツ所有者はインターフェースに隠れて視認性が失われることを危惧している。しかし、このインターフェースは、たとえ表から見えなくても、メタデータやコンテンツを幅広く公開することにより大きな利益をもたらす可能性がある。そのようなインターフェースをWebインターフェースやその他のエンドユーザサービスの基礎として注目すれば、最大の出力を持つ開発の焦点を1点に絞ることが可能になる。

## A5.3 インターフェースの選択肢

この節では、リポジトリがコンテンツを公開するマシンインターフェースを提供する際に使用できる様々な標準や仕様、プロトコルを検討する。本研究で検討の対象となるインターフェースを表<sup>5</sup>に示した。ある目的にとって何がベストであるかについては未だはっきりしていないことは同意されているが、引き続きサービスやエンドユーザに公開すべきものを明確にするため、また、実装を簡単にするために特定のインターフェースに焦点を絞る必要性があることも認識されている。

インターフェース標準	備考
OAI-PMH	リポジトリがメタデータ（と適切にパッケージ化されればコンテンツ）を公開することより、ハーベスティングやアグリゲーション、サービスプロバイダによる配信を可能にする。 <a href="http://www.openarchives.org/">http://www.openarchives.org/</a> リポジトリによるRSSリーダーなどのアプリケーション・サービスに対するメタデータのク

RSS	ントロールされた配信を可能にする。 <a href="http://en.wikipedia.org/wiki/RSS_file_format">http://en.wikipedia.org/wiki/RSS_file_format</a>
ATOM	RSSと同じようにメタデータのコントロールされた配信を可能にするが、Base64エンコーディングすることによりコンテンツの配信も可能にする。 <a href="http://www.atomenabled.org/">http://www.atomenabled.org/</a>
SRW/U	リポジトリやリポジトリのアグリゲーションに対して対象を絞った検索を可能とする。SRWはSOAPを使った標準であり、SRUはRESTを使った標準である。 <a href="http://www.loc.gov/standards/sru/">http://www.loc.gov/standards/sru/</a>
Z39.50	SRW/Uの先駆者であり、Web自体より前から存在する。Z39.50も対象を絞った検索を可能とする。 <a href="http://www.loc.gov/z3950/agency/">http://www.loc.gov/z3950/agency/</a>
OpenURL	様々なリソースの間で文脈依存のリンクを可能とする。リポジトリの外部を参照するリンクも、内部を参照するリンクも含めることができる。 <a href="http://www.niso.org/standards/standard_detail.cfm?std_id=783">http://www.niso.org/standards/standard_detail.cfm?std_id=783</a>
SQI	学習オブジェクトリポジトリ間の相互運用性を促進するために開発されたシンプルなクエリインターフェース。 <a href="http://ariadne.cs.kuleuven.ac.be/vqwiki2.5.5/jsp/Wiki?LorInteroperability">http://ariadne.cs.kuleuven.ac.be/vqwiki2.5.5/jsp/Wiki?LorInteroperability</a>

表 5: リポジトリで使用するためのインターフェース標準

#### A5.3.1 OAI-PMH

オープン・アーカイブズ・イニシアティブは、1999年のサンタ・フェ会議に起源を持ち ([Van de Sompel, 2000](#))、メタデータの共有を容易にする方法を記述することにより、研究者の間でe-プリントの共有を促進することを追及した。これらの議論から登場したメタデータハーベスティングプロトコルは、その後、元来の対象であったe-プリントを超えて多くの状況で使用されるようになってきているが、このプロトコルの利用は依然としてオープンアクセスリポジトリによるものが中心であり、オープンアクセスプロセスの一部となっている ([Lagoze, 2003](#))。OAI-PMHの実装は比較的容易であり、おそらくこの理由により、Z39.50の使用で経験したものと同一シナリオをたどることになった。つまり、一定の水準を超えた実装を行えばこの標準を利用することはできるが、標準とその機能を完全に使いこなすにはさらなる努力が必要となる。たとえば、OAIセットを使えば、おそらく、より洗練されたハーベストを行うことにより、より対象を絞ったサービスが可能になる。しかし、英国におけるリポジトリの実装の現状について調査したSHERPAプロジェクトの最近の研究によるとセットはあまり広く実装されていない。

OAIモデルはデータプロバイダとサービスプロバイダの存在に依存している。データプロバイダはリポジトリとしてサービスプロバイダがハーベストするための情報を公開する。このプロセスは単方向性である。すなわち、サービスプロバイダがリポジトリからメタデータを引き寄せる (pull)。データプロバイダとサービスプロバイダの役割が逆転し、リポジトリがサービスからハーベストするようにならない限り、このプロセスを反対に進めることはできない。OAI-PMHを双方向に使用したり、表に示した他の標準を使用したりすることは有益であり、A5節で示した更新時のギャップを埋める可能性がある。また、OAI-PMHは、主にe-プリントで広く使用されていることから、ハーベストにはシンプルダブリンコアレコードが主に使用されている。この対象の限定は、おそらく相互運用性という面では価値があったが、プロトコルの持つ潜在的な有用性を制限するものであった。OAI-PMHを他の資料に使用することはもちろん、特に、別のメタデータ標準と共に使用することは役に立つだろう。プロトコルの国際的な標準化や抽象化はこれを支援するだけでなく、転送プロトコルとしてHTTPを使用するなど、OAI-PMHに残っているその他の制約の解決にも役立つだろう。

サービスプロバイダは、様々なリポジトリからハーベストを行えば、その結果得られるアグリゲーションを、Webインターフェースを通じて直接アクセスを提供することも、他のエンドユーザサービスに公開することもできる (そして、OAI-PMHの役割はエンドユーザサービスの背後に置かれることがベストであろう)。たとえば、アグリゲーションをSRW/Uのターゲットとして公開することができる (たとえば、最近OAIsterがSRUターゲットとなったように<sup>111</sup>)。ビールフェルト大学のBASEサーチエンジン<sup>112</sup>は、収集したメタデータとコンテンツに対するFASTベースのサーチエンジンを設け、正式な標準の利点とWebサーチエンジン技術の利点の両方を利用している。一般に、エンドユーザサービスのためのプラットフォームとして、OAI-PMHを使ってメタデータとコンテンツを同時に収集することは、分散型の検索

モデルにおいて受け入れられていると考えられるが、その根拠を追加できるのはさらなるエンドユーザサービスの開発だけである。BASEやユトレヒト大学のOmegaメタデータデータベース<sup>113</sup>にとっては、対象となるユーザーが利用できるより良いエンドユーザサービスが存在するところにメタデータとコンテンツをもたらすことに価値がある。アグリゲーションアプローチは検索の他に、報告書本文の第2章や第3章で明らかにしたようなその他のサービスにも応用することができる。

#### A5.3.2 RSS

RSSは、実装は比較的容易ではあるが、シンプルな1つの標準ではなく、時間の経過と共に開発された一群の標準である。RSS 0.9と1.0はRDFに基づいたRSSを表し、0.91、0.92、2.0はRDFモデルの使用を強制しない別のバージョンを表している。すべてのバージョンは比較的広く使用されており、主要なRSSリーダーは万全を期すためにすべてのバージョンに対応することになる。RSSという略語の意味も、RDF Site Summary、Rich Site Summary、Really Simple Syndication、Rela-time Simple Syndication（最後の2つは主にRSS 2.0で使用されている）というようにバージョンにより異なっている。

これらすべてのバージョンに共通の要因はRSSの役割、すなわち、リポジトリなどの情報の所有者が情報を読みたいと願う者に、RSSリーダーやブラウザ、ポータルなどその他のWebアプリケーションを通じて、公開情報の抜粋を押し付ける（push）できるようにすることである。この幅広い実装により、出版社がメタデータの配布方法としてOAI-PMHに代わるものとしてRSSの使用に興味を持つという、学術コミュニティが監視をしなければならないような傾向をもたらしていることは注目に値する。

RSSの主な利用方法であるイベントやニュースの通知以外にも、ブログエントリや検索更新要求にも配信を使うことができる（たとえば、SCRANの検索はRSSに埋め込むことができ、定期的に再実行することによりエンドユーザは最新結果の通知を受けることができる）。RSSのより広い応用例も登場している。たとえば、AmazonのOpenSearchサービスは、検索結果をHTML、RSS、ATOMで返すことができ、表示や別利用に柔軟に対応することができる<sup>114</sup>。

#### A5.3.3 ATOM

ATOM配信標準は、RSS 2.0が登場しないのではないかと思われていたことに起源を持つ。現在はIETFで正式に標準化作業中である。RSSほど広くは使用されていないが、知名度は増加中である。その理由の1つは多くのGoogleサービスで使用されていることによる。上の表で示したように、Base64エンコーディングを使うことによりメタデータだけでなくコンテンツも配信することができる。これは、MPEG-21 DIDLパッケージングフォーマットでそのようなエンコーディングが使われていることと同じである。このように、ATOMは配信とパッケージ化の両目的で使用される可能性を持っている。

#### A5.3.4 SRW/U

SRW/Uは、米国議会図書館によるZING（Z39.50 International: Next Generation）イニシアティブに起源を持つ。SRW（Search/Retrieve Web service）は本格的なWebサービスで、クエリと結果の受け取りにSOAPメッセージを使用する。また、検索式の作成にはCQL（Common Query Language）<sup>115</sup>を使用する。SRU（Search/Retrieve via URL）もCQLを使用するが、クエリの作成と送信にURLを使用するのでSRWより実装が容易である。検索結果はSRWと同じようにSOAPメッセージとして返される。両標準のうち最近ではSRUに重点が置かれるようになってきたことが注目に値する。SRUはNISO MetaSearchイニシアティブ<sup>116</sup>の議論において、（リポジトリ横断など）複数のターゲットを横断する効果的で一貫した検索を可能とする公約数的な機能の第一候補となっている。このような分野で潜在的な利点はあるものの、SRWやSRUのような標準を使用した分散検索のスケラビリティに関しては以前から懸念されており、未だに解決されていない。また、European Library（[van Veen, 2004](#)）など、Z39.50をSRW/Uに変換するゲートウェイは数多く稼動しているが、SRW/Uターゲットの数は多くない。

#### A5.3.5 Z39.50

Z39.50標準は様々なコンピュータシステムの相互運用性を可能にする技術として1980年代に登場した。この標準は、相互運用性に関して数多くの能力を持っているが、もっとも広く使用されたのは図書館や情報分野におけるデータベースの横断検索であった。Z39.50はコンピュータシステム間のコミュニケーションを可能とする共通のシンタックスを提供するが、これに準拠するためのレベルの幅が非常に大き

く、Z39.50のクライアントやターゲットの設定が比較的複雑なこともあり、広範囲の利用という意味での成功に影響を及ぼした。また、この標準はWebより前に登場したので、Web上で使用されるようには設計されていない。今では、新規開発にはSRW/Uの使用が多く選択されている。広く使用されているオープンソースの機関リポジトリソフトウェアパッケージにおいて、Z39.50インターフェースを提供しているものは1つも無いのに対して、SRW/Uは少なくとも2つのシステムで利用できることは注目に値する。Z39.50が主に図書館や情報分野で広く使用されていることも、今では、より広い相互運用性を旨とするに当たっての制約になっていると思われる。しかしながら、Z39.50は継続的に利用されており、これを完全に無視することはできない。

#### A5.3.6 OpenURL

OpenURLは、1990年代にベルギーのアントワープ大学のHerbert Van de Sompel等により行われた研究とそれに続く、SFXサービスの開発につながったEx Librisにおける作業に起源を持つ。[\(Van de Sompel, 2001\)](#)。OpenURLはサーチャーが情報資源を探す際の文脈に依存した情報資源間の動的なリンク機能を提供する。元来、書誌的資料のために設計されたものであるが、NISO OpenURL 1.0 Z39.88-2004標準は、ほとんどあらゆる種類のオブジェクト間のリンクを容易にするように設計されており、リポジトリがOpenURLのソースとターゲットの両役割を果たすことができる可能性を提供している。リポジトリからのOpenURLリンクはクランフィールド大学のDSpaceでテストされている。また、他の場所からリポジトリへリンクする可能性も存在する。広く認識されている問題の1つは、OpenURLにおける引用情報の書式である。リポジトリはこれを可能にするためにできるだけ粒度の細かいフォーマットでデータを持つ必要がある。

#### A5.3.7 SSI

SSIは、学習オブジェクトリポジトリ間の相互運用性を可能とするために、EU CEN/ISSS Learning Technologies Workshopイニシアティブから現れた。この技術は、学習オブジェクトリポジトリのオランダ全国ネットワークであるLOREnet[117](#)での使用によりテストされている。

### A5.4 相互運用性に対する軽量なアプローチ

表に挙げた標準のほとんどが持っている主な利点の1つに、少なくとも基本的なレベルの実装は比較的簡単なことが挙げられる。これはインタビューを行った者の間で最も要望の高い特徴であった。Z39.50は、多かれ少なかれその価値を証明してきたとはいえ、もはや人気を得ることはないだろう。一方、必要に応じて再利用や再公開を可能とするそのオープンさによりOAI-PMHやRSSが広く使用されている。これらのアプローチは軽量であるが、結果として相互運用性をより簡単に得ることができる。OCKHAMイニシアティブはこの軽量アプローチに焦点を絞り、様々な標準やインターフェースを組み合わせることで最大の効果を得る方法を示している[\(Xiang, 2005\)](#)。OAI-PMHによりハーベストされたコレクションは、更新の度にRSSを使って配信することができ、具体的な検索のためにはSRW/Uターゲットとなり、他の場所で発見されたアイテムの所在を教えるためのOpenURLターゲットの役割も果たすこともできる。標準を組み合わせることで1つにまとめることもできる。DLF Aquiferプロジェクト[118](#)は、OAIでハーベストされるメタデータに実行可能なURLを含むRSS断片を埋め込むことにより、次のアクションを起こしたり、エンドユーザに実行させたりする「アセットアクション」を実験している。

軽量アプローチは、希望するインターフェースをサポートするのに最もふさわしい標準を調査するための方法論として役に立つ。複数のインターフェースを提供することにより、ある目的にどのアプローチが最適であるかをリポジトリはテストすることができる。広く受け入れられるためには、標準を組み合わせるための共通のアプローチが必要になるが、インターフェースAPIが明確に定義されオープンである限り、第三者はそれらを採用し、それらを足場にするに違いない。

標準の中には、メタデータとコンテンツのより大きなアグリゲーションを使うことにより付加価値を提供する力が増すものがある。たとえば、RSSはより対象を絞った関連の更新を提供することができるし、OpenURLリンクはリンクの対象となるオブジェクトの数が多いほど結果を得る可能性が高くなると思われる。

### A5.5 リポジトリ仕様

これまでに説明したインターフェース標準で実現する軽量アプローチは相互運用性への明確な経路を提供する。各標準が（一部重なる可能性はあるにせよ）主に特定の分野のみに機能を絞っていることには妥協が必要である。たとえば、OAI-PMHは公開されたメタデータの収集、SRW/Uは検索の実行、RSSは通知の提供である。すべての標準は限定的にも拡張的にも使用することができるが、各々の持つ第一の目的は明確に定義されている。各標準はリポジトリとある特定のレベルで相互作用することを許可しているが、多くのリポジトリがこのレベルに対応することができるので、相互運用性は比較的容易に達成することができる。

リポジトリとの相互作用をもっと広範囲に行い、複数の機能を相互運用性を持って実行したい場合は、リポジトリが提供する必要があるインターフェースはもっと複雑で、様々なリクエストを扱うことができる必要がある。表6に示したように、より深いレベルの相互運用性を試み、可能とするための様々なレベルの数多くのリポジトリ仕様が開発されてきた。

仕様	備考
IMS Digital Repositories Interoperability (IMS DRI)	リポジトリ共通機能の高レベルの相互運用性のための勧告を提供するために設計された。一連のガイドライン以上のものは策定されなかった。 <a href="http://www.imsglobal.org/digitalrepositories/">http://www.imsglobal.org/digitalrepositories/</a>
JSR 170/283	JSR 170とその後継仕様であるJSR 283は、Javaコミュニティプロセスの仕様である。リポジトリとアプリケーションの間でコンテンツの双方向アクセスを詳細なレベルで実装独立な方法により実現するコンテンツリポジトリAPIを提供する。 <a href="http://www.jcp.org/en/jsr/detail?id=283">http://www.jcp.org/en/jsr/detail?id=283</a>
Open Knowledge Initiative Digital Repository Open Service Interface Definition (OKI DR OSID)	OSIDは一般にソフトウェアシステム間の規約を提供する。DR OSIDはリポジトリ資産のストレージと検索の問題に取り組んでいる。 <a href="http://www.okiproject.org/">http://www.okiproject.org/</a>
eduSource Communication Layer (ECL)	カナダにおけるIMS DRIの実装であり、LionShareピアツーピアプロジェクトで使用された。 <a href="http://ecl.iat.sfu.ca/">http://ecl.iat.sfu.ca/</a>

表 6: リポジトリインターフェース仕様

これらのリポジトリ仕様が策定されているにもかかわらず、インタビューを行った者の多くはこれを知らなかった。JSR 170/283以外はすべて学習オブジェクトリポジトリの分野を起源としており、エンドユーザーサービスと学習オブジェクトの詳細な相互運用性や相互作用を促進することを目指したものであり、その役割の多くは学習オブジェクトに特有のものであった。たとえば、オランダのLOREnetはオランダにおける学習オブジェクトリポジトリ間の相互運用性を実現するためにDS OSIDの使用を試みている。しかし、オープンアクセスの分野では、望んでいるのは単に発見のためにリポジトリコンテンツを公開することであり、必要とする相互運用性の程度ははるかに低いものであり、既存のニーズを満たすためにリポジトリ仕様は不要であった。また、高い相互運用性は上で述べた軽量なアプローチの実装を超えるさらなる作業を必要とするが、多くのオープンアクセスリポジトリの所有者はこれを提供することができない。

5.3節で説明したように、Z39.50やおそらくはOAIのように、詳細な仕様には、対象となるすべてのリポジトリで詳細さにおいて同程度の実装がされないかもしれないというリスクが存在する。従って、コミュニケーションや合意が得られやすい管理されたコンソーシアムでの実装がより適しているだろう。仕様が提供できるものをより良く理解できるように援助が必要である。あるレベルにおいて、仕様はリポジトリシステムの構築を支援するために使用することができる。仕様はリポジトリが従うべき標準的なサービスを提供するからである。しかし、保存計画のためのOAIS<sup>119</sup>のような参照モデルがおそらくより良い開始点となるだろう。参照モデルは（例えば、JSR 170/283とJavaの関係のような）特定の技術との関係がなく、また、リポジトリに組み込むべきものより抽象的な記述を提供するからである。

JSR 170/283は複合型オブジェクトデータモデルを使用しており興味深い。MPEG-21 DIDLのように、リポジトリとそのコンテンツにとって価値があると思われる明確なシンタックスを提供する機会を提供するからである。メタデータ標準あるいはパッケージング標準としての相互運用性はシンタックスに限られており、セマンティックについては別問題である。また、相互運用性は問題を隠蔽することによってもおそら

く達成することができるだろう。DR OSIDは、より深いレベルでの相互運用性を必要とすることなく、また、おそらくその深いレベルにおけるニーズや問題をあいまいにすることなく相互運用性を達成できる機能を持つラッパーを提供する。このアプローチは従来のシステムでうまく機能するので、リポジトリ仕様は広範囲な連合にそのようなシステムを含めるための貴重な方法を提供するだろう。

一般に、また、学習オブジェクトリポジトリの分野においてさえ、これらの仕様の実装率は低いことは注目に値する。相互運用性が効果を発揮するためには、仕様の幅広い採用が必要である。この仕様は比較的複雑であり、その実装と保守には努力が必要となるので、全世界的というよりローカルの定義され管理されたシナリオにより適している。これらの仕様はその役割を持っているが、その役割は、実装に必要な努力を正当化するするために厳密に定義される必要がある。

## A5.6 Webサービスインターフェース

リポジトリで利用できるインターフェースに関してこれまで検討してきたが、Webサービスについては特に言及しなかった。しかし、これはリポジトリ横断型サービスを提供するための重要な構成要素であり、既に使用されている。これまでに説明したインターフェース標準の多くはレベルに違いがあるとはいえWebサービスであると考えることができる。特に、SRW/Uの開発は、Web上でZ39.50の機能を利用可能にするものであり、両者（SRWとSRU）は共にWebサービスである。SRWはメッセージの転送にSORPを使用しており、SRUはRESTベースのWebサービスである。RSSとATOMもWebサービスプロトコルであり、Webサービスの生産者と消費者の間でXMLメッセージを交換するよう設計されている。OAI-PMHも同じ理由でWebサービスと考えることができるが、ある種の定義に厳密には適合しないかもしれない。リポジトリ仕様はWebサービス環境において使用することができる。

これらのインターフェースが提供する機能は対象が限定されており、主に単方向性であるので、Webサービスとしての能力には限界がある。たとえば、RSSは通知サービスにはふさわしいが、これに付随するサービスを付加するにはより複雑な相互作用が必要となる。たとえば、双方向コミュニケーションを可能にするSimple Sharing Extensionsの使用である。また、Webサービスをうまく稼働させるためには高品質のメタデータが不可欠である。しかし、柔軟なサービス指向の方法によるより複雑な相互作用を容易にするために、まずWebサービスを軽量な基礎の上に構築することから始めることができるだろう。柔軟なアプローチは、極めて重量級であると考えられるリポジトリ仕様の目的と対立する可能性がある。両アプローチの利点と欠点を完全に評価するためにはWebサービス環境において様々なインターフェースを使ったさらなる実験が必要である。

Webサービスは、あるシステムから別のシステムへのコンテンツの真のプッシュ機能を実現することができる。RSSはプッシュメカニズムであると考えられているが、アグリゲータやエンドユーザーサービスから開始する必要がある。一方、Webサービスは、自動的に起動される別の基準に基づいてコンテンツをプッシュできるようにすることができる可能性がある。リポジトリ仕様と同じように、リンクの両端（リポジトリとサービス/アプリケーション）に、複雑な相互作用が効果的に行われる条件が知らされていれば、Webサービスは役に立つ。Webサービスプロトコルはまだあまり成熟していないので、管理された環境以外に幅広く適用することはできない。たとえば、カリフォルニアデジタルライブラリはWebサービスを使用しているが、今のところ主に内部的な利用に留まっている。また、Webサービスを使用する場合も、技術が変化してもサービスが有効であることをできるだけ保証するためにSORPに依存することはできるだけ避けている。

リポジトリソフトウェアとの相互作用を支援するさらなるWebサービスが登場している。DSpaceは、投稿、検索、撤回機能を持つ「軽量ネットワークインターフェース」と呼ばれる一連のRESTベースのWebサービスを開発している<sup>120</sup>。これは研究者が資料をまずリポジトリに投稿し、次に、部局のWebページなど別の場所に資料を表示するためにメタデータを抽出できるように設計されている。リポジトリがコンテンツ所有者に提供できるものを使ってコンテンツ所有者が自ら付加価値を与える気になるようにインターフェースは使用されている。Fedoraデジタルリポジトリシステムはリポジトリとのすべての相互作用を4つの異なるWebサービスAPIインターフェース（API-M、API-Aとそれらの簡易版）に頼っている<sup>121</sup>。また、Webサービスはリポジトリとその他のエンタープライズシステムとの間のコミュニケーションを図るためにも使用することができる。エジンバラ大学はIRRAプロジェクト<sup>122</sup>と共同で、大学のDSpaceリポジトリと研究情報システムとの相互作用を容易にするWebサービスの開発に取り組んでいる。また、同様な方法でリポジトリを大学の人事システムへリンクすることも検討している。

これらの具体的なWebサービスに加えて、軽量アプローチを使用してWeb上やWebとデスクトップの間で

大きな相互作用を可能とするWeb 2.0アプローチが最近流行している。

## A5.7 Web 2.0

Web 2.0 (O'Reilly, 2005) は、1つのアプローチとして、Web上で動的に相互作用する能力とコンテンツを様々な用途に再利用可能にする能力を重視する。これらの仕組みは、Web 2.0という名前で新たにまとめられているが、以前から存在しているものであり、本報告書で既に説明した多くの開発の基礎となっている。また、幅広い利用を得るためにコンテンツを公開したいというオープンアクセスコミュニティの要望にも合致している。上に挙げたインターフェース標準はすべて動的な相互作用とコンテンツの再利用を可能にする。たとえば、OpenURLは動的なリンクを可能とし、OAI-PMHとRSSは様々な目的に再利用できるメタデータのアグリゲーションを可能にする。アグリゲーションはエンドユーザサービスを構築するためのプラットフォームを提供する。2大サービスと呼ばれているGoogleとAmazonはこれを理解し、自らが所有する情報の大規模なアグリゲーションを使用して付加価値サービスを提供するだけでなく、独自のAPIを公開することにより他の人にも同じことができるようにしている。これらの洗練された接続ポイントは柔軟な組み合わせを可能としている。

OAI-PMHやOpenURLなどの軽量インターフェースの潜在能力を最大化することには、これらを組み合わせで最大の効果を得ることも含まれる。Web 2.0の名の下に実装されている主要な技術はAJAXである (Garrett, 2005) が、これ自体独立した既存のプロトコルの組み合わせである。RSSリーダーアプリケーションにRSSをフィードするように、Webとデスクトップの間で相互作用が可能になると、エンドユーザにとって最も役に立つような相互作用も容易になる。OpenURLリンクを自動的に生成するためにCOinS<sup>123</sup>をWebページに埋め込むことは、既存の標準を柔軟に使用してエンドユーザとの相互作用を容易にする方法のもう1つの例である。

相互作用ができると、自分の見つけたメタデータやコンテンツに情報を付加する機会をエンドユーザに提供することになる。タグ付けやアノテーションは個々のエンドユーザが資源を後で再利用することを助けるが、de.icio.usやAmazonなどのサービスを通じて、他の人にも利益を与えることができる。この追加メタデータの作成には、(この新たなサブコンポーネントの識別を必要とする) 明確に定義された方法で記録し、保管する機能を必要とするが、他の(おそらくはパーソナル化した) エンドユーザサービスを構築するための基礎の役割を果たすことができるよりリッチなりポジトリを導くことにもなる

Web 2.0アプローチは多くの機会を提供する。しかし、このアプローチが約束する素晴らしいプレゼンテーションやエンドユーザサービスを、特に学術環境において実現するには、そのための頑強な基礎が必要となる。そこにはスキルの問題、エンドユーザサービスがアクセシビリティの要件を満たすことを保証する必要、学術情報やコンテンツの複雑性を「進化した超単純な世界」へ移行する明らかな方法を見つける必要がある。また、Web 2.0は、我々が何年も前から持つてはいたが完全には活用していなかった要素を使ってできることを再考することでもある。

## A6 アーキテクチャ

メタデータとインターフェースに関する具体的な問題とそれらがエンドユーザサービスの構築に与える影響については既に検討したが、「リポジトリからエンドユーザサービスに至る」連鎖を実現し持続するためにそれに関わる様々な構成要素をまとめる方法についても検討する必要がある。リポジトリの全体的なアーキテクチャの具体的な例が最近提案されており、特にaDOReとCORDRAは注目に値する。リポジトリとリポジトリ横断型のサービスを計画する際には、アーキテクチャの一般的側面と並行してこれらも検討するべきである。

### A6.1 アーキテクチャの選択肢

リポジトリからエンドユーザサービスへの連鎖における様々な構成要素を組織化する際に採用可能なアーキテクチャ上のアプローチは数多く存在する。表7はこれらを簡単にまとめたものである。

アーキ  
テク  
チャモ  
デル

備考

集中型	コンテンツとメタデータをすべて1つの場所に収集し、エンドユーザサービスは直接そこを利用する。
分散型	コンテンツとメタデータは分散して保管され、エンドユーザサービスは必要に応じてその都度それらと相互作用する。
ハーベ スト型	メタデータと可能であればコンテンツをプル技術を使って定期的に1箇所に収集し、エンドユーザサービスにアクセスを提供する。エンドユーザはまずこのアグリゲーションを利用し、必要に応じて情報元のリポジトリに導かれ詳細な情報を得る。
プッ シュ型	メタデータと可能であればコンテンツをプッシュ技術を使って定期的に1箇所に収集し、エンドユーザサービスにアクセスを提供する。エンドユーザはまずこのアグリゲーションを利用し、必要に応じて情報元のリポジトリに導かれ詳細な情報を得る。
ピア ツー ピア 型	個々のリポジトリが互いに、またエンドユーザサービスと対等に相互作用する。

表 7: リポジトリからエンドユーザサービスへの連鎖を組織化するアーキテクチャモデル

ピアツーピア型を除いて、他の4つのアプローチは程度の差はあれ、その組織化において、最上層に中央アグリゲータあるいはエンドユーザサービスを持つ階層性を持っている。ピアツーピア型では、階層的組織を持つことはなく、すべての構成要素は対等であり互いに相互作用する。

どのアプローチにも利点と欠点がある。検討する必要がある2つの主要な問題は、更新遅延とスケーラビリティである。全体的なアーキテクチャにおいてリポジトリの連合レベルが高くなると、コンテンツの更新が遅延する可能性が高くなる。これは、アーキテクチャの問題であると同時に連合をどのように組織化するかという問題でもあるが、スケーラビリティも影響するかもしれない。連合はまたスケーラビリティやシステムのパフォーマンスにも影響を与える。

アー キテ ク チャ モデ ル	更新遅延	スケーラビリティ
集中 型	中央に位置する階層の組織に依存する。理論的には小さいが、実際には大きい可能性がある。	中央に保管するレコードのサイズに関連するスケーラビリティの問題が発生する可能性がある。すなわち、データ量が大きくなると相互作用に時間がかかるかもしれない。
分散 型	リポジトリは必要のあるときにしか接続されないし、常に最新の情報を持っているので更新遅延は小さい。	分散検索の経験からスケーラビリティの問題があることが示されている。たとえば、SRWが最大の効果を発揮するのはおそらく10ターゲット以下である。対象を絞った検索に向いている。
ハー ベ ス ト 型	前回のハーベスト後にリポジトリが更新されることがあるので更新遅延の問題が発生する可能性がある。ただし、ハーベストの間隔を短くすればこの問題を減少させることができる。	メタデータだけをハーベストしている場合はスケーラビリティの問題は小さい。コンテンツもハーベストする場合は、この問題は大きくなる。
プッ シュ 型	リポジトリは情報が更新されるたびにプッシュするので更新遅延の問題は考えられないが、アグリゲータやエンドユーザサービスがプッシュされた情報を取り込むのが遅い場合は更新遅延の問題が発生する可能性がある。	メタデータだけがプッシュされている場合はスケーラビリティの問題は小さい。コンテンツも含まれている場合は、この問題は大きくなる。
ピア ツー ピア 型	対象となるノードが多い場合はノード間に情報が広がるために時間がかかるので、更新遅延の可能性が大きい。	スケーラビリティの問題については議論が続いている。未知の要因がある場合は、さらなる研究を必要とする。

表 アーキテクチャモデルにおける更新遅延とスケーラビリティの影響

ここで注意すべき重要な要因は、これらのモデルにおいて当該システムの保守作業がどこで行われるかということである。集中型とハーベスト型は中央の負荷が相対的に高くなり、プッシュ型、分散型、ピアツーピア型の負荷はリポジトリ側でより高くなる（ただし、後者の場合、各ノードが決められた作業を行うように中央の調整が必要になることに注意が必要である）。作業負荷がかかる場所ではエンドユーザサービスが適切に提供されることを保証するために必要な調整作業を行わなくても済むようにバランスを取る必要がある。

各アプローチは自らの役割と置き所を持っている。集中型アプローチは、保存やコンテンツの品質向上などコンテンツ自体を使った多様な活動を可能にする。一方、ハーベスト型アプローチは、コンテンツもハーベストしていればこれも可能であるが、メタデータにより重点を置き、エンドユーザサービスを構築するための基礎となるアグリゲーションを作成する軽量な方法を提供する。プッシュ型アプローチはハーベスト型のプル機構とは反対の機構を提供するが、公開するメタデータをリポジトリ側でより大きくコントロールすることが可能である。分散型アプローチは主に発見サービスを対象としている。ピアツーピア型は、発見サービスや関連のエンドユーザサービスも支援するが、潜在的には、特にバックアップなどのコンテンツを使った一連の活動を実現する際に大いに効果を発揮する。また、ピアツーピア型アプローチは、もしかすると基礎となるアーキテクチャに最も堅牢な実装を必要とするかもしれない。LionShareピアツーピアシステム<sup>125</sup>を実装したSPIREプロジェクト<sup>124</sup>の経験から、遭遇するかもしれないいくつかの問題が明らかになっている。

オープンアクセスリポジトリに関して言えば、ハーベスト型アプローチがこれらのリポジトリが所有するメタデータとコンテンツへのアクセスとその管理の両者を支援するための最も幅広い機能を提供する。プッシュ型も代替アプローチとして優れているが、現状の標準ではサポートされていない。プッシュ型であると考えられているが技術的にはプル型で実装されているRSSは、ほとんど同じ機能を提供するものである。オープンアクセスリポジトリ側でより大きなコントロールが必要となる場合は、ピアツーピア型も代替アプローチとなる可能性がある。実際、より幅広い公開のためにはハーベスト型やプッシュ型による公開が、また、より（詳細に）管理された公開のためにはピアツーピア型の公開が利用できるようになれば、リポジトリにとっては非常に役に立つだろう。Web検索エンジンで採用されている集中型アプローチも可能であるが、これを効果的に運用するためには商用検索エンジンが提供できるような大量の資源を必要とする。分散型アプローチはオープンアクセスリポジトリ横断型のエンドユーザサービスを支援するための単独のアプローチとしては最適なものではない。しかし、これらのリポジトリを他の情報資源と共により広い情報環境に置く場合は検討する必要がある。このような環境においては、分散型検索が存在する意義が十分にあるだろう。しかしながら、最初にオープンアクセスリポジトリのメタデータをアグリゲートすれば、分散型アプローチが持つスケーラビリティの問題を制限することができるだろう。

## A6.2 中間的共通基盤サービス

リポジトリ、アグリゲータ、エンドユーザサービスは、リポジトリとエンドユーザを結ぶ情報連鎖における<sup>3</sup>つの主要な環であると考えられる。これらの環の仲介者として機能し、必要に応じて共有することができる追加サービスにより、三者の間の相互作用を支援することができる。このようなサービスの多くは本報告書で既に言及しているが、この節では、これらのサービスが相互作用に追加することができる価値を具体的に説明する。そのようなサービスは数多く提案されているが、その一覧を表<sup>9</sup>に示した。

中間サービス	役割	価値
サービスレジストリ	リポジトリとアグリゲータで利用可能なサービス（インターフェース）のリスト	相互作用のためにリポジトリが公開しているものと相互作用の仕方をアグリゲータやエンドユーザサービスが発見できるようにする。
コレクションレジストリ	リポジトリが所有するコレクションに関する情報のリスト	リポジトリにどんなコレクションが所有されているかをアグリゲータやエンドユーザサービスが発見できるようにし、これらとの相互作用を調整する。
フォーマット	オブジェクトに関するフォーマットのリストとそれ	ファイルフォーマットに関する情報とこれを管理する方法をアグリ

レジストリ	の関連情報	ゲータやエンドユーザサービスに提供する。
ライセンスレジストリ	オープンアクセスなどの環境で使用できるライセンスのリスト	リポジトリがメタデータやコンテンツと標準的なライセンスを関連付けられるようにし、アグリゲータやエンドユーザサービスがメタデータやコンテンツを使用できる、あるいはできない条件を知ることができるようにする。
メタデータスキーマレジストリ	メタデータスキーマと存在する場合はアプリケーションプロファイルのリスト	リポジトリがローカルで使用する適切なメタデータスキーマやアプリケーションプロファイルを識別できるようにする。場合によっては、リポジトリ間の相互運用性を容易にするスキーマ間のマッピングを可能にする。アグリゲータやエンドユーザサービスがリポジトリとの相互作用を構築ために使用することができる。
メタデータクロスウォークサービス	検索式の構築や拡張を支援するためにメタデータスキーマ間のマッピングを容易にするサービス	リポジトリ間の相互運用性を促進する。また、適切なマッピングによりアグリゲータやエンドユーザサービスがリポジトリとの相互作用を構築できるようにする。このサービスはスキーマレジストリに含まれることになるかもしれない。
専門用語/主題典拠サービス	検索式の構築や拡張を支援するために専門用語間のマッピングをしたり主題語の典拠の役割を果たしたりする専門的なサービス	メタデータクロスウォークサービスと同じように、専門用語サービスは、様々な主題語彙間のマッピングや統制語による検索を行うことにより、リポジトリとの相互作用（主に発見）を構築したり、拡張したりすることを可能にする。
著者名典拠サービス	コンテンツの管理やアクセスを支援するために所有する著者名典拠リストに基づいてマッピングを行うサービス	専門用語サービスと同等なサービスで、著者名を対象を絞ったものである。しかるべき名前情報の配置場所により集中型サービスか分散型サービスのいずれかになる。
識別子解決サービス	オブジェクトの所在場所の特定を支援するために識別子の解決をするサービス	識別子を解決することによりオブジェクトの実際の配置場所を知ることができる
関係情報サービス	分散されたオブジェクト間の関係を確立するサービス	オブジェクト間の関係に関する情報を提供することにより、アグリゲータやエンドユーザサービスがリポジトリとの相互作用ができるようになる。
フォーマット変換サービス	オブジェクトのフォーマットを変換するサービス	オブジェクトのフォーマットを配信の都度その場で、あるいは保存プロセスの一部として変換できるようにする。
メタデータ生成サービス	リポジトリやアグリゲーションをリッチにするためにメタデータの自動生成を支援するサービス	エンドユーザサービスを通じた提供を容易にするためにリポジトリやアグリゲータが所有するメタデータを高度化することができるようにする。
正規化サービス	発見を容易にするためにリポジトリが所有するメタデータの規格化を支援するサービス	エンドユーザサービスを通じた提供を容易にするためにアグリゲータがアグリゲートしたメタデータを統一したり組織化したりできるようにする。
ランク付け/アーノーションサービス	既存のレコードにユーザーが作成したメタデータを追加できるようにするサービス	エンドユーザが相互作用しているオブジェクトにメタデータを追加し、オブジェクトと関連させて格納できるようにする

表 9: オープンアクセスリポジトリ横断型のエンドユーザサービスを支援する中間サービスの候補者リスト

これらの中間的共通基盤サービスは主にリポジトリとエンドユーザサービスを結ぶ連鎖における<sup>3</sup>つの主要構成要素の間に位置し、これらとマシンツーマシン・インターフェースにより相互作用する。既存の開発から、たとえば、Information Environment Service Registry (IESR) [126](#)へのユーザインターフェースのように、中間サービスへのユーザインターフェースをさらに提供することが有用であることを示唆されている。ただし、そのようなユーザインターフェースがそれ自体でエンドユーザサービスとして有効であるのか、あるいは、中間サービスがどのように機能するかを「知る」ことができるから必要とされるのかは依然として明らかではない。後者のニーズはそれ自体で有効であり、中間サービスがリポジトリとアグリゲータまたはエンドユーザサービスとの間でどのように機能するかを立証することができる可能性がある。このような立証により、マシンインターフェースを提供するために使用される相互運用可能なインターフェース標準は望ましい機能を発揮することができ、有益であるという信頼を形成することができる。識別子リゾルバの具体的な例であるOpenURLリゾルバは、JISC情報環境におけるプレゼンテーション層に位置し、中間サービスがこの役割をどのように果たすことができるかを示す一例であり、また、エンドユーザサービスとなる中間サービスの一例でもある。リゾルバはユーザーに（OpenURLを管理したいかどうかという）選択要素を提供するが、このユーザーによる選択が有用である場合に、中間サービスをエンドユーザサービスとして位置付けることが有効となる。

仲介者として機能する際、これらの共通基盤サービスはリポジトリ相互やリポジトリとアグリゲータまたはエンドユーザサービスとの間の橋渡しを支援することができる。（たいていは利用できる構成要素を知らないだけであるが）そのような連携を確立することがそれほど容易でないかもしれない場合、サービスレジストリやコレクションレジストリの役割は特に重要である。リポジトリとアグリゲータまたはエンドユーザサービスとの相互作用に中間サービスを介在させることにより、より効果的な関係を作り出すことが可能になる。しかし、その安定性や考えられるサービスに対する依存性を監視し、注意深く開発する必要がある。なぜなら、エンドユーザが期待するレベルの機能を妨げているものが彼らの知ることができない要素であったとしたら、エンドユーザはサービス障害に耐えられないからである。

中間サービスを1つのサービスとして効果的に利用するためには、これらの間の関係をさらに研究する必要もある。これは、同じ種類のサービス（IESRとOpenDOARといった異なるサービスレジストリなど）だけでなく、異なる種類のサービス（サービスレジストリとライセンスレジストリ、あるいは、専門用語サービスと典拠サービスなど）の関係にも適用すべきである。そのような連携が同じサービスに組み込まれるかもしれない。たとえば、IESRはサービスレジストリとコレクションレジストリの2つの役割を持っている。

中間サービスは、それら相互、あるいはレジストリやアグリゲータとの間の相互運用性を最大にするために一般に認められている標準を使って構築する必要がある。しかし、実際には、中間サービスの有用性はほとんどテストされていないので、オープンアクセスリポジトリとそのようなサービスを組み合わせた研究が必要とされる。

### A6.3 認証と承認

先の節で示したように、オープンアクセス環境においては原則として認証と承認は必要ではない。しかし、制限が必要な場合やある程度アクセスをコントロールするための仕組みが必要な場合もある。また、オープンアクセスリポジトリをより広い情報環境に置く場合やオープンアクセス情報資源とクローズドアクセス情報資源との間で相互作用が生じる可能性についても検討する必要がある。

アクセスに制限を設ける際には、制限を適用する粒度を検討する必要がある。役割に基づいて個人レベルでアクセスを認めるのか、あるいは、グループレベルでアクセスを管理するのか。このアプローチはリポジトリのオブジェクトにも適用することができ、オブジェクトにより異なるアクセスレベルを持つことも可能である。Shibboleth [127](#)は両レベルの承認を実装する可能性を提供するが、機関はこの利点を最大に生かし、初期認証を提供するためにしかるべき適当なID管理システムを持つ必要がある。アクセス同様、機関リポジトリでは特に、投稿や削除などの管理業務にも認証は必要である。関連する認証・承認システムを計画する際には、システムが要件を満たすことを保証するために関連するモデルによりこれらを根拠つけることが重要である。セキュリティやアイデンティティのモデルは存在するが、要件を満たすほどこれらの適用範囲が詳細であるかは明らかではない。

オープンアクセス環境において認証と承認が必要であろうと欠落していようと、これらの要因がオープン

アクセス配信にどのように影響を及ぼすかをより良く理解するために、一般的な信頼と来歴に関する問題とモデルを研究する必要がある。オープンアクセスはこれらが無くては機能しないとは言わない——もちろん、機能している証拠は非常に多い——が、これらは、追加情報やデータに対するさらに高度なオープンアクセスサービスを開発するためにも、オープンアクセス自体が時間が経っても完全に信頼できるものであることを保証するためにも、考慮に入れる必要がある問題である。

#### A6.4 サービス指向アーキテクチャ

一般に、システムの計画に関する最近の関心のほとんどは、サービス指向アーキテクチャ (SOA) の使用とこれが推進する個々のサービスコンポーネントの組み合わせにある。SOAアプローチは、リポジトリ横断型サービスの提供を考える上で、それに向けた努力が価値を持つようなレベルの柔軟性を提供する。リポジトリ、アグリゲータ、エンドユーザサービス、中間サービス、これらすべては、それぞれが異なるサービスインターフェースを提供する個々のサービスコンポーネントであると考えられることができる。これらの間の柔軟な相互作用は、エンドユーザのニーズをより満足させるようにリポジトリのメタデータやコンテンツを公開することを約束する。6.1節で示したアーキテクチャ選択肢を考えると、SOAアプローチはすべてに採用される可能性もあるし、どれにも採用されない可能性もある。本研究の結果は、SOAは長期的には非常に好ましいことを示している。

しかし、SOAは単に実装できるようなアーキテクチャではなく、システムが実現することを期待するものを分析するために必要なアプローチであり、どんなサービスコンポーネントが必要になるかを決めるための設計段階である。したがって、長い時間をかけてそれに向かって努力することが効果的なアーキテクチャであり、アプローチである。このアプローチは、すべてのコンポーネントが関連する標準を（最大の効果を得るには厳密に）順守する必要があること、および、機関をまたぐ連合環境においては調整が必要となることにより影響を受ける。

調整が必要であることは、現在開発中のe-Framework for Education and Research<sup>128</sup>を導いた。このイニシアティブは、英国のJISCとオーストラリアのDESTにより支援され、アメリカとカナダ、ニュージーランドが参加している。この多国的アプローチは、SOAに向けた移行とこの移行を実現することができる調整の枠組みを提供することがどちらも時間がかかることを認識している。

SOAは、相互運用性の可能性を増加させる一方で、個々のコンポーネントを柔軟に置き換えることにより大きな持続可能性も提供する。SOAは、リポジトリコンテンツとサービスが「相対的には不変な配管の上に緩やかに配置されることにより、軽量でユーザーに合ったコレクションや投稿、組織化機能を支援する」<sup>129</sup>というシナリオに向かっている。そのプロセスは緩やかなものであってもSOAは実装される必要があり、現在この方向に向かっている組織も存在する。California Digital Libraryは、時間の経過と共に必要に応じて機能を追加するプラグインプレイアプローチを確立した。SOAは、タスマニアで行われた学習と教育を対象とするLeAPイニシアティブ<sup>130</sup>でも取り上げられている。従来、学習と教育目的ではモノリシックなシステムを使用してきたが、LeAPは、はるかに細かい粒度を可能にする一連のコンポーネントシステムに置き換えることによりこの考えを打破した。その後、この粒度はユーザーのニーズに合わせてシステムを組み合わせる方法に大きな柔軟性を与え、特に同一のコンテンツに対し様々なエンドユーザビューを可能にした。

LeAPアーキテクチャには数多くのリポジトリが存在するが、この粒状アーキテクチャは単一の中央貯蔵庫の存在を排除するものではない。このアーキテクチャは中央リポジトリ貯蔵庫を生み出す複数のサービスをもつ逆ピラミッド型であるかもしれない。実際、コンテンツを1回格納し、何回も使用するという利点が存在する (WORMアプローチ)。アグリゲーションもまたそこから多くのサービスが現れる単一の貯蔵庫である。しかし、粒状アプローチは、個人的なコンテンツ貯蔵庫を含む全体的なアーキテクチャの中に他のリポジトリを取り込むことを可能にする。

SOAは、その性格がまさにWeb 2.0であるが、有用な長期的目標であり、LeAPはこれが教育という文脈において実用的であることを示している。様々なコンポーネントが分離され、アーキテクチャの階層の数が増加しているので、円滑かつ効果的な相互作用を可能にするためには、これらの層がどのように相互作用するのか、および、各層のサービスをどのように統合するかについて理解する必要がある。この分野のさらなる理解は必要であるが、これは、エンドユーザニーズに合わせて必要に応じて効果的に改造することができるアーキテクチャを得ることにもなる。

#### A6.5 リポジトリサービス層

インタビューを行った者の多くは、全体的なアーキテクチャがリポジトリやそれを設定する方法に影響を及ぼすとは考えていなかったと述べた。リポジトリはコンテンツやメタデータの貯蔵庫の役割を果たすが、リポジトリが提供するインターフェースは、理想的にはオープンな標準に基づいているべきであるが、他のコンポーネントと相互作用する公開されたサービスの役割を果たす。このインターフェースは、理想的にはアグリゲータやエンドユーザサービスからリポジトリを隠すものであるべきである。多くの場合、インターフェース標準は個々のリポジトリレベルで実装されている（たとえば、OAI-PMHインターフェースはリポジトリソフトウェアに組み込まれている）。しかし、リポジトリインターフェース層あるいはリポジトリサービス層として別個の階層として扱えば、これらをそれとはわからない方法で様々なリポジトリにサービスを提供する中間サービスとして使用することができる可能性がある。

このようなリポジトリサービス層は、ジョン・ホプキンス大学のSayeed ChoudhuryらによるTechnology Analysis of Repositories and Servicesプロジェクト<sup>131</sup>の中で汎用アーキテクチャアプローチとして提案されている。このリポジトリとアグリゲータまたはエンドユーザサービスの分離は、両グループがサービス層との相互作用に焦点を絞ることができることを意味する。共通のサービス層を持つことにより、そして、理想的には少数の適切なAPIを採用することがベストであるが、リポジトリを横断する相互運用性を確立することが可能となる。また、各コンポーネントがサービス層と通信することができれば、必要に応じてコンポーネントの交換や置換ができるSOAアプローチも可能となる。DS OSIDなどのリポジトリ仕様は、OAI静的リポジトリゲートウェイ<sup>132</sup>が行っているように、この方法で機能することができる。これらの仕様により、リポジトリは管理しなければならない相互作用を仕様それ自体との相互作用だけに限定することができ、アグリゲータやエンドユーザサービスとの連絡を管理する必要がなくなる。すなわち、仕様が両者の間の仲介者の役割を果たすことになる。

このように注意を向ける対象を絞り込むことにより、リポジトリはコンテンツへのアクセスではなくコンテンツの管理に注力できるようになる。リポジトリは、エンドユーザサービスが持っている要求や要件を満たすためにこれらを知る必要があるが（エンドユーザサービスはリポジトリがそれを提供しなければ何も構築することができない）、アクセスの管理は中間サービス層に任せることができる。このリポジトリとその他のコンポーネントの分離は互いの依存性も減少させる。オブジェクトレベルでは、オブジェクト自体のフォーマットにまでこれを拡張することができる。ただし、通常これは変更の有無とは関係なくすべての相互作用に組み込まれている。Fedoraは、保管と配信の両目的のために複数のフォーマットにその場で変換する機能を持っているが、そのためにはリポジトリを構築する際に注意深い計画と設定作業が必要となる。aDOReが採用したPathways InterDisseminatorは、オブジェクトのタイプ依存性を排除し、エンドユーザの要求に応じて柔軟にフォーマットを提供する可能性を提供するもう1つのメカニズムである。

エンドユーザサービスをその基礎となるコンポーネントから分離することは、リポジトリだけでなくアグリゲータにも適用できる。OAIsterへのSRUインターフェースの追加は、同様なアグリゲータサービス層はアグリゲータがそのタスク（アグリゲーション）の管理に集中することを可能にし、一方で、管理された方法でそのコレクションを公開するという意味で有用であるかもしれないことを示唆している。

#### A6.6 ローカルインフラストラクチャへのリポジトリの統合

リポジトリは単独で考えることはできず、より広いインフラストラクチャの文脈に置かなければならない。理論レベルでは、この分野には既にいくつかの活動が存在し、そのような統合を容易にするための指針を提供しているOAIsterやDFL Services Framework<sup>133</sup>などの参照モデルが存在する。機関における実践レベルでは、リポジトリを、たとえば、コンテンツ管理システムや仮想的学習研究環境、現行の研究情報システム、機関ポータルなどの文脈に位置付けることを含んでいる。リポジトリの役割は必要となる統合や相互作用のレベルに影響を及ぼすことになる。たとえば、オープンアクセスリポジトリは研究成果のアーカイブとなることができるが、これは研究プロセスの最終段階でのみ使用されるものであり、このレベルで必要とされる統合の度合いは低いものである。リポジトリを複数のコンテンツ種別のために使用したり、日々の作業ツールとして使用したりする場合は、リポジトリがこの役割を果たせるようにするためにより密接な統合が必要になる。エンドユーザが相互作用を行うプレゼンテーションレベル（ポータルに検索や投稿画面を表示するなど）やシステム間で情報を交換するデータレベルに統合の焦点を絞ることもできる。

ローカルレベルでリポジトリの使用を推進するためには、ユーザーにリポジトリを発見してもらうのではなく、リポジトリをユーザーの元へ届けることが重要である。これは様々なシステムにリポジトリへのアクセスを埋め込むことにより実現することができる（これは同時にリポジトリが単なる慣れるべき「もう1つシステム」と見なされることを防ぐことにもなる）。機関ポータルや仮想的学習環境、図書館目録

を通じてリポジトリサービスを表面化することは、たとえば、ポートレットやWebサービスを使うことにより可能である (Awre, 2005) が、望ましい方法であるかもしれない。また、リポジトリが、Wordなどの一般に使用されている編集ソフトウェアからアクセスでき、情報の取り込み、編集、保存、削除が簡単にできるようになれば便利であろう。

機関で採用されている全体的なアーキテクチャはリポジトリの統合の難易に大きな関係を持つことになる。アーキテクチャがSOAアプローチなどを使用することにより階層化されており、これらの間に明確に定義されたインターフェースが構築されていれば、リポジトリのような追加コンポーネントを必要に応じて組み込むことが可能である。しかし、逆に、他のシステムからリポジトリ機能の設計や公開方法が影響を受けるかもしれない。たとえば、受入機能の設計は、学習オブジェクトを受け入れて分解するという仮想的学習環境のニーズにより決められる可能性がある。

様々なシステムのリポジトリに対する要求はそれぞれ異なっているので対処がより複雑になる。先の節で説明したオープンな標準を使ったインターフェースを、システムがそのようなインターフェースを利用できると仮定すれば、管理システムなどとの相互作用に使用することができる。これらは必要となる相互作用の種類を考えると軽量に過ぎるかもしれないが、より詳細で構造的なりポジトリ仕様 (5.3節参照) を使用することでより大きな機能性を提供することができる。この状況は、リポジトリ仕様が提供する独立したリポジトリサービス層が両者の仲介者としての役割を果たし、各コンポーネントが処理しなければならない相互作用を単純化するという事例を助長する。

リポジトリが幅広いコンテンツを扱う場合、リポジトリがその内部的・外部的な基礎となるシステムやサービスへコンテンツを供給する役割を果たすことができれば有用である。効率的な相互作用にはコンテンツの明確な識別が不可欠である。原則的には、識別子を解決する識別子解決サービスが存在すれば、明確な識別によりコンテンツをほとんどどこにでも置くことが可能になる。これが利用できるようになるまでは、リポジトリはコンテンツを一緒に持ち込むことにより利益をもたらすことができ、機関の様々な役割の間のクロスオーバーを助長することができる。たとえば、従来は識別されていなかったクロスオーバーである教育と研究の両目的でコンテンツを使用することが容易になる。

## 参考文献

- Awre, C., Waller, S., Allen, J., Dovey, M.J., Hunter, J. and Dolphin, I. Putting the Library into the Institution: Using JSR 168 and WSRP to Enable Search within Portal Frameworks. Ariadne, October 2005, Issue 45. Available at <http://www.ariadne.ac.uk/issue45/awre/>
- Bekaert, J., Hochstenbach, P. and Van de Sompel, H. Using MPEG-21 DIDL to represent complex digital objects in the Los Alamos National Laboratory Digital Library. D-Lib Magazine, November 2003, 9 (11). Available at: <http://www.dlib.org/dlib/november03/bekaert/11bekaert.html>
- Bibliographic Services Task Force. Rethinking how we provide bibliographic services for the University of California: final report. December 2005. Available at <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>
- Blanchi, C. DVIA Registry Architecture. Presentation given at the 2006 Defense Technology Information Center Conference.
- Campbell, D. ARROW Discovery Service Harvesting Guide. June 2005. Available at <http://arrow.edu.au/docs/files/harvesting.pdf>
- Duke, M. Delivering OAI records as RSS: an IMesh Toolkit module for facilitating resource sharing. Ariadne, October 2003, Issue 37. Available at <http://www.ariadne.ac.uk/issue37/duke/>
- Foster, N.F. and Gibbons, S. Understanding faculty to improve content recruitment for institutional repositories. D-Lib Magazine, January 2005, 11 (1). Available at <http://www.dlib.org/dlib/january05/foster/01foster.html>
- Garrett, J.J. Ajax: a new approach to Web applications. 2005. Available at <http://www.adaptivepath.com/publications/essays/archives/000385.php>
- Godby, C.J., Young, J.A. and Childress, E. A repository of metadata crosswalks. D-Lib Magazine, December 2004, 10 (12). Available at: <http://www.dlib.org/dlib/december04/godby/12godby.html>
- Halbert, M. DLF Aquifer Study on Institutional User Services.
- Heery, R. and Patel, M. Application profiles: mixing and matching metadata schemas. Ariadne, September 2000, Issue 25. Available at: <http://www.ariadne.ac.uk/issue25/app-profiles/>  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html>
- Jerez, H., Manepalli, G., Blanchi, C. and Lannom, L.W. ADL-R: the first instance of a CORDRA registry. D-Lib Magazine, February 2006, 12 (2). Available at <http://www.dlib.org/dlib/february06/jerez/02jerez.html>

- Lagoze, C. and Van de Sompel, H. The making of the Open Archives Initiative Protocol for Metadata Harvesting. Library Hi Tech, 21 (2): 118-128.
- Lynch, C.A. Institutional repositories: essential infrastructure for scholarship in the digital age. ARL Bimonthly Report 226, February 2003. Available at <http://www.arl.org/newsltr/226/ir.html>
- Lyon, L. eBank UK: building the links between research data, scholarly communication and learning. Ariadne, July 2003, Issue 36. Available at <http://www.ariadne.ac.uk/issue36/lyon/>
- Manepalli, G., Jerez, H. and Nelson, M.L. FeDCOR: an institutional CORDRA registry. D-Lib Magazine, February 2006, 12 (2). Available at <http://www.dlib.org/dlib/february06/manepalli/02manepalli.html>
- McCown, F., Liu, X., Nelson, M.L. and Zubair, M. Search engine coverage of the OAI-PMH corpus. IEEE Internet Computing, March/April 2006, 10 (2): 66-73. Preprint available at <http://library.lanl.gov/cgi-bin/getfile?LA-UR-05-9158.pdf>
- Nixon, W.J., Drysdale, L. and Gallacher, S. Search services at the University of Glasgow: PKP Harvester and Google. DAEDALUS project report, 2005. Available at <https://dspace.gla.ac.uk/handle/1905/425>
- O'Reilly, T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. 2005. Available at
- Powell, A., Day, M. and Cliff, P. Using simple Dublin Core to describe eprints, version 1.2. 2003. Available at <http://www.rdn.ac.uk/projects/eprintsuk/docs/simpledc-guidelines/>
- Sanderson, R., Young, J. and LeVan, R. SRW/U with OAI: expected and unexpected synergies. D-Lib Magazine, February 2005, 11 (2). Available at <http://www.dlib.org/dlib/february05/sanderson/02sanderson.html>
- Tourte, G and Powell, A. Encoding full-text links in the eprint jump-off page, version 1.0. 2004. Available at <http://www.rdn.ac.uk/projects/eprintsuk/docs/encoding-fulltext-links/>
- Van de Sompel, H. and Beit-Arie, O. Open linking in the scholarly information environment using the OpenURL Framework. D-Lib Magazine March 2001, 7 (3). Available at: <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>
- Van de Sompel, H. and Lagoze, C. The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine, February 2000, 6 (2). Available at <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>
- Van de Sompel, H., Bekaert, J., Liu, X., Balakireva, L. and Schwander, T. aDORe: a modular, standards-based digital object repository. The Computer Journal, 2005, 48 (5): 514-535. Preprint available at <http://arxiv.org/abs/cs.DL/0502028>
- van Veen, T. and Oldroyd, B. Search and retrieval in The European Library: a new approach. D-Lib Magazine, February 2004, 10 (2). Available at <http://www.dlib.org/dlib/february04/vanveen/02vanveen.html>
- van Westrienen, G. and Lynch, C.A. Academic institutional repositories: deployment status in 13 nations as of mid-2005. D-Lib Magazine, September 2005, 11 (9). Available at <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>
- Waters, D.J. The metadata harvesting initiative of the Mellon Foundation. ARL Bimonthly Report 217, August 2001. Available at <http://www.arl.org/newsltr/217/waters.html>
- Xiang, X, and Lease Morgan, E. Exploiting "light-weight" protocols and open source tools to implement digital library collections and services. D-Lib Magazine, October 2005, 11 (10). Available at <http://www.dlib.org/dlib/october05/morgan/10morgan.html>

## 脚注

1. arXiv, <http://arxiv.org/>
2. Securing a Hybrid Environment for Research Preservation & Access (SHERPA), <http://www.sherpa.ac.uk/>
3. White Rose Consortium ePrints Repository, <http://eprints.whiterose.ac.uk/>
4. BioMed Image Archive, <http://www.brisbio.ac.uk/index.html>
5. DSpace@Cambridge, <http://www.dspace.cam.ac.uk/>
6. Edinburgh Research Archive (ERA), <http://www.era.lib.ed.ac.uk/index.jsp>
7. Targeting Academic Research for Deposit and Disclosure (TARDIS), <http://tardis.eprints.org/>
8. Data-providers for Academic E-content and the Disclosure of Assets for Learning, Understanding and Scholarship (DAEDALUS), <http://www.lib.gla.ac.uk/daedalus/index.html>
9. Harvesting Institutional Resources in Scotland Testbed (HaIRST), <http://hairst.cdli.strath.ac.uk/>
10. Theses Alive!, <http://www.thesesalive.ac.uk/>
11. Registry of Open Access Repositories (ROAR), <http://archives.eprints.org/>
12. EPrints, <http://www.eprints.org/>

13. DSpace, <http://www.dspace.org/>
14. Organisation for Economic Co-operation and Development (OECD) Declaration on Access to Research Data From Public Funding, January 30, 2004  
[http://www.oecd.org/document/0,2340,en\\_2649\\_34487\\_25998799\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html)
15. AHDS, <http://www.ahds.ac.uk/>
16. ESDS, <http://www.esds.ac.uk/>
17. BADC, <http://badc.nerc.ac.uk/home/>
18. PubMed, <http://www.pubmed.gov/>
19. eBank UK project, <http://www.ukoln.ac.uk/projects/ebank-uk/>
20. Citation, Location, And Deposition in Discipline and Institutional Repositories (CLADDIER),  
<http://claddier.badc.ac.uk/>
21. Scoping a Geospatial Repository for Academic Deposit and Extraction (GRADE),  
<http://edina.ac.uk/projects/grade>
22. Repository for the Laboratory (R4L), <http://r4l.eprints.org/>
23. Submission, Preservation and Exposure of Chemistry Teaching and Research Data (SPECTRA),  
<http://www.lib.cam.ac.uk/spectra/>
24. Source-to-Output Repositories (StORe), <http://jiscstore.jot.com/WikiHome>
25. DILIGENT project, <http://www.diligentproject.org/>
26. EDINA Sound and Picture Studio, <http://www.edina.ac.uk/multimedia/>
27. CLIC study,
28. JISC user requirements study for a moving pictures and sound portal, [http://www.jisc.ac.uk/index.cfm?name=project\\_study\\_picsounds](http://www.jisc.ac.uk/index.cfm?name=project_study_picsounds)
29. JISC Exchange for Learning Programme, [http://www.jisc.ac.uk/programme\\_x4l.html](http://www.jisc.ac.uk/programme_x4l.html)
30. iLumina digital library of sharable undergraduate teaching materials for chemistry, biology, physics, mathematics, and computer science, <http://www.ilumina-dlib.org/>
31. SMETE Digital Library of teaching and learning materials, <http://www.smete.org/smete/>
32. Multimedia Educational Resources for Learning and Online Teaching (MERLOT),
33. DARC, <http://www.surf.nl/en/projecten/index2.php?oid=106>
34. Email from Les Carr to JISC-REPOSITORIES mailing list, 9th March, <http://www.jiscmail.ac.uk/cgi-bin/Webadmin?A2=ind0603&L=jiscrepositories&T=0&O=A&X=77C61153BB025CA0DC&Y=c.awre%40hull.ac.uk&P=3300>
35. SPARC Open Access Newsletter, December 2005, <http://www.earlham.edu/~peters/fos/newsletter/12-02-05.htm#topstories>
36. Directory of Open Access Journals, <http://www.doaj.org/>
37. Creative Commons, <http://creativecommons.org/>
38. University of Glasgow Library RSS feeds, <http://www.lib.gla.ac.uk/rss/>
39. Griff Richards, personal email communication, March 2006
40. Larry Lannom, personal communication, March 2006
41. OAister SRU configuration, <http://oaister.umdl.umich.edu/o/oaister/sru.html>
42. d+, <http://www.jisc.ac.uk/index.cfm?name=dplus>
43. Web Services for Remote Portlets (WSRP), [http://www.oasisopen.org/committees/tc\\_home.php?wg\\_abbrev=wsrp](http://www.oasisopen.org/committees/tc_home.php?wg_abbrev=wsrp)
44. Fedora authorization with XACML policy enforcement,  
<http://www.fedora.info/download/2.1b/userdocs/server/security/AuthorizationXACML.htm>
45. JISC CETIS Conference Repositories session, Heriot-Watt University, November 2005, <http://www.e-framework.org/events/conference/programme/repositories>
46. JISC Digital Repositories Programme Support Team deposit API activity,  
[http://www.ukoln.ac.uk/repositories/digirep/index/Deposit\\_API](http://www.ukoln.ac.uk/repositories/digirep/index/Deposit_API)
47. Augmenting interoperability across scholarly repositories, <http://msc.mellon.org/Meetings/Interop/>
48. SRW Update, <http://srw.cheshire3.org/docs/update/>
49. WebDAV, <http://www.Webdav.org/>
50. EThOS project, <http://www.ethos.ac.uk/>
51. Grokker, <http://www.grokker.com/>
52. Van de Sompel, H. Lessons in Cross-Repository Interoperability learned from the aDORe effort. Presentation at the 4th CERN Workshop on Innovations in Scholarly Communication (OAI4), 20th-22nd October 2005,  
<http://oai4.Web.cern.ch/OAI4/>
53. CORDRA, <http://cordra.net/>
54. EPrints "Request eprint" button, [http://www.eprints.org/news/features/request\\_button.php](http://www.eprints.org/news/features/request_button.php)

55. DSpace Request Copy Add-on documentation, <http://wiki.dspace.org/RequestCopy>
56. NISO OpenURL 1.0 Z39.88-2004 standard, [http://www.niso.org/standards/standard\\_detail.cfm?std\\_id=783](http://www.niso.org/standards/standard_detail.cfm?std_id=783)
57. National Centre for Text Mining (NaCTeM), <http://www.nactem.ac.uk/>
58. OAIS, <http://nssdc.gsfc.nasa.gov/nost/isoas/>
59. DSpace's Lightweight Network Interface is being used to support this, <http://wiki.dspace.org/LightweightNetworkInterface>
60. IRIScotland project, <http://www.iriscotland.lib.ed.ac.uk/index.html>
61. NEREUS consortium, <http://www.nereus4economics.info/index.html>
62. Dublin Core, <http://www.dublincore.org/>
63. MARC, <http://www.loc.gov/marc/>
64. Metadata Object Description Schema (MODS), <http://www.loc.gov/standards/mods/>
65. VRA (Visual Resources Association) Core version 3.0, <http://www.vraWeb.org/vracore3.htm>
66. NISO Metadata for Images in XML (MIX), <http://www.loc.gov/standards/mix/>
67. Encoded Archival Description (EAD), <http://www.loc.gov/ead/>
68. IEEE Learning Object Metadata (LOM), <http://ltsc.ieee.org/wg12/>
69. UK LOM Core, <http://www.cetis.ac.uk/profiles/uklomcore>
70. Text Encoding Initiative, <http://www.tei-c.org/>
71. Categories for the Description of Works of Art (CDWA), [http://www.getty.edu/research/conducting\\_research/standards/cdwa/index.html](http://www.getty.edu/research/conducting_research/standards/cdwa/index.html)
72. ISO 19115:2003 metadata for geographic information, <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35>
73. HE/FE profile, <http://go-geo.data-archive.ac.uk/ProfileIndex.htm>
74. Federal Geographic Data Committee, <http://www.fgdc.gov/>
75. Electronic Thesis and Dissertation Metadata Standard (ETD-MS), <http://www.ndltd.org/standards/index.en.html>
76. Networked Digital Library of Theses and Dissertations (NDLTD), <http://www.ndltd.org/>
77. UK Metadata Core Set for Electronic Theses and Dissertations, <http://www2.rgu.ac.uk/library/guidelines/metadata.html>
78. Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA), <http://cordra.net/>
79. Information Environment Metadata Schema Registry (IEMSR), <http://www.ukoln.ac.uk/projects/iemsr/>
80. See the homepages of these standards for further information
81. Fedora XML (FOXML), <http://www.fedora.info/download/2.0/userdocs/digitalobjects/introFOXML.html>
82. Web Ontology Language (OWL), <http://www.w3.org/TR/owl-features/>
83. SIMILE project, <http://simile.mit.edu/>
84. Topic Maps, <http://www.topicmaps.org/>
85. JSTOR/Harvard Object Validation Environment (JHOVE), <http://hul.harvard.edu/jhove/>
86. Digital Record Object Identification (DROID), <http://www.nationalarchives.gov.uk/aboutapps/pronom/droid.htm>
87. National Library of New Zealand Metadata Extraction Tool, <http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction88>
88. Kea automatic keyword extraction tool, <http://www.nzdl.org/Kea/>
89. National Centre for Text Mining information extraction tools list, <http://www.nactem.ac.uk/software.php?software=infoextraction>
90. Open Language Archives Community metadata, <http://www.languagearchives.org/OLAC/metadata.html>
91. OAI Sets, <http://www.openarchives.org/OAI/2.0/guidelines-repository.htm#Sets>
92. Directory of Open Access Repositories (OpenDOAR), <http://www.opendoar.org/>
93. Scopus Author Identifier, <http://info.scopus.com/etc/authoridentifier/>
94. Archive Ingest and Handling Test, . See also articles in the December 2005 issue of D-Lib Magazine at <http://www.dlib.org/dlib/december05/12contents.html>
95. MPEG-21 standard, <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>
96. Metadata Encoding & Transmission Standard (METS), <http://www.loc.gov/standards/mets/>
97. IMS Content Packaging specification, <http://www.imsglobal.org/content/packaging/index.html>
98. ATOM, <http://www.atomenabled.org/>
99. XML Formatted Data Unit (XFDU), <http://sindbad.gsfc.nasa.gov/xfdu/>
100. Los Alamos National Laboratory Research Library, <http://library.lanl.gov/lww/>
101. Portico Electronic Archiving Service, <http://www.portico.org/>
102. Resource Aggregation Model for Learning, Education and Teaching (RAMLET), <http://ieeeltsc.org/wg11CMI/ramlet/>

103. Akamai, <http://www.akamai.com/>
104. Lots of Copies Keeps Stuff Safe (LOCKSS), <http://www.lockss.org/lockss/Home>
105. Microsoft Simple Sharing Extensions for RSS and OPML, <http://msdn.microsoft.com/xml/rss/sse/>
106. SHERPA-DP: Creating a persistent preservation environment for institutional repositories, <http://ahds.ac.uk/about/projects/sherpa-dp/index.html>
107. PRESERV: Preservation Eprint Services, <http://preserv.eprints.org/>
108. Hybrid Archives project, <http://ahds.ac.uk/about/projects/hybrid-archives/index.htm>
109. SRW Update, <http://srw.cheshire3.org/docs/update/>
110. Web-based Distributed Authoring and Versioning (WebDAV), <http://www.Webdav.org/>
111. OAister SRU target details, <http://oaister.umdl.umich.edu/o/oaister/sru.html>
112. Bielfeld Academic Search Engine (BASE), [http://base.ub.uni-bielefeld.de/index\\_english.html](http://base.ub.uni-bielefeld.de/index_english.html)
113. Utrecht University Library Omega metadatabase, <http://omega.library.uu.nl/>
114. OpenSearch, <http://opensearch.a9.com/>
115. Common Query Language (CQL), <http://www.loc.gov/standards/sru/cql/index.html>
116. NISO MetaSearch Initiative, [http://www.niso.org/committees/MS\\_initiative.html](http://www.niso.org/committees/MS_initiative.html)
117. LOREnet, <http://www.lorenet.nl/en/page/page.view/home.page>
118. Digital Library Foundation Aquifer Initiative, <http://www.diglib.org/aquifer/>
119. Open Archival Information System Reference Model, <http://nssdc.gsfc.nasa.gov/nost/isoas/>
120. DSpace Lightweight Network Interface, <http://wiki.dspace.org/LightweightNetworkInterface>
121. Fedora Access and Management Web Services, <http://www.fedora.info/definitions/1/0/api/>
122. Institutional Repositories & Research Assessment (IRRA) project, <http://irra.eprints.org/>
123. OpenURL COinS: a convention to embed bibliographic metadata in HTML, <http://ocoins.info/>
124. Secure Personal Institutional and Inter-Institutional Repository Environment (SPIRE) project, <http://spire.conted.ox.ac.uk/cgi-bin/trac.cgi>
125. LionShare P2P project, <http://lionshare.its.psu.edu/main/>
126. Information Environment Service Registry (IESR) Web search interface, <http://www.iesr.ac.uk/service/iesrsrch?type=new>
127. Shibboleth, <http://shibboleth.internet2.edu/>
128. e-Framework for Education and Research, <http://www.e-framework.org/>
129. Jerry Persons, Stanford University, personal communication
130. See report at [http://www.jisc.ac.uk/index.cfm?name=altilab\\_papers](http://www.jisc.ac.uk/index.cfm?name=altilab_papers)
131. A Technology Analysis of Repositories and Services, <https://wiki.library.jhu.edu/display/RepoAnalysis/ProjectRepository>
132. OAI Static Repositories specification, <http://www.openarchives.org/OAI/2.0/guidelines-staticrepository.htm>