

# NDLデジタルアーカイブポータル(PORTA)の 現状と将来



平成21年1月29日  
国立国会図書館関西館  
電子図書館課  
柴田 昌樹

# 1. PORTAの位置づけ

## 電子図書館サービスの目標

(国立国会図書館電子図書館中期計画2004から)

- 国のデジタルアーカイブの重要拠点となる
- 日本のデジタル情報全体へのナビゲーション総合サイトを構築する

### ① デジタル・アーカイブの構築

- NDL所蔵の資料の電子化の推進
  - 図書等のデジタル化
- インターネット情報資源の収集と保存
  - Webアーカイブとデジタル・デポジット
  - 長期保存対策

### ② 情報資源に関する情報の充実

- レファレンス情報、解題情報、統制語辞書類

### ③ デジタル・アーカイブのポータル機能

- デジタルアーカイブの統合検索、ワンストップナビゲーション

アクションプラン



# 「知識はわれらを豊かにする」 をデジタルアーカイブの観点からみて

国立国会図書館60周年を迎えるに当たってのビジョン(長尾ビジョン)

(2) 日本の知的活動の所産を網羅的に収集し、国民の共有資源として保存します

収集・保存すること

(3) 利用者が求める情報への迅速で的確なアクセスまたは案内できるようにします

(4) 利用者がどこにいても、来館者と同様のサービスが受けられるように努めます

提供可能にすること

情報資源

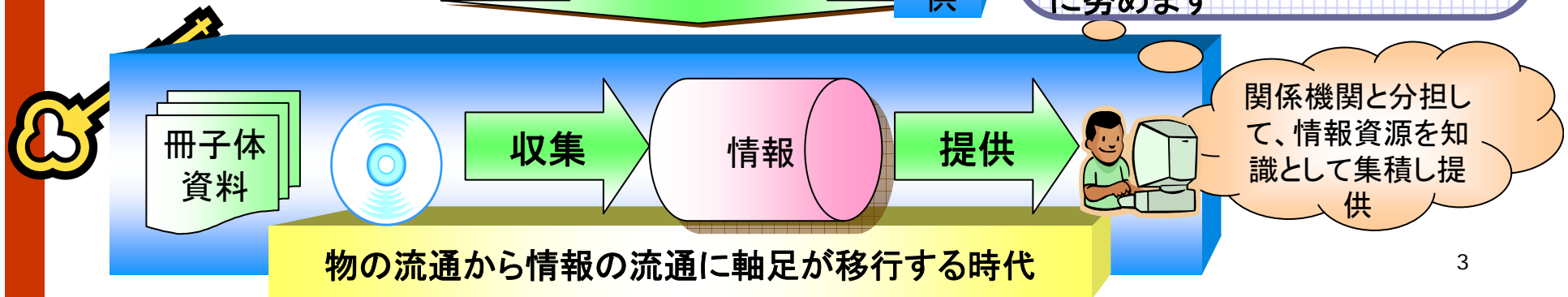
関係機関と協力して収集・提供

(1) 国会に対するサービスをより高度なものとし、立法補佐機能をさらに強化します

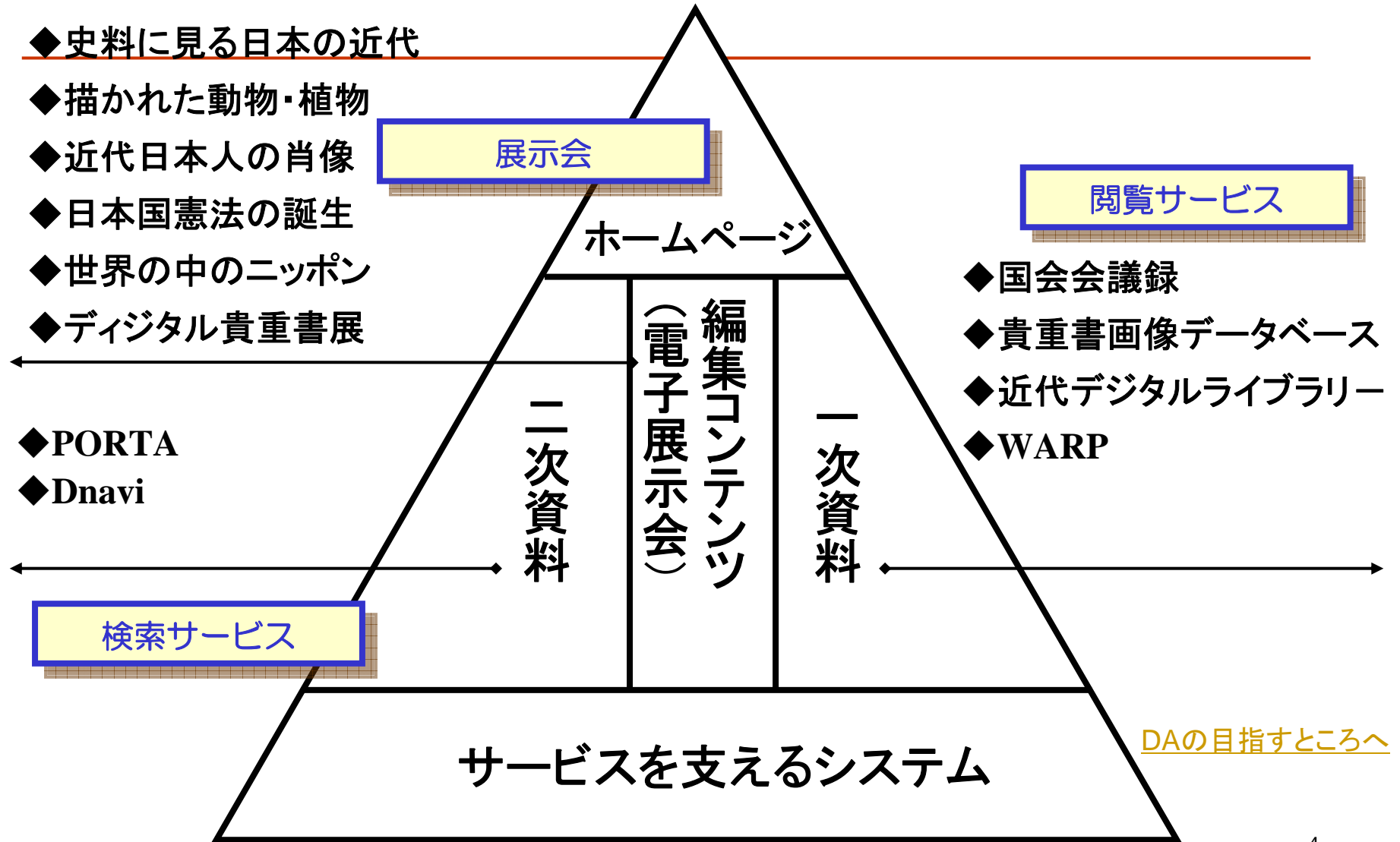
(5) 社会に多様で魅力的なサービスを提供し、国立国会図書館の認知度を高めます

(6) 公共図書館をはじめとする国内の各図書館とより密接な連携・協力を進めます

(7) 海外の図書館との密接な連携を行い、情報の共有・交換に努めます



# 現状の電子図書館サービス



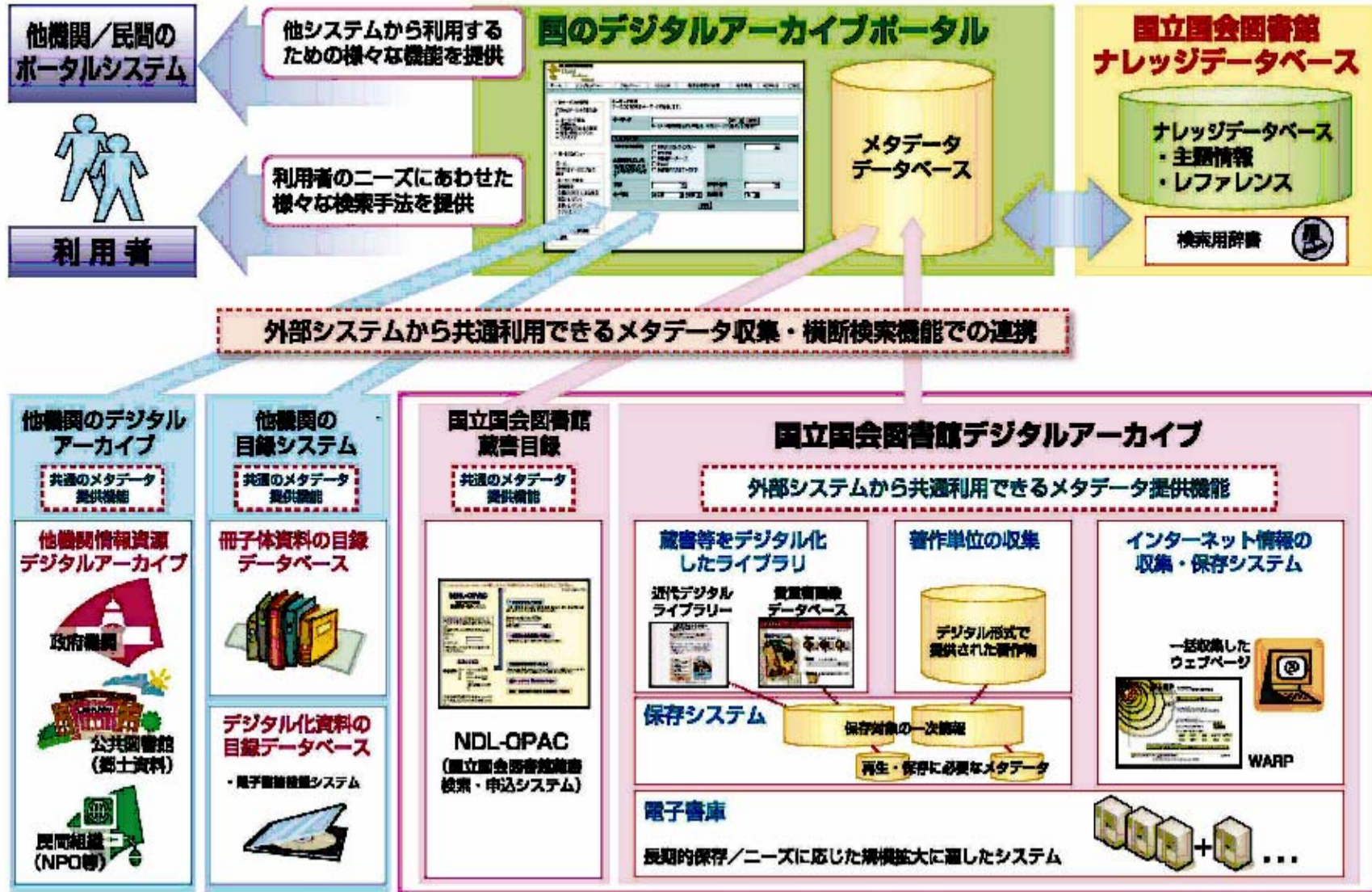
# NDLのデジタルアーカイブ事業 の目指すところ

---

- 国としてのデジタルアーカイブの構築を目指して
  - 国としてのデジタルアーカイブの構築と提供を目指す。
  - 世界規模のデジタルアーカイブの構築に向けて、日本の実施主体としての一翼を担う。
- 各組織でデジタルコンテンツを作成、提供
  - デジタル化した紙資料、ボーンデジタルをウェブで公開
- 分担して、デジタルアーカイブで保存
  - 後世に残すために、デジタルアーカイブを構築
  - 各機関のアーカイブの集合で国のデジタルアーカイブを構成
    - NDLだけでなく、分野、地域、業種毎に
- **どこに保存されていても一元検索と提供**
  - 各デジタルアーカイブを、統合利用できる環境の整備
  - 各分野、地域等毎のポータル構築
  - 各ポータルをあわせて、国の**デジタルアーカイブポータル**を提供



# NDL電子図書館サービスの全体像



# PORTAの構築コンセプトと狙い

## コンセプト (2004年)

当館が保有するデジタルコンテンツに限らず、広く国のデジタル情報全体を、利用者が所蔵場所を意識せず、ワンストップで適切に利用できる総合的なポータルサイトを構築する。

## 目的

利用者の問題解決ニーズに答えるために、所蔵機関の壁を越えて、異なる種別の情報と情報の関連を辿れるようにし、利用者が必要とする情報そのものを的確に入手できるようにする。(単に所在を示すのではなく、可能な限り必要な情報を利用できるように)  
(統合検索は第一歩として、埋もれた情報を可視化するもの。)

## プロトタイプ (2005年~)

・今後提供するサービスイメージを提示し、ポータルの実現に向けて、システムの有用性及び適用する技術の妥当性を検証することを目的とする。

## 意義

- データ提供者にとって
- ・PORTAでの検索により、今までに訪問しなかった人が、各データベース内にナビゲートされる。
- ・訪問してきた人が他のコンテンツへもナビゲートできるようにすることにより、データ提供者サイト全体の利用が増える。
- ・また、実物を閲覧するための来館利用のきっかけにもなる。
- 利用者にとって
- ・Google等の検索では見つけられなかったコンテンツを見つけ出すことができるようになる
- ・DBを手当たり次第に検索する必要がなくなる

## 本格システム (2007年~)

- プロトタイプでの意見を反映
- 実運用規模を想定
- ・大量アクセス、大量データ、大量ユーザに対応
- ・拡張容易性、障害時運用継続性、環境変更容易性、及び、直感的操作性を確保する。
- ・可能な限り、先進技術の適用を目指す。

## 統合検索対象の拡大策

- ・統合検索の普及に不可欠なサイトは、個別対応をしてでも統合検索できるようにする  
(標準インタフェースを実装していなくても)
- ・統合検索の利用が促進されると標準インタフェースを実装するサイトが増えてくる
- ・⇒ある程度普及すれば、加速度的に統合検索先が増える

## 2. PORTAの現状

### 構想からプロトタイプ、正式提供までの経緯

年月	イベント	概要
2004年2月	電子図書館中期計画2004策定	ポータルは、3つの柱のうちの1つ。
2004年10月	プロトタイプ開発開始	複数台のPCに、SOA指向で機能分散した形で設計・構築。各種標準プロトコルを実装。 OSSのみで構築。 NII高野先生にアドバイスを受ける。
2005年7月	プロトタイプ試験公開	サービスの有用性、適用技術の妥当性を検証。
2005年10月	正式版設計・開発開始	実運用規模を想定。 ・大量アクセス、大量データ、大量ユーザ対応 ・拡張容易性、障害時運用継続性、環境変更容易性 ・直感的操作性 ・可能な限り、先進技術の適用を目指す。
2007年10月	正式版公開	バグ対応のため公開が遅れる。 依然として、レスポンスが遅く、性能向上、バグ対応中...
2008年10月	機能強化開始	・検索機能の改善 ・データプロバイダ登録機能の改善
2009年6月頃	DAシステムインフラで稼動開始	NDLデジタルアーカイブシステムとして共通のシステム基盤で提供



# データプロバイダ追加の経緯(1)

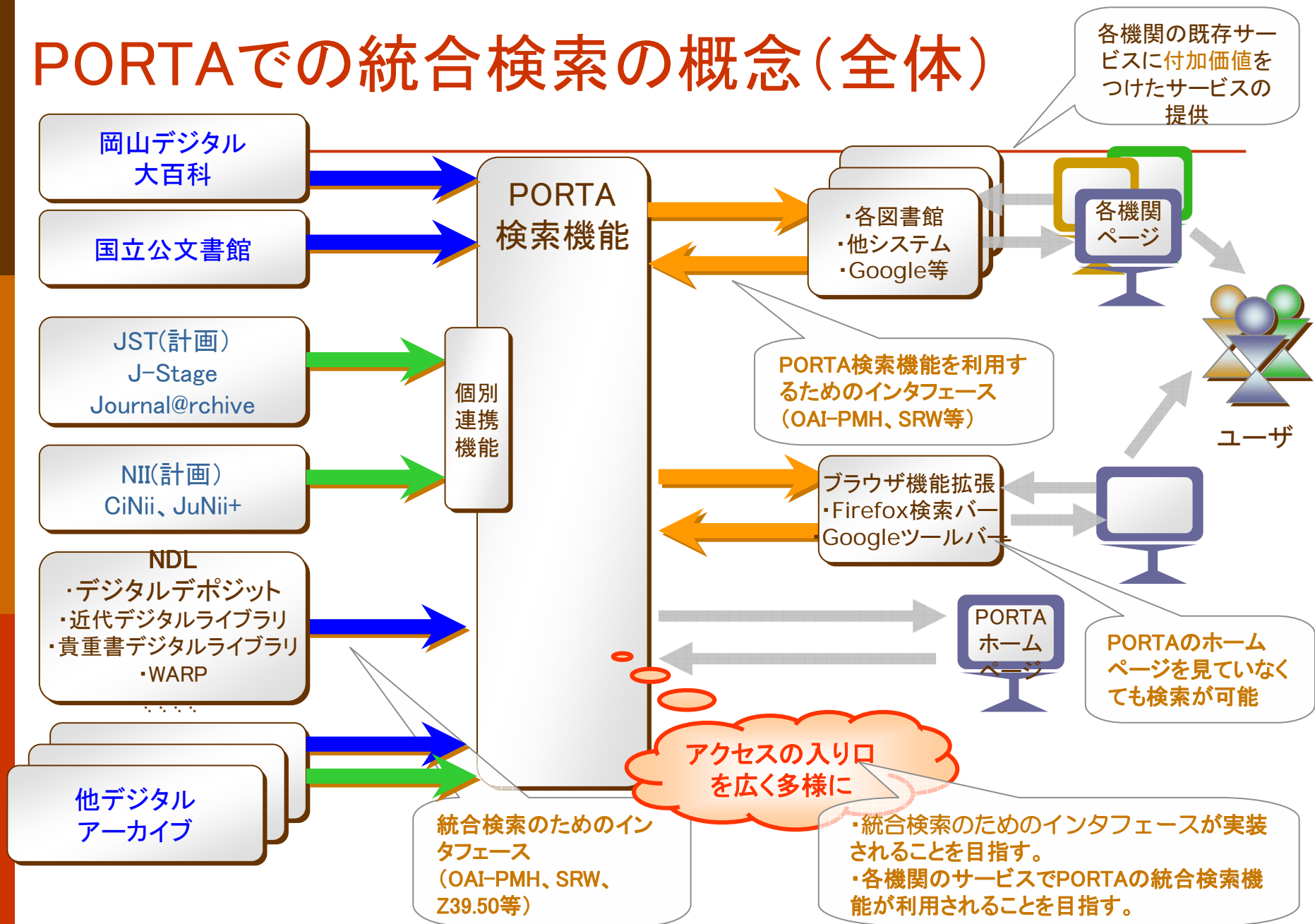
公開年月	外部データベース		NDLデータベース	
2005年7月	青空文庫	ファイル転送	NDL近代デジタルライブラリー	OAI-PMH
	府省所管のアーカイブサイト情報	OAI-PMH	Dnavi	OAI-PMH
2005年12月	デジタル岡山大百科	Z39.50	貴重書画像データベース	OAI-PMH
	新書マップ(NPO法人連想出版)	RSS1.0	NDL蔵書目録(和図書・和雑誌)	ファイル転送
			NDLプランゲ文庫	OAI-PMH
2006年2月			レファレンス協同データベース	RSS2.0
2006年4月	国立公文書館デジタルアーカイブ・システム	SRU/SOAP	カレントアウェアネス	RSS2.0
			NDL雑誌記事索引	ファイル転送
2006年10月	一橋デジタルアーカイブス	OAI-PMH		
2007年3月	アジア歴史資料センター	SRU/SOAP		
2007年10月	秋田県立図書館(アーカイブ・記事・索引)	ファイル転送	児童書デジタルライブラリ・児童書総合目録 WARP	スクレイピング
2008年3月	日本ペンクラブ電子文藝館	ファイル転送		
	農林水産関係試験研究機関総合目録	OpenSearch		
	一橋大学附属図書館(HERMES-IR)	OAI-PMH		
	京都大学学術情報リポジトリ+貴重資料画像	OAI-PMH		

## データプロバイダ追加の経緯(2)

公開年月	外部データベース		NDLデータベース	
2008年10月	奈良女子大学附属図書館所蔵資料画像	メタデータ作成		
2008年11月	茨城県立図書館デジタルライブラリ	メタデータ作成	電子展示会	メタデータ作成
	神戸大学附属図書館 新聞記事文庫	SRU/SOAP		
2008年12月	hon.jp	Webサービス		
	九州大学デジタルアーカイブ	ファイル転送		
2009年1月	国立美術館(国立美術館4館の所蔵作品)	OAI-PMH		
	奈良県立図書情報館	ファイル転送		
	ACADEMIC RESOURCE GUIDE(ARG)	RSS2.0		
	実践女子大学Directory of Open Access Journals in Japan(DOAJJ)	ファイル転送		
準備中	筑波大学附属図書館、早稲田大学古典籍、東京大学大学院情報環、インプレス、日経Biz、科学技術振興機構、国立情報学研究所、人間文化研究機構			

最新状況は「[検索対象アーカイブ一覧](#)」

# PORTAでの統合検索の概念(全体)



# PORTA検索機能の詳細

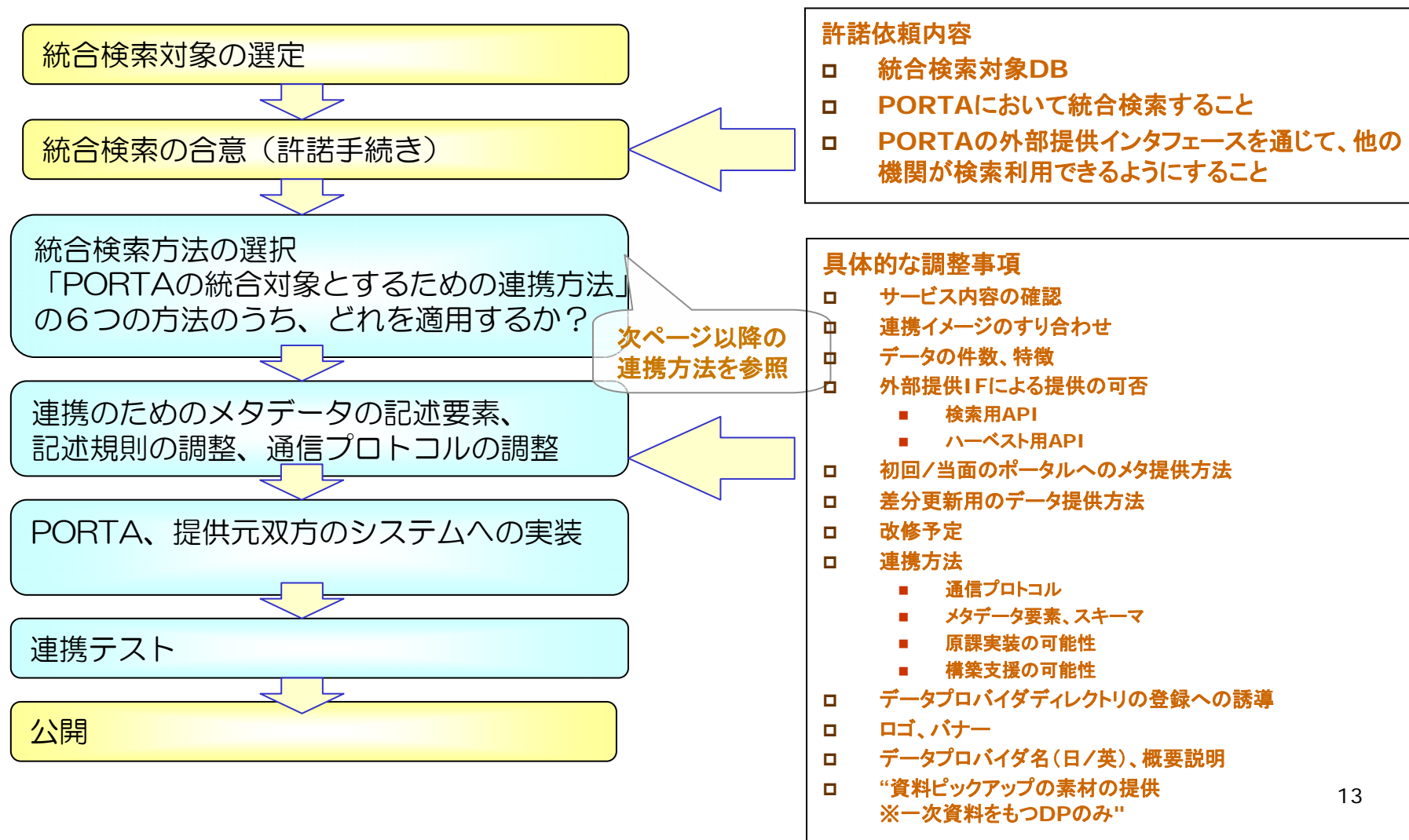
## 検索の種類

- 簡易検索、詳細検索
- 分類検索 — NDC/NDLC分類ごとに表示 (NDC推定も行う)
- 連想検索 — 入力した文章から連想される資料を検索  
(連想検索エンジンGETA:NI高野教授作成)

## 検索補助

- 辞書検索 — 辞書データを検索 (NDLSHによる統制語等)
- 辞書による検索支援 — 辞書データを使って、関連する資料を検索
- 関連情報リンク — 関連するサイト、サービスを検索  
(Amazon,BK1等非公的機関も含む)

# 統合検索の実現までのプロセス



# PORTAの検索対象とするための連携方法

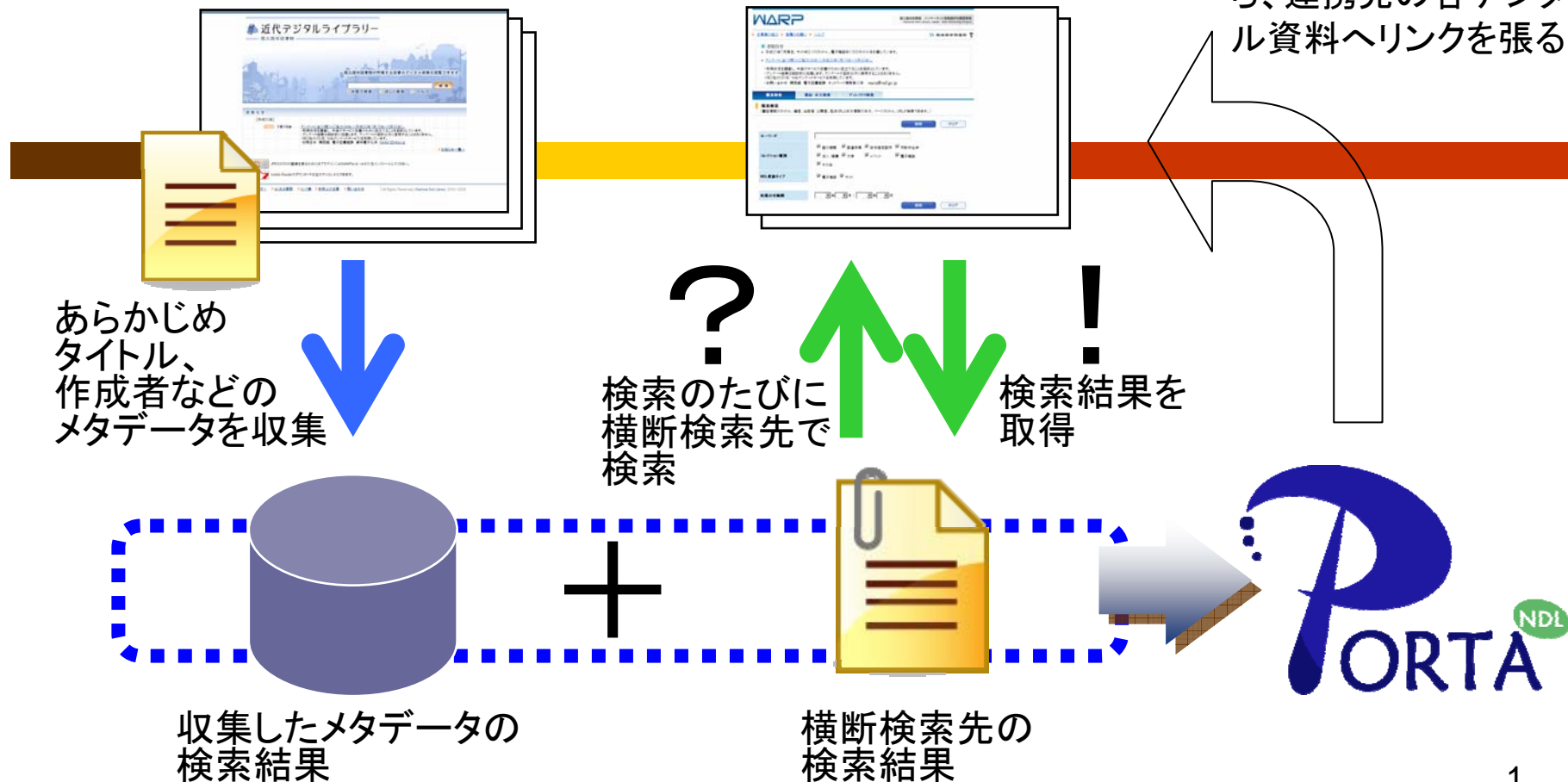
(PORTAホームページから)

大きく分けて、以下の(A)(B)いずれかでの連携になります。

## (A)メタデータの収集

## (B)横断検索

PORTAの検索結果から、連携先の各デジタル資料へリンクを張る





## 具体的な方法

具体的な方法としては、以下の手段を用意しています。

区分	具体的方法	説明 (P4以降)
(A)メタデータの収集	メタデータをOAI-PMHで配信	①
	メタデータをRSSで配信	①
	メタデータファイルをFTPで転送	②
	メタデータファイルをメール等で送付	③
	メタデータを国会図書館で作成	④
(B)横断検索	SRW	⑤
	Z39.50	⑤
	OpenSearch	⑤
	スクレイピング	⑥

## 必要なデータ

個々のデジタル資料ごとに、最低限「タイトル」「リンク先URL」が必要ですが、

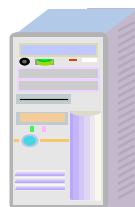
- ・他に「著者(作成者)」「出版者(公開者)」「出版年(公開年)」「ISBN」「ISSN」「分類(NDC/NDLC)」「Web公開の有無」「有償/無償」等のデータもあれば、より詳しい検索・表示ができます。
- ・前ページ④⑥の方法の場合は、連携先システムの画面から、国会図書館側でデータを切り出します。それ以外の場合は、連携先機関でデータを用意していただく形になります。
- ・Dublin Coreに準拠したメタデータがあれば、基本的にそのまま使用可能です。
- ・PORTA内部ではDublin Coreを拡張した「dcndl\_porta」フォーマットを使用しています。[\(共通仕様及び連携に関するガイドライン 2.1参照\)](#)  
それに準拠したメタデータを用意していただければ、完全にそのまま使用できます。
- ・「リンク先URL」は、できれば個々のデジタル資料に直接遷移できるURLが望ましいですが、無理ならば上位の目次ページや、トップページのURLでも構いません。

## 個々の方法についての説明

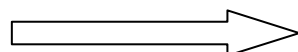
### ①メタデータ配信のインタフェースを実装 (OAI-PMH、RSS)

連携先のシステムに、OAI-PMHまたはRSSのインタフェースを実装していただく必要があります。

PORTAからはそのインタフェースを使って、定期的にメタデータを収集します。



OAI-PMH  
またはRSS



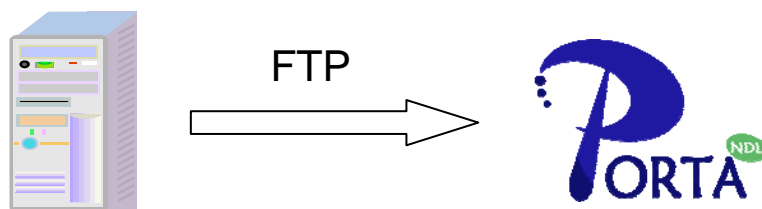
連携先システム

- ・ OAI-PMHで連携する場合の詳細な仕様は[共通仕様及び連携に関するガイドライン](#)3.1を、RSSは3.3をご参照ください。
- ・ RSSよりもOAI-PMHの方が、より詳細なメタデータを記述することができます。

## ②メタデータファイルをFTPで転送

メタデータをTSV、XML等の形式で、FTPによってPORTAシステムに転送していただきます。

FTPの向きは連携先からputする形、PORTAからgetする形、どちらも可能です。



連携先システム

・転送頻度、件数に制限はありません。

### ③メタデータファイルをメール等で送付

メタデータをExcel、TSV、XML等の形式で、国会図書館宛てにメール、媒体郵送等で送付していただきます。

国会図書館では受け取ったファイルを、PORTAに投入します。



・データの追加・更新が少ない場合に限りです。（3ヶ月に1回程度）

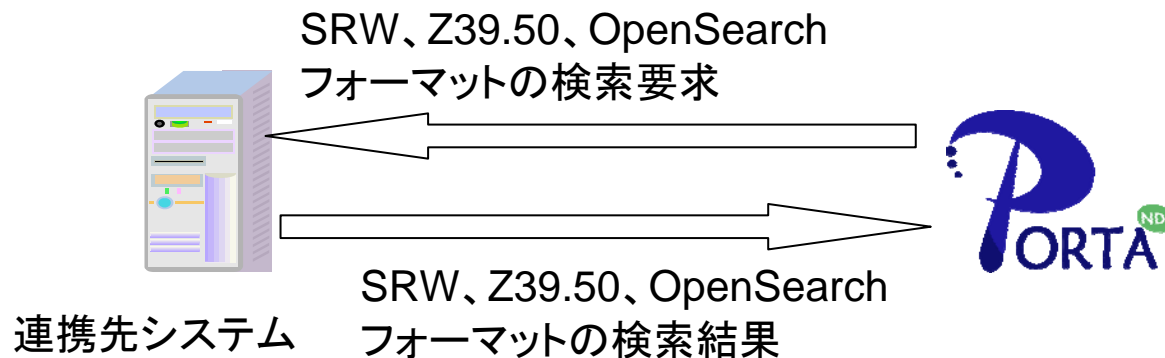




## ⑤横断検索のインタフェースを実装 (SRW、Z39.50、OpenSearch)

連携先のシステムに、SRWまたはZ39.50またはOpenSearchのインタフェースを実装していただく必要があります。

PORTAからはそのインタフェースを使って、利用者からの検索の度に検索要求を投げ、結果を受け取ります。



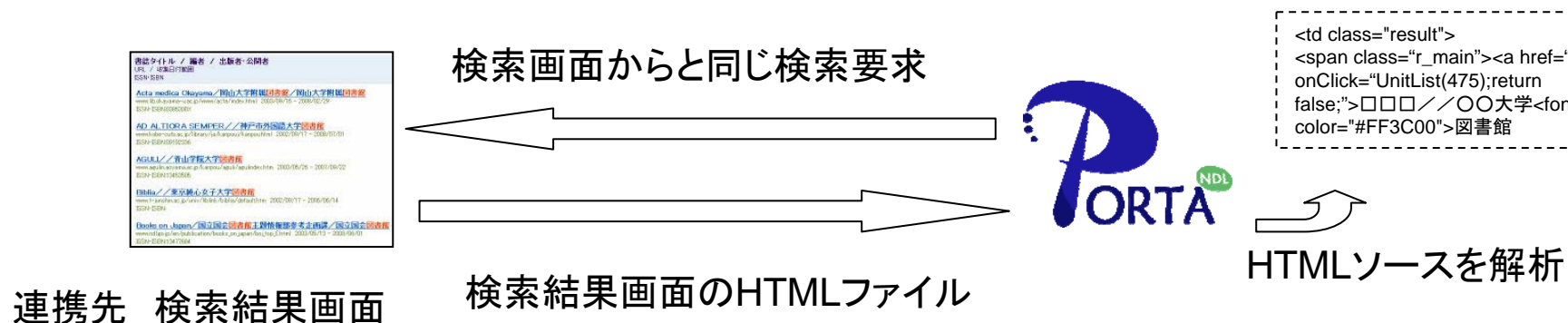
- ・ SRWで連携する場合の詳細な仕様は[共通仕様及び連携に関するガイドライン3.2](#)を、Z39.50は3.4を、OpenSearchは3.5をご参照ください。

## ⑥連携先システムの画面をスクレイピング

利用者の検索の度に、『連携先の検索画面から検索した場合と同様の要求』をPORTAから投げます。

連携先の検索結果画面を解析（スクレイピング）し、タイトル、URL等を取り出して、利用者に表示します。

連携先の機関では、特に作業をしていただく必要はありません。

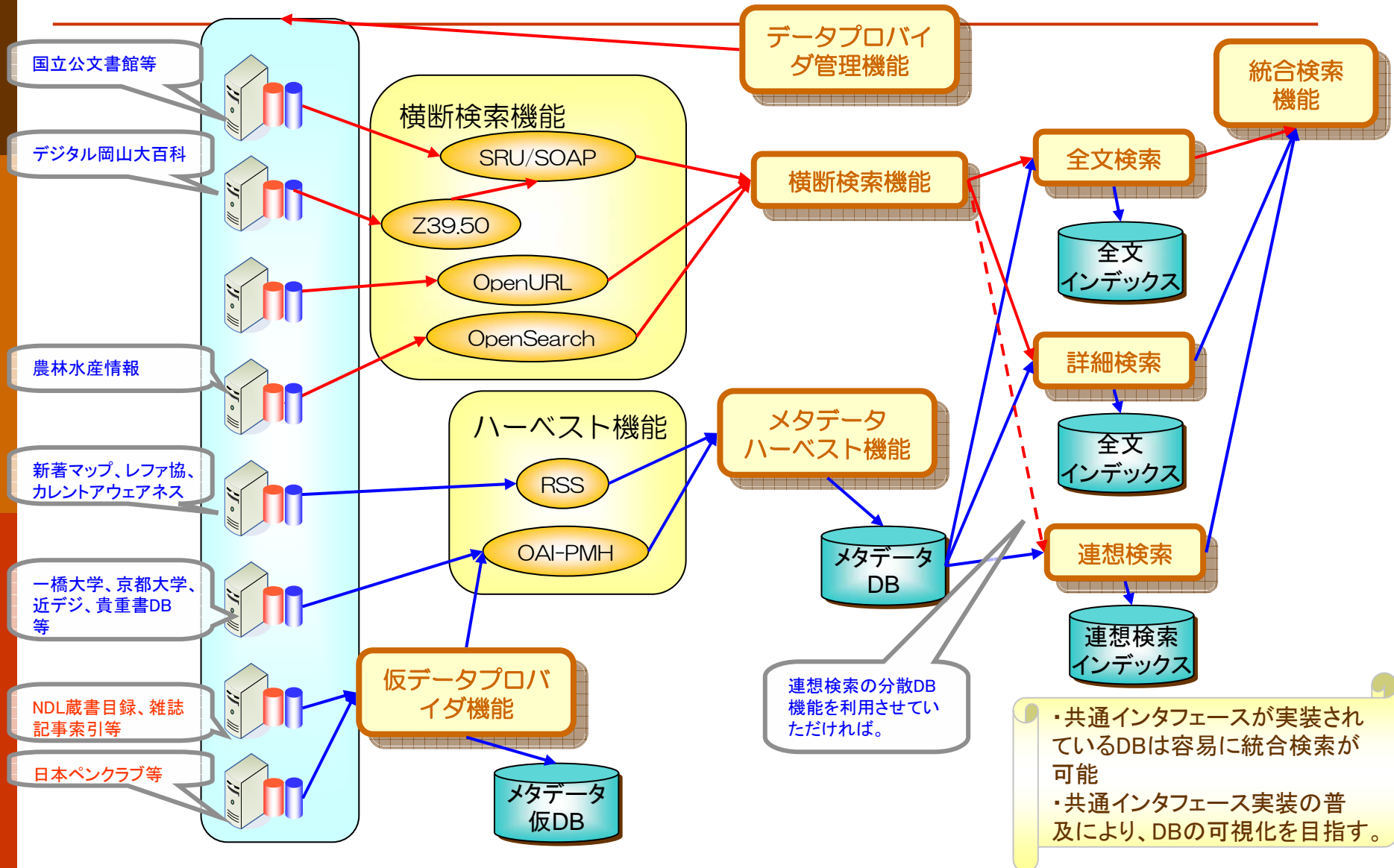


・連携先での新規実装は不要なものの、検索結果画面のフォーマットが一行変わっただけでも解析できなくなることで、画面から取得できる範囲しかメタデータを取れないことから、他に方法がない場合の最後の手段とさせていただきます。

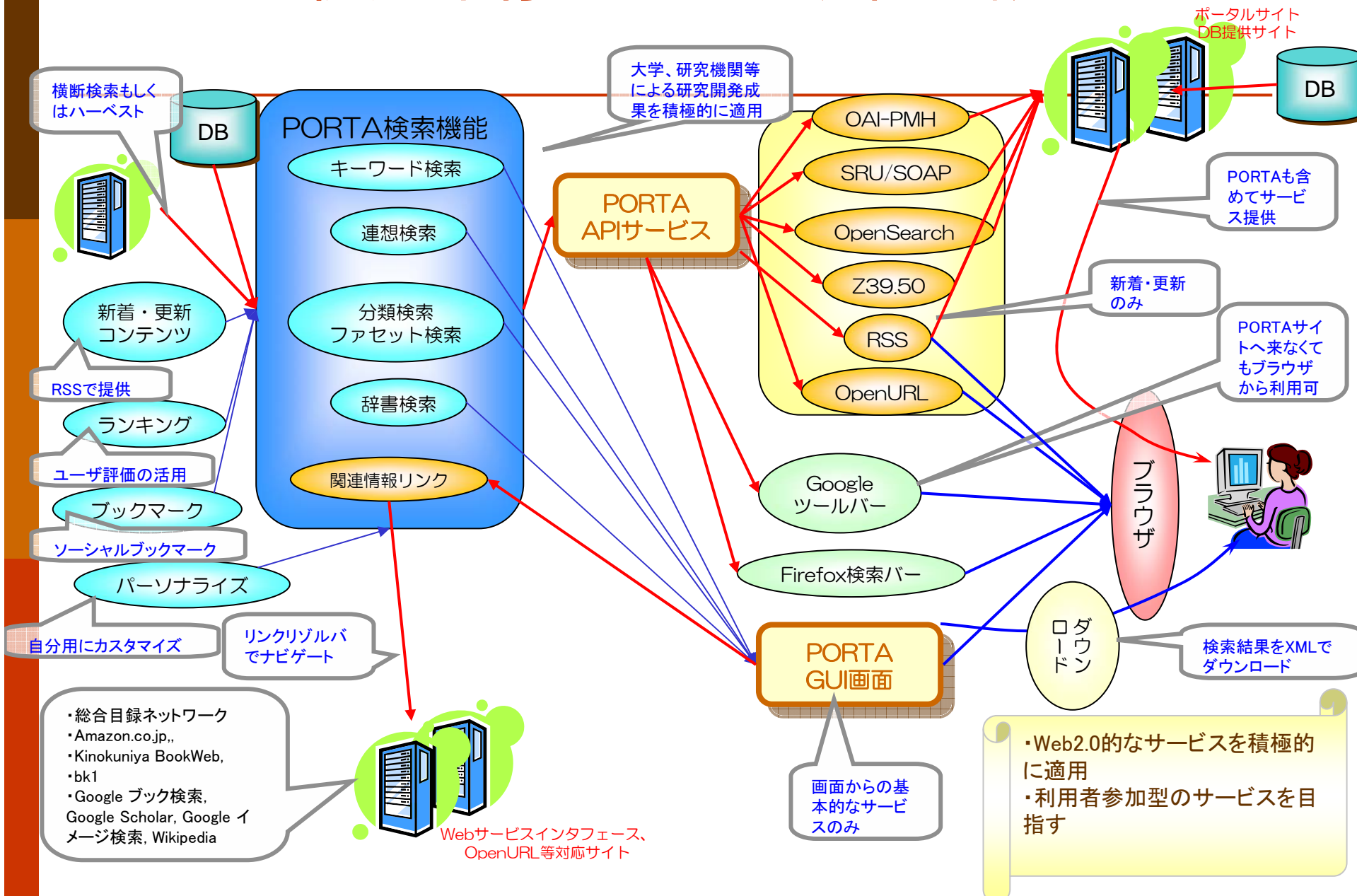
・この方法で連携後、連携先の検索結果画面や検索要求のフォーマットに変更があった場合は、国会図書館担当者にご連絡ください。

・Flash、Ajax等、単純なHTMLではない検索結果画面は解析できません。

# メタデータ収集・横断検索の概念



# 検索・閲覧サービス提供の概念



# PORTA連携のための共通仕様

		メタデータ形式
コ ン テ ン ツ	DC系	DC-NDL
		Junii, Junii2
		RSS2.0にDC-NDL要素を拡張
		oai_dc
	RSS系	RSS1.0形式
		RSS1.0にDublin Coreモジュールを追加
		RSS2.0形式、Podcast形式
		OpenSearch形式
		Atom0.3形式
	MODS系	NDL-DAメタデータスキーマ(METS,MODS準拠)
	MARC系	JAPAN/MARC(M)、JAPAN/MARC(S)
	その他	SRW、Z39.50等に準拠した仕様
		標準仕様での提供が困難な場合は、データプロバイダ独自仕様も考慮

## 通信プロトコル

メタデータ記述要素

メタデータ記述規則

通信プロトコル	
ハーベスト	OAI-PMH
	RSS
横断検索	SRW(SRU/SOAP)
	Z39.50
	OpenSearch
リンクリゾルバ	OpenURL
サービス連携	WebサービスAPIによるシステム連携
	UDDI

### 一方針

- 市場で広く普及している、もしくは、普及が見込まれる複数の仕様を可能な限り受け入れる。
- (最低限のメタデータであっても受け入れて利用する)

# APIの公開

## 【検索用API】

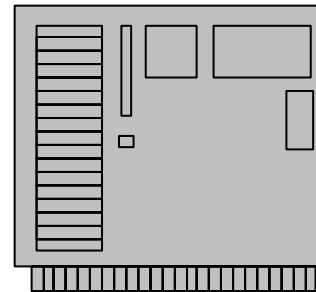
(平成20年3月末  
公開)



## 【ハーベスト用API】

(平成20年12月  
公開)

SRW  
OpenSearch  
OpenURL  
Z39.50



## 【検索用APIの利用】

- ・ 広島市立図書館「広島文学資料」
- ・ 東京大学OPAC Plus “言選Web”
- ・ 明治大学オンラインナレッジ検索 等

付加価値をつけたサービスの提供



## 【ハーベスト用APIの利用】

- ・ WebcatPlusと  
近代デジタルライブラリーの連携
- ・ 想—IMAGINE Book Search
- ・ hon.jp(電子書籍検索サイト) 等



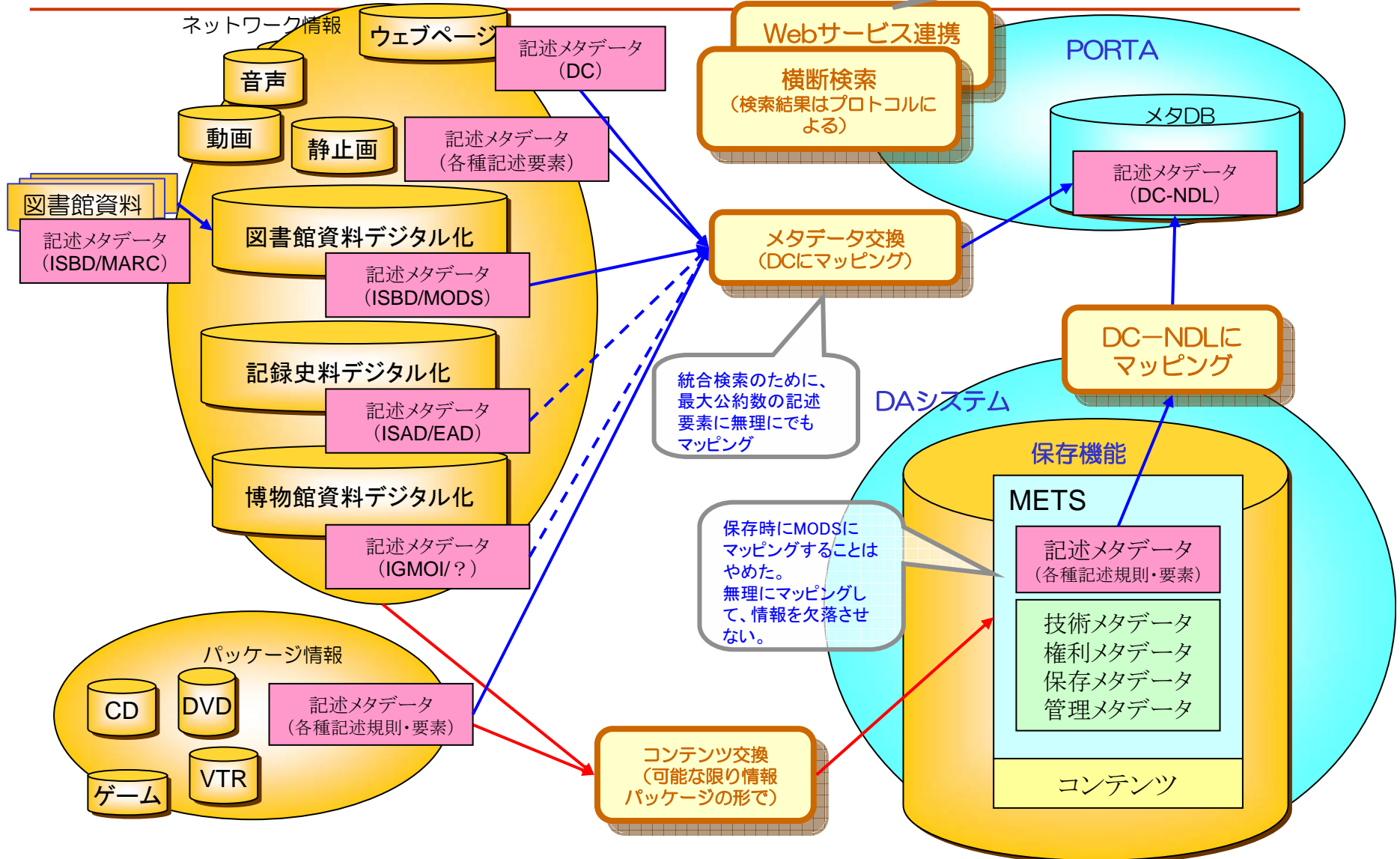
# メタデータマッピング

(DC-NDLを拡張)

## DCNDL\_PORTA

(1)Title	(11)Source
(2)Creator	(12)Language
(3)Subject	(13)Relation
(4)Description	(14)Coverage
(5)Publisher	(15)Rights
(6)Contributor	
(7)Date	(16)Annotation *
(8)Type	(17)Thumbnail *
(9)Format	* はPORTAでの拡張項目
(10)Identifier	

# PORTA/DAシステムでの メタデータの仕様の関係



# 利用状況

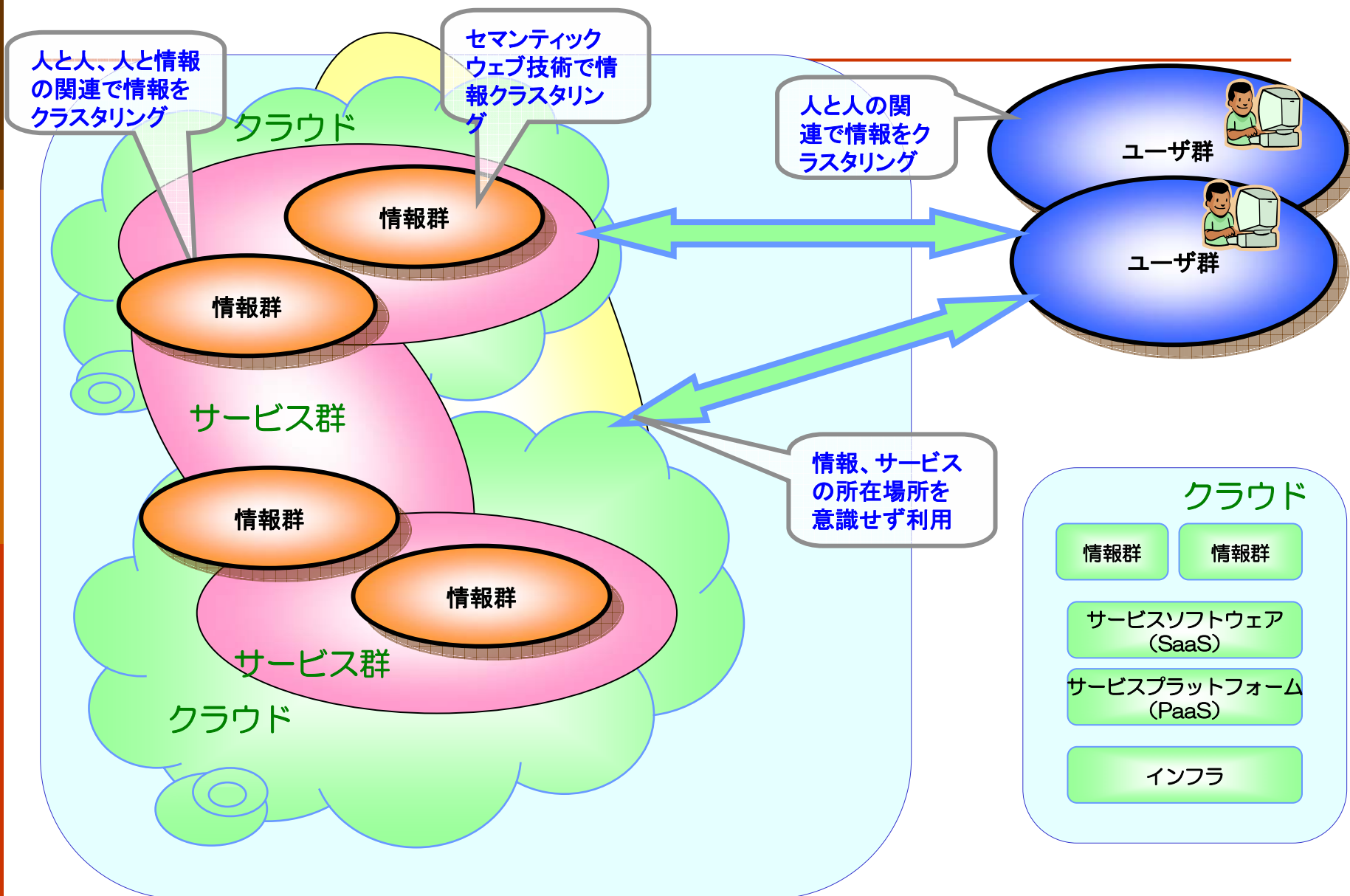
		平成20年 12月分	平成20年度 平均 (4月～12月)
累積データ(件)	ハーベストしたメタデータの 検索	7,197,937	—
新規データ(件)	ハーベストしたメタデータの 検索	74,126	67,216
トップページのアクセス(件) [日本語版]		19,590	18,260
トップページのアクセス(件) [英語版]		350	293
データへのアクセス(件)		2,310,882	574,642
RSSへのアクセス(件) [日本語版]		48,951	27,561
RSSへのアクセス(件) [英語版]		368	133
検索(件)		59,181	35,659
	うち、外部提供インターフェー ス(API)経由	38,794	19,191

## 3. PORTAの今後の方向性

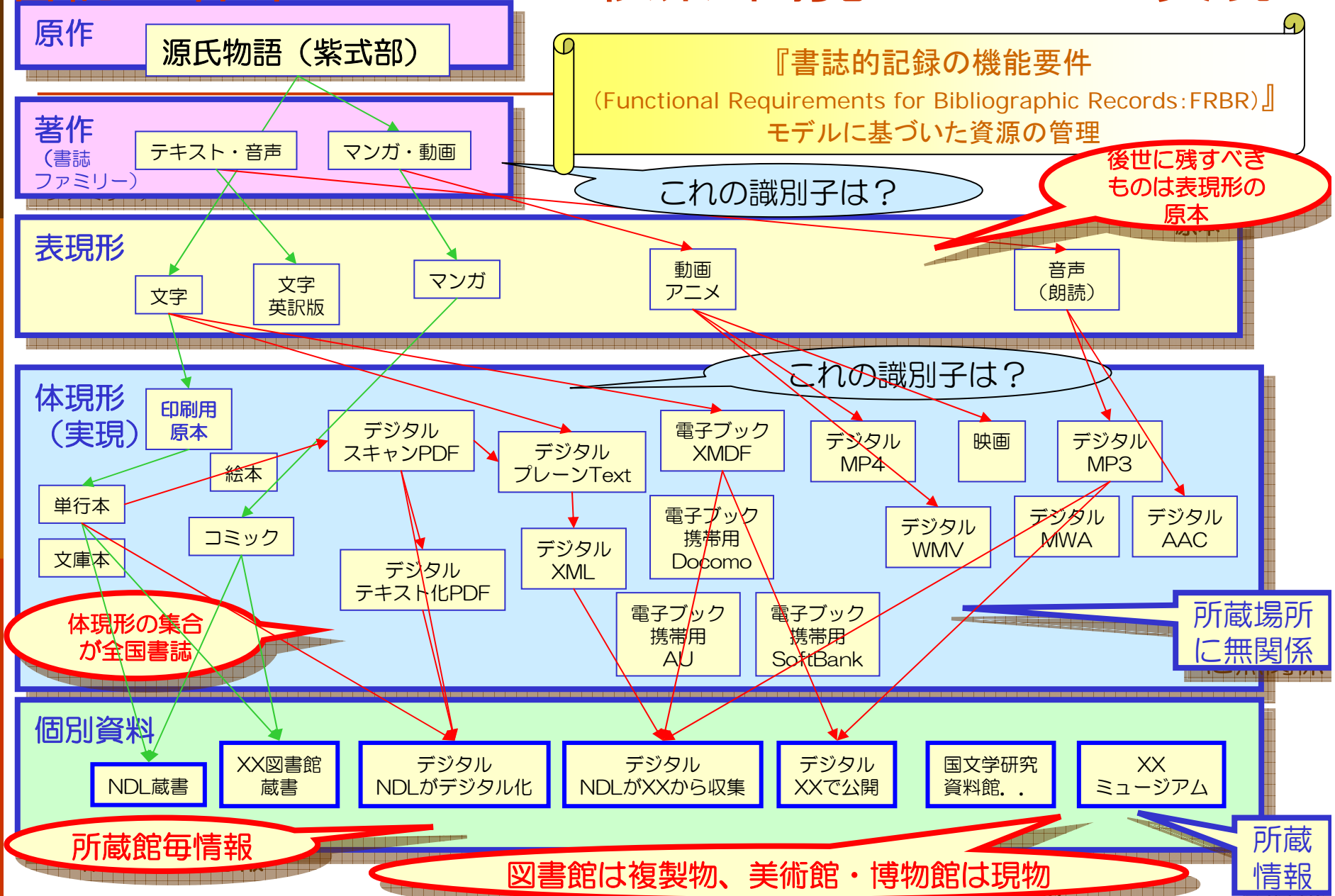
---

- 継続して、**デジタルコンテンツの可視化**を目指す
  - 各データプロバイダに検索やサービス連携のためのAPIが実装されることの普及啓発活動を継続
  - PORTAの外部インターフェースを利用した、様々なサービスの出現を期待
  - 各データプロバイダが共通インターフェースを実装したら、インキュベーションの役割を終える
- 利用者の**情報探索行動を体系的に支援**
  - コンテンツの体系的な検索・閲覧サービスの実現
  - レファレンス情報や各機関の蔵書目録と統合した検索
- 関係機関と協力して、**統合利用環境**を構築
  - 分散デジタルアーカイブ、統合検索ポータルの実現に向けた協力関係
  - 今後は、単なるメタデータ交換ではなく、セマンティックWebサービスでの連携を目指す

# 次世代は、 人と情報が関係付けられたサービスの連携（イメージ）

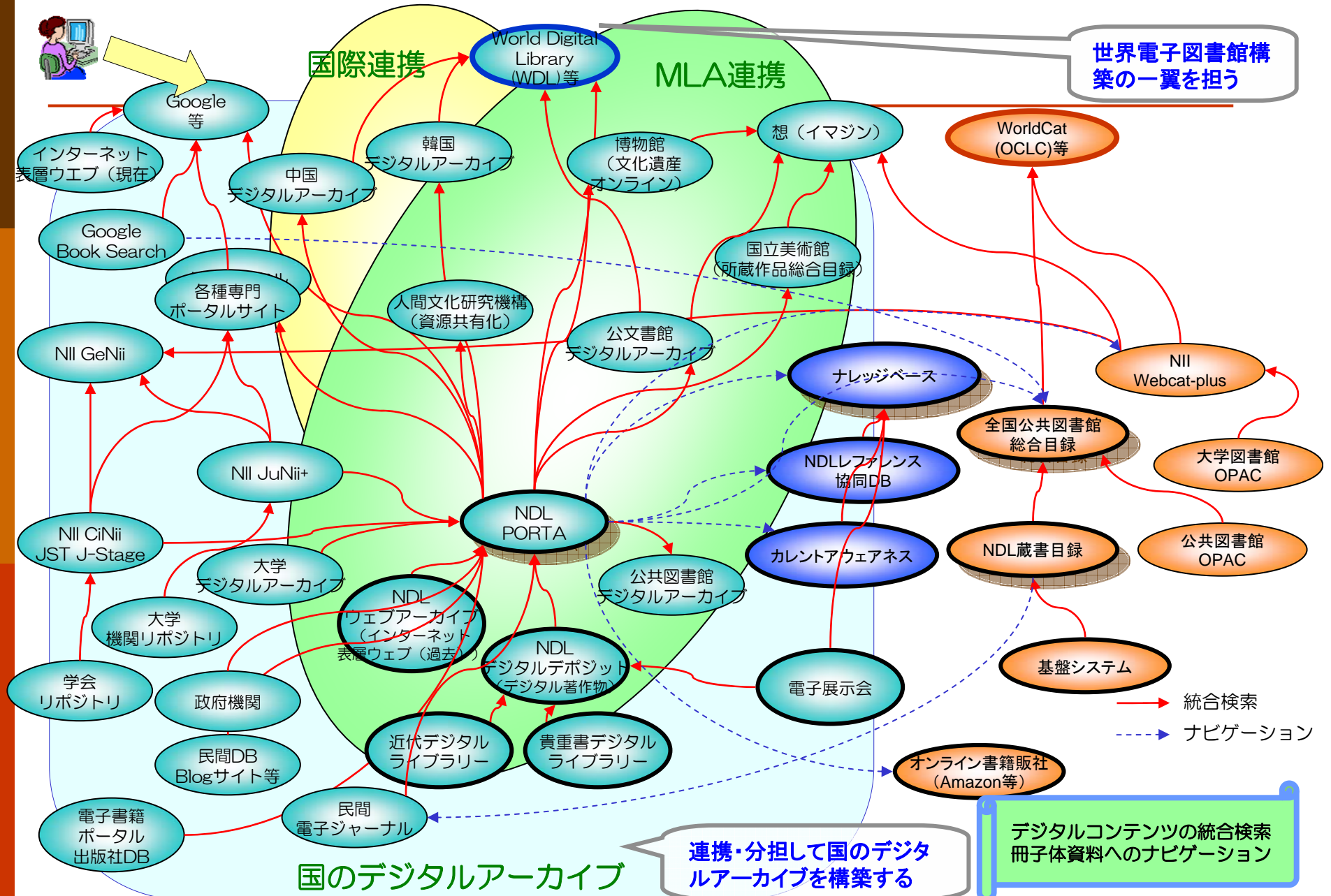


# 書誌的体系に基づいた検索・閲覧サービスの実現





# 統合まずは、連携による組織を越えた統合検索



世界電子図書館構築の一翼を担う

連携・分担して国のデジタルアーカイブを構築する

デジタルコンテンツの統合検索  
冊子体資料へのナビゲーション

# デジタルアーカイブシステムの構築に向けた主な連携協力

## 政府の施策

内閣官房

- 知的財産戦略本部
- IT戦略本部

## 国際協力

- World Digital Library
- LC
- IIPC
- 日中韓連携

## 分散アーカイブシステム構築

国の研究開発プロジェクト

- 各大学研究室
- デジタルインフォテック（慶應義塾大学）
- 文部科学省、経済産業省、総務省
- AIST、NICT、IPA. . .

共通仕様・技術提供

- 日本規格協会
- 電子情報技術産業協会（JEITA）

民間の技術開発、製品開発

- インターネットサービスベンダー
- 図書館パッケージベンダー



# NDLのデジタルアーカイブ事業に関連する 政府等の施策(2008年度)

## 「デジタルアーカイブの推進に向けた 申入れ」(平成20年3月12日)の要約 自民党デジタルアーカイブ小委員会

平成16年6月の「国立デジタルアーカイブ構  
想」の提言に沿って、わが国のデジタルアー  
カイブの総合ポータルとなるPORTAが開設  
された。

- 国立国会図書館のウェブアーカイブの  
本格実施のための法制度の実現
- 全国図書館のデジタルアーカイブの統  
合化
- 国立公文書館と国立国会図書館が協力
- 世界最先端のデジタルアーカイブ技術  
への対応

## 知財計画2008(6月18日)知的財産戦略本部

国立国会図書館のデジタルアーカイブ化と図書館資料  
の利用を進める

- デジタル化やインターネット情報資源等を収集保存し、  
一般ユーザーの利用に供する取組みの促進。
- このため、権利者の経済的利益や出版ビジネスとの関  
係を考慮
- NDL蔵書のデジタル化の推進に必要な法的措置を20  
08年度中に講ずる
- 国立国会図書館と他の図書館等との連携や図書館等利  
用者への資料提供の在り方については、関係者間の協  
議を促進し。(113/157ページ)

<http://www.ipr.go.jp/sokuhou/2008keikaku.pdf>

## 重点計画2008(8月20日)IT戦略本部

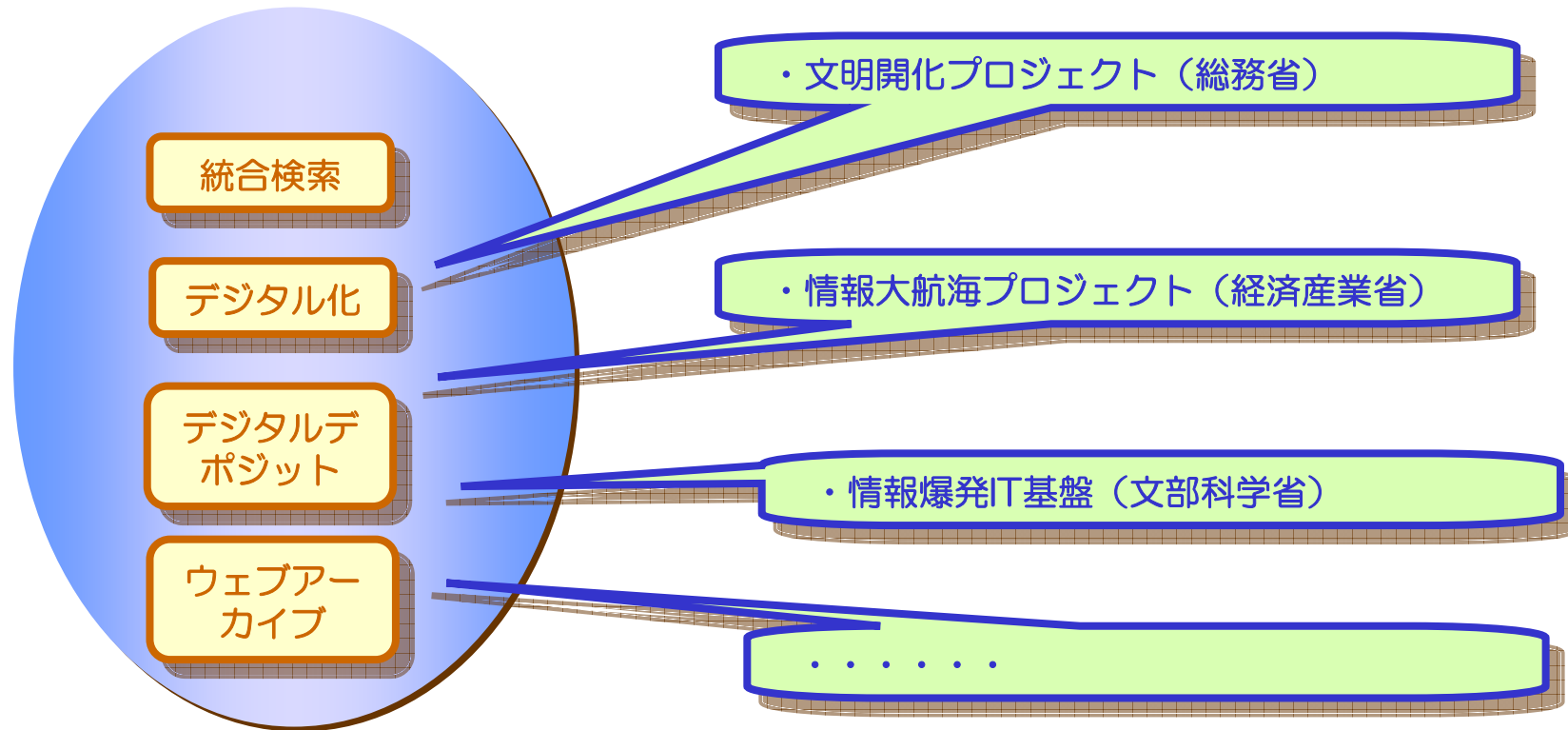
「デジタル文明開化プロジェクト」の推進(総務省及び関係省庁)

- 「ICT 成長力強化プラン」33に基づき、国立国会図書館、国立公文書館、他  
府省庁、地方公共団体、図書館・博物館・美術館、大学等との連携を図り、産  
学官を挙げてデジタル化を推進、日本中の知的財産を総デジタル化してつなぐ
- 「デジタル文明開化プロジェクト」を実施する。

<http://www.kantei.go.jp/jp/singi/it2/juuten2008/pubcom.pdf>

経済財政改革の  
基本方針

# 政府の情報利活用プロジェクト



# デジタルコレクションの構築に向けた主な連携協力(想定)

## ナショナルデジタルコレクションの構築

### 内閣官房

- 知的財産戦略本部
- IT戦略本部

### 学術情報

- NII、JST
- 大学図書館
- 電子ジャーナル出版者

### 公共図書館

- 都道府県立図書館
- 政令指定都市立図書館
- 市町村立図書館
- 専門図書館

### 美術館、博物館、文 書館、図書館連携 (MLA連携)

- 国立公文書館
- 東京国立博物館
- 国立美術館
- 各美術館、博物館、  
公文書館、図書館

### 出版者等

- 日本電子出版協会  
(JEPA)
- 日本印刷技術協会  
(JAGAT)
- 書籍出版者
- 電子書籍出版者
- インターネット書  
籍販売者

### 民間デジタルアーカイブ • 歴史的音盤アーカイブ 協議会 (HiRAC)

### 商用ポータルサイト • Google、Yahoo等

### 民間・個人サイト

- Blog、Wiki、SNSサイト
- ソーシャルブックマークサイ  
ト

### 政府情報

- 総務省行政管理局 (e-GOV)
- 各府省支部図書館
- 国立印刷局

# デジタルアーカイブの今後の構築計画(想定)

項目	平成21年度	平成22年度	平成23年度	平成24年度	平成25～26年度	平成27～28年度	平成29年度～	
世界電子図書館の構築 (World Digital Library)	サンプルコンテンツ公開	各国のコンテンツを地域、時系列、テーマ等で統合検索できるように		世界規模の一つのデジタルアーカイブの構築と提供				
日本のデジタルコレクションの提供 (分散DBの統合利用環境)	MLA連携の合意(相互運用性等)	MLA・学術サイト連携		商用サイト連携		各機関のDA構築支援・デジタル化支援		
NDLデジタルコレクションの構築 (デジタル化・コンテンツ収集)	貴重書デジタル化	大正期刊行物・学位論文・雑誌記事等デジタル化		政府サイト・学術サイトの制度的収集(ダークアーカイブであっても)		民間サイトを含めて選択的収集		
PORTA・DAシステムの環境の構築	複数分散デジタルアーカイブの統合に関する研究開発成果の適用		SemanticWebサービス、Ontology, TopicMap等の技術の適用		多角的な利活用に関する研究開発成果の適用			
	収集精度・効率を高める技術の適用 IIPCの成果を活用してバージョンアップ				OAIに準拠した保存システム(保存計画を含む)の実装		エミュレーション・マイグレーション機能の実装	
	磁気ストレージ増強		半導体ストレージ導入		恒久的保存デバイスによる保存用ストレージ導入 ▼平成28年頃			