

デジタルリソースの 長期保存に関する概観

筑波大学図書館情報メディア系

杉本重雄

sugimoto@slis.tsukuba.ac.jp

目次

1. はじめに…デジタル保存とは
 - 基本スタンス、ことば
 - 記録文書のライフサイクルとアーカイブ
 - デジタル化の意味すること
2. デジタル保存に関する基本的な話題
 - 保存について考える
 - 関連する表儒、プロジェクト、技術など
 - アーカイブシステムとデジタル保存を俯瞰する
3. デジタル保存におけるメタデータ
 - よく出てくる標準…OAIS, PREMIS, METS
4. まとめ

1. はじめに…デジタル保存とは

- この講演の基本スタンス
 - デジタル保存（Digital Preservation）に関連する単語や概念
 - 記録文書のライフサイクルとアーカイブ
 - デジタル化（デジタルコンテンツ）の意味すること
 - 保存について考える
-
- ここでは技術的な観点から述べることとし、社会制度（法制度やアーカイブ機関の問題等）、人材養成などは、重要な問題ではあるがここでは扱わない

はじめにー基本スタンス

- 保存に関する立ち位置
- 技術的な視点から議論する
- 無意味な質問の例（自分にとっては）
 - 「何年間保存できればよいか？」
 - ⇒ 実際的な疑問かもしれないが、「長期間」というのが多くの場合の答え
 - ⇒ 「保存は次世代につなぐこと」ととらえる
 - 「和紙は1000年持つが、デジタルも1000年持つか？」
 - ⇒ 和紙であっても適切に維持管理されなければ残らない
- 保存を俯瞰的にとらえたいー領域固有の要件と領域間共通の要件
- 保存されるリソースに応じた組織化、提供、保存の技術と環境
- 領域にまたがって共通の保存機能を提供するインフラ
- 領域間で情報を連携させるための環境

ことば…

Digital Archive/Curation/Preservation

- Digital Archive

- 日本では広く使われる。文化財を中心としてデジタルコンテンツを収集、蓄積、提供するサービス
- 欧米ではDigital Library、Digital Curationの方が伝わりやすい印象
- Archiveということばには保存・保管という意味が含まれるので、長期にわたる保存は暗黙に含まれていると考える

- Digital Curation

- 価値あるデジタルリソースを収集、蓄積、保存、提供するサービス
- 長期利用を可能にすることで価値を付加することを目的*
- Curationの語源にはcare (cure) といった意味合い
- Data Curationということばもある…最近のData Science、Research Dataの収集・長期利用といったことばとつながる

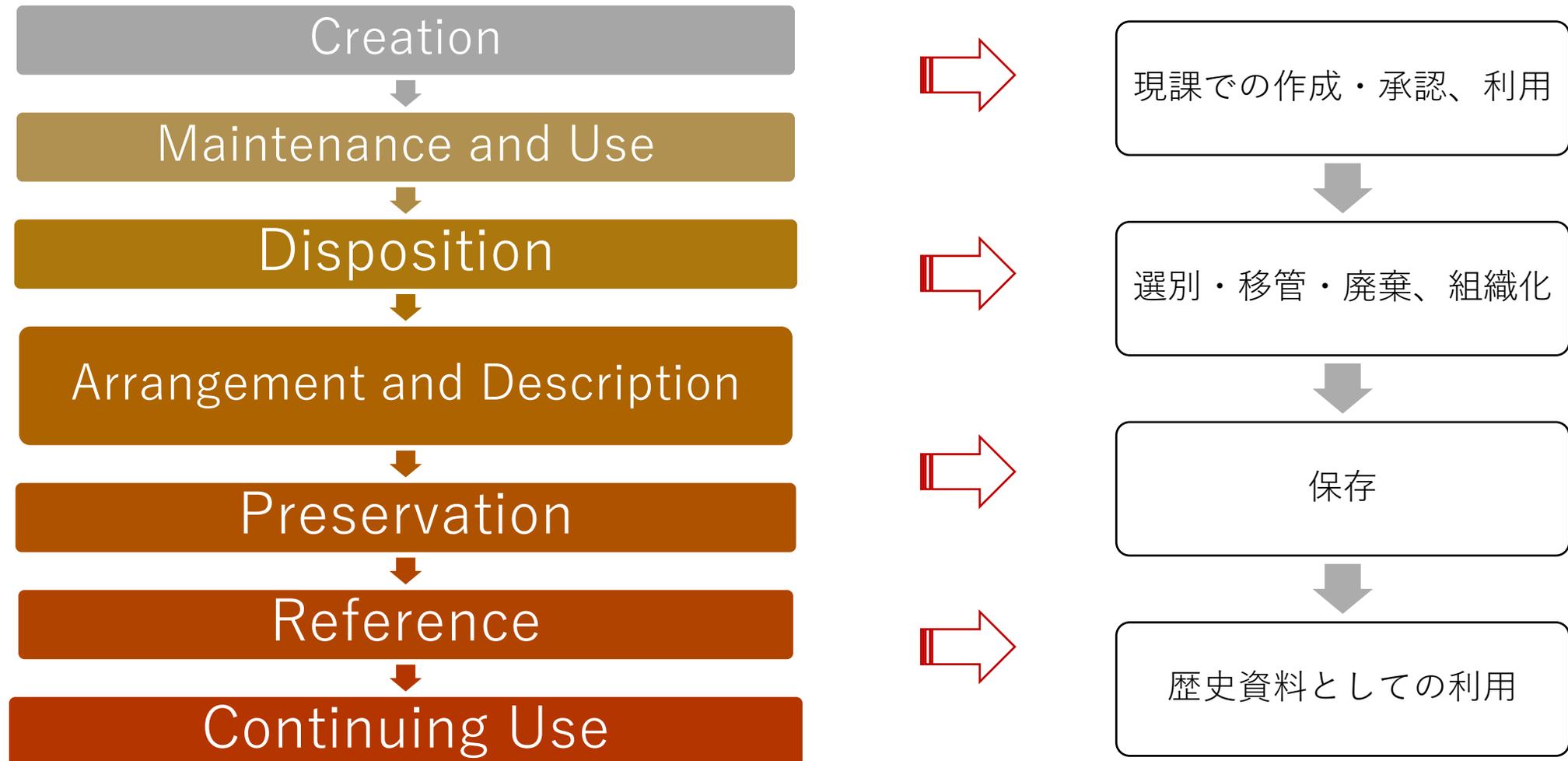
* Digital Curation Centre(<http://www.dcc.ac.uk/>)より

ことは…

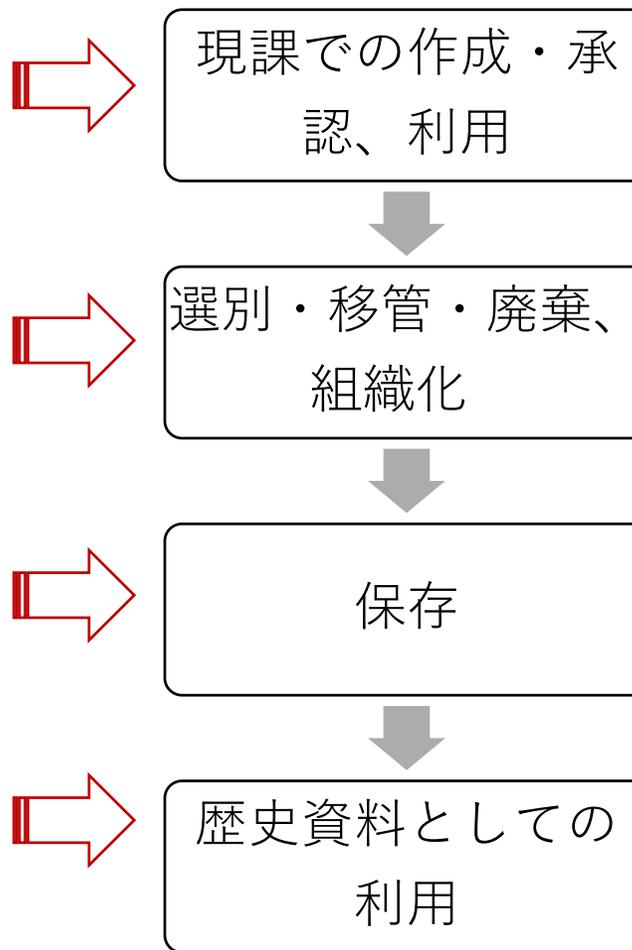
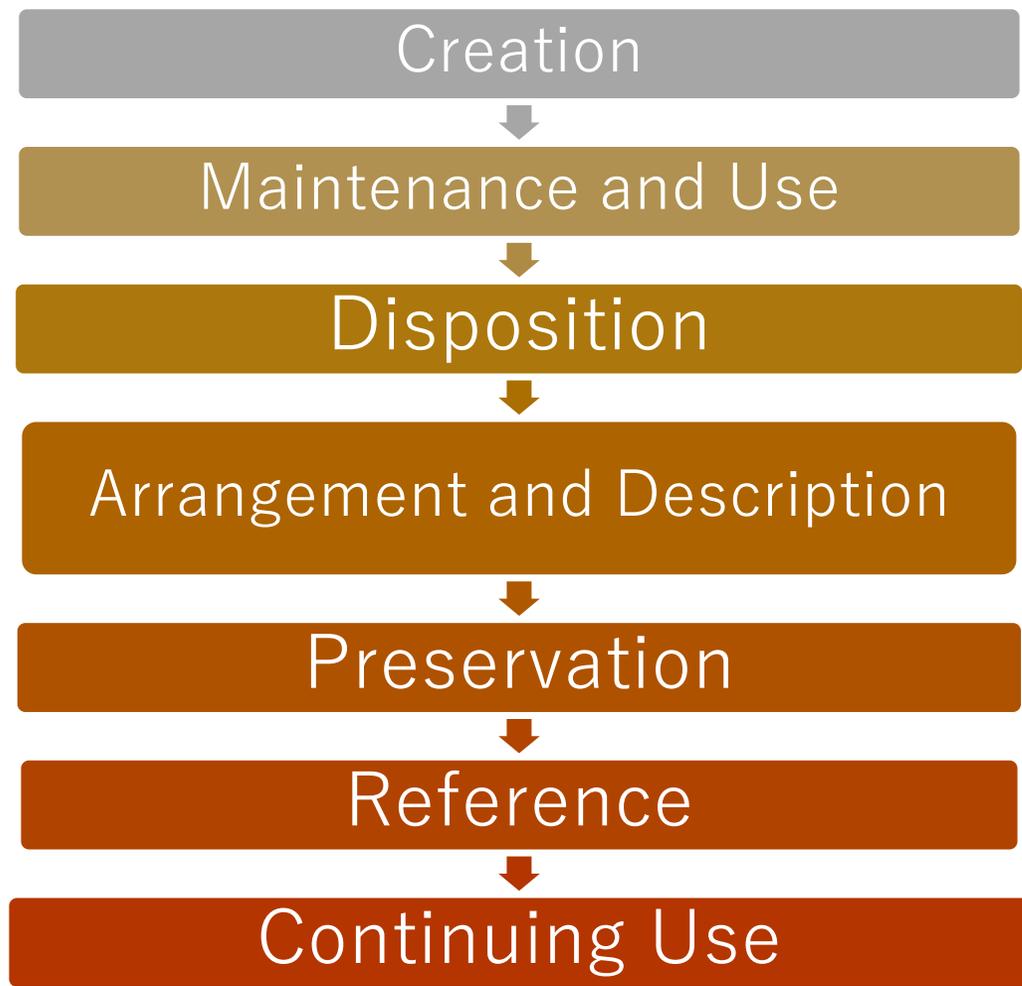
Digital Archive/Curation/Preservation

- Digital Preservation…デジタルリソースの保存
 - デジタルリソースはさまざま…デジタル化して作られた資料から、Born Digital 資料、データベースまで
 - CD等のパッケージに入れたリソース（パッケージ型リソース）と特定のパッケージに入れずにネットワーク等を介して提供するリソース（ネットワーク型リソース）
 - 保存と長期利用はほぼ同じ意味
 - なぜか、 Digital Conservationという語はデジタル保存では耳にしない
- 他に、関連する用語
 - Digital Library (Archives, Museum)： 図書館、文書館、ミュージアムの機能をネット環境に実現したもの（サービス）
 - Research Data, Digital Humanities等

記録文書のライフサイクルとアーカイブ



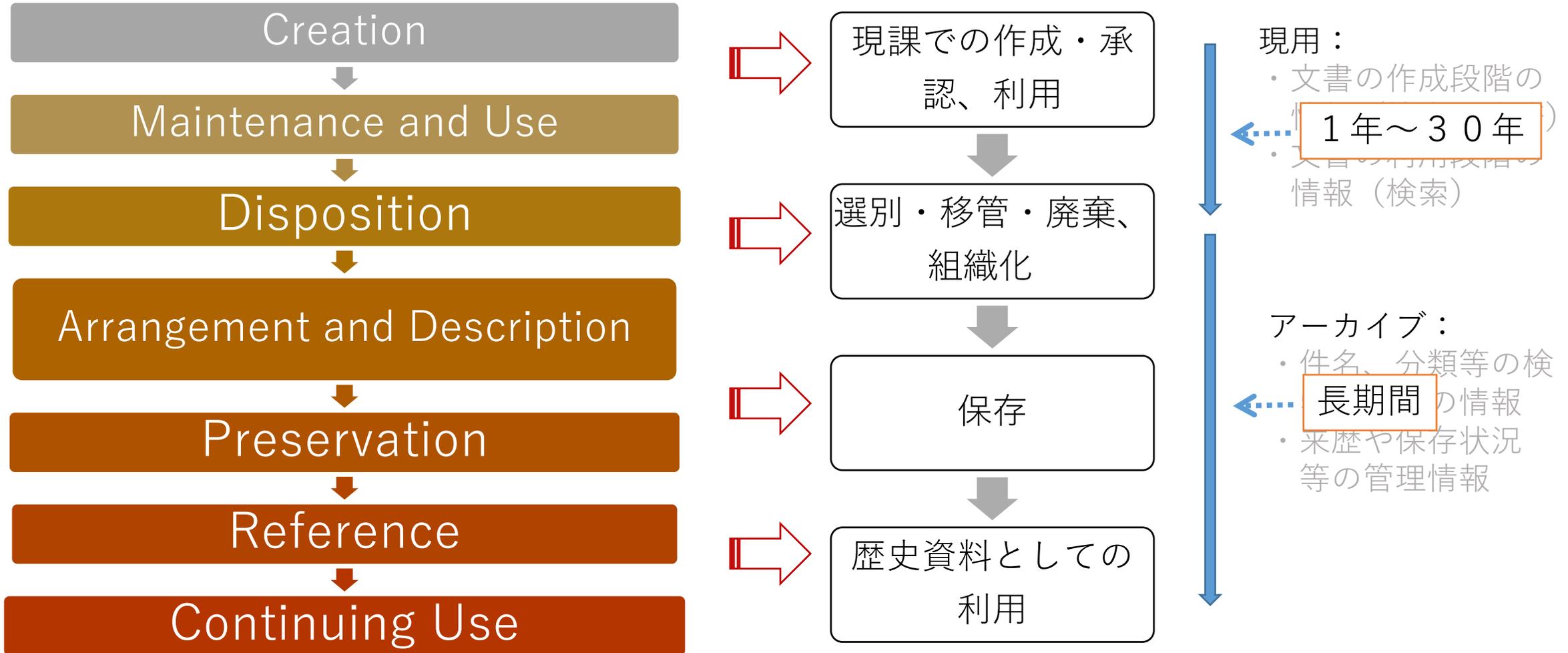
記録文書のライフサイクルとアーカイブ



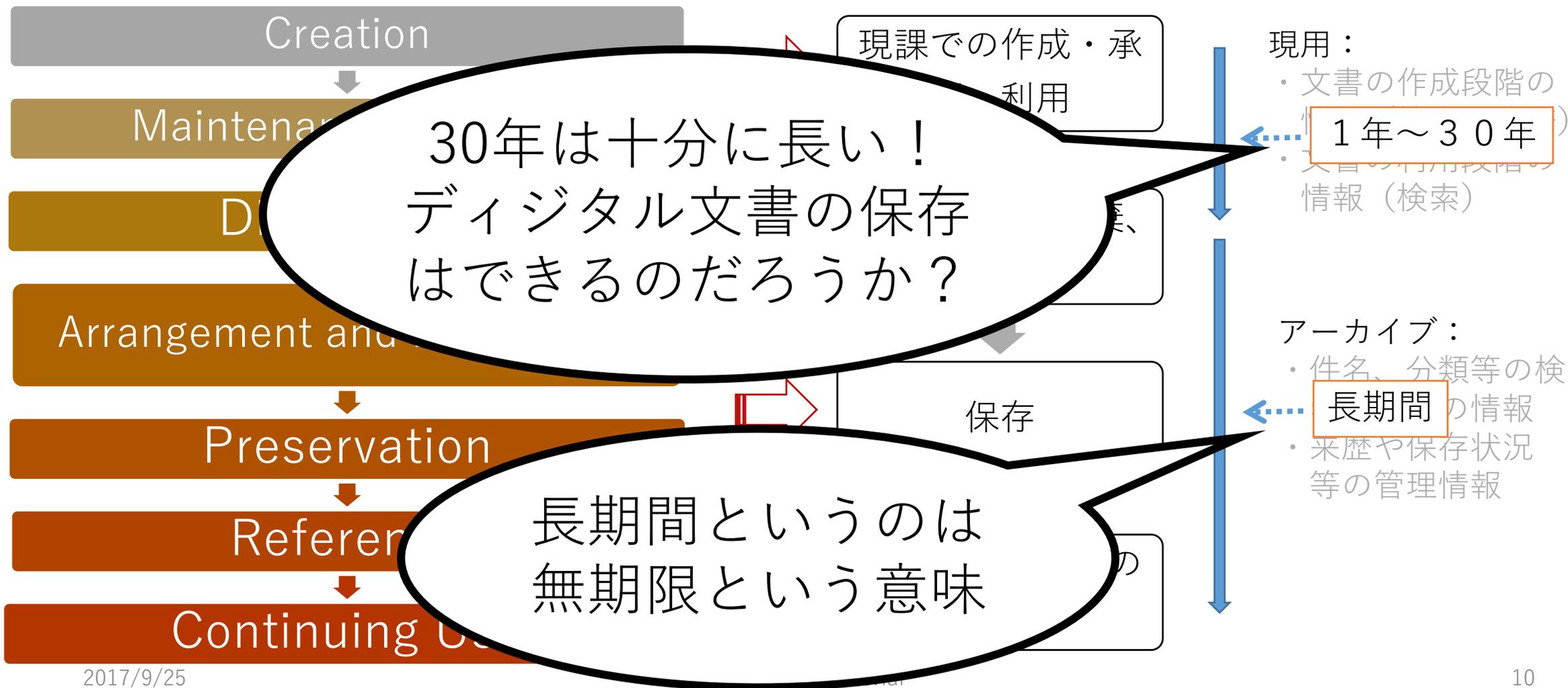
現用：
・文書の作成段階の情報（検索、決済）
・文書の利用段階の情報（検索）

アーカイブ：
・件名、分類等の検索のための情報
・来歴や保存状況等の管理情報

記録文書のライフサイクルとアーカイブ



記録文書のライフサイクルとアーカイブ



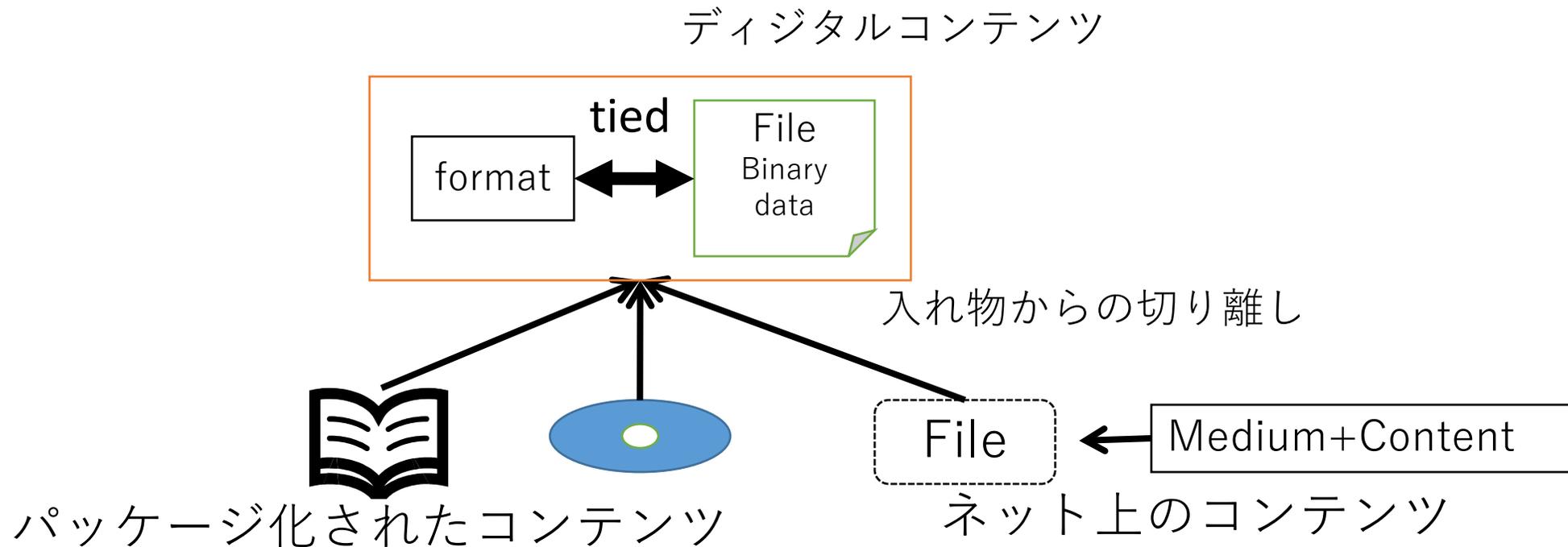
デジタル保存の前に デジタル化の意味することを考える

- コンテンツを運ぶ媒体である本や雑誌は物理的な実体。電子ファイルであっても、コンピュータの記憶装置（ディスク）上におかれたファイルという入れ物と一体化した「モノ」



デジタル化の意味すること 入れ物からの中身の分離

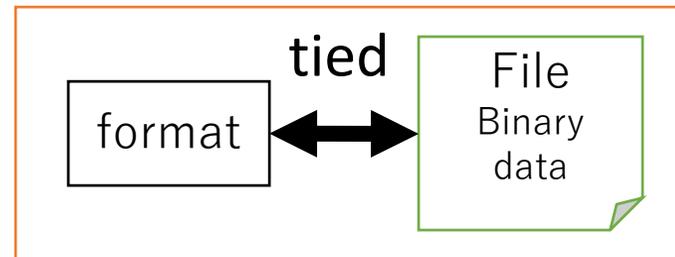
- ネットワーク環境では、中身（content）は入れ物から分離（uncouple）され、利用される。



デジタル化の意味すること 入れ物からの中身の分離

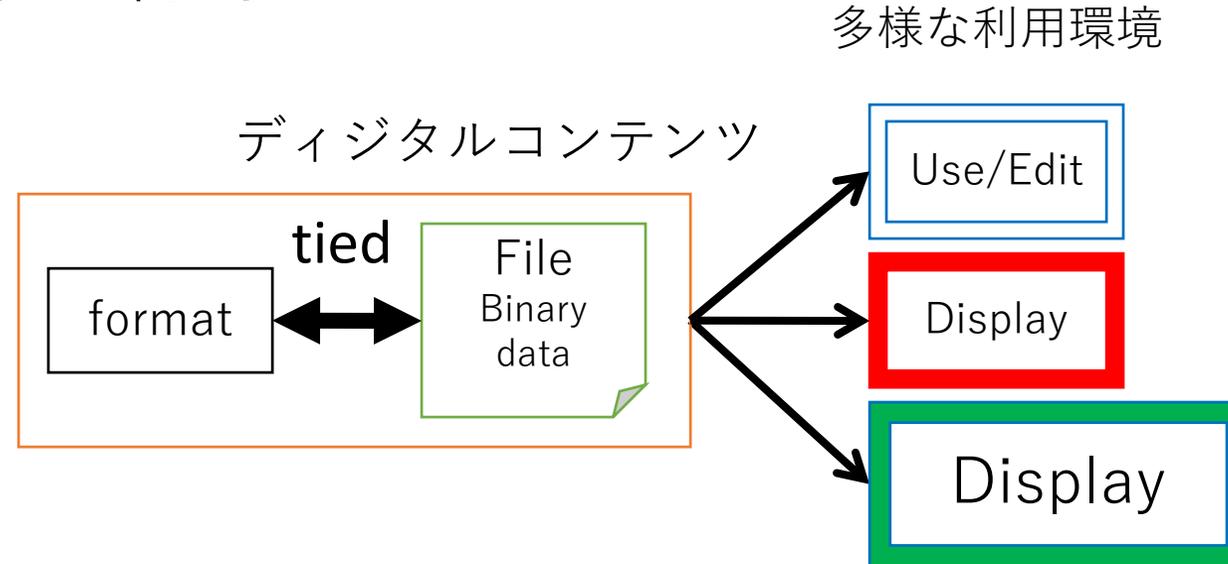
- 入れ物が見える必要はない。
- 中身が使えれば良い。

デジタルコンテンツ

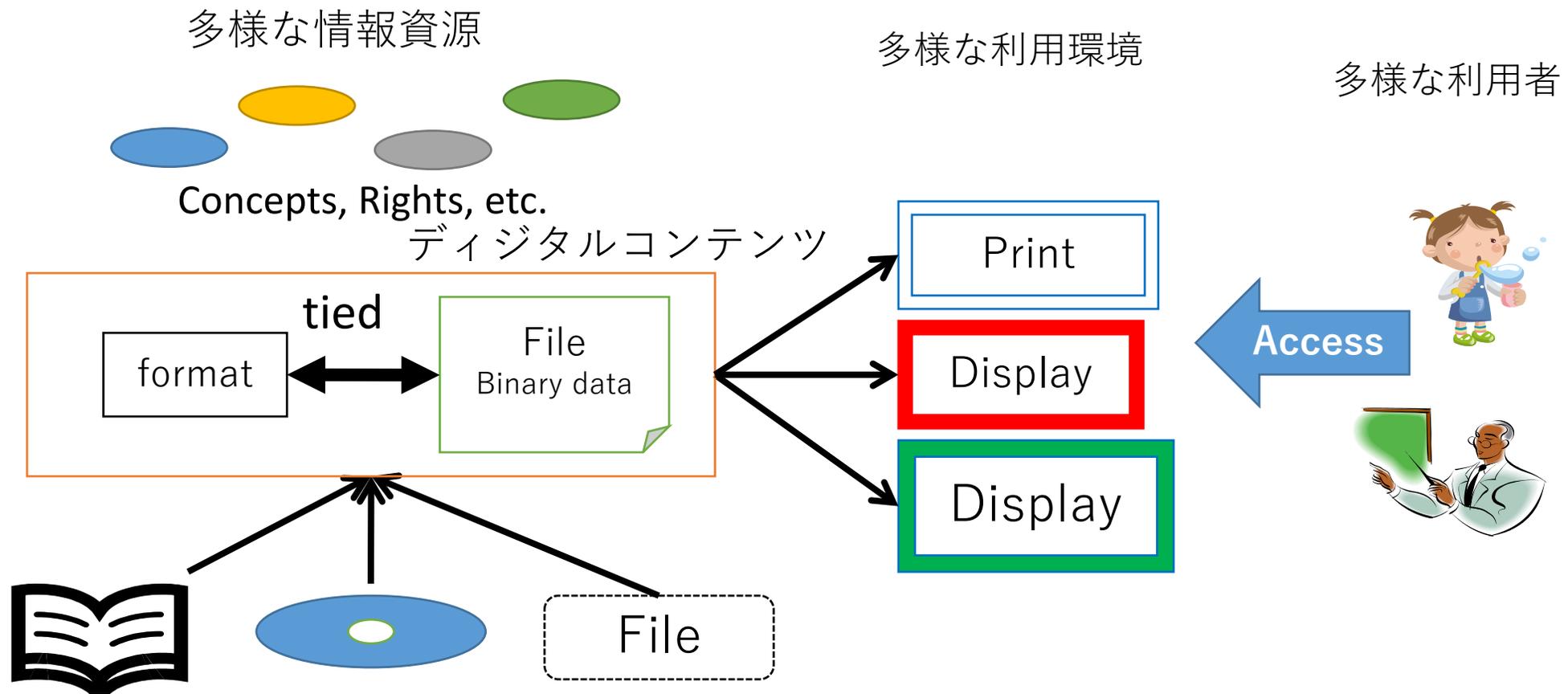


デジタル化の意味すること 入れ物からの中身の分離

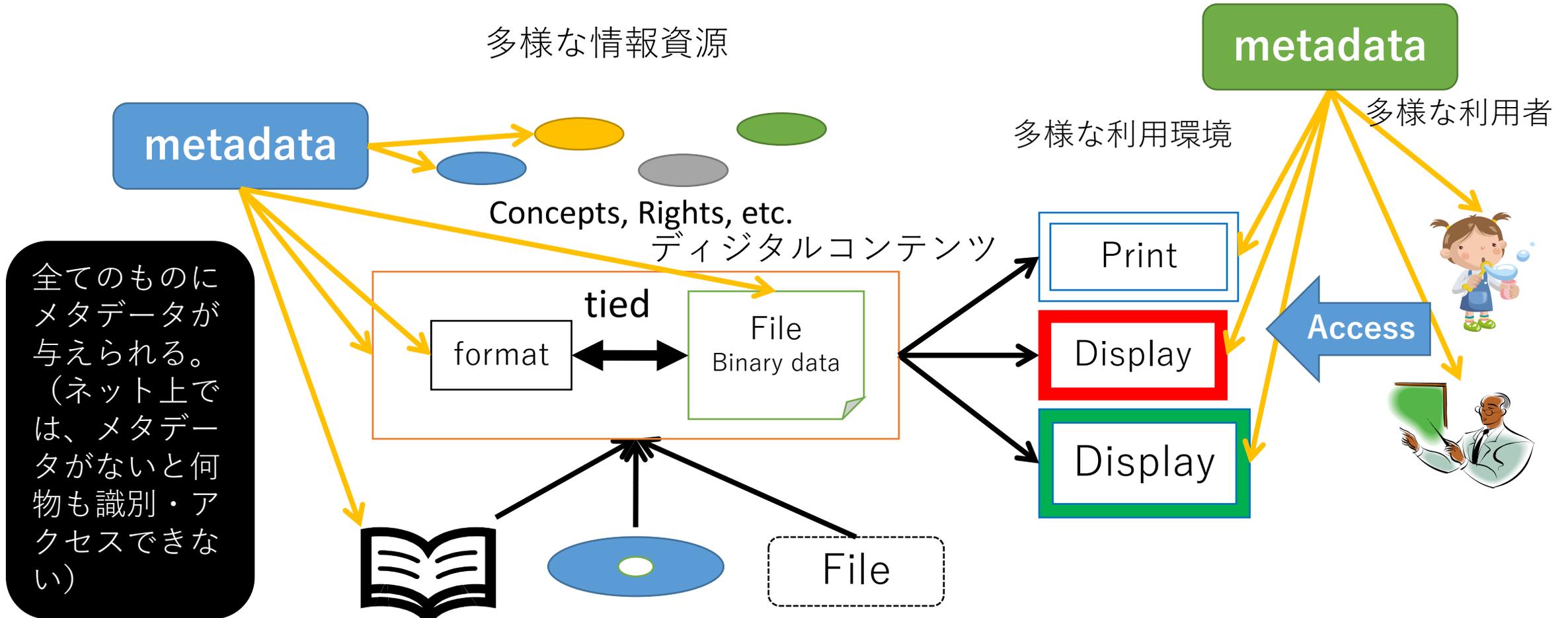
- 中身は、利用者の特性や好み、仕事の内容、利用環境の特性に応じて、いろいろな形で利用される



デジタル化の意味すること 入れ物からの中身の分離

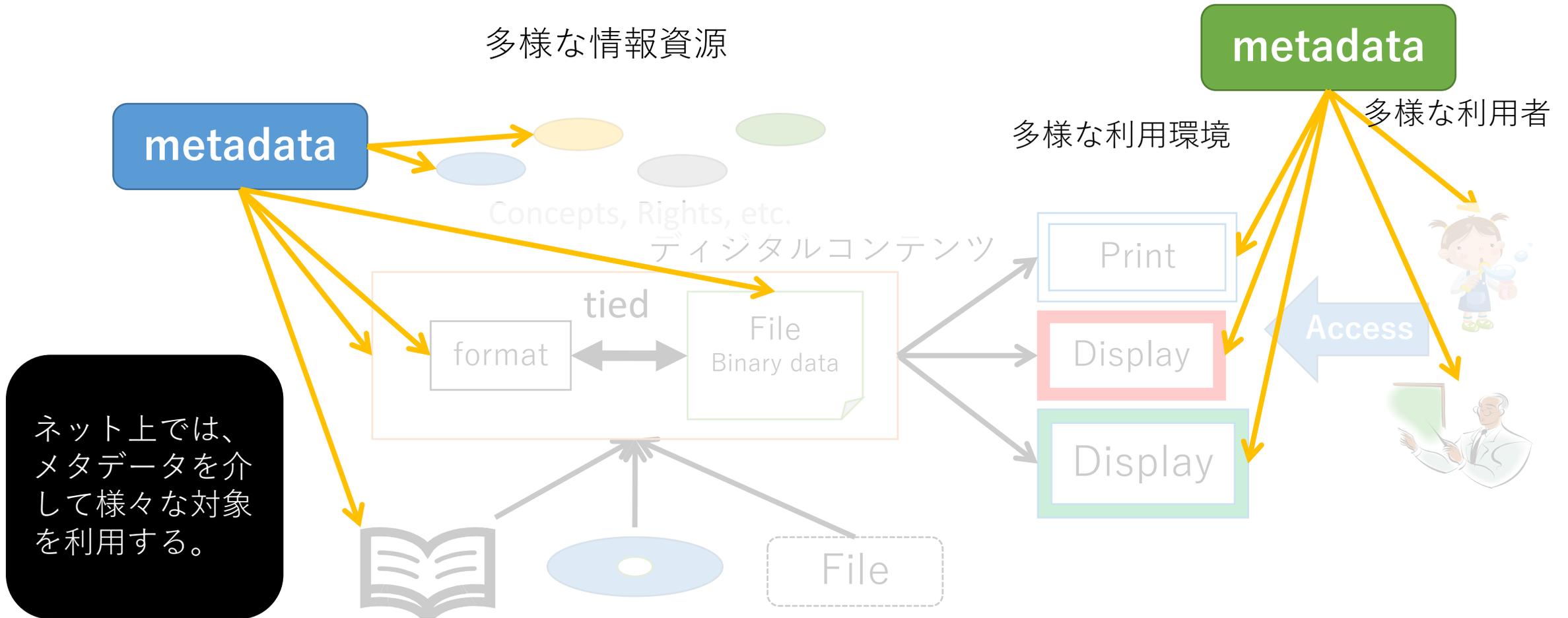


入れ物からの中身の分離とメタデータ

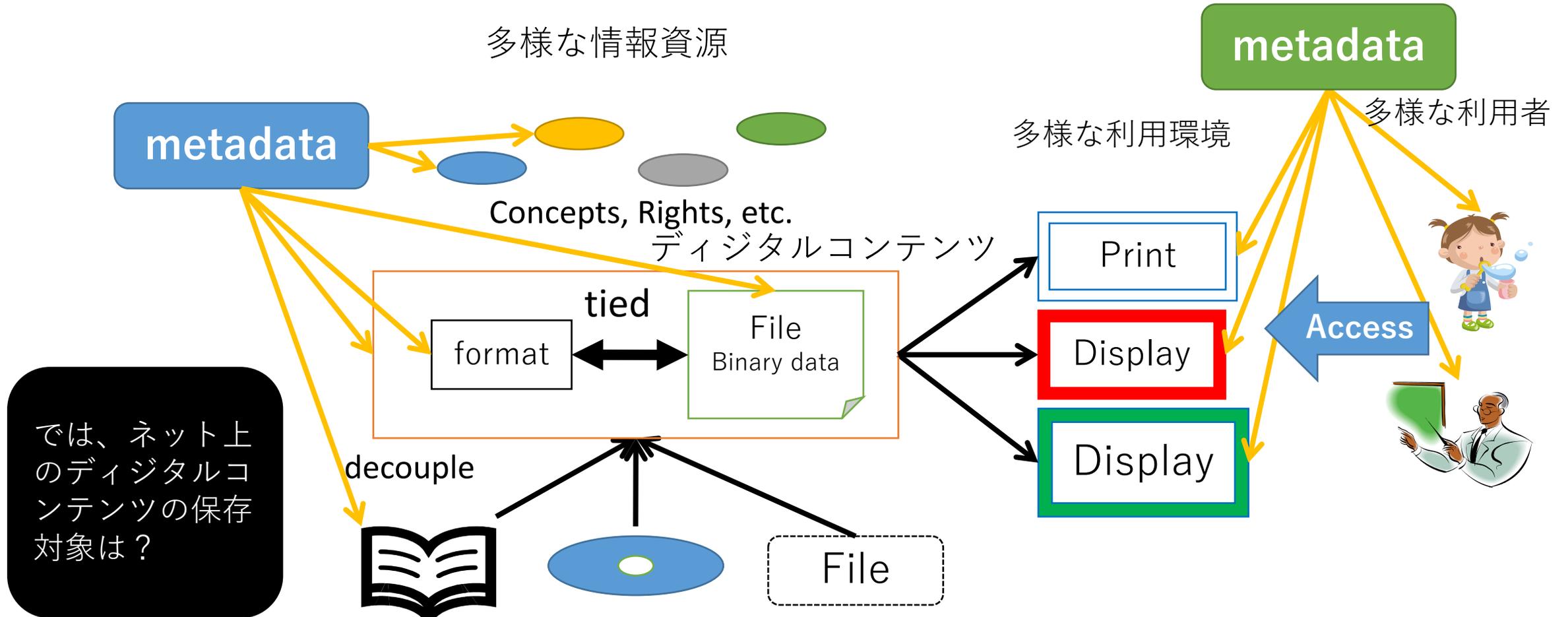


全てのものにメタデータが与えられる。(ネット上では、メタデータがないと何物も識別・アクセスできない)

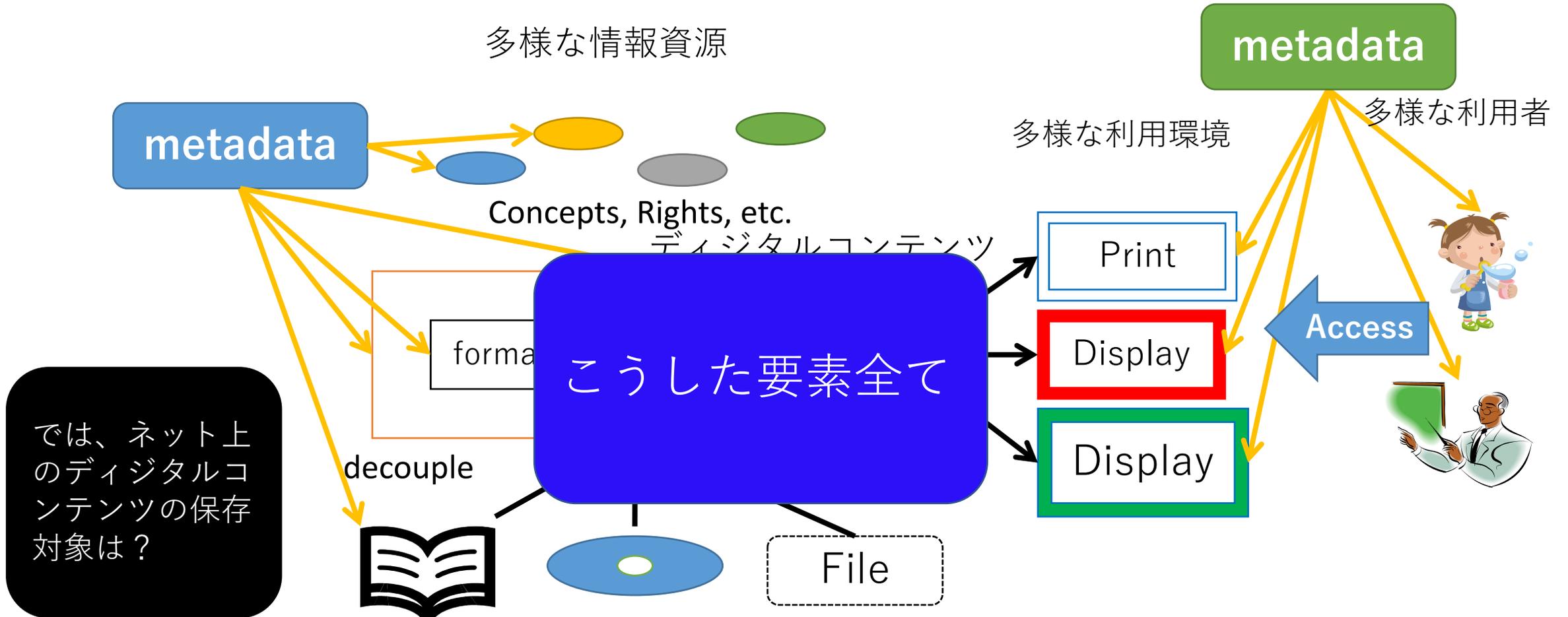
入れ物からの中身の分離とメタデータ



入れ物からの中身の分離とメタデータ



入れ物からの中身の分離とメタデータ



デジタル化の意味すること 入れ物からの中身の分離

デジタル保存の対象

デジタル形式で表現されたすべてのオブジェクトを
対象とすると考える。

- デジタルコンテンツとそのメタデータ
- 利用環境とそのメタデータ
- 利用者のメタデータ
- メタデータに関するメタデータ
- (デジタルコンテンツを格納したパッケージ)

デジタル化の意味すること 入れ物からの中身の分離

デジタル保存の対象は？
(加えて)

- コンテンツに関わる様々な権利、制約がわからないと実際に利用できなくなる
- コンテンツに関する様々な権利や制約等を示した記述（すなわち、権利や制約等に関するメタデータ）を維持管理していかなければならない

保存について考える

- 保存は、保存対象を利用可能な状態に保つための維持管理作業
 - 「使うために残すこと」と「（たまたま）残ったものを使うこと」は同じではない
 - 「和紙は1000年、デジタルは…」は、「残すこと」か「残ったもののこと」か？
- 保存はモノの保存とは限らない
 - モノの保存…冊子体書籍、映画フィルム、アニメのビデオテープ、土器、絵画等
 - 中味の保存…書籍の内容、映画そのもの、アニメそのもの
 - 機能の保存…ページをめくりながらの読書、動画像（映画、アニメ）の視聴
 - 技術の保存…編集・印刷・製本の技術、映像を再生するための技術
 - 文化の保存…書籍や映画、アニメの中味や技術を理解するのに必要な情報の保存
- たとえば、蒸気機関車（モノ）の保存
 - 静態保存…どのような形、大きさであったかを実感させてくれる
 - 動態保存…実際の動き、機関車の機能・性能を体験できる
 - 動態保存のためには維持管理するための技術、動かすための技術の保存も必要

保存について考える

- デジタル保存
 - 基本的な考え方は同じ→利用可能な状態を保つための維持管理
 - 領域、保存目的、利用方法によって保存要件は異なる
 - 分野や目的に特化した要求からビットレベルの高信頼性・長期保存機能のような他分野に共通の要求までさまざま
 - 「何年もてばよいか」ではなく、「次世代につなぐにはどのようにすればよいか、何が必要か」と考えるべき
 - 環境コストを含めた保存コストの議論は重要
- デジタルリソース
 - デジタルとアナログの基本的な違いは、中味と入れ物の分離可能性
 - 入れ物メディアに依存するデジタルコンテンツはある
 - デジタルコピーでは、ビットレベルで全く同じ複製ができる

保存について考える — よく出る話題

- ボーンデジタルコンテンツ (Born Digital) とデジタル化コンテンツ (Digitized)
 - 文化財のデジタルアーカイブはデジタル化コンテンツが多い
 - 電子文書、ゲーム、デジタル写真・ビデオ等はボーンデジタル
 - センサを使って採取したデータから作る3DCGコンテンツは？
- Born Digitalコンテンツ保存の困難さが言われるが、それは何を意味しているのか？
 - PDFのように保存を意識したフォーマットを持つ標準もある。
 - 今や、文書も写真もビデオもBorn Digital
 - フォーマットの陳腐化はBorn Digitalに限った問題ではない
 - 動的コンテンツの保存の問題ととらえればよいであろう

保存について考える — よく出る話題

- 保存の仕組み

- エミュレーション (emulation) : デジタルコンテンツを再生するためのシステム環境を別のシステム上に作る
 - 例えば、家庭用ビデオゲームのエミュレータをPC上に作る
 - エミュレータがOS依存だとすれば、エミュレータの長期利用性も課題
- マイグレーション (migration) : 新しい環境にコンテンツを移し替える
 - ひとつひとつのコンテンツの移植に対応しなければならない
- リフレッシュ (refresh) : ファイルの内容 (ビット列) を新たにコピーしなおす

保存について考える — よく出る話題

- 保存対象の種類
 - パッケージに入ったコンテンツの保存
 - コンテンツの取り出しに機器を必要とするものの共通の問題
 - VHSテープ, CD, Floppy, MO, etc.
 - メディアの劣化、再生機材の消滅の問題
 - ビット列としてのデジタルコンテンツ
 - ファイルフォーマットとは無関係にビットデータとして保存する
 - ワープロ等、再生ソフトウェア依存の静的コンテンツ
 - 保存対象コンテンツが持つすべての機能性（テキスト検索、レイアウトなど）を残したいのか、限定しても構わないのか
 - リンクを含むコンテンツの場合、リンクの種類に応じた保存
 - プログラム、スクリプト、それらを含む動的なコンテンツ
 - ビット列としてのプログラム（動作するもの）の保存
 - ソースコード、実行環境や環境要件に関する文書（これは複雑）の保存
 - VRを用いた高品位な文化財デジタルアーカイブの保存

保存について考える – 電子文書を例に

- 紙の文書であれば、物体としての文書を保存するのが基本
 - マイクロフィルム化して保存することも昔から行われてきた
 - ⇒ 「文書の保存 = 中身の保存」を意味する
 - ⇒ 紙の文書が持つ機能（ページめくり等）は保存の対象外を意味する
- 電子文書の保存も同様に考える
 - 文書の編集・閲読に利用するワープロのバージョン変更を認めるか？
 - レイアウトに影響することがある
 - 文書に含まれるハイパーリンクを有効な状態で残さねばならないか？
 - テキスト検索機能等の電子文書が持つ機能を残さねばならないか？
 - 字体の変更、外字等への対応は？
- 何を保存するかは保存方針の問題であり、方針に基づいて適切な技術を選び、維持管理するための環境を作るべき

2. デジタル保存に関する基本的な話題

- 関連する標準、プロジェクト、技術などについて
 - ここでは、いくつかの事例を示すのみ
- アーカイブシステムの視点から見たデジタル保存
- デジタル保存の機能を俯瞰する

関連する標準、プロジェクト、技術など

- 標準…OAIS, PREMIS, METS
 - OAIS …保存システム（アーカイブシステム）の参照モデル
 - デジタル保存でよく参照される標準…デジタルコンテンツをパッケージに入れて倉庫に預けるイメージで考えると良い
 - PREMIS…デジタル保存のための記述のメタデータ
 - デジタル保存に特化して作られたメタデータ標準
 - METS…デジタルアーカイブのための記述のメタデータ
 - デジタル化して作った文化的コンテンツのアーカイブ（Making of America II）での経験をベースにして作られたメタデータ
 - いくつかのファイルをひとまとめにパッケージ化する
- この話題については、後で詳しく

関連する標準、プロジェクト、技術など

- プロジェクト：90年代終わり頃からいくつかの取り組み
 - CEDARS イギリス 1998
 - PANDORA 1996 オーストラリア国立図書館 … Web archive
 - NEDLIB オランダ国立図書館 1998 （ヨーロッパの国立図書館連携）
 - NDIIPP 米国議会図書館 2000
 - Nestor ドイツ 2003
 - 国立国会図書館における電子資料保存
 - パッケージ系とネットワーク系に分けた議論は90年代から
 - 電子資料保存に関する調査研究・報告書（2003～）
 - Web archive (WARP)は2002年より

関連する標準、プロジェクト、技術など

- 電子ジャーナル（電子出版物）の保存
 - 90年代にデジタルライブラリ開発がすすめられた当時は図書館側での電子化（特に遡及的電子化）が多くあったが、出版社側の電子ジャーナル化が進むことにより、論文の保存に関する不安が議論された
 - LOCKSS Lots Of Copies Keeps Stuff Safe (<https://www.lockss.org/>)
- ゲームの保存
 - 動的コンテンツの代表としてのビデオゲームは大きな話題（代表的な話題）
 - 立命館大学ゲーム研究センターでの取り組み

関連する標準、プロジェクト、技術など

- Web archiving
 - Web上のコンテンツはその時代を表す重要な資料であることの認識
 - クローラーによる収集
 - 網羅的収集と選択的収集
 - Internet Archive (<https://archive.org/>)
 - NDLインターネット資料収集保存事業 (WARP, <http://warp.ndl.go.jp/>)
 - International Internet Preservation Consortium (IIPC) : 国際的な連携
 - クローラーやメタデータツールの共有<https://netpreserve.org/web-archiving/tools-and-software/#tools--software>

関連する標準、プロジェクト、技術など

- 公文書・業務文書の電子保存
 - US National Archives and Records Administration (NARA)
 - ERA: Electronic Records Archives
 - Digital Preservation Strategy:
<https://www.archives.gov/preservation/electronic-records.html>
 - 公文書の電子化と電子保存は、Australia, UK, Singapore他、多くの先進国で進む
 - 国立公文書館
 - デジタルアーカイブ： 電子化文書と電子公文書の保存・提供
 - InterPARES：公的な記録文書の電子保存に関する国際プロジェクト
 - University of British Columbiaが中心のプロジェクト
(<http://www.interpares.org/>)

関連する標準、プロジェクト、技術など

- デジタル保存一般に関する組織・サービス（重要な情報源）
 - Digital Curation Centre (DCC) : <http://www.dcc.ac.uk/>
 - イギリスのJISC（高等教育機関の情報環境に関する組織）の下で作られた組織
 - Int'l Conf. on Digital Curation、種々のプロジェクトを進める
 - Digital Preservation Coalition (DPC) : <http://www.dpconline.org/>
 - 多くの図書館等の組織の連携によってつくられた非営利組織
 - プロジェクトを進めるほか、ハンドブックやいろいろなツール等の紹介
 - 米国議会図書館（LoC）デジタル保存サイト : <http://www.digitalpreservation.gov/>
 - NDIIPPから続くLoCでのデジタル保存に関するサイト
 - デジタル化のガイドラインやLoC推奨フォーマット等、種々のリソースの提供

関連する標準、プロジェクト、技術など

- URIの長期利用性
 - Web上でのリンク切れの問題
 - URI (Uniform Resource Identifier)
 - URL (Uniform Resource Locator)とURN(Uniform Resource Name)
 - 永続性を持つ識別子
 - DOI (Digital Object Identifier)
 - PURL (Persistent URL)
 - 永続性を持つ識別子は本当に永続的か？
 - URIの解釈は識別子を解釈し、リンク先につなぐリゾルバ(resolver)による
 - そのため、永続的識別子を扱うリゾルバが永続性を持たねばならない

関連する標準、プロジェクト、技術など

- Web Archiveされた文書に含まれるリンクからのアクセス
 - 論文の中での参考文献では「いつアクセスしたものか」をつけるルールはあるが、リンクが切れることやリンク先の内容が変わることは保証できない
 - Web Archiveで、ページ収集時点の内容を残すにはリンク先も収集時点のものでなければならない
 - 現実にはページ収集のタイムスタンプをもとに、それに近い時点で収集されたりリンク先ページにつながざるを得ない
 - MEMENTO(<http://timetravel.mementoweb.org/>) …過去のURLを使った過去のページへのアクセス
- スタイルシートやScriptを含むWebページの保存
 - スタイルシートやScriptが解釈可能な状態で保存できるかといったソフトウェア保存の問題

関連する標準、プロジェクト、技術など

- ツール等 (1/2)
 - Digital Preservation Handbook (Digital Preservation Coalition)
 - デジタル情報資源の長期維持管理とアクセスに関する国際的に信頼でき、かつ実際的な指針を示す
 - <http://dpconline.org/handbook>
 - PRONOM (UK National Archives)：ファイルフォーマットのレジストリ
 - 文化・歴史・ビジネス分野の電子記録の長期アクセス支援のためのファイルフォーマット、ソフトウェアプロダクト、並びに他の技術的要素に関する信頼のおける情報の利用者のためのオープンレジストリ
 - <http://www.nationalarchives.gov.uk/PRONOM/>
 - JHOVE：JSTOR/Harvard Object Validation Environment
 - ファイルフォーマットの同定、バリデーション、フォーマット依存の特性の識別
 - <http://jhove.sourceforge.net/>

関連する標準、プロジェクト、技術など

- ツール等 (2/2)
 - Drambora : Digital Repository Audit Method Based on Risk Assessment
 - Risk Managementの視点からのデジタル保存支援
 - <http://www.repositoryaudit.eu/>
 - Plato : Preservation Planning Tool
 - デジタル保存プランの作成支援
 - <http://www.ifs.tuwien.ac.at/dp/plato/intro/>

関連する標準、プロジェクト、技術など

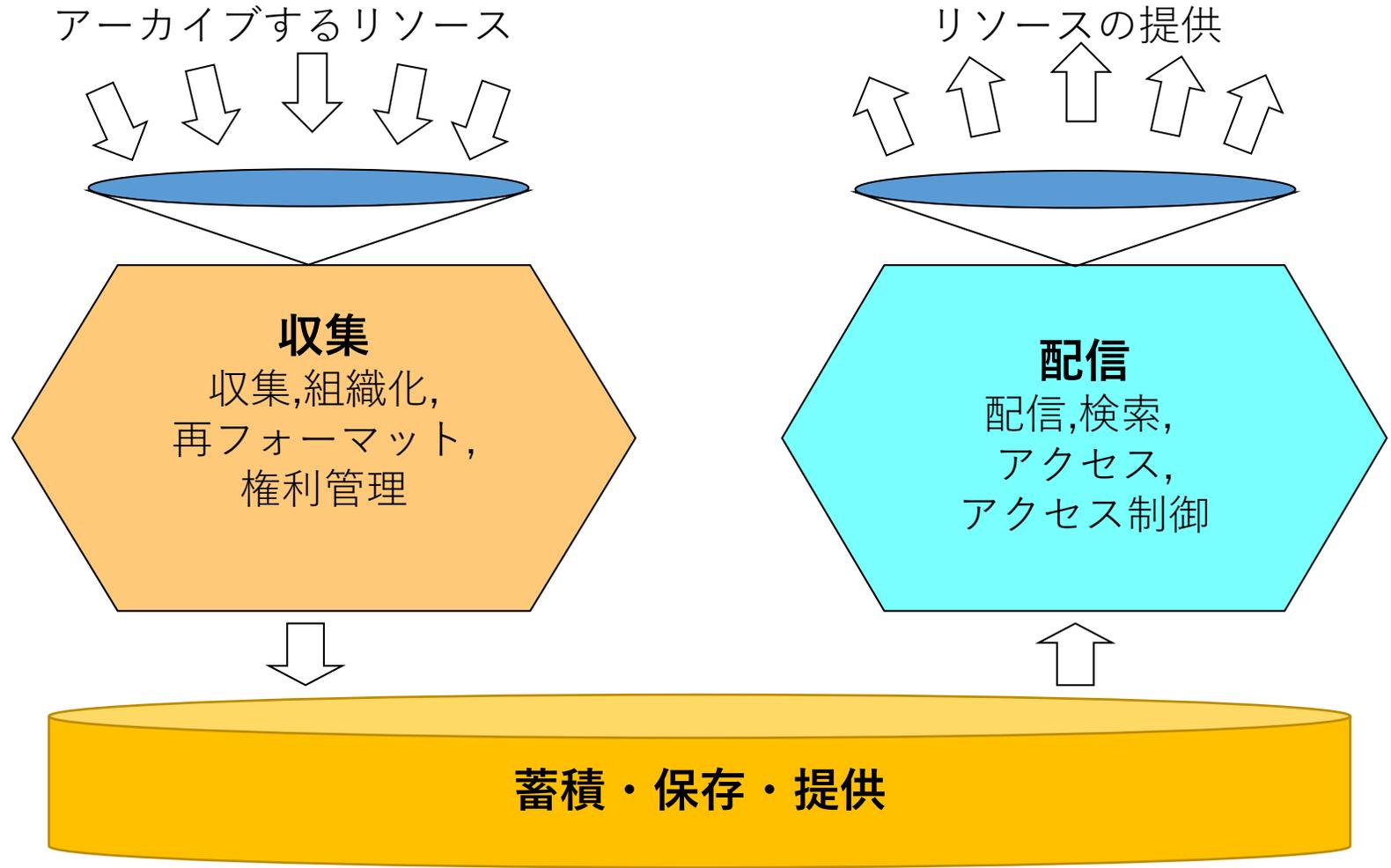
- メタデータの保存・長期利用
 - デジタルコンテンツの長期利用のためには、そのメタデータが長期利用できないといけない
 - メタデータの長期利用のためには、そのメタデータ（＝メタデータスキーマ）が長期利用できないといけない
 - メタデータスキーマの長期利用のためには、そのメタデータ…
- メタデータスキーマ（表現形式、構造的定義、メタデータ語彙等）の長期維持管理は、伝統的にはスキーマ定義の文書を保存すればよかったが、Webの進化とともに、スキーマ自体がコンピュータ処理可能かつネット上で共有できる形式で記述されるようになってきており、それに対応する長期の維持管理技術が必要とされている

関連する標準、プロジェクト、技術など

(その他)

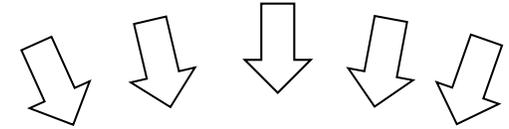
- ビットデータの長期保存：ビットレベルの長期保存は可能（保存クラウド）
- メタデータレジストリ：メタデータの長期利用・相互運用のためのインフラ
 - メタデータを記述するための統制語彙の保存
 - メタデータの構造の定義の保存
 - メタデータの作り方・解釈方法の記述の保存

アーカイブシステムの視点から

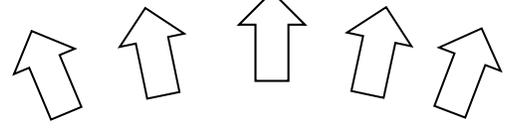


アーカイブシステムの視点から

アーカイブするリソース



リソースの提供

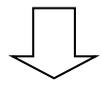


収集

収集, 組織化,
再フォーマット,
権利管理

配信

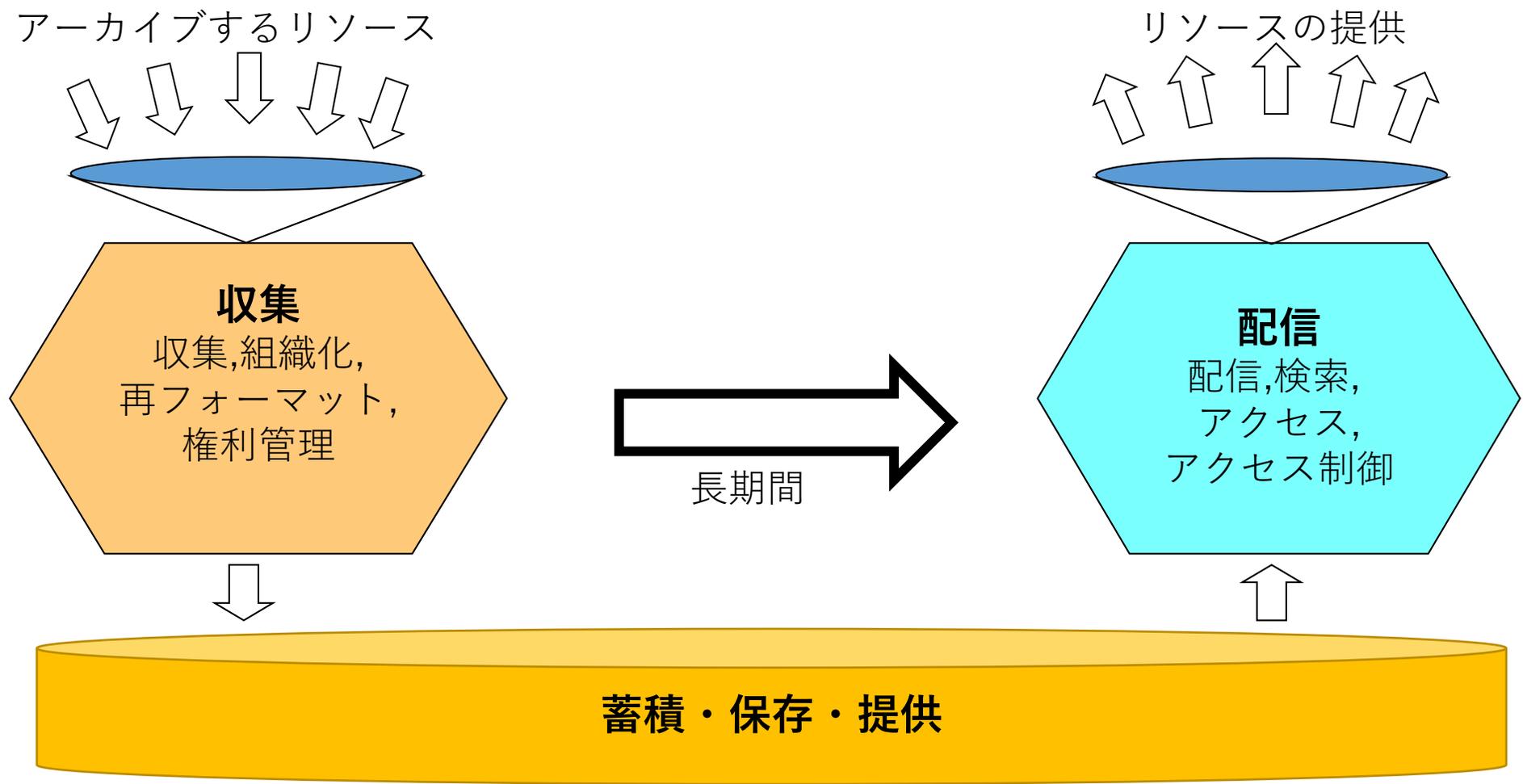
配信, 検索,
アクセス,
アクセス制御



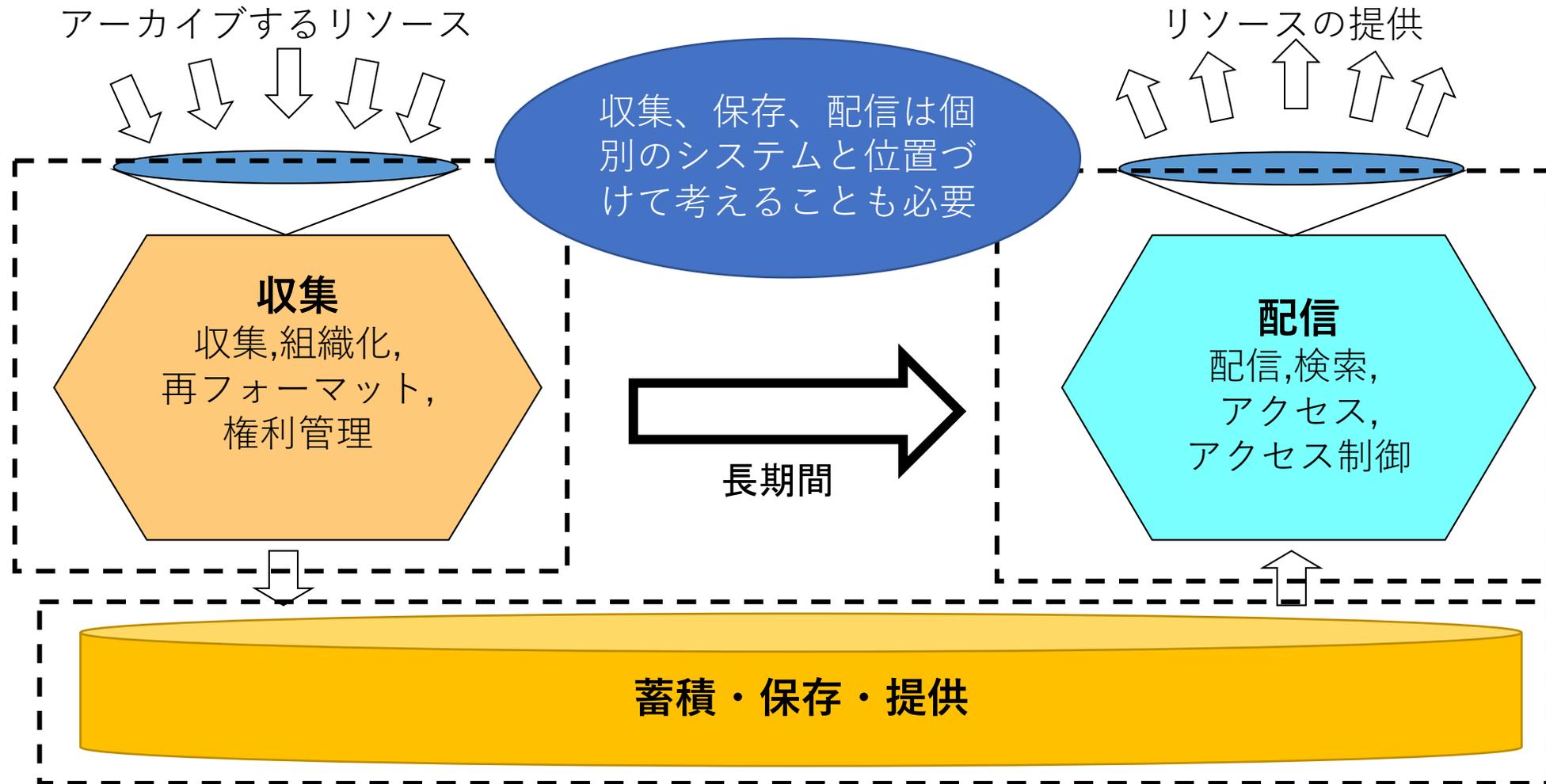
蓄積・保存・提供

ひとつの統合されたシステムとしてのデジタルアーカイブ

アーカイブシステムの視点から



アーカイブシステムの視点から



アーカイブシステムの視点から

デジタル保存 (Digital Preservation)

- デジタルコンテンツを、長期間、利用可能な状態で提供できるようにすること
- デジタルコンテンツの保存と、コンテンツを格納した記録メディアの保存は同義ではないことに注意が必要
- デジタルコンテンツの保存には、コンテンツに関連する多様な情報を利用可能な状態で残すことも求められる
- 「保存」部分だけを考えれば、種々の異なるコンテンツの保存を受け持つ「保存倉庫」サービスととらえることができる
- 収集・提供・保存を一体のシステムとしてコンテンツを管理する場合（既存の多くのデジタルアーカイブ）、データベースの長期利用・保存ととらえられる

保存

デジタル保存の機能を俯瞰する

- 以下は、デジタル保存を「デジタルコンテンツの長期間の維持管理」ととらえた場合に必要と考えられる機能（私見）
- 保存のためのシステムに要求される機能
 - 高信頼性ビットレベルデータ保存環境、低環境負荷のデータ保存環境
 - 保存対象に関わるメタデータの長期利用性の保証
 - ファイルフォーマット情報、メタデータスキーマ情報等の保存
 - 保存ポリシー、権利情報等の保存管理情報の保存、など
- コンテンツの維持管理の視点から
 - 保存上のリスク要因を収集し、リスク発見を支援する機能
 - コンテンツ間の依存関係を管理し、変更の影響範囲を管理する機能
 - 権利管理等の社会的要因を管理し、時間経過や制度の変更による影響範囲を管理する機能

デジタル保存の機能を俯瞰する

- ステークホルダーの視点
 - 図書館等の機関(MLA)、行政機関、研究機関、教育機関、出版放送等の機関、文化財の所有者、コンテンツ制作者、一般市民他さまざま
 - 現在ある知識資産とこれから作られる知識資産を集め、将来の利用者のために保存する
- インフラと個別領域機能の視点
 - 個別機関（特に中小規模機関）で頑健な保存環境を作ることは難しい
 - デジタル保存のためのクラウド（のような環境）が欲しい
 - 技術的共通機能は多いが、保存方針は個々の領域や機関に依存
 - 共同で使える技術基盤や保存に必要な情報の収集と蓄積・提供はインフラとして整備すべき

図書館・博物館・
美術館・文書館

ステークホルダー

行政
自治体・企業

コンテンツ産
業（出版・放
送・映画等）

一般市民

Packaged
Resource
Preservation
パッケージ型資
源（CD, カセッ
ト等）の保存

Digital
Conservation
Factory
デジタル資源
の補修サービス

Digital
Archives

Long-Life Active
Data

Digital Preservation
Planning (Assistance)

デジタル保存に関連する機能を 俯瞰する

メタデータを長期間利用するための基盤

OAIS-based Information
Package Repository
OAIS標準に基づき情報
パッケージを維持管理、保
存する

Trusted Bit-Level Preservation Repository
ビットレベルでの保存を保証する

機能要素

ステークホルダー

図書館・博物館・
美術館・文書館

行政機関・自治体・企業

研究機関・大学

コンテンツ産業（出版・放送・映画等）

一般市民

Packaged Resource Preservation
パッケージ型資源（CD, カセット等）の保存

Digital Conservation Factory
デジタル資源の補修サービス

File Format Registry
ファイルフォーマット情報を収集し保存提供する

Digital Archives
多様な分野の文化資源のデータベース

Long-Life Active Datasets
長期間利用するデータ
Open Data, Research Data, そのほか

Digital Preservation Planning (Assistance) Services
保存プラン作成を支援する

Metadata Schema Information Services (for Digital Resource Longevity)
デジタルコンテンツの長期利用のためのメタデータの共有と

Digital Preservation Management (Assistance) Services
保存マネジメント（コンテンツの維持管理）を支援する

Metadata Schema Standard Registry
メタデータを長期間利用するための基盤

OAIS-based Information Package Repository
OAIS標準に基づき情報パッケージを維持管理、保存する

Rights Management Services
権利管理

Trusted Bit-Level Preservation Repository
ビットレベルでの保存を保証する

黒、灰色のところは国レベルで取り組むことでスケールや高い信頼性によるメリットがあるもの

機能要素

ステークホルダー

図書館・博物館・
美術館・文書館

行政機関・自治体・企業

研究機関・大学

コンテンツ産業（出版・放送・映画等）

一般市民

Packaged Resource Preservation
パッケージ型資源（CD, カセット等）の保存

Digital Conservation Factory
デジタル資源の補修サービス

ファイルフォーマット等の情報の長期利用基盤

長期間利用するデジタルアーカイブデータベース

メタデータの共有と長期利用のための基盤

メタデータを長期間利用するための基盤

デジタルアーカイブとそのコンテンツを長期間に渡って利用できるようにするための環境、技術基盤

社会制度

ビットデータとして保存するための高信頼基盤

OAIS標準に基づき情報リソースを維持管理、保存する

黒、灰色のところは国レベルで取り組むことでスケールや高い信頼性によるメリットがあるもの

機能要素

ステークホルダー

図書館・博物館・
美術館・文書館

行政機関・自
治体・企業

研究機関・大学

コンテンツ産
業（出版・放
送・映画等）

一般市民

デジタルアーカイブを俯瞰する

ステークホルダーと機能要素

Pa
Re
Pres
パッ
源（
ト等

Con
F
ディ
の補

ights
gement
vices

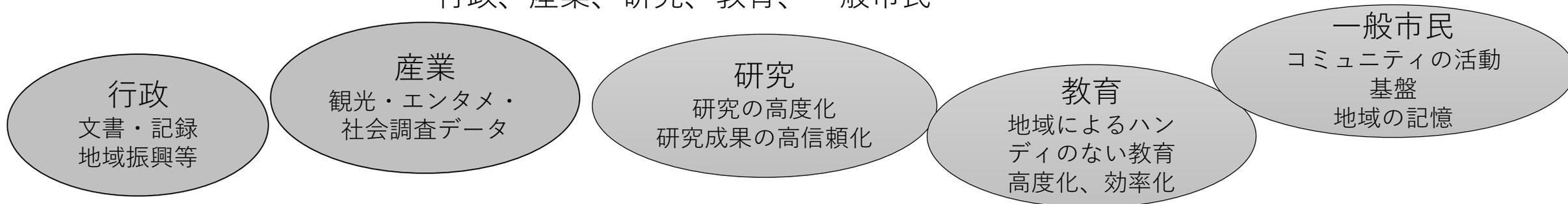
管理

のところ
ルで取り
とでスケー
い信頼性に
よるメリットがあ
るもの

ビットレベルでの保存を保証する

デジタル保存のための機能要素

利用コミュニティ
行政、産業、研究、教育、一般市民



デジタルコンテンツの収集・蓄積・提供のレイヤー
ーデジタルアーカイブ、研究データ、オープンデーター



デジタル保存の基盤レイヤー

デジタル
アーカイブ・
保存に関する
法制度、社会
基盤

インフラとして考えることのできる機能

- 保証されたビットレベルでの保存
- ファイルフォーマットの収集と提供（長期）
- メタデータの長期利用のための基盤（スキーマ情報）
- 権利管理情報の提供、など

規模のメリット
が大きい領域

利用コミュニティ

行政、産業、研究、教育、一般市民

行政
文書・記録
地域振興等

産業
観光・エンタメ・
社会調査データ

研究
研究の高度化
研究成果の高信頼化

教育
地域によるハン
ディのない教育
高度化、効率化

一般市民
コミュニティの活動
基盤
地域の記憶

デジタルコンテンツの収集・蓄積・提供のレイヤー
ーデジタルアーカイブ、研究データ、オープンデーター

多様な領域で作られるデジタルアーカイブ・データベース
利用者コミュニティ、利用方法、権利管理の考え方、応用向け技術はさまざま

デジタル保存の基盤レイヤー

デジタル
アーカイブ・
保存に関する
法制度、社会
基盤

インフラとして考えることのできる機能

- 保証されたビットレベルでの保存
- ファイルフォーマットの収集と提供（長期）
- メタデータの長期利用のための基盤（スキーマ情報）
- 権利管理情報の提供、など

規模のメリット
が大きい領域

少し立ち止まってデジタル保存を考える

- 「コンテンツの維持管理」
 - 「コンテンツ」はデジタルデータそのもの（入れ物と中味の分離）
 - パッケージごと保存しなければならないものの存在は否定しない
 - パッケージ内のデータの読み取り装置だけではなく再生装置全体を考える必要
- 「コンテンツを生かし続けるための管理技術」
 - 低エネルギー消費かつ長期間安定したビットデータ保存技術は非常に重要
 - その一方、ソフトウェア・データベースの性質上、維持管理のために保存フォーマットが変化していくことは前提と考えざるを得ない
 - マイグレーション、エミュレーション
 - 維持管理のためには管理対象のことをよくわかるようにしなければならない
- 「デジタルコンテンツの長期維持管理のための情報」
 - メタデータが必要
 - メタデータも長生きでないといけない

3. デジタル保存におけるメタデータ

- メタデータの定義（共通理解のために）
- デジタル保存のためのメタデータ関連標準
 - OAIS
 - PREMIS
 - METS

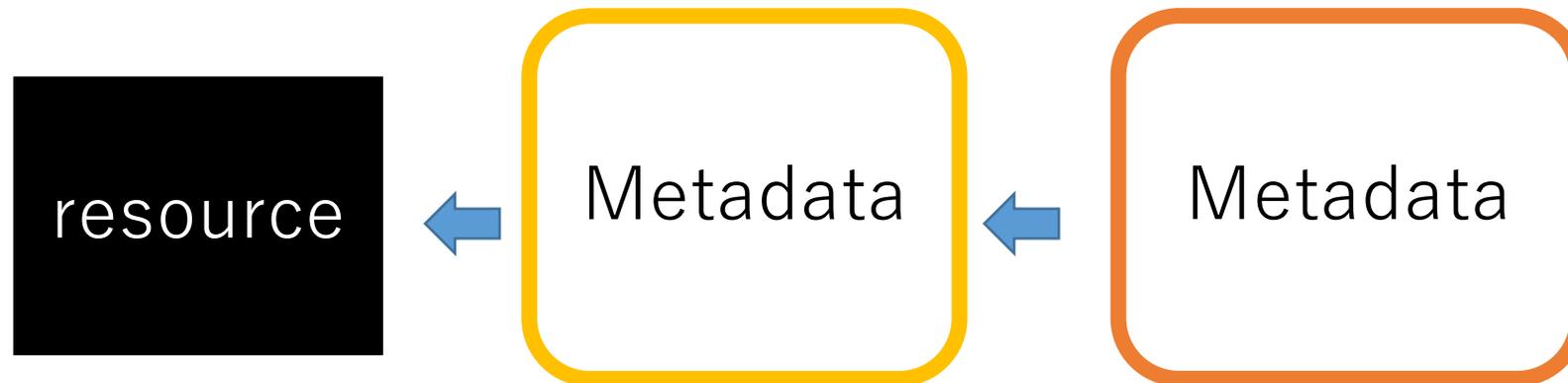
Metadata – 単純化したモデル

- メタデータ (Metadata)は、
データに関する (構造化された) データ
(Structured) Data about Data
 - 記述対象リソースに関して、何らかの視点、何らかの目的の下になされる記述



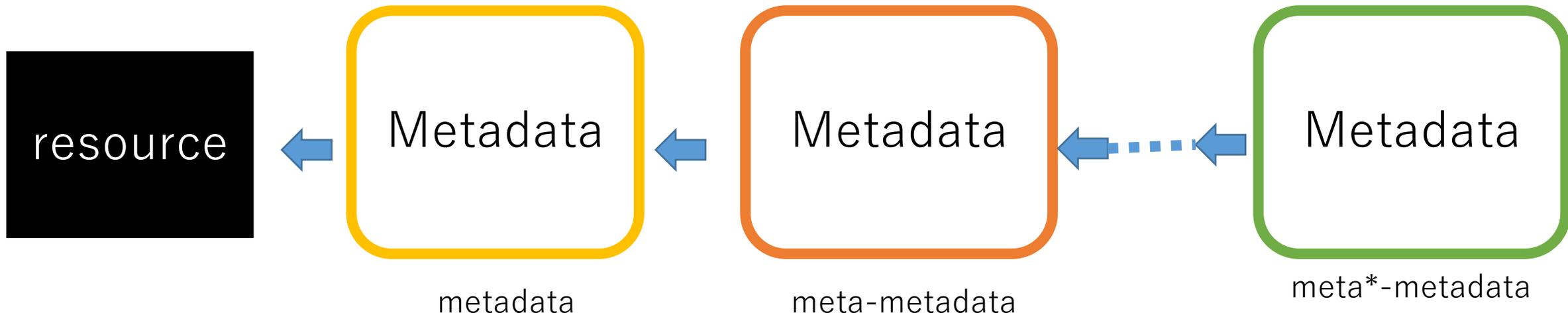
メタデータとメタメタデータ

- メタメタデータ：メタデータのメタデータ
- 例えば、メタデータスキーマはメタデータのメタデータ



メタデータとメタメタデータ

- メタデータスキーマのスキーマは？
 - ⇒ メタ・メタ・メタデータといっても構わない
- 忘れてならない点は、これらすべてがデジタルオブジェクトであり、かつ保存の対象と考えるべきであること



デジタル保存によく出てくる標準

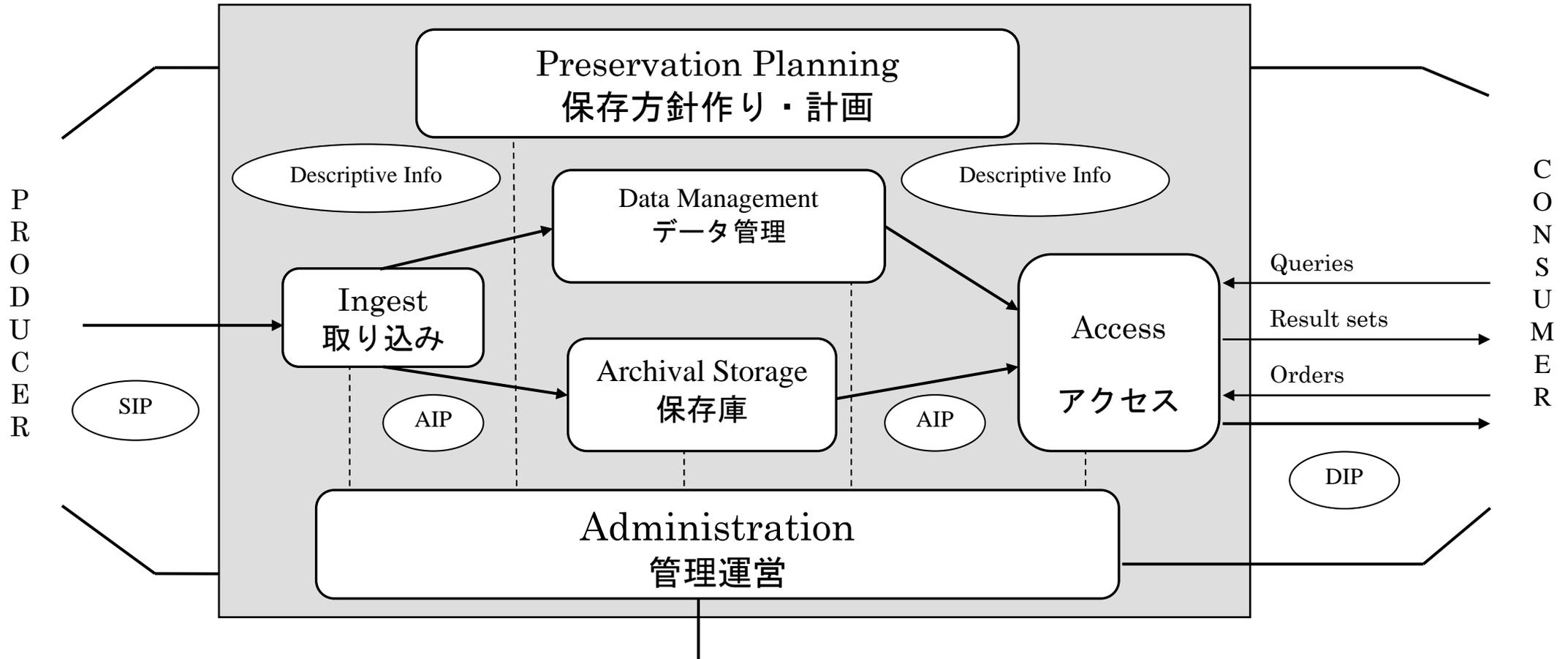
- OAIS (Open Archival Information System)
 - デジタル保存のシステムのための参照モデル
 - モノを箱に入れて保管する倉庫のイメージ
 - Information Package ← Information Object + Preservation Description Information
 - Information Object ← Data Object + Representation Information
- PREMIS (PREservation Metadata: Implementation Strategies)
 - デジタル保存のためのメタデータ標準
 - Data Dictionaryとして記述項目等を定義・提供
- METS (Metadata Encoding Transmission Standard)
 - Container (Information Package)の役割

デジタル保存の標準モデル…

Open Archival Information System (OAIS)

- OAIS は保存システムの参照モデルの国際標準
- 保存システムでは、保存プランに従った実際の保存作業
- 保存対象はパッケージに入れる (xIP) = 段ボール箱のイメージ
 - 箱に保存したいものを入れて、倉庫に保存する
 - パッケージには、外側に中身がわかる情報を書いておく
 - パッケージには保存の作業のために必要な情報を書いておく → PDI
- デジタルコンテンツを入れる箱
 - 保存のために送り込むときのパッケージ → SIP
 - 保存システムの中でのパッケージ → AIP
 - 保存システムから送り出すときのパッケージ → DIP

OAIS: システムの機能要素



SIP: Submission Information Package

MANAGEMENT

AIP: Archival Information Package

DIP: Dissemination Information Package

OAIS参照モデル 情報オブジェクト (Information Object)



- データオブジェクト（ビット列）と、それを表現し利用可能な形にするための情報（表現情報）が加わったものを情報オブジェクトとする。
 - 表現情報:たとえば、ファイルフォーマットに関する情報、イメージデータを再生するために必要な情報など

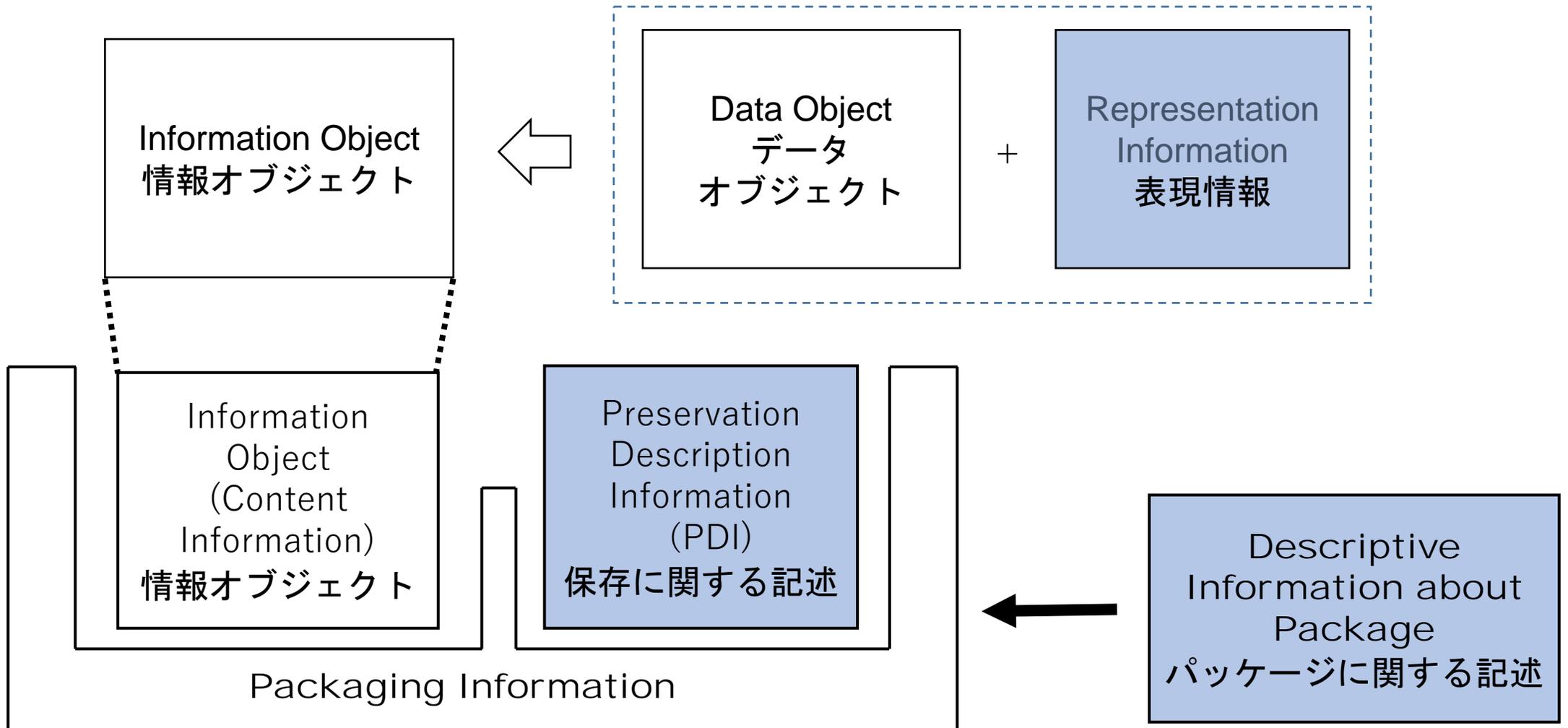
OAIS参照モデル:Information Package



PDIの記述内容

- 来歴 (Provenance)
- コンテキスト (Context)
- 参照 (Reference)
- 不変性 (Fixity)
- 権利管理情報 (Access Rights)

OAIS参照モデル:Information Packageのメタデータ



デジタル保存とメタデータ

- 表現情報、PDI、IPの記述的情報はすべてメタデータ
 - メタデータが解釈不能になればコンテンツは失われる
- メタデータの保存
 - データ実体であるメタデータの保存
 - メタデータの構造や意味を定義するメタデータスキーマの保存
- メタデータ実体の保存
 - 従来的な環境では、メタデータのデータベースの長期維持
 - LOD環境では、メタデータはサイト間で転送・共有可能な実体
- メタデータスキーマの保存
 - (従来的環境) スキーマを書いた文書の保存
 - LOD環境では、スキーマもサイト間で転送・共有可能なデータ実体
- メタデータ保存に関する新しい考え方が必要

OAISについて

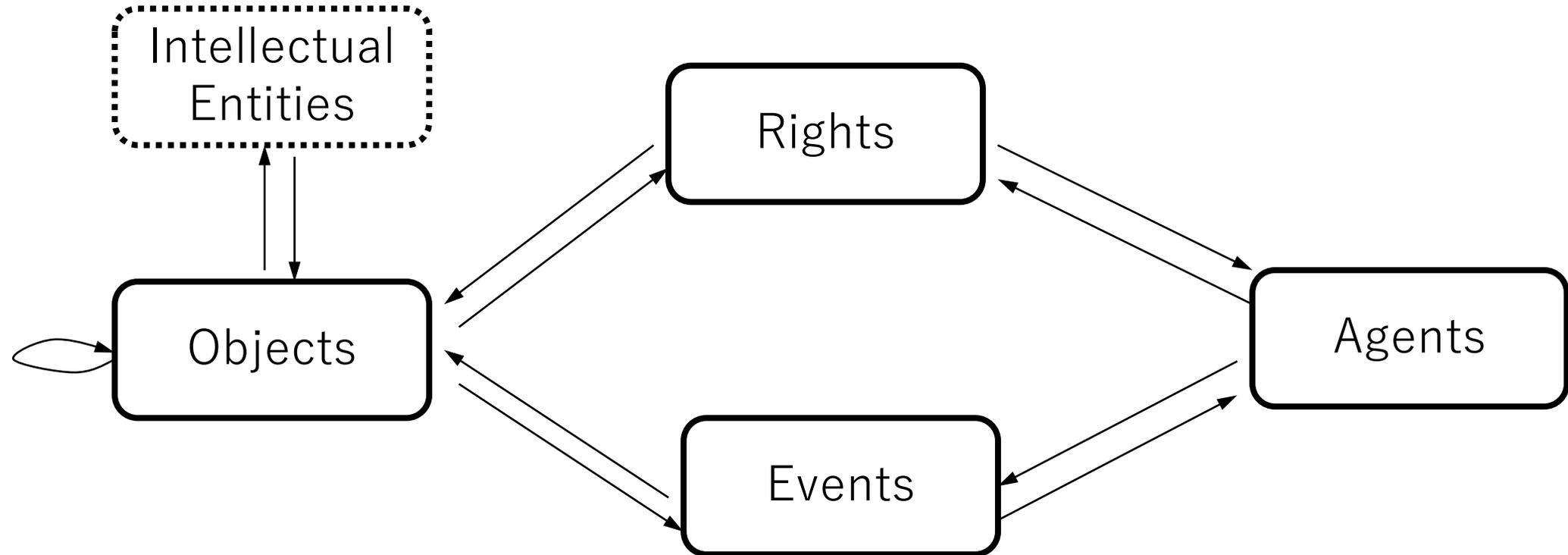
- 倉庫のイメージでとらえるとわかりやすい
 - PDIとして記述される情報は保存のためには重要
 - 時間経過やそれに伴う作業によってInformation Packageの中味は変更されることがある
- OAISの位置づけ：デジタル保存システムの参照モデルとしてとらえ、必要とされる機能を理解するために使えばよい
 - 多くのデジタルアーカイブは、コンテンツをリアルタイムに提供し続けているので、IPの出し入れとはフィットしない
 - ビットレベルでの高信頼性データ保存環境は必須である一方、保存システム（サービス）間で共有可能である。しかしながら、OAISの枠組みの中ではうまくとらえられない → 保存クラウドをどうとらえるか

PREMIS

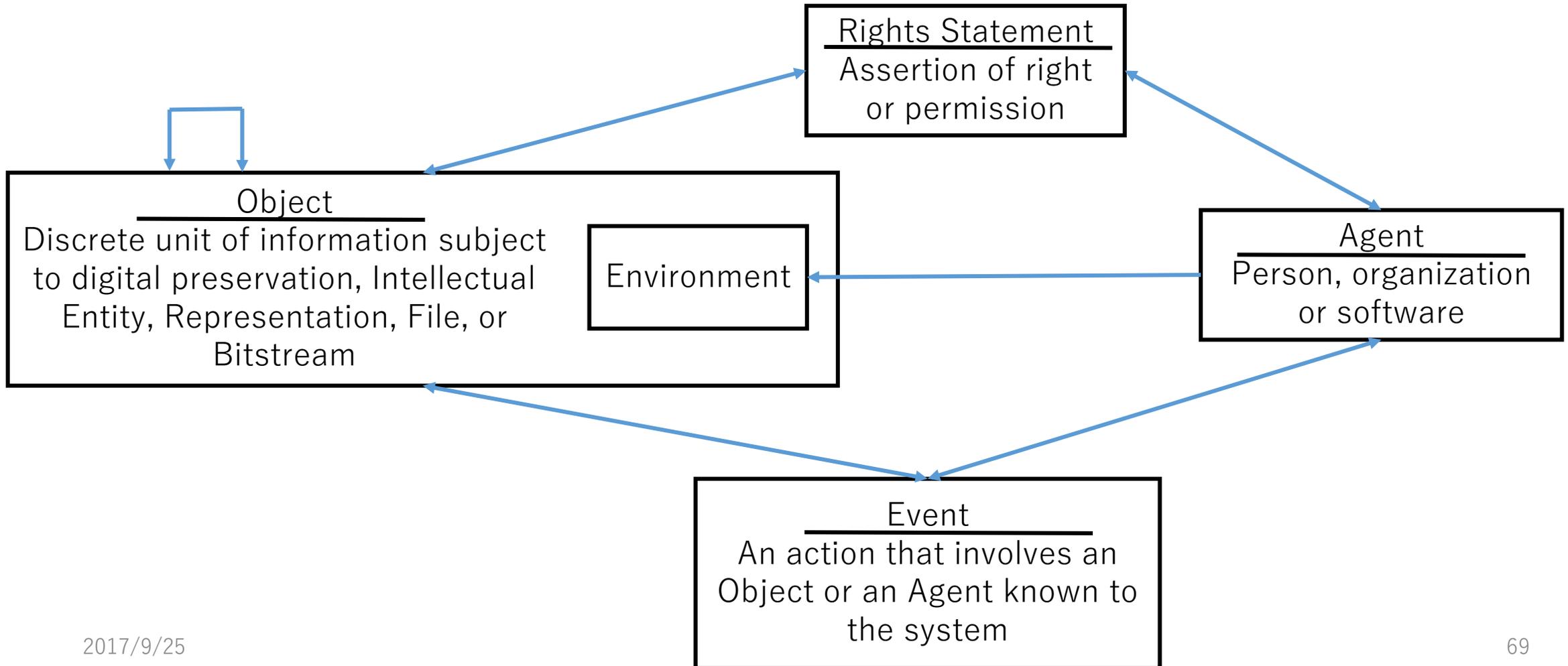
- Preservation Metadata: Implementation and Strategies
 - <https://www.loc.gov/standards/premis/>
- PREMISのデータモデル
 - PREMIS Data Dictionary 現在バージョン (version 3.0, 2015.6)
<https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>
 - 4種の実体 (Entity) からなる
 - (Digital) Object : デジタル保存の対象となる情報の単位
 - Environment: Digital Objectを支える何らかの技術 (ソフトウェア、ハードウェア)
 - Event : ObjectやAgentを含む、あるいは影響を及ぼす動作・作業
 - Agent : Event、Rights、Environment Objectに関連付けられ人・組織・ソフトウェア・システム。
 - Rights : ObjectあるいはAgentに関連する権利や許諾
 - 記述項目は実体毎にData Dictionaryに定義

PREMIS – Data Model (version 2)

- PREMISのデータモデル：どんな要素を記述対象としているか



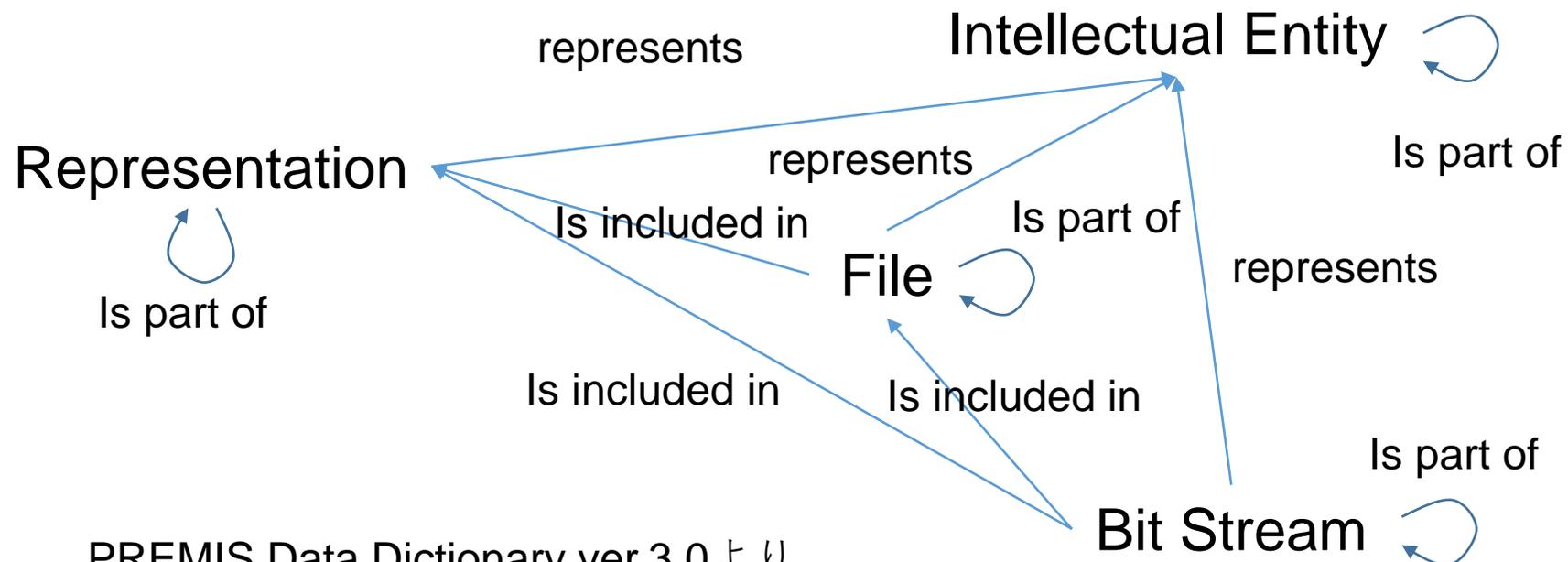
PREMIS – Data Model (version 3.0)



PREMIS – Data Model (version 3.0)

- Objectの4カテゴリ間の関係

- Intellectual Entity：デジタル保存対象に関係すると認められる知的内容実体
- Representation：Fileの集まりであり、再生のための構造メタデータを含む
- File：オペレーティングシステムが扱うことのできる名前を与えられたバイト列
- Bit Stream：保存の対象となるファイルに含まれるデータ



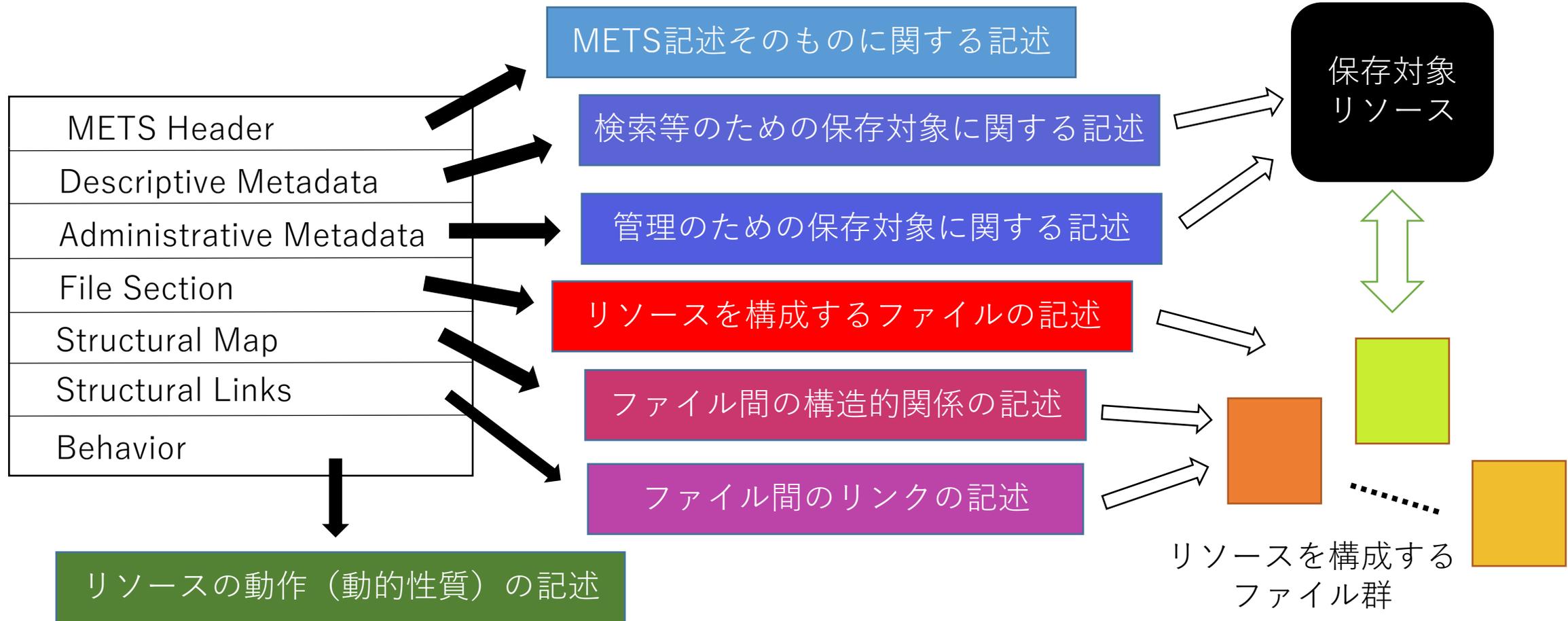
METS: Metadata Encoding and Transmission Standard

- METS home: <http://www.loc.gov/standards/mets/>
- Making of America IIでの経験をもとにDigital Library Federationで開発されたメタデータ標準
- デジタル化された図書のように構造を持つデジタルオブジェクトの記述
 - OAISのInformation Packageへの利用
“Depending on its use, a METS document could be used in the role of Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) within the Open Archival Information System (OAIS) Reference Model.” METS overview (<https://www.loc.gov/standards/mets/METSOverview.v2.html>) 参照

METS: Metadata Encoding and Transmission Standard

- METS記述(METS document)は7つのセクションからなる
 - METS Header：METS記述そのものに関する記述。たとえば、METS記述作成者
 - Descriptive Metadata：記述対象リソース（デジタルオブジェクト）に関する記述的メタデータ。検索等に用いる
 - Administrative Metadata：記述対象リソースに関する管理的メタデータ。リソースの作成・蓄積、知的財産権、デジタル化元リソースのメタデータ、来歴情報等
 - File Section：記述対象リソースを構成するすべてのファイルのリスト
 - Structural Map：記述対象リソースの階層的構造を表現する。リソースを構成する要素ファイルとそれらのメタデータを結ぶ
 - Structural Links：Structural Mapに示される要素ファイルを結ぶハイパーリンクの存在に関する記述
 - Behavior：リソースの動作（リソースの動作、動的性質等）の記述。個々の動作記述には、動作内容を（抽象的に）表現するinterface definition elementと、動作を実現するExecutable Codeを表すmechanism elementを持つ

METS: Metadata Encoding and Transmission Standard



4. まとめ

- デジタル保存は、デジタルコンテンツを利用可能な状態で保持すること
 - そのために、メディアの保存から技術の保存まで、種々の側面がある
 - 基本はコンテンツの維持管理
- デジタル保存に求められること
 - モノの保存から中味の保存へ視点の切り替えが必要
 - 保存対象、要求要件、方針によって保存方法は異なる
- デジタルコンテンツを支えるシステム環境の各階層で保存について考えねばならない
 - ビットレベルの保存から、利用方法の保存まで
 - メタデータはキー ⇒ メタデータの長期利用性も課題

おわりに

- 「何を残すことができれば保存できたことになるのか？」
 - 同じ対象であっても、保存機関によって答えは異なる
 - 国立公文書館での調査の中で出会ったことば「文書のエッセンスを残す」
 - ⇒ 文書の保存はモノの保存と同じではないことを意味する
 - 紙の文書をマクロフィルム化すると紙の機能を失うことになる
 - 電子文書を印刷して残すと電子文書の機能を失うことになる
 - 文書を電子化し、電子テキスト化すると、検索機能を加えることになる
- 「保存は維持管理」 ⇒ 「デジタル保存はデジタルコンテンツを使い続けられるようにするための維持管理」
 - 維持管理のための技術の開発
 - システム環境変化に伴い生じるリスク、保存計画への影響を見つけ保存に活かす技術
 - 維持管理に必要な共通基盤の開発

おわりに

- コンピュータの保存、技術の保存…ミュージアムの視点
 - Living Computer Museum (シアトル)
 - 動態保存されているマイクロコンピュータやミニコン
<http://www.livingcomputers.org/Discover/At-The-Museum/VintageComputers.aspx>
 - ミュージアムが関わるのは、文化財としてのコンピュータやソフト等
 - Digital Conservation(とでもいうべき技術)?
- 国レベルでの取り組みへの期待
 - デジタル保存には個別の機関・組織の取り組みが必要ではある一方、インフラとして共通化できる要素もある
 - 環境負荷に関する研究の必要性
 - 紙やフィルムの保存も環境負荷ゼロではない