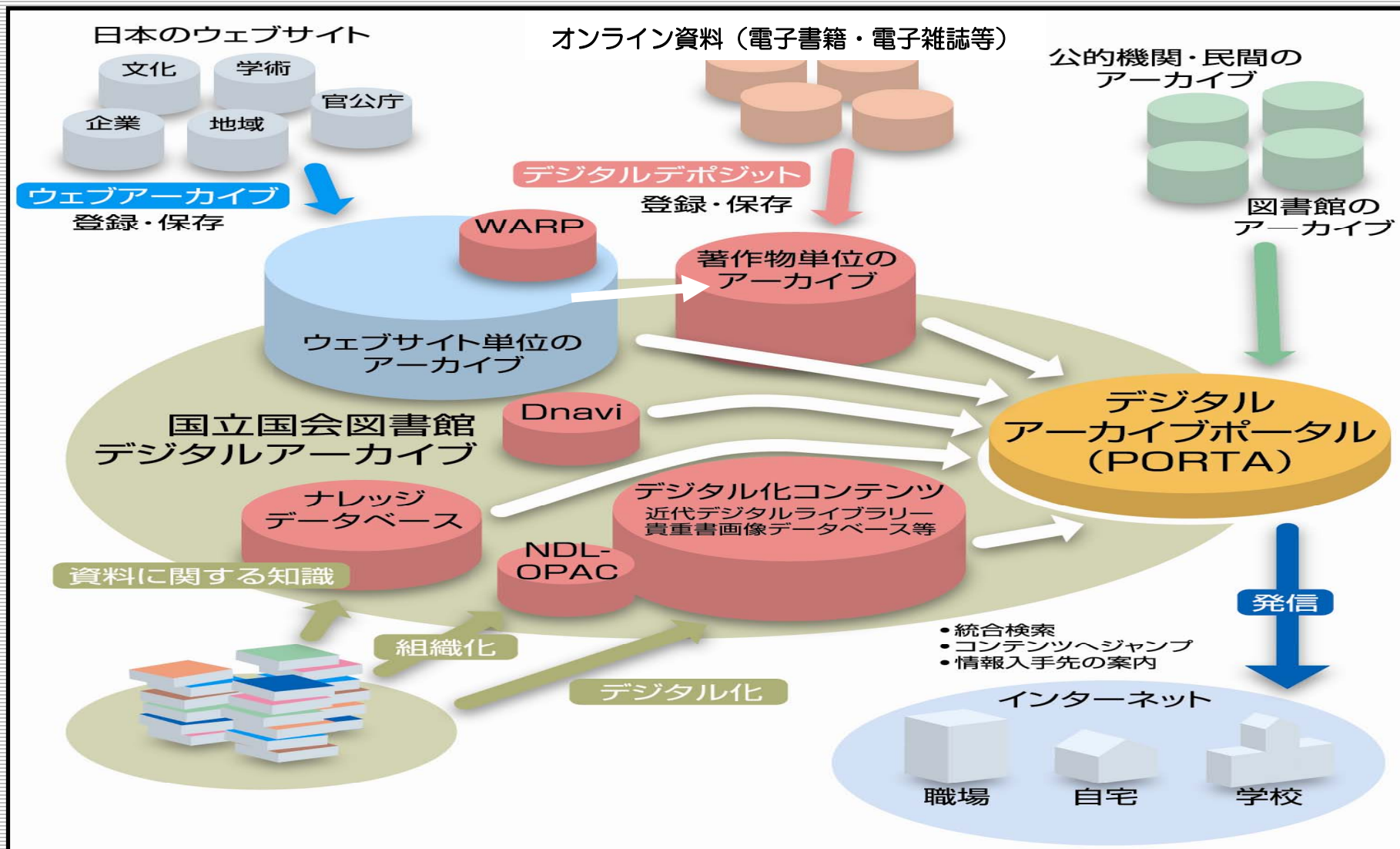


デジタル・アーカイブの標準化に向けて ～国立国会図書館の取組状況～

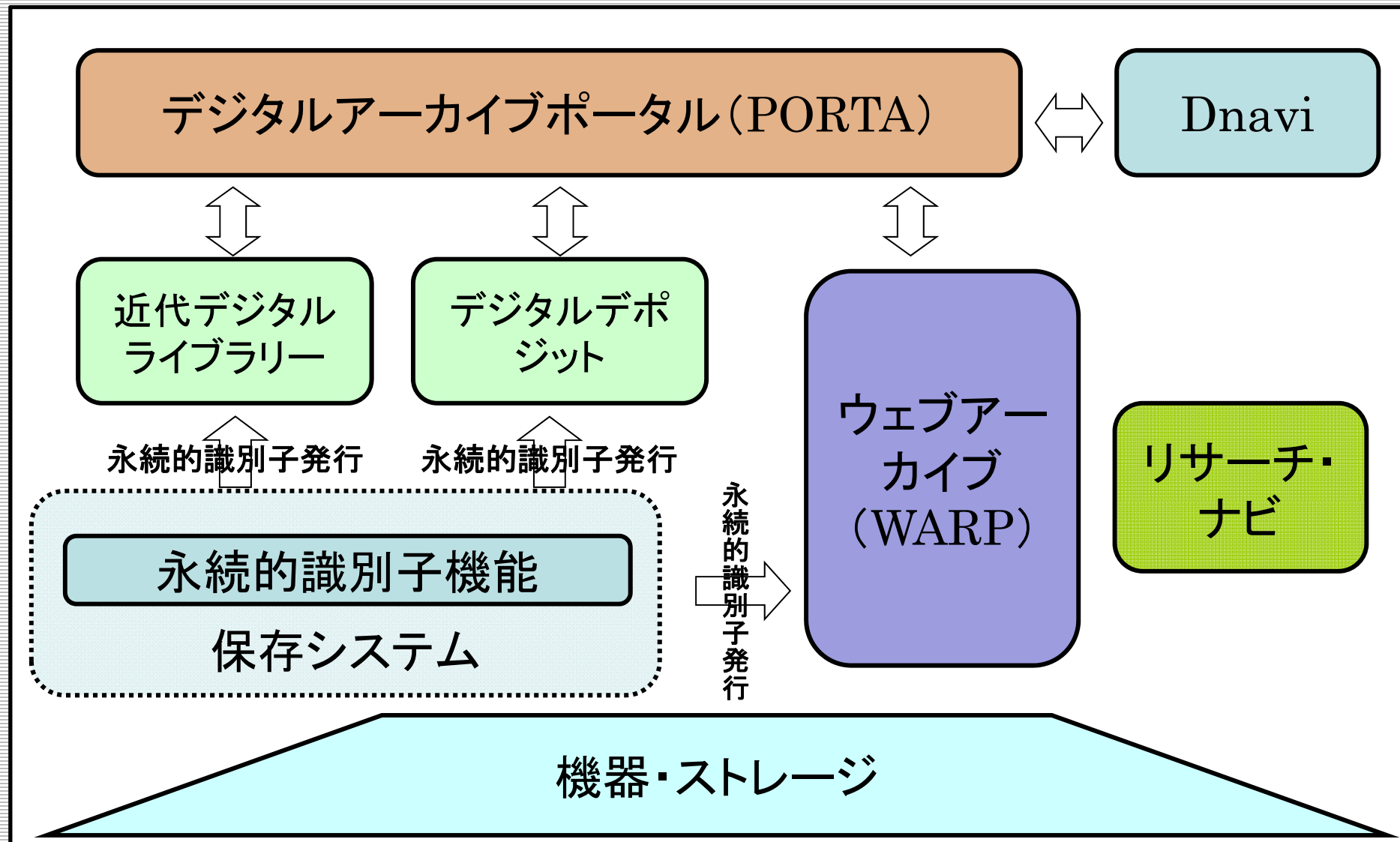
平成22年4月15日

国立国会図書館総務部企画課長 田中 久徳

国立国会図書館のデジタル・アーカイブ事業



デジタルアーカイブ (DA) システム



標準化に関する基本方針

技術的問題の解決に向けた関係機関との協力

(ア) デジタル・アーカイブを共有するための相互運用性の確保

デジタル・アーカイブの共有を目的として、国際的な共通仕様に留意し、当館と関係機関間での相互運用性を確保する。とりわけメタデータや長期保存・提供に関する国内及び国際的な技術標準を重視する。任意の機関が、さまざまなデジタル・アーカイブを用いて、その利用目的に応じて付加価値を高めたコンテンツ提供が可能となるように、手続きの共通化を図る。

(イ) メタデータ収集における共通仕様の確保

ポータル構築の基盤として、さまざまな情報提供機関からメタデータを収集し利用する共通の枠組みを確保し、実装・利用の推進を行う。

(出典) 「国立国会図書館電子図書館中期計画2004」

データの標準化① メタデータ

(1) 書誌データ(MARC)

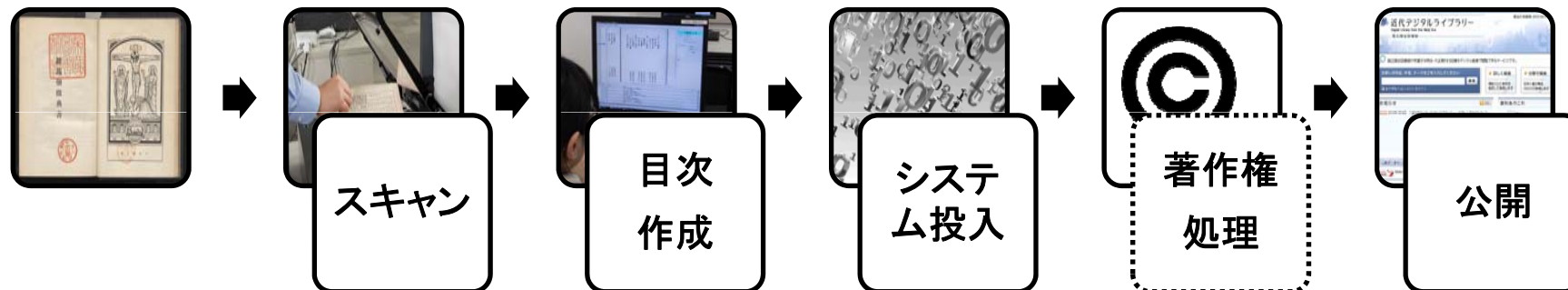
- 国際的提供や交換のためにフォーマット仕様を変更
JAPAN-MARC→MARC 21(2012年1月予定)
- 文字コードの変更 JIS→Unicode(2012年1月予定)

(2) DC-NDL

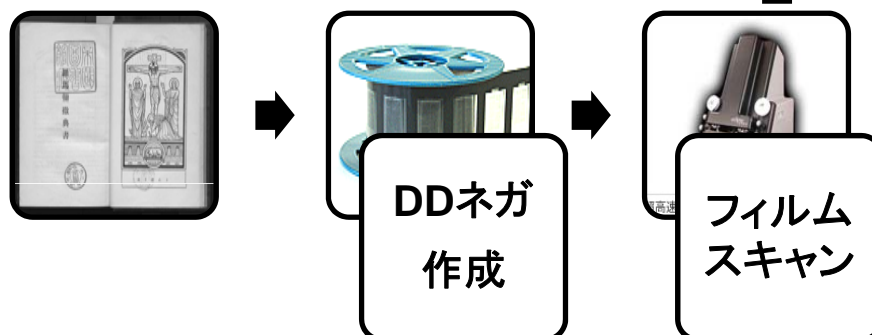
- 2001年3月:The Dublin Core Metadata Element Setを採用した「国立国会図書館メタデータ記述要素」(NDLメタデータ)を発表
- 2007年5月:改訂し「国立国会図書館ダブリンコアメタデータ記述要素」(DC-NDL)と改称
- 現在、改訂作業中

データの標準化② 資料デジタル化(作業フロー)

原資料から



資料を撮影したフィルムから



マイクロフィルム、大判・中判・35mmフィルム等に撮影したフィルムを使用

データの標準化② 資料デジタル化(主な仕様1)

「国立国会図書館資料デジタル化の手引き」に準拠。デジタル化の対象、提供方法等の要件に応じて選択可能なオプションを定めている。

(画像フォーマット)

- 保存用画像: JPEG2000
- 提供用画像: JPEG2000、JPEG
- サムネイル画像: JPEG

(解像度)

- 保存用・提供用画像: 300～400dpi
- 表紙・標題紙サムネイル画像: 256×256ピクセル
(プロパティ等により拡大・縮小して表示)
- 本文サムネイル画像: 100×100ピクセル以内
(自動作成)



データの標準化② 資料デジタル化(主な仕様2)

(色階調)

- ①白黒2値 ②8bitグレー ③24bitカラーのいずれか
(今後は②又は③)

(その他)

- 原本からのスキャンはオーバーヘッド方式スキャナ
(デジタルカメラも可、自動ページめくり機能は不可)
- 目次データをテキスト化
- メタデータ作成
- テキスト化については、調査・試行の段階

課題： 配信等のデジタルデータに係る権利対応

データの標準化② 全文テキスト化の課題

(OCR)

- 印刷状態の悪い資料、手書き資料、旧字体資料は認識精度が低い
- OCRのみでは視覚障害者等のニーズを満たす精度の実現は困難
- ベンダー独自の辞書フォーマット

(校正・構造化)

- 費用が高い
- 作成ルール(表、図、数式、外字等の扱い)

(フォーマット)

- 何にするか(プレーンテキスト、構造化etc)

データの標準化③ データの流通に関するプロトコル

(統合検索と横断検索)

➤ 統合検索関連プロトコル(メタデータ収集)

✓ OAI-PMH

✓ RSS

➤ 横断検索関連プロトコル

✓ SRW/SRU、Z39.50

✓ OpenSearch

✓ OpenURL

(新着情報配信)

➤ RSS

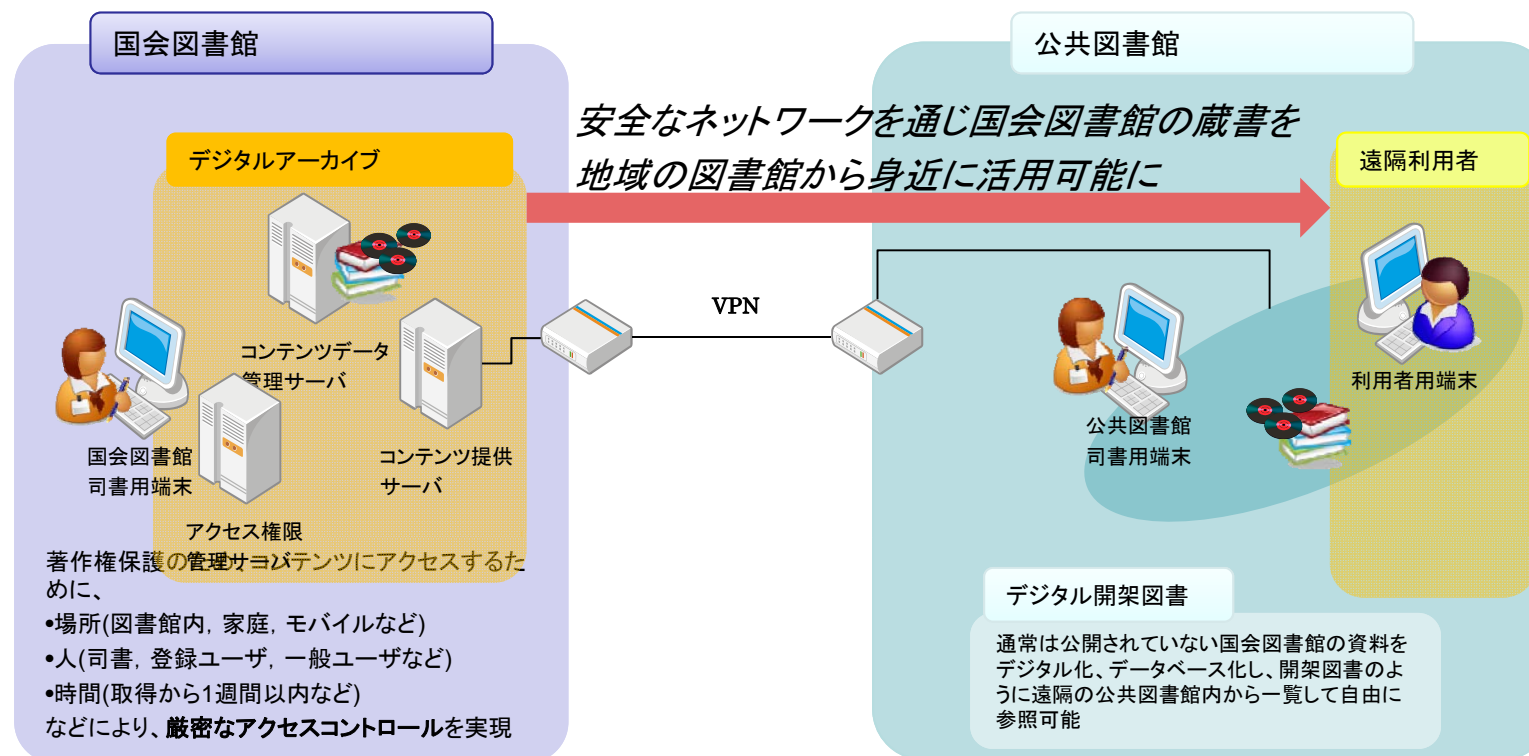
オンライン資料の収集に係るフォーマットの問題

- 電子書籍には多種多様なフォーマットが存在
 - AZW(Kindle用)、XMDF、ドットブック、PDF、DAISY、ePUB …
- 版面デザインの確保
 - 紙資料と同じレイアウトを保持する必要があるか
- 長期保存の観点からの要請
 - 内容面の保存
- アクセシビリティの確保
 - テキストデータと構造化の対応
- DRMの解除の必要、改変等の権利上の課題

(参考) 図書館内利用に限定した配信技術

○総務省ユビキタス特区実証事業(平成21年度)

○地域の公共図書館において、ネットワーク技術・セキュリティ技術などを活用し、国会図書館のデジタルアーカイブのデータを、著作権を保護しつつネットワークを通じて閲覧できるサービス



(参考) 閲覧期間の限定による貸出技術

国会図書館の蔵書を電子化して利用者の認証と閲覧期間の制限を施したオンライン貸出しサービスの可能性を実験する。

