

# Rosetta

第 1.0 版

平成 22 年 3 月 31 日

---

保存システムの構築に係る要素技術に関する調査 1 式

**履歴**

版数	改版日	内容	備考
1.0	2010/03/31	(初版)	

## 目 次

1. Rosetta について .....	1
1.1. 種別 .....	1
1.2. ライセンス .....	1
1.3. 開発・運用主体 .....	1
1.4. 概要 .....	1
1.5. 開発状況 .....	2
1.6. システム要件 .....	2
2. 機能・構成 .....	3
2.1. 機能概要 .....	3
2.2. 設計 .....	6
2.2.1. コンポーネント構成 .....	6
2.2.2. 外部インタフェース .....	6
2.2.3. データモデル .....	9
2.3. ソースコード .....	9
3. 性能 .....	10
4. まとめ .....	11
5. 参考文献 .....	12

## 1. Rosetta について

特に記述がない限り、本資料の調査対象は、Rosetta Version 1.0 である。

### 1.1. 種別

Rosetta はデジタルリポジトリシステムである。

### 1.2. ライセンス

商用ライセンスである。

### 1.3. 開発・運用主体

#### (1) 開発者

ExLibris 社

開発パートナー：ニュージーランド国立図書館

(National Library of New Zealand：以下、NLNZ)

#### (2) 現在の運用者

ExLibris 社

### 1.4. 概要

#### (1) 概要

Rosetta は現在 NLNZ にて運用されている。

商用のため、ライセンスを購入しないと使用することはできない。

Rosetta には以下の特徴がある。

OAIS 準拠・・・SIP の受入、データ管理、DIP 作成から保存計画<sup>1</sup>モジュールまで、OAIS モデルに準拠した全てのソリューションを 1 つのシステムで対応可能。

拡張性・・・データベースに Oracle を採用しており、増加し続けるデジタルファイルに対応した拡張性の高い大規模システムを構築可能。

オンラインアーカイブ・・・オンライン上で持続可能なアーカイブを提供。

#### (2) 開発の経緯・目的

ニュージーランドは小規模国家であり、国全体でデジタル保存に取り組むという理念がある。その理念に基づき、デジタル情報の収集・提供・恒久的な保存を使命とする国家デジタル遺産アーカイブプログラム (NDHA) プログラムが 2004 年に設立され、2006 年に Rosetta の開発が始まった。

Rosetta の目的は、真正性<sup>2</sup> (authenticity) が確認されたコンテンツを正確に表示することである。

<sup>1</sup> 保存計画機能は Rosetta Version 2.0 で実装予定

<sup>2</sup> デジタルコンテンツが長期間経過後も、同じように利用 (閲覧) できること

## 1.5. 開発状況

- ( 1 ) 開発進捗状況  
Rosetta Version 2.0 が 2010 年 4 月に稼動予定である。
- ( 2 ) 最新安定バージョン  
Rosetta Version 1.0
- ( 3 ) 最新開発バージョン  
Rosetta Version 2.0
- ( 4 ) ロードマップ
  - 2008 年 10 月 2 年の開発期間を経て、NLNZ にて Rosetta Version 1.0 稼動開始
  - 2009 年 1 月 Rosetta Version 1.0 システム公開
  - 2009 年 12 月 Rosetta Version 2.0 のベータ版公開予定[1]  
保存計画 ( preservation planning ) も実装予定<sup>3</sup>
  - 2010 年 4 月 NLNZ で Rosetta Version 2.0 稼動予定

## 1.6. システム要件

データベースとして Oracle Database が必要である。

<sup>3</sup> 保存計画、Version 2.0 に関する詳細な情報は不明。

## 2. 機能・構成

### 2.1. 機能概要

#### (1) 主な機能

Rosetta は OAIS 参照モデルに準拠して設計されている。

各機能の対応関係は図 1 の通り。なお、OAIS 参照モデルの保存計画 (Preservation Planning) に対応する Preservation は、Rosetta Version 2.0 に実装予定の機能であり、Rosetta Version 1.0 には実装されていない。

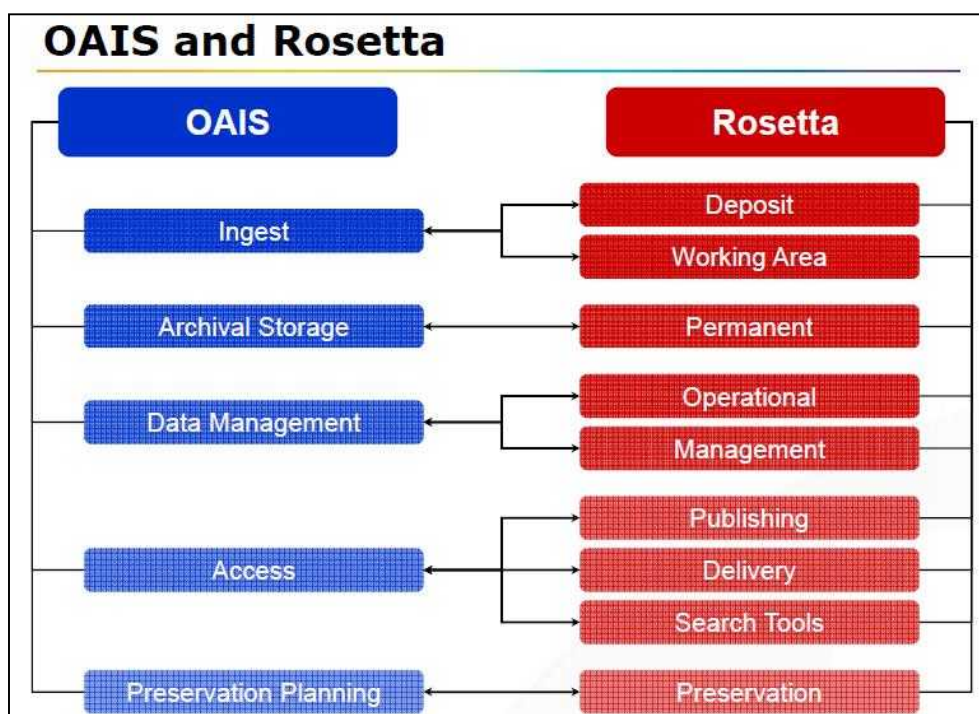


図 1 OAIS と Rosetta の対応関係 [1]

Rosetta の構成と各機能は以下の通りである。

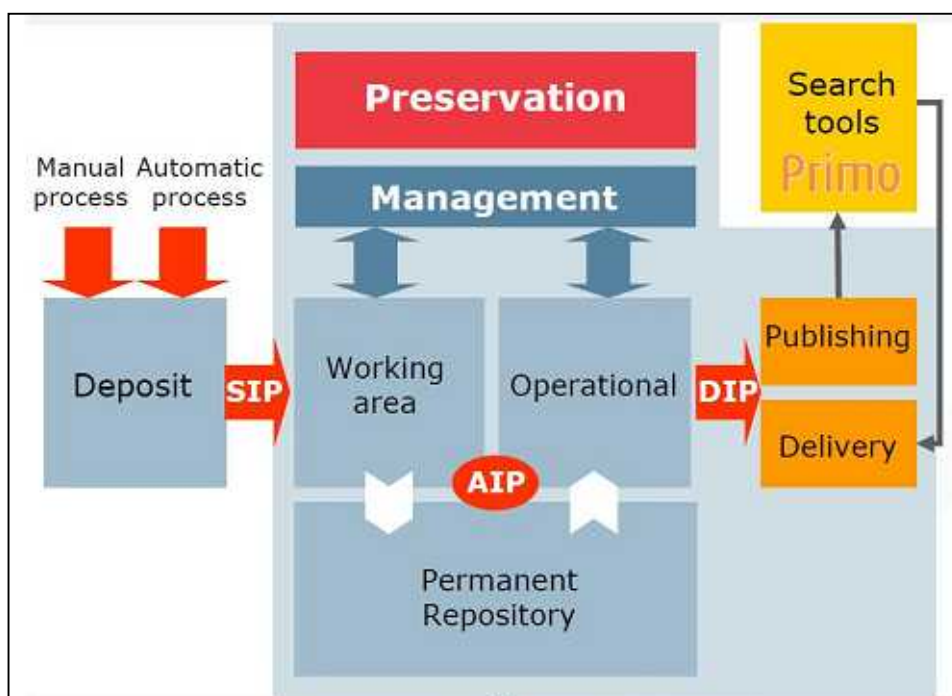


図 2 Rosetta 構成図 [1]

- Deposit
 

ユーザーによる不定期の手動登録から、商用・学術出版物の大量自動登録に至るまで、さまざまな形式で登録されるコンテンツの受入を行うモジュール。SIP の作成もここで行われる。
- Working area
 

Permanent Repository に登録する前に、実際に受け取った SIP を加工して AIP に変換するための中間領域。SIP として受け取った単位を分割したり、まとめたりすることもある。

以下の 3 ステップで作業が行われる。

  - JHOVE、Metadata Extraction Tools、DROID 等、外部ツールによる識別（Rosetta 自体は識別機能を備えていない）
  - 保存するコンテンツの選別（判定は人が行う）
  - メタデータ生成やファイルコピーの作成等（object enrichment）
- Permanent Repository
 

加工された AIP を保存するリポジトリ。将来的に新システムへの移行が必要になった場合にアップロードを可能にするため、標準的なディスクに保存されている<sup>4</sup>。
- Operational
 

Rosetta 内のオブジェクト管理を行い、AIP から DIP を生成する。
- Management
 

Working area と Operational 上にあり、Rosetta 内のインフラ管理機能を担う。
- Preservation
 

保存計画（Preservation Planning）機能。Rosetta Version 2.0 で開発中。
- Publishing・Delivery
 

エンドユーザーが DIP を使い、様々なビューアでコンテンツを閲覧するための機能。

<sup>4</sup> 何のデータをアップロードするか、また、標準的なディスクが何かは詳細不明。

- SearchTools  
Rosetta 外部の検索ツール<sup>5</sup>を使用し、他のアーカイブシステムとの統合検索を行うことができる。

(2) 保存関係の機能

- SIP 受入  
Deposit で受入れたコンテンツを SIP として Working area に渡す。
- フォーマット識別  
Working area にて JHOVE 等によるフォーマット識別が行われる。
- 記述情報等メタデータ作成・管理  
Working area にて記述情報や技術情報に関するメタデータの作成が行われる。
- AIP 作成・管理  
Working area にて AIP が作成され、Permanent Repository に保存される。
- 陳腐化フォーマット情報の取得・管理  
保存計画機能であり、Rosetta Version 1.0 には実装されていない。
- 陳腐化リスクの高い AIP の自動検知  
保存計画機能であり、Rosetta Version 1.0 には実装されていない。
- フォーマット変換  
保存計画機能であり、Rosetta Version 1.0 には実装されていない。
- SIP 再受入  
Rosetta Version 1.0 には実装されていない。

<sup>5</sup> 図 2 の"Primo"は ExLibris 社が開発した検索ツール。その他 Google 等の検索ツールも使用可能。

## 2.2. 設計

### 2.2.1. コンポーネント構成

「2.1 機能構成（1）主な機能」参照。

### 2.2.2. 外部インターフェース

（1）他システムと連携して利用することの可否

Rosetta は、フォーマット識別に JHOVE 等のツールを使用し、他システムとの連携が行われているが、新たなシステムを組み込んで連携することはできない。

（2）実際の利用例

Rosetta の外部インターフェースのうち、受入（Deposit）画面例を下記に示す[2]。

#### 【Deposit】

コンテンツの入力

The screenshot shows a web browser window displaying the 'ExLibris Preservation Deposit' form. The form is titled 'Descriptive Information' and contains several input fields: Creator (Frank Gehry), Title (Walt Disney Concert Hall - images), Description (The Walt Disney Concert Hall at 111 South Grand Avenue in Downtown Los Angeles, California is the fourth hall of the Los Angeles Music Center. Bounded by Hope Street, Grand Avenue, 1st and 2nd Streets, it seats 2,266 people and serves), Content Type (Manuscripts), Creation date (DD/MM/YYYY), Publisher's name, and Notes. Navigation buttons 'Back', 'Cancel', and 'Next' are visible at the bottom.

図 3 コンテンツ入力画面 [2]

（入力項目）

- ・ Creator・・・作成者
- ・ Title・・・タイトル
- ・ Description・・・説明
- ・ Content Type・・・コンテンツタイプ
- ・ Creation date・・・コンテンツ作成日
- ・ Publisher's name・・・公開（出版）者名
- ・ Notes・・・注記

## アクセス権の設定

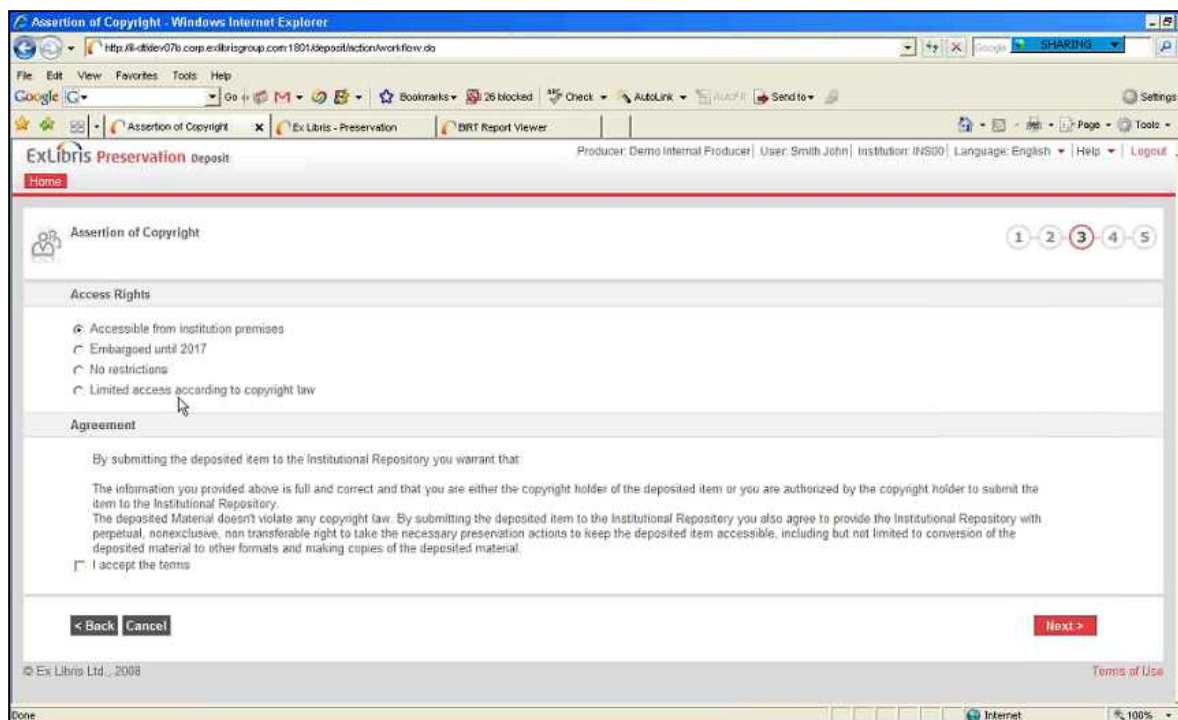


図 4 アクセス権設定画面 [2]

(アクセス権)

- ・ 組織（機関）内からアクセス可
- ・ 2017 年までアクセス禁止
- ・ アクセス制限なし
- ・ 著作権法によるアクセス制限あり

ファイル入力

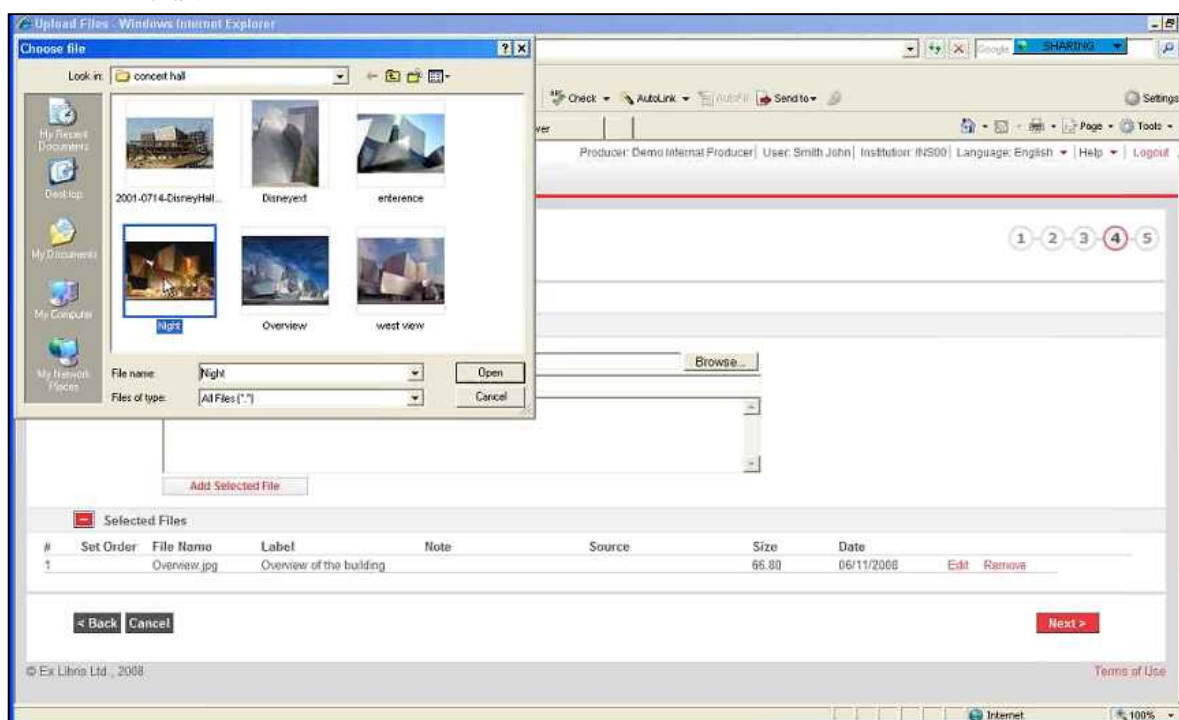


図 5 ファイル入力画面 [2]

入力情報確認画面

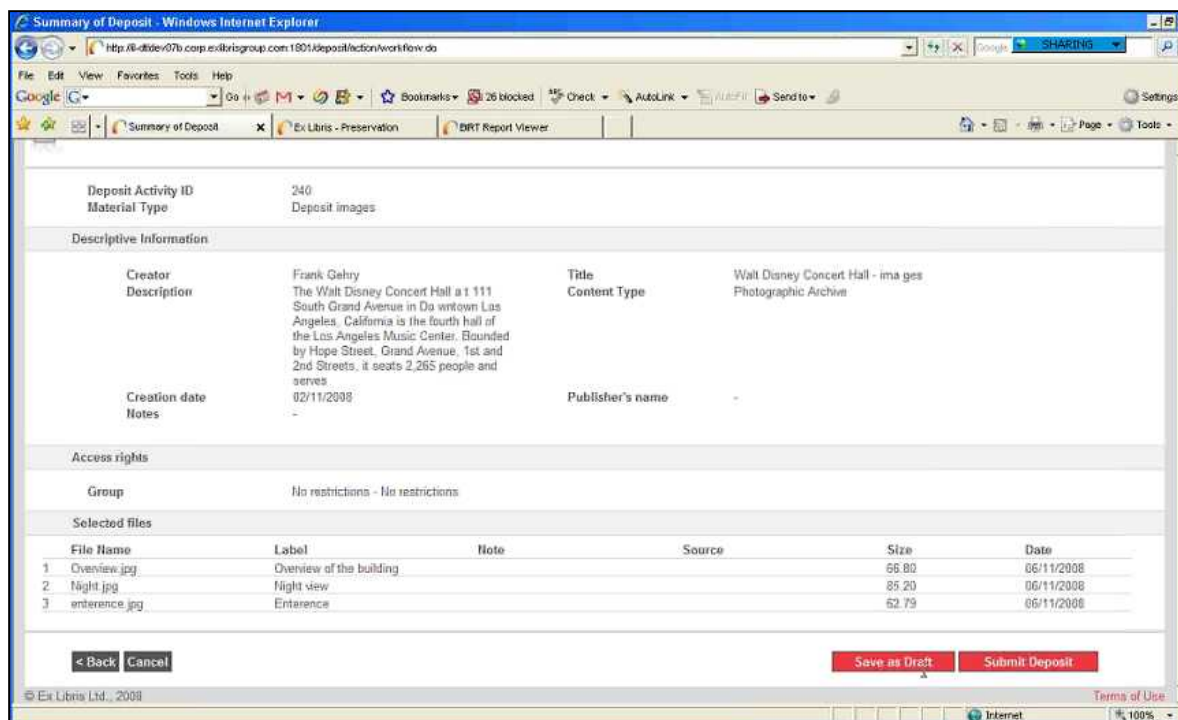


図 6 入力情報確認画面 [2]

### 2.2.3. データモデル

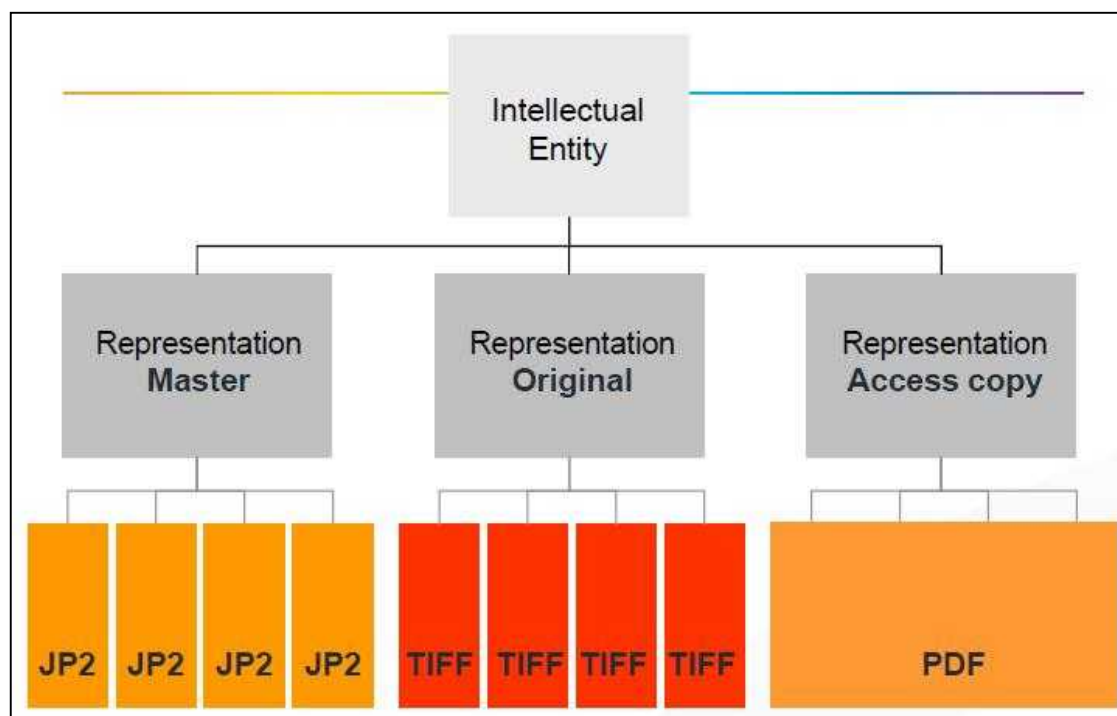


図 7 Rosetta のデータモデル例 [1]

Rosetta のデータモデルは以下の通りである（図 7 参照）。

- 一つの Intellectual Entity（知的エンティティ：以下、IE）に対して、複数の Representation（表現形）があり、各 Representation には複数のファイルがある。
- たとえば、一つの IE に対して、オリジナル（Original）、マスター（Master）、提供用（Access Copy）の 3 つの Representation が存在し、オリジナルは複数の TIFF ファイル、マイグレーションしたマスターは複数の JPEG2000（JP2）ファイル、エンドユーザーへの提供用は複数の JPEG（JPG）ファイル又は複数のオリジナルファイルを 1 つに統合した PDF ファイル、といった構造でデータを持つことができる。
- Representation のうち、提供用は Permanent Repository には保存されていないため、簡単に削除ができる。

### 2.3. ソースコード

商用のため公開されていない。

### 3. 性能

Rosetta の性能に関する情報は公開されていない。

#### 4. まとめ

Rosetta は OAIS に準拠しており、デジタルリポジトリとしての基本的な機能を備えている。保存計画 (Preservation Planning) に関しては、2010 年 4 月にリリース予定の Version 2.0 で実装される予定であり、現在のバージョンには実装されていない。

## 5. 参考文献

- [1] A preservation system Challenges & practical experience  
<http://www.schk.sk/fileadmin/docs/digilib/2009/Yaniv-LEVI.pdf>  
(最終アクセス日 2009年12月3日)
- [2] The Ex Libris Digital Preservation System  
<https://sunmeetings.webex.com/ec0600l/eventcenter/recording/recordAction.do;jsessionid=WDqFK36K1ysN8NdtvZvXKKGt6rgLJ7QL07lsrktypvkNJw0jKnNK!-1006933448?theAction=poprecord&actname=%2Feventcenter%2Fframe%2Fg.do&apiname=lsr.php&renewticket=0&renewticket=0&actappname=ec0600l&entappname=url0106l&needFilter=false&&isurlact=true&entactname=%2FnbrRecordingURL.do&rID=28425317&rKey=FA6F8AF564821AFD&recordID=28425317&rnd=8441549799&siteurl=sunmeetings&SP=EC&AT=pb&format=short>  
(最終アクセス日 2009年12月3日)

# DAITSS

第 1.0 版

平成 22 年 3 月 31 日

---

**履歴**

版数	改版日	内容	備考
1.0	2010/03/31	(初版)	

## 目 次

1. DAITSS について .....	1
1.1. 種別 .....	1
1.2. ライセンス .....	1
1.3. 開発・運用主体 .....	1
1.4. 概要 .....	1
1.5. 開発状況 .....	2
1.6. システム要件 .....	3
2. 機能・構成 .....	4
2.1. 機能概要 .....	4
2.2. 設計 .....	8
2.2.1. コンポーネント構成 .....	8
2.2.2. 外部インタフェース .....	9
2.2.3. データモデル .....	9
2.3. ソースコード .....	10
3. 性能 .....	11
4. まとめ .....	12
5. 参考文献 .....	13

## 1. DAITSS について

特に記述がない限り、本資料の調査対象は、DAITSS (Dark Archive<sup>1</sup> In The Sunshine State<sup>2</sup>) [1]のバージョン 1.2.6 である。

### 1.1. 種別

DAITSS はデジタルリポジトリシステムである[1]。

### 1.2. ライセンス

GPL[1]

### 1.3. 開発・運用主体

#### (1) 開発者

フロリダ図書館自動化センター<sup>3</sup> (Florida Center for Library Automation : 以下、FCLA) の DAITSS 開発チーム<sup>4</sup>[1]。

#### (2) 現在の運営者

FCLA

### 1.4. 概要

#### (1) 概要

DAITSS は、FCLA により実運用されており、以下 11 の州立大学のリポジトリをバックアップしている。

- ・フロリダ工科大学
- ・フロリダ国際大学
- ・フロリダ州立大学
- ・フロリダ大学
- ・フロリダ A&M 大学
- ・フロリダ・アトランティック大学
- ・フロリダ・ガルフ・コースト大学
- ・フロリダ・ニューカレッジ
- ・セントラルフロリダ大学
- ・北フロリダ大学
- ・西フロリダ大学

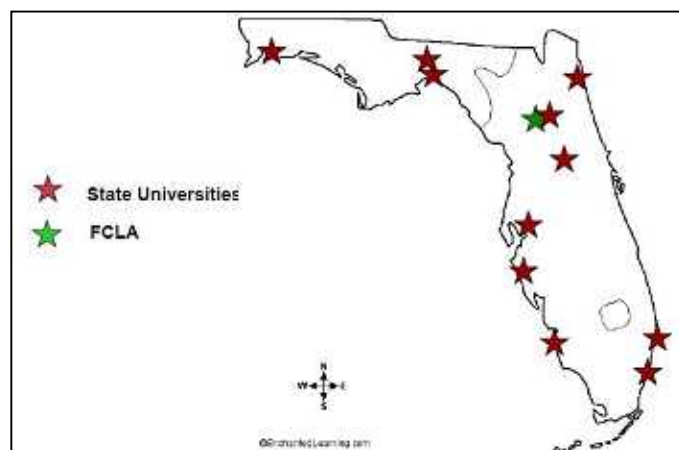


図 1 フロリダ州の州立大学と FCLA

DAITSS はデジタル保存リポジトリシステムであり、受入、データ管理、提供というリポジトリ機能に加え、フォーマットの正規化<sup>5</sup>、大量データの一括又は要求に応じたマイグレーションという保存計画機能に必要な機能も備えている。

DAITSS は、デジタルライブラリのアプリケーションや機関リポジトリのソフトウェアなどのシステム向けバックエンド・システムとしての使用を目的とした「ダーク・アーカイブ」。

<sup>1</sup> 保存することに特化し、利用を前提としないアーカイブ。

<sup>2</sup> 米国フロリダ州の愛称。

<sup>3</sup> 公的資金により、フロリダ州の大学図書館を支援する自動化サービスを提供している。

<sup>4</sup> チームは以下の 5 人で構成されている。

Priscilla Caplan : ディレクター Randy Fischer : チームリーダー Franco Lazzarino : 開発者

Carol Chou : 開発者・フォーマットの専門家 Emmanuel Rodriguez : 開発者

<sup>5</sup> ファイルがデジタル保存に適したフォーマットになっていない場合、適したフォーマットに変換する。

イブ」である。

(2) 開発の経緯

2000 年初頭、デジタル化プロジェクトで作成された論文やマスタファイルを長期保存するための方法について調査を開始した。当時利用可能であったコンテンツ管理やデジタルライブラリ、機関リポジトリソフトウェア等を調査した結果、フォーマットのマイグレーション等の保存戦略を実行できる既存の製品がなかったため、独自にリポジトリソフトウェアを開発することとなり、DAITSS の開発計画が始まった。

## 1.5. 開発状況

(1) 開発進捗状況

DAITSS 2.0 が 2009 年第 4 四半期に完成予定であったが、2010 年第 2 四半期に延期されている[1]。

開発履歴：

2002 年	米国博物館・図書館サービス振興機構 (IMLS) <sup>6</sup> による 3 年間の助成が認められ、フォーマットに関する専門家が開発チームに加わる。
2005 年 11 月	コンテンツ (ファイル) の提供機能及びメタデータ抽出機能を除いた初期バージョン (Florida Digital Archive <sup>7</sup> 用) の製造を開始[1]。
2006 年 12 月	全ての機能を備えた最初のバージョンが完成。
2007 年	GPL ライセンスの OSS としてリリース。
2007 年 4 月 12 日	DAITSS 1.2.1 リリース。
2007 年 4 月 27 日	DAITSS 1.2.2 リリース。
2007 年 6 月 29 日	DAITSS 1.2.3 リリース。
2007 年 9 月 13 日	DAITSS 1.2.4 リリース。
2007 年 11 月 8 日	DAITSS 1.2.5 リリース。
2007 年 12 月 7 日	DAITSS 1.2.6 リリース。
2008 年 8 月 29 日	DAITSS 2.0 のブリーフィング及びマイルストーン[2]を発表。

(2) 最新安定バージョン

DAITSS 1.2.6

(3) 最新開発バージョン

DAITSS 2.0 (2010 年第 2 四半期完成予定)

(4) ロードマップ[3]

2008 年第 3 四半期	ストレージサービス製造 受入サービスの設計
2008 年第 4 四半期	記述子、アクションプラン及びマイグレーションサービスのテスト
2009 年第 1 四半期	提供サービスを設計 受入サービスのテスト
2009 年第 2 四半期	フォーマット識別サービスのテスト PREMIS 形式にデータベースを変更
2009 年第 3 四半期	XML 形式のデータ出力サービス、データベース、登録サービスのテスト

<sup>6</sup> 米国の博物館・図書館サービス法により設置されている助成機関。

<sup>7</sup> FCLA が運営する保存リポジトリ

2009 年第 4 四半期	保管サービスを「スーパーサーバー」に拡張するテスト 提供用パッケージの再受入、受入マネージャーのテスト
2010 年第 1 四半期	DAITSS1 から DAITSS2 へのデータベース変換テスト 提供サービスのテスト 既存モジュールの製品化
2010 年第 2 四半期	データベース変換機能稼働 DAITSS2 リリース予定

DAITSS1 から DAITSS2 への変更点は以下の通りである[2]。

- 全コンテンツの保存レベル<sup>8</sup>が FULL に統一される。
- 必要に応じて様々なフォーマットで DIP を生成できる。
- 重複パッケージ検出メソッドを実装する。
- 受入を拒否されたパッケージの情報を記録する。
- 新モジュール：管理インタフェース

## 1.6. システム要件

DAITSS は Java 5 以上のバージョンが必要であり、Unix 又は Linux 環境で動作する。RHEL 4 及び Ubuntu 6.10 LTS Linux ディストリビューション<sup>9</sup>において、インストールに関するテストが行われている。

インストール・スクリプトには、bash シェル<sup>10</sup>を必要とし、標準 Unix ユーティリティである、tar<sup>11</sup>、gz<sup>12</sup>、sed<sup>13</sup>、awk<sup>14</sup>を用いる[4]。

### 最低限必要なソフトウェア

- MySQL バージョン 4.x 以上で、InnoDB<sup>15</sup>データテーブルをサポートするもの。  
---DAITSS と同じホストにインストールする必要がある
- Sun Java 仮想マシン (JVM) バージョン 5 以上。  
---JVM は Sun JRE、J2SE、J2EE のいずれかがインストールされていれば、存在する
- Sendmail<sup>16</sup>や Postfix<sup>17</sup>などのメール転送プログラム

<sup>8</sup> DAITSS には全体保存 (FULL) と部分保存 (BIT) の 2 段階の保存方法がある。部分保存では、ストレージ確保、データ監視、複数マスタの生成が行われ、全体保存ではさらに、フォーマットの正規化、ローカライズ、マイグレーション処理が行われる。

<sup>9</sup> Linux を一般利用者が利用できる形にまとめ上げたもの (頒布形態)。LTS は Long Term Support (長期間サポート) を意味する。

<sup>10</sup> Unix で使用するシェル (ユーザと OS を仲介するための基本ソフトウェア) の一つ。

<sup>11</sup> ファイルフォーマットの一つ。ファイルのアーカイブに用いられ、多数のファイルを一つのファイルにまとめることができる。

<sup>12</sup> gzip (GNU zip) によって圧縮されたファイルの拡張子。gzip とは、汎用のデータ圧縮プログラム。

<sup>13</sup> 入力ストリームに対してテキスト変換などのデータ処理を行うために使用されるプログラム。

<sup>14</sup> C 言語に似た文法を持つスクリプト言語。テキストファイルの加工に使用。

<sup>15</sup> MySQL のためのデータベースエンジン。配布されている全ての MySQL バイナリパッケージに標準搭載されている。

<sup>16</sup> UNIX で古くから使われてきたメールサーバソフトウェア。

<sup>17</sup> Sendmail との操作上の互換性を確保しつつ、管理・設定が容易で、高速・安全であることを指向して開発されている FOSS。

## 2. 機能・構成

### 2.1. 機能概要

#### (1) 機能

DAITSS は、以下の機能を有する[4]。また、OAIS と DAITSS の機能比較については、表 1 参照。

- **Preprocessing** (Prep, 事前処理)
  - Ingest (受入) 用に SIP を準備。  
高度なチェックやパッケージの変更が必要でなければ、この処理はスキップ可能。
  - Preprocessing の一般的な機能は以下の通り。
    1. SIP が単一コンテンツから成るファイルかどうかを確認する (拡張子が.zip の場合)。
      - a) 単一ファイルの場合、.zip ファイルのデジタル署名を調べ (デジタル署名があれば) 認証された提供者のソースであること、SIP が署名後に改竄されていないことを確認する。
      - b) ファイルを解凍する。
    2. SIP に SIP 記述子<sup>18</sup>があるかどうかを確認する。  
SIP 記述子がないか、Prep 設定ファイルにより SIP 記述子を生成することが指定されている場合、METS 形式の SIP 記述子を生成する。  
SIP 記述子がなく、生成もできない場合は SIP を拒否する。
    3. SIP に SIP 記述子と、少なくとも 1 つのコンテンツファイルが存在することを確認する。
    4. SIP 記述子で参照されていないファイルを削除する。
    5. Ingest (受入) 処理のために、SIP を ingest/in/ ディレクトリに書き出す。
- **Ingest** (受入)
  - 受入処理が始まると、最初に ingest/in/ ディレクトリ内の SIP リストが生成される。
  - SIP リストの各エントリーに対して以下の処理が行われる。
    1. SIP が存在し、書き込み及び読み取りが可能であることを確認する。
    2. SIP 記述子を確認する。METS スキーマに準拠しているかどうかをチェックする。
    3. SIP 記述子のメタデータを、後で使用するために出力する。このメタデータには知的オブジェクトに関するメタデータや、所有するアカウント、サブアカウント、プロジェクトコードの情報も含まれる。
    4. SIP 内の各ファイルのデータファイルオブジェクトを生成する。
      - a) 各ファイルフォーマットの主要な特徴との照合を繰り返し、ファイルタイプを判定する。まず、SIP 記述子の MIME タイプを照合し、次に拡張子による照合を行う。識別可能フォーマット(表 2 参照)と合致しない場合、Unix の “file” 機能を使用し、仮のフォーマットを返す。

<sup>18</sup> SIP 内の各ファイルについて記述された METS 文書のこと

- b) ウィルスチェックやファイル及びビットストリームに関する技術メタデータの抽出・作成等を行い、一意の ID を付与したデータファイルオブジェクトを生成する。
  - c) ファイルに外部リンクがある場合、リンク先ファイルをダウンロードし、ダウンロードしたファイル用データファイルオブジェクトを生成する。この処理は全てのリンクに対して行われる。
  - d) 全体保存 (FULL) と指定されたファイルについて、そのフォーマット用のマイグレーションメソッド<sup>19</sup>が存在する場合、マイグレーションを実施したバージョンのデータファイルオブジェクトを生成する。
  - e) 全体保存と指定されたファイルについて、SIP 内の他のファイルへの参照がある場合、内部ファイル名への参照の代わりにローカライズ<sup>20</sup>したデータファイルオブジェクトを生成する。
  - f) 全体保存と指定されたファイルについて、そのフォーマット用の正規化メソッドが存在する場合、正規化したデータファイルオブジェクトを生成する。
  - g) 知的エンティティ及び全てのデータファイルオブジェクトに関するメタデータを作成する( 関連性や影響を及ぼしたイベントに関する情報を含む )。
  - h) AIP を生成する ( AIP 記述子を生成し、ingest/out/ディレクトリに書き出す )。
    - i) /in ディレクトリの SIP を削除する。
    - j) AIP 用に受入情報を XML 形式にフォーマットし、データ提供者及びアーカイブにメールを送信する。
- **Storage Maintenance** (ストレージ管理)
    - Fixity checker (不変性チェッカー)・・・n 個のマスタファイル<sup>21</sup>それぞれに対して、2 つのメッセージダイジェストを別のアルゴリズムで生成し、マスタデータベースと照合。破損したファイルは、正常版に置き換えられる。
  - **Access** (アクセス)
    - Activity reports (アクティビティレポート)・・・「SIP が受入れられた」「AIP が削除された」など特定のアクションが完了したときに発行される。
    - Repository statistics (リポジトリ統計値)・・・使用領域 (GB)、知的エンティティ数、データファイル数、その他の統計値をリアルタイム又はバッチレポートで確認できる。

<sup>19</sup> 実際にどのような処理が行われるのか、詳細は不明である。

<sup>20</sup> データファイルが他のファイルへのリンクを持つ場合、リンク先ファイルをダウンロードして取り込むこと。

<sup>21</sup> DAITSS では AIP のマスタコピー数を指定できる。デフォルトは n=3。

- **Dissemination (提供)**

- 認証されたユーザに DIP を出力する。

DIP はオリジナルの SIP とファイルにより構成されている。

AIP はストレージに残ったままとなる (Withdrawal の手続きでしか削除できない) が、Dissemination の Re-Ingest (再受入) プロセスで AIP を更新することができる。

- 以下、Dissemination の手順を記す。

1. Request authentication (リクエスト認証)
2. Package assembly (パッケージの集合)・・・AIP 内にある全てのファイルのコピーが Re-Ingest (再受入) キューに配置される。Dissemination がイベントとして記録される。
3. Re-Ingest (再受入)・・・前回のパッケージ受入以降、DAITSS でフォーマット識別やフォーマットプロセスが更新されていることが考えられる。最新の情報を利用するために、パッケージに対して再度、Ingest (受入) 処理を行う。再受入処理は、全てのパッケージ内コンテンツに対して行われる。
  - a) データベースから知的エンティティを検索し、知的エンティティのメタデータを取得する。
  - b) パッケージ内の各ファイルの AIP 記述子を取得する。メタデータを出力し、ビットストリームを生成するためにデータファイルの構文解析を行う。  
ファイルタイプが識別対象外の場合 (表 2 参照) MIME タイプを確認する。  
ファイルに対応するフォーマット情報に変更があった場合、新しいデータファイルオブジェクトを生成し、メタデータに記録する。  
データファイルにマイグレーション、正規化、ローカライズが必要な場合、それぞれ適したバージョンを生成する。
  - c) 不要なファイル (古い AIP 記述子とマイグレーション実施前のファイル) を特定し、陳腐化したファイルとして記録する。
  - d) 新しい AIP 記述子を生成する。AIP 記述子と新しく生成されたファイルをストレージに書き出す。陳腐化したファイルを削除する。管理データベースを更新する。
4. DIP の書き出し・・・DIP 記述子を生成する。DIP 内のファイルを ZIP ファイルにまとめ、メッセージダイジェストを生成する。ZIP ファイルとメッセージダイジェストを Dissemination ディレクトリに配置する。
5. 希望する機関に DIP が利用可能であることを知らせるため、Dissemination Report をメールで送信する。

- **Withdrawal (削除)**

- AIP を削除する時に使用する。
- データ提供者が所有するコンテンツをリポジトリから削除することを要請した場合、データ提供者がリポジトリの使用を停止した場合又は著作権のステータスが変わった時に行われる。

表 1 OAIS と DAITSS の比較

	DAITSS
	<b>Ingest (受入) 機能</b>
SIP の受け取り SIP の品質を保証 AIP の生成 アーカイブのデータベースに登録するための 記述情報を AIP から抽出 Archival Storage と Data Management の同 期をはかる	SIP の受け取り SIP の品質を保証 AIP の生成 アーカイブのデータベースに登録するための 保存情報を AIP から抽出又は作成 Archival Storage と Data Management の同 期をはかる ローカライズ、正規化、マイグレーションし たバージョンのファイルを必要に応じて生成
	<b>Archival Storage (アーカイブストレージ) 機能</b>
AIP を Ingest から受け取り、恒久的ストレ ージに追加 エラーチェックを実行 要求された AIP を Access に提供 ストレージの階層管理 情報が追加されたメディアを更新 ディザスタリカバリ機能を提供	AIP を Ingest から受け取り、恒久的ストレ ージに追加 エラーチェックを実行。管理データベースと ストレージの不変性、完全性のチェック。 要求された AIP を Access に提供 (未対応) 左記の 3 機能は DAITSS の外部で提供される べき機能
	<b>Data Management (データ管理) 機能</b>
データベース更新 データベースでクエリを実行し、結果セット を生成 結果セットからレポートを作成 スキーマとビューの定義及び参照整合性を維 持	データベース更新 データベースでクエリを実行し、結果セット を生成 結果セットからレポートを作成 左記の機能は関連データベースを直接扱う職 員が実行
	<b>Administration (運用統括) 機能</b>
提出に関する契約の要請や交渉 提出物が基準に従っているか検査 基準、方針の運用 カスタマーサポートの提供 その他の類似機能	左記の機能は DAITSS の外部で職員が実行す べき機能 DAITSS では、提出に関する取り決めについ ての情報をシステムに取り入れることが可能

## (2) 保存関係の機能

- DAITSS の陳腐化したファイルの検知は、上記 (1) Dissemination (提供) 処理時に自動で行われる。
- DIP 要求 Re-Ingest (再受入) 再受入時に最新のフォーマット情報を元にマイグレーション等必要な処理が行われる。
- 検知単位は拡張子単位である。
- マイグレーションに必要な情報は、変換ツールと対象ファイルフォーマットを設定ファイルに追加する形で管理される。そのツールの対象となるすべてのファイルがマイグレーション可能であることを確かめた上で、設定ファイルに反映される。

## 2.2. 設計

### 2.2.1. コンポーネント構成

DAITSS のコンポーネントは、OAIS 機能モデルを基に構成されている[5]。

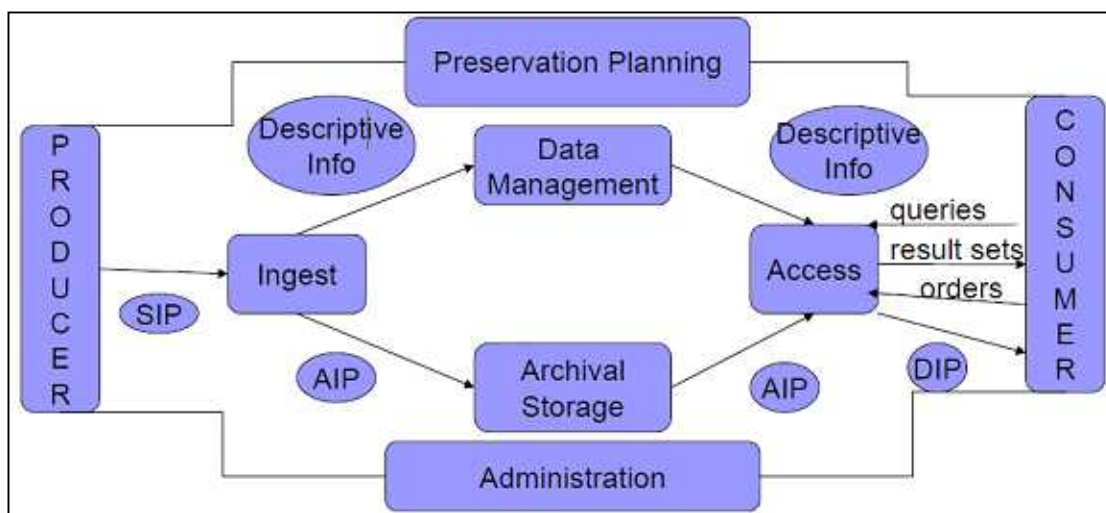


図 2 OAIS 機能モデル

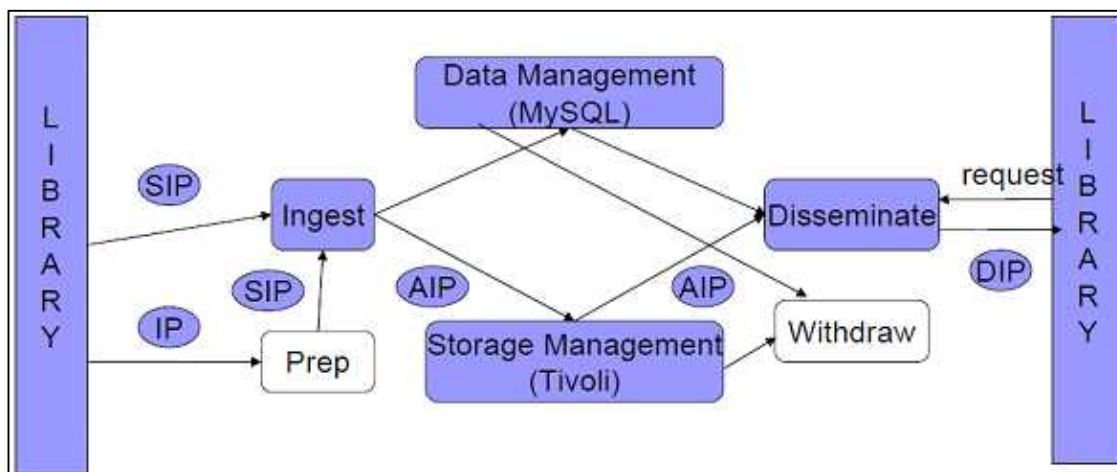


図 3 DAITSS の構成

(データ管理に MySQL、ストレージ管理に Tivoli を使用)

## 2.2.2. 外部インタフェース

公開されていない。

## 2.2.3. データモデル

### (1) オブジェクトデータモデル[4]

DAITSS では、デジタルオブジェクトに関する情報を「知的エンティティ」「データファイル」「ビットストリーム」の3つの階層で管理する。

- 知的エンティティ・・・ひとつの単位として記述、説明されるひとまとまりのコンテンツ。  
例：1冊の本、1枚の地図、1枚の写真
- データファイル・・・単一名のデジタルファイル。  
例：PDF ファイル、TIFF ファイル、XML ファイル
- ビットストリーム・・・保存を目的に共通属性として付加されるビット配列であり、1つ以上のビットストリームがデータファイルに埋め込まれている。  
例：AVI ファイルは音声のビットストリームと映像のビットストリームで構成されている。

### (2) 情報パッケージ[4]

DAITSS は、記述子として METS スキーマ[6]を採用しており、OAIS に準拠した以下のパッケージが生成される。

- SIP・・・ データ提供者は、Ingest（受入）に適切な SIP を作成する必要がある。システム側から強制されるものではないが、各 SIP は単一の知的エンティティとして構成されることが理想的である。  
SIP には、パッケージの内容を記述した SIP 記述子（METS 文書）と知的エンティティを構成するデータファイルが含まれなければならない。
- AIP・・・ AIP はオリジナルの SIP 記述子と、保存レベルが「部分」または「全体」である SIP からの全てのデータファイルにより構成されている。  
さらに、SIP に含まれていないものの、SIP 内のファイルからリンクされているファイルが含まれている場合もある。  
例えば、XML スキーマへのリンクを持つ XML ドキュメントがスキーマそのものを含んでいない場合、DAITSS はそのスキーマをダウンロードし、AIP に取り込む。  
AIP は、知的エンティティに関するメタデータを記述した AIP 記述子及びそれぞれのデータファイルで構成される。
- DIP・・・ DIP はオリジナルの SIP 内のすべてのファイルと追加されたリンクファイルにより構成される。さらに DIP は、あらゆるファイルの「最新かつ最良」バージョンで構成される。なお、これらのファイルは受入時とは異なる場合もある。  
DIP 記述子は AIP 記述子の全てのメタデータで構成される。

### 2.3. ソースコード

ソースコードは Subversion (SVN) で管理されており、メンバーしかダウンロードできない。しかし、ウェブブラウザ上でソースを確認することは可能である[1]。

### 3. 性能

#### (1) 処理可能なフォーマット数

実際にテストは行っていないが、DAITSS では、12 種類のフォーマットの識別に対応しており、拡張子単位で処理を行う（表 2）[7]。また、今後対応予定と考えられるフォーマットは、下記の 21 種類である（表 3）[7]。

表 2 DAITSS 識別可能フォーマット [7]

		フォーマット詳細
.AVI	video/avi	Audio-video Interleaved File Format
.CSV	text/csv	Comma Separated Values
.DTD	application/xml-dtd	Document Type Definition
.JP2	image/jp2	JPEG 2000 part 1, ISO/IEC 15444-1
.JPG, .JPEG	image/jpeg	JPEG File Interchange Format
.MOV	video/quicktime	Quicktime File Format
.PDF	application/pdf	Portable Document Format
.PRO	text/x-pro	PrimeOCR PRO Format
.TIF, .TIFF	image/tiff	Tagged Image File Format
.TXT	text/plain	Plain Text File
.WAV, .WAVE	audio/x-wave	Waveform Audio
.XML	application/xml	Extensible Markup Language

表 3 DAITSS 識別対応予定フォーマット [7]

		フォーマット詳細
.BMP	image/bmp	Device Independent Bitmap for Windows
.BZ2, .TBZ2	application/x-bzip2	BZIP2 Compressed Archive
.DOC	application/msword	Microsoft Word File Format
.EXE	application/x-dosexec	Windows Executable
.GIF	image/gif	Graphics Interchange Format
.GZ, .TGZ	application/x-gzip	GZIP Format
.HTML, .HTM	text/html	Hypertext Markup Language
.JPX	image/jpx	JPEG 2000 part 2, ISO/IEC 15444-2
.MID, .MIDI	audio/x-midi	MIDI Audio
.MP2	audio/mpeg	MPEG 1/2 Audio Layer 2
.MP3	audio/mpeg	MPEG 1/2 Audio Layer 3
.MPG, .MPEG	video/mpeg	MPEG 1/2 Video Format
.PBM	image/x-portable-bitmap	Portable Bitmap (PBM)
.PGM	image/x-portable-greymap	Portable Graymap (PGM)
.PPM	image/x-portable-pixmap	Portable Pixmap (PPM)
.PNG	image/png	Portable Network Graphics
.PS	application/postscript	Postscript
.RAR	applicaton/x-rar-compressed	RAR Format
.RTF	text/rtf	Rich Text Format
.XLS	application/vnd.ms-excel	Binary Exchange File Format / Microsoft Excel
.ZIP	application/zip	ZIP Format

#### 4. まとめ

DAITSS は、OAIS に準拠し、受入れた SIP から AIP を作成し、DIP を発行することができる。DAITSS では、DIP の要求があった時にアーカイブから取り出したパッケージを再度受入れることで自動的に陳腐化したファイルを検知する機能を有し、必要に応じてマイグレーションを行っている。

ただし、陳腐化検知機能を持った特別なモジュールが存在するわけではない上、サポートしているフォーマットも少なく、日本独自フォーマットの識別にも対応していない。また、識別可能なフォーマットの追加機能も備えていない。

## 5. 参考文献

- [1] DAITSS - Trac  
<http://daitss.fcla.edu/>  
(最終アクセス日 2009年12月3日)
- [2] DAITSS 2.0 Project: Briefing and Milestones  
<http://daitss.fcla.edu/browser/daitss/trunk/docs/d2-project-summary.pdf?format=raw>  
(最終アクセス日 2009年12月3日)
- [3] DAITSS 2.0 Architecture and Timeline, Revised August 2009  
<http://daitss.fcla.edu/browser/daitss/trunk/docs/d2-ingest-architecture-and-milestones.pdf?format=raw>  
(最終アクセス日 2009年12月3日)
- [4] DAITSS 1.x OVERVIEW  
<http://www.fcla.edu/digitalArchive/pdfs/DAITSS.pdf>  
(最終アクセス日 2009年12月3日)
- [5] Ingest and Dissemination with DAITSS  
[http://www.ils.unc.edu/digccurr2007/slides/fischer\\_slides\\_8-4.pdf](http://www.ils.unc.edu/digccurr2007/slides/fischer_slides_8-4.pdf)  
(最終アクセス日 2009年12月3日)
- [6] DAITSS METS Document Profile for Submission Information Packages  
[http://www.fcla.edu/digitalArchive/pdfs/DAITSS\\_METS\\_SIP\\_Profile.pdf](http://www.fcla.edu/digitalArchive/pdfs/DAITSS_METS_SIP_Profile.pdf)  
(最終アクセス日 2009年12月3日)
- [7] FCLA-Digital Archive:Format Information  
<http://www.fcla.edu/digitalArchive/presents/recognizedFormats.xls>  
(最終アクセス日 2009年12月3日)