

# デジタルアーカイブポータル のコンセプトと今後の方向性



平成18年5月26日

国立国会図書館 総務部 企画課  
電子情報企画室

# 電子図書館サービスの目標

- 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



アクションプラン

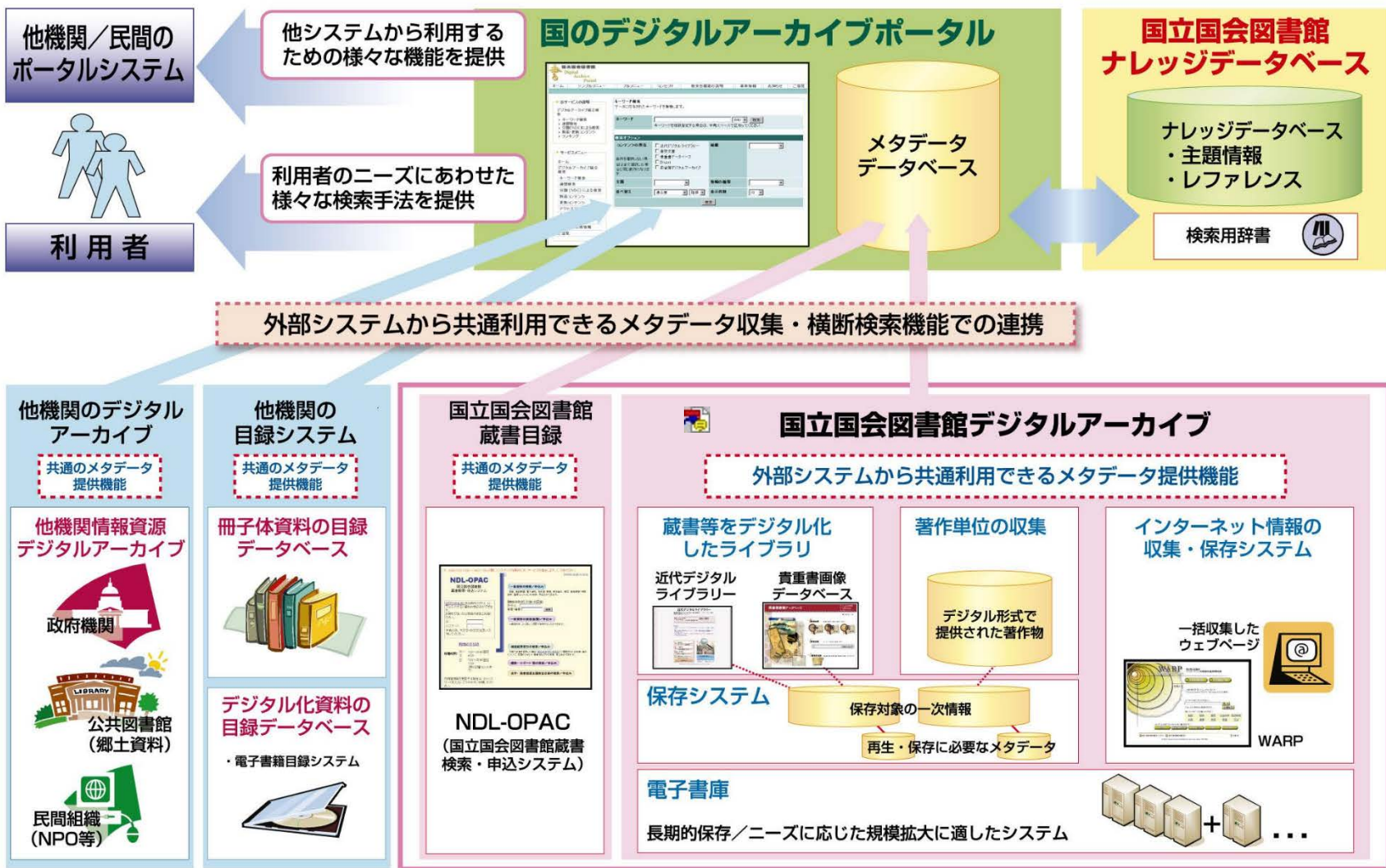
## ① デジタル・アーカイブの構築

- 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

## ② 情報資源に関する情報の充実

## ★ ③ デジタル・アーカイブのポータル機能

# NDL電子図書館サービスの全体像



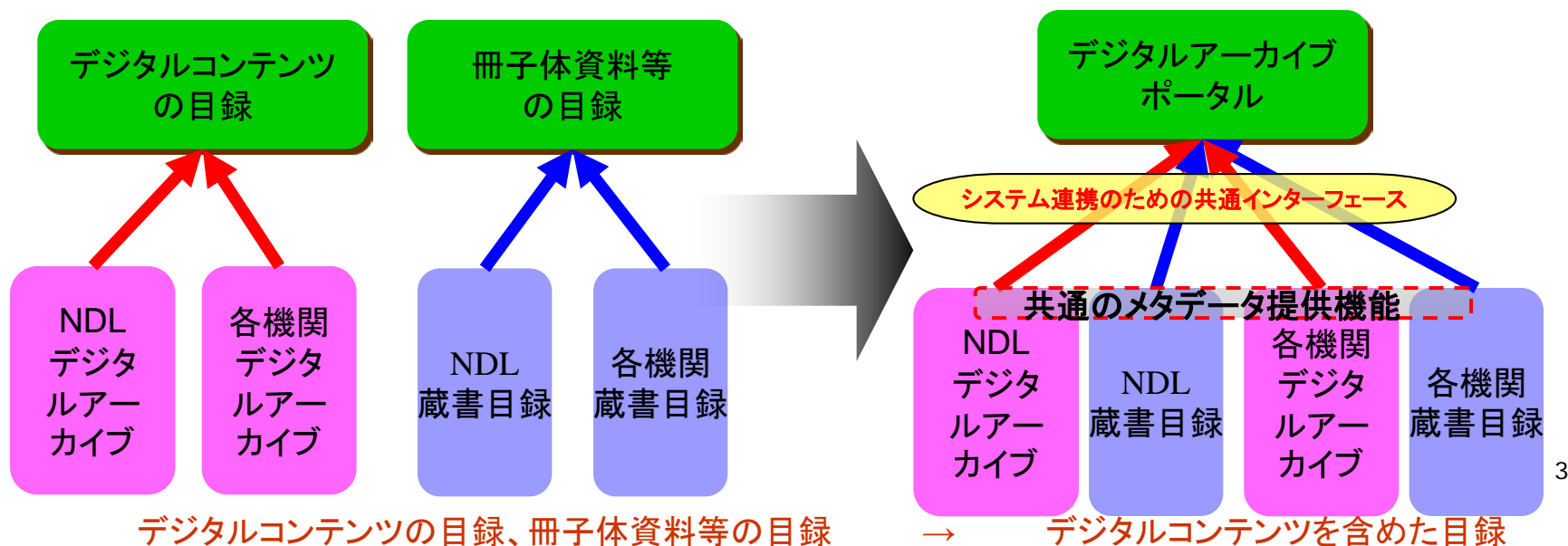
# デジタルコンテンツも含めた目録検索と閲覧

## 目的

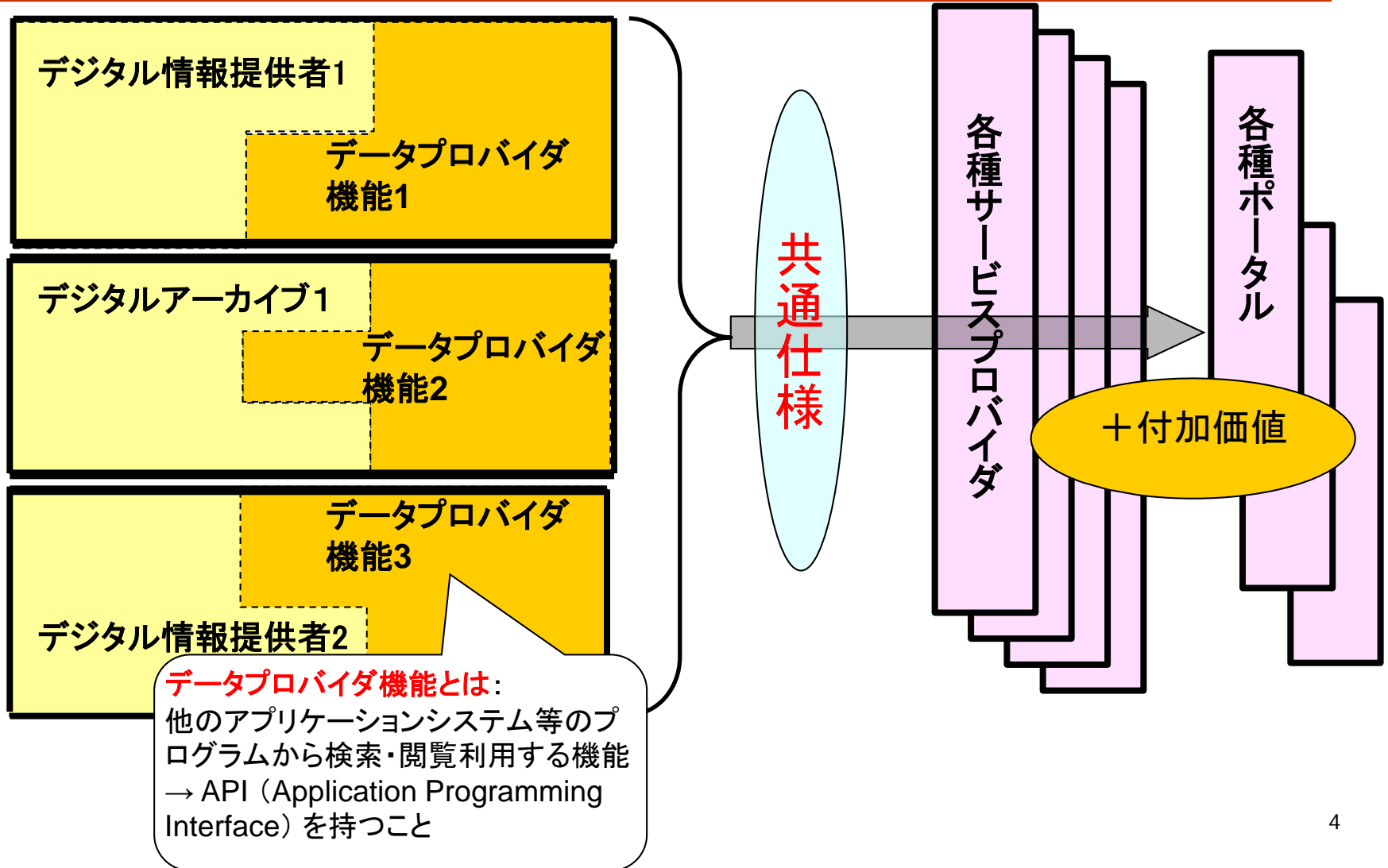
- 国のデジタル情報全体へのナビゲーションとしての総合的なポータルサイトの提供
  - 当館が保有するデジタルコンテンツ
  - 広くデジタルコンテンツを作成し提供する機関と協力・連携

## 目標

- ワンストップで的確に閲覧利用
- 様々な利用者、利用形態でのニーズに対応
- 紙媒体等の情報資源に関しても所蔵情報へ案内

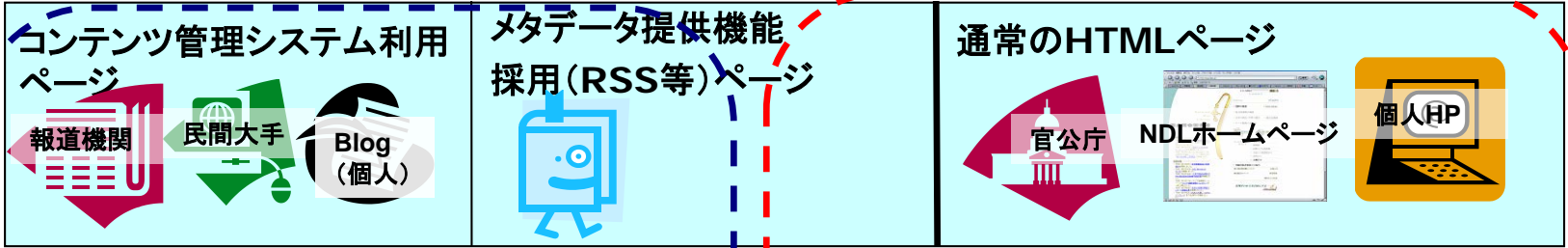


# データ提供側で共通インターフェースを !!

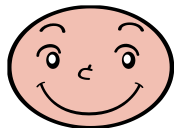
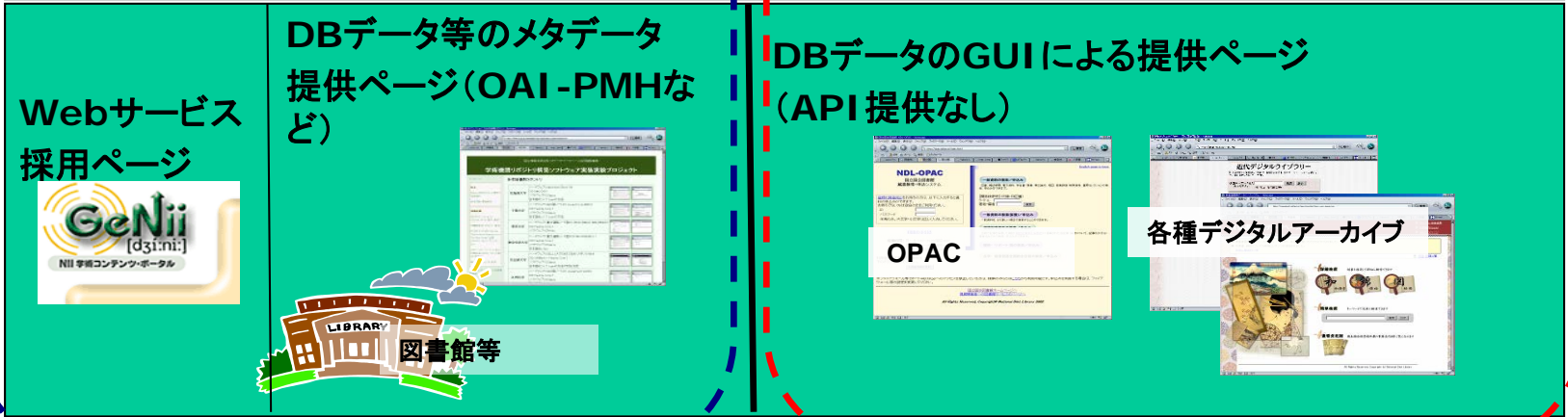


# 焦点はメタデータの収集： ワンストップ実現のため

表層



深層



メタデータの自動収集が可能



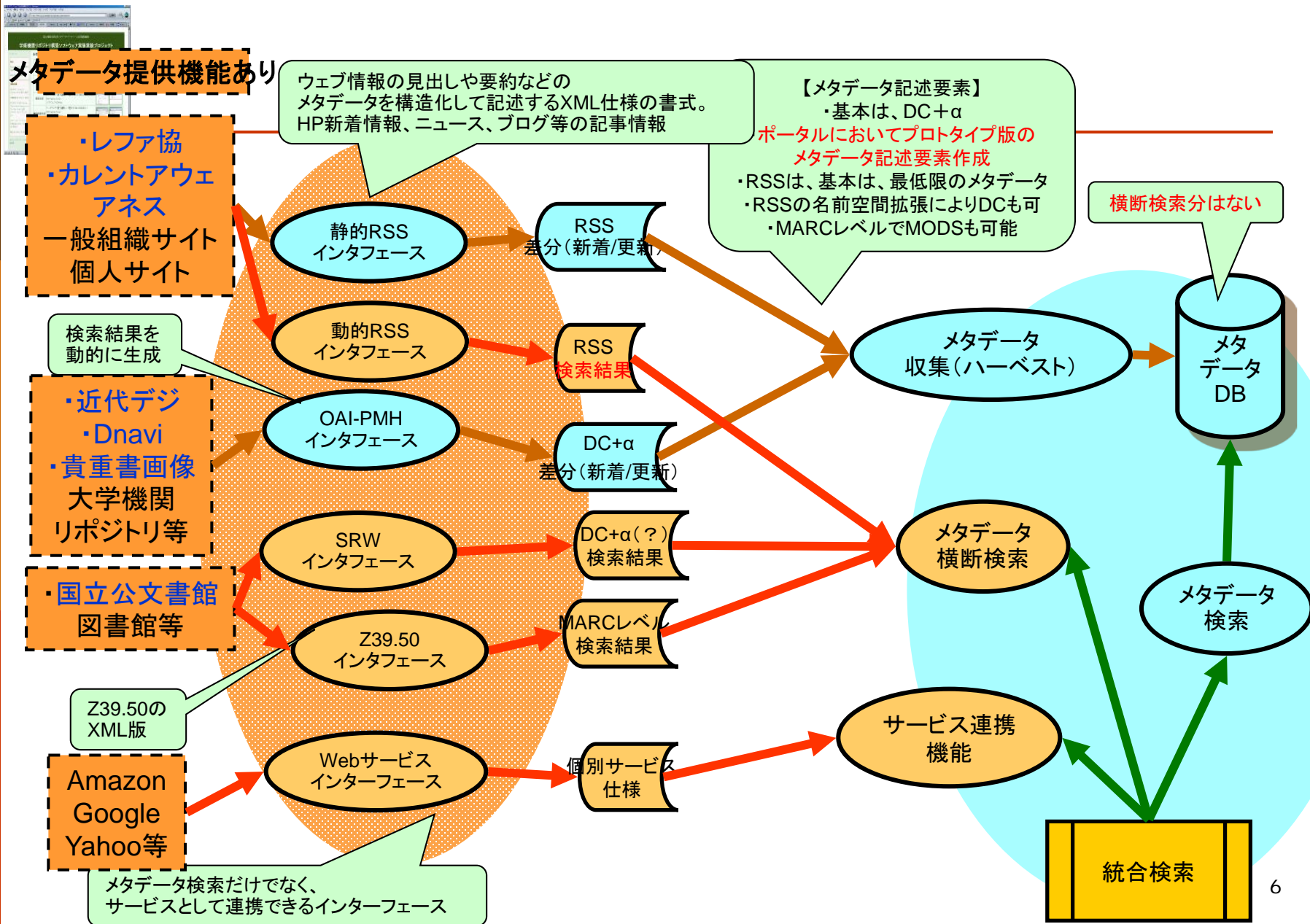
ワンストップポータル



メタデータを収集するための仕組みが必要

- ・半自動メタデータ生成
- ・メタデータ提供機能の実装

# メタデータ提供機能がある場合の収集



# メタデータ提供機能がない場合の収集

メタデータ提供機能なし

・国、政府機関等  
従来型ホームページ  
従来型データベース



ブックマーク作成機能  
【手動】

RSS作成機能  
【半自動】

仮データプロバイダ  
機能【自動】

メタデータ

メタデータの集合

RSSファイルの集合

RSS  
データ

OPML  
データ

RSSアグリゲータ  
機能

RSSファイルの内容に従って  
メタデータを収集する機能

メタデータ  
収集(ハーベスト)

メタデータ  
横断検索

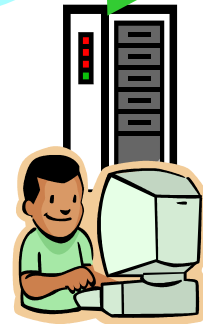
サービス連携  
機能

収集メタデータ生成機能

メタデータ  
DB

メタデータ  
検索

統合検索



ユー



# 商用ポータル、オンライン書店の動き

---

- Google, Yahoo, goo, amazon等の現状
  - Google:
    - Google Book Search, Google Scholar, Google Sitemaps, WebサービスAPI提供
  - Yahoo:
    - Yahoo!ディレクトリ検索, My Yahoo!
  - goo:
    - 日本語自然文検索、日本語概念フィルタリング
    - 次世代型ブログ検索実験
  - amazonドットコム:
    - なか見！検索, オンデマンド出版、
    - ユーザレビュー、評価、
    - WebサービスAPIの提供

# 商用ポータル、オンライン書店の動き 利用者環境の変化

## □ 商用ポータル等の今後

- 他社システムを有機的に組み入れ自社ポータルを強化
- 次世代ウェブ技術、手法を活用して、加速的に進化していく
- 検索機能は、関連する情報も有機的に含めて
- 検索対象は、表層ウェブから、深層ウェブ、さらに書籍もデジタル化して
- 操作性も、利用者のニーズに合わせたカスタマイズがブラウザの上で自由に
- 次世代ウェブ技術の積極的な適用

## □ 利用者環境の変化

- iPod、PSP、DS等のモバイル端末の普及
- モバイルインターネット利用者がPCを初めて追い抜く  
——総務省(2006年05月19日)

# NDLポータルがやるべきこと

---

- 公的機関を対象とした統合検索
- 大量の蔵書、収集コンテンツ、各分野の専門的知識等の豊富な資源を活かしたサービス
- 商用ポータル、オンライン書店等の資源（機能、コンテンツ）を積極的に活用
- NDLポータル自身も、他のポータルのデータプロバイダに

# メタデータ収集及び提供実施の観点

---

## □ メタデータ収集（統合検索対象）

- 収集対象機関
- ハーベストと横断検索
- メタデータ記述要素の詳細度

## □ メタデータの提供（他システムへ提供）

- ハーベストと横断検索
- 図書館としての新たな公共サービス

# 関係機関と協力することが必要性

- NDLのポータルとアーカイブ構築は電子図書館事業の車の両輪
  - NDLが全てのコンテンツを収集することは不可能。
  - 他機関のコンテンツとの横断的利用によって利便性が向上。
    - 短期的アクセスを保証 → デジタルアーカイブポータル
  - 利用者の必要な情報資源の収集と長期的保存。
    - 長期的アクセスを保証 → デジタルアーカイブ
    - ポータルが収集するメタデータ情報は個別収集の起点としての利用可能



## □ RSSとは、

- ウェブ情報の見出しや要約などのメタデータを構造化して記述するXML仕様の書式

## □ RSSの経緯

- 1999年3月Netscape、2000年以降blogで普及
- 現在では、blogのみならず新聞社、雑誌社などでのニュースサイトの更新情報
- 最近では、放送、音楽サイトでのiPodcast等でのコンテンツURL提供

## □ RSSの仕組み

- 更新情報の提供者は、RSS形式のファイルをWebサイトに置き、利用者はこれをRSSリーダーなどを用いて参照→Webサイトの更新情報を得ることができる
- 主な要素
  - Channel要素でサイト情報、item要素で、見出し、概要、リンク先、日付等
- RSSの拡張
  - 名前空間(モジュール)を追加して要素を拡張することが可能
  - PRISM, DublinCore等の要素、さらにFOAF、ジャーナルに掲載された記事のDOIも

## □ RSSリーダー(RSSアグリゲータ)

- 利用者が登録したRSSを定期的に取り得して、更新された情報の概要をまず把握し、その内容を取捨選択した上でWebサイトにアクセスすることが可能<sup>13</sup>

# RSSの利用の現状



- 図書館等での利用イメージ
  - お知らせ:休館日、サービス変更のお知らせの通知
  - 速報情報:雑誌の目次情報や図書館での新着受入情報などの情報の配信
  - 信頼ある書誌データを公共財として、検索結果を機械可読な形で提供
- 国内外の事例
  - カナダのアルバータ大学図書館
    - 新着図書について学内の図書館ごとに、またサブジェクトごとに分類したRSSを提  
National Library of MedicineのPubMedでは
      - データベースの検索結果をRSSのリンクとして生成。
      - このURLをRSSリーダーに登録すれば定期的に検索結果が配信され、最新情報を追うことができる
  - 京都大学図書館機構、一橋大学附属図書館
- NDLデジタルアーカイブポータルでの事例
  - メタデータ収集
    - 新書マップ、レファレンス協同データベース、カレントアウェアネス
  - メタデータ提供
    - 新着・更新コンテンツ
    - お知らせ、ニュース

# 今年度の実施事項

---

- 19年度稼動を目指した本格システムの設計・開発
- 本格システムで必要な機能の検証を、引き続きプロトタイプシステムで行う
- データプロバイダ立上げ支援
  - 共通仕様の提示、普及・啓発活動
  - データプロバイダ化の実装支援



# 本格システムで実現したいこと

---

## □ メタデータ収集・組織化

- データプロバイダの拡大
- データプロバイダの自動登録・管理(UDDI)
- メタデータの推定登録、半自動抽出

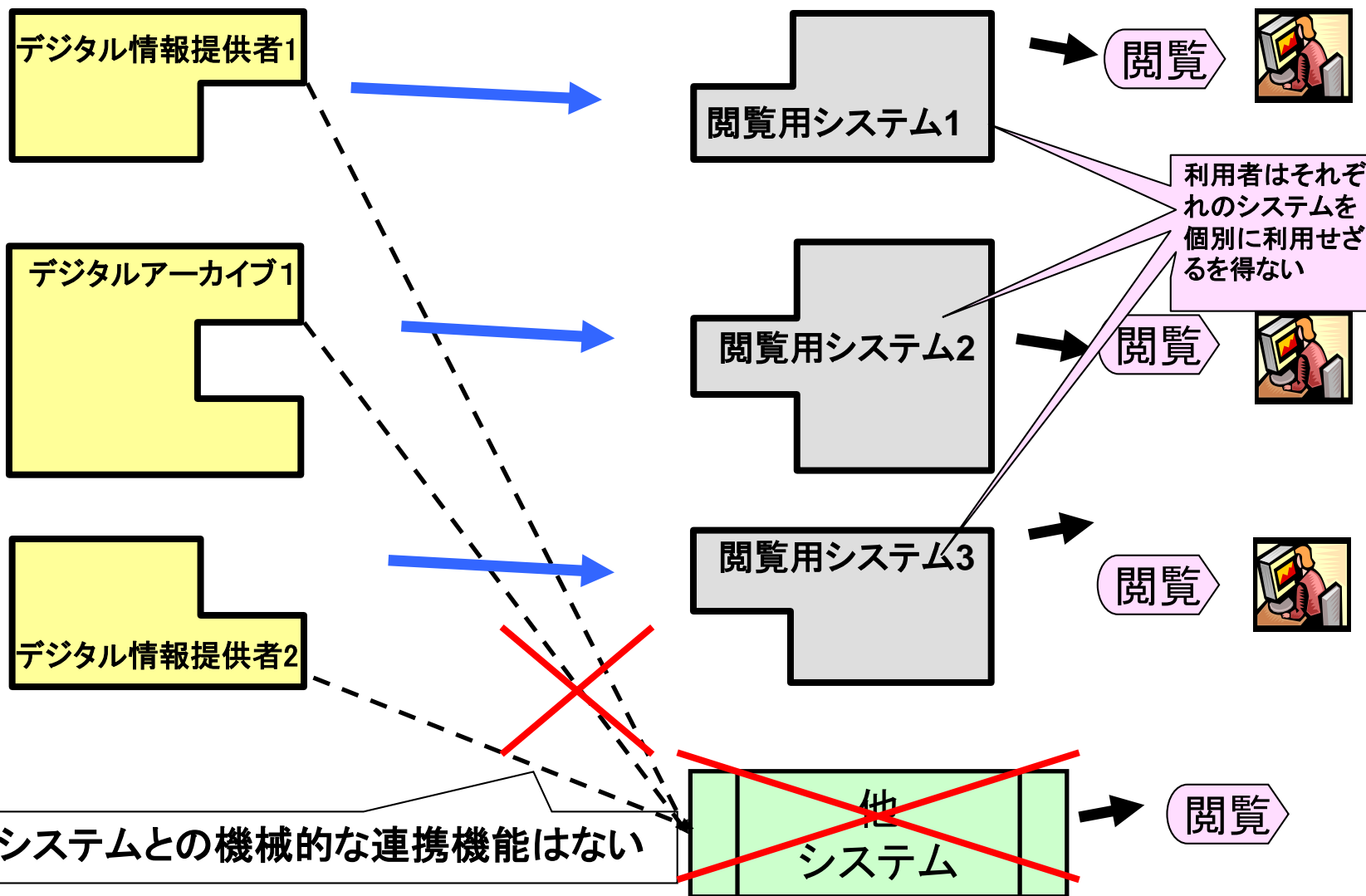
## □ 提供

- 画面操作性、Look & Feelの改善
- 様々な辞書データを取り込んだ辞書検索
- きめ細やかな検索
- マイポータル機能
- データを外部システムへ機械的に提供(APIの提供)

## □ インフラ

- 大量のデータ管理、アクセス、ユーザ管理が可能なシステム基盤

# 個別アプリケーションへは個別アクセスが一般的



# デジタルアーカイブポータル構築計画

---

## □ 16年度は

- 必要な機能の検証、適用すべき技術の検証
  - 見た目は容易に変更可能なように
  - ベースシステムは、将来性のあるオープンソース
    - 統合検索機能をモジュール化して実装
- コンテンツはサンプル
  - 著作単位、サイト単位の統合検索の試行

## □ 17年度は

- 19年度に必要な機能の評価と検証
- 検証に必要なコンテンツの追加
- 大規模対応のインフラシステムの設計
- 共通仕様の策定

## □ 18年度は

- 本格稼働システムの設計・構築
- 共通仕様の実装の啓発活動

# 開発状況(別添資料)

---

## □ 館内報告

- 電子図書館中期計画2004(平成16年2月17日)
- 平成16・17年度の業務・システム構築作業計画(平成16年12月27日)
- プロトタイプシステムの構築について(平成16年9月14日)
- プロトタイプシステムの平成17年度開発状況について(平成17年11月29日)

## □ 外部との意見交流

- 平成15年電子図書館全国連絡会議連絡会「日本のデジタルアーカイブポータルをどのように構築すべきか」(平成16年3月19日)
  - 内閣官房、総務省行政管理局、Yahoo!、Google、三菱総研)
- ポータル検索機能の強化に関する意見聴取(平成17年3月23日)
  - NTTレゾナント、Yahoo!、富士通

## □ プロトタイプシステムの概念図

- 機能情報関連図(16FY構築分)等
- 機能情報関連図(17FY構築分)等

# デジタルアーカイブ拡大策

## □ 最低限の共通仕様の策定を推進

- メタデータの作成ルール(項目及び内容記述等)
  - 最低限Simple DC(15要素)
- 機械可読なインターフェースの共通仕様の選定
  - RSS、OAI-PMH、Webサービス等



## □ 共通仕様の普及啓発

- 共通仕様の実装支援ツール類の試作及び提供
- 登録されたサイトの自動リンク



## □ 将来的な共通仕様の普及啓発活動

- 一次情報の機械可読化(ウェブサイトのセマンティックWeb化)
- インターネット上でのシステム連携(デジタルアーカイブのXML Webサービス化)

# 他のポータル機関の動向とNDLとの関係

## <国のポータル>

### □ NII

- GeNii (Webサービス化)
- 学術機関レポジトリ (OAI-PMH)
- 文化遺産オンライン構想
- NDLとの関係???
  - 共通仕様作り
  - 相互に統合検索実験?

### □ JST

- 科学技術ポータルを計画中??
- NDLとの関係???
  - 科学技術ポータルを協力して構築??

### □ 公文書館

- デジタルアーカイブ (Webサービス化、OAI-PMH)
- NDLとの関係???
  - 相互に統合検索実験?

### □ e-GOV

- 現在はインデキシングデータ受領不可能
- 次期e-GOVに期待

## <民間ポータル>

### □ Yahoo!

- 総合ポータル、MyYahoo
  - RSSリーダー
- ウェブ全文検索、ディレクトリ検索
- NDLとの関係???
  - Yahooキャッシュデータの利用
  - ホスティングスペース
  - YahooJapanのデジタルデータの利用
  - YahooJapanからデジタルアーカイブへの誘導
  - NDLポータル検索窓のモジュール追加?

### □ Google

- ウェブ全文検索 (深層ウェブもそろそろ)
- GooglePirnt
  - デジタル化とメタデータ収集も
- [Google Personalized Home](#)開始
- Google Sitemapsベータ開始
  - 収集希望urlリスト (RSS、OAI-PMHも)
  - 登録手続き
- NDLとの関係???
  - 蔵書目録のインデキシング
  - 国会会議録のインデキシング
  - GooglePrintは???

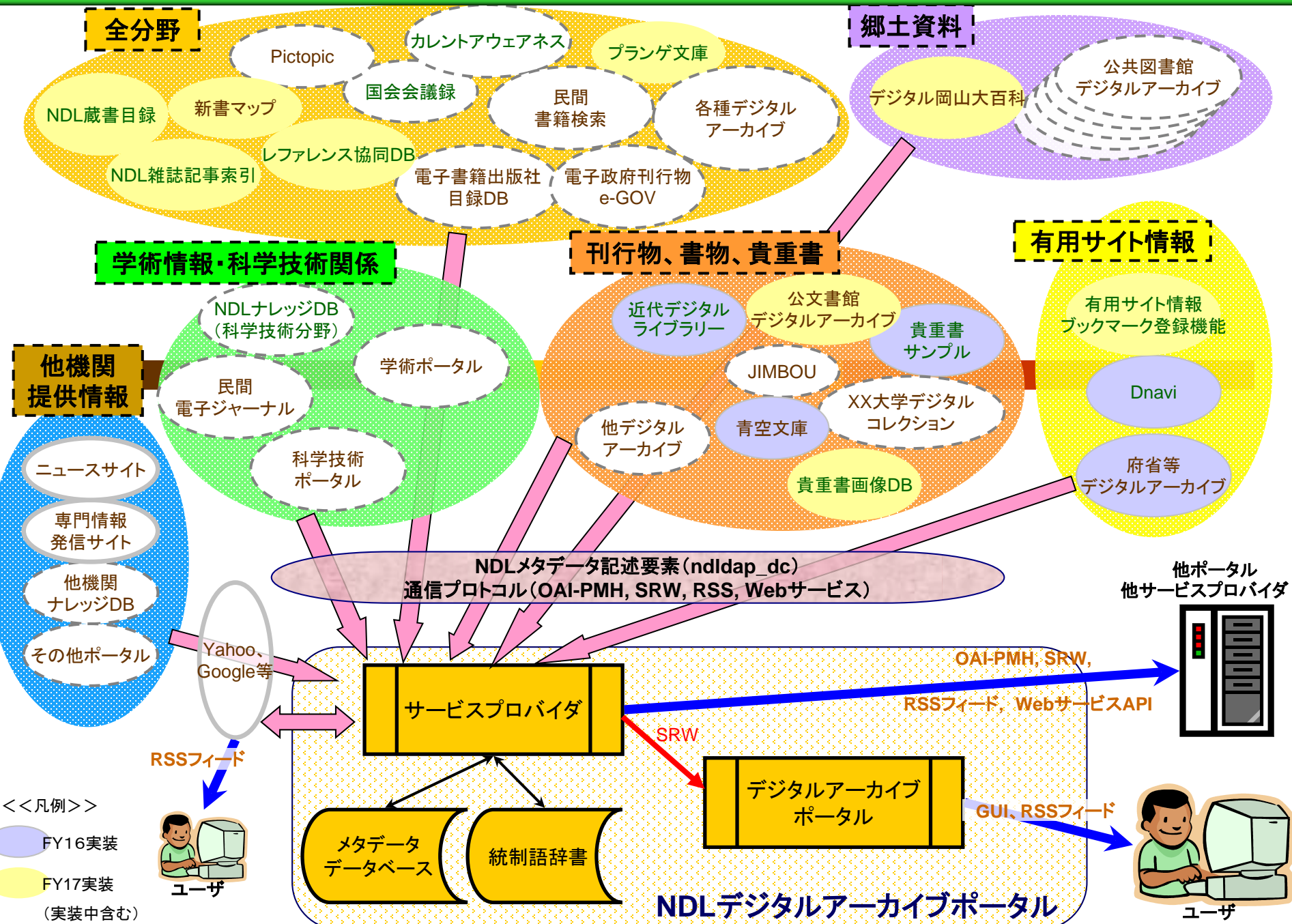
### □ goo

- Googleを利用してWeb全文検索
- 各種付加価値のついた検索機能
- NDLとの関係???
  - 検索に関する各種実証実験の成果 (プロダクト等)の提供

# NDLデジタルアーカイブポータルが求めるメタデータ 交換のための共通仕様

- **メタデータ記述要素**
  - DCLレベル(ダブリンコア基本15要素または、DC+a)
  - MARCレベル(MODS等)
  - 必要最低限レベル(Title、Description、date、URL程度)
  - ポータルプロトタイプでのメタデータ記述要素(ndldap)
    - DC+aの要素に、MARCの要素の中から、メタデータ検索のために必要と思われる項目をマッピング
    - プロトタイプ版として18年3月に策定
    - 今後、DCLレベルでのNDLメタデータ記述要素として、改版し、NII等の関係機関と調整する
- **通信規約**
  - OAI-PMH → DCLレベル
  - SRW/SRU → DCLレベルからMARCレベル
  - Z39.50 → MARCレベル
  - RSS → 必要最低限レベルから、DCLレベル、MARCレベル
  - Webサービス連携 → メタデータ検索だけでなく、サービスとして連携できるレベル(amazon.comのようなサービス)
- **共通仕様の実装が困難な場合**
  - 共通仕様実装支援ツール類の提供
  - ポータル側でゲートウェイ機能の実装

# 【NDLデジタルアーカイブポータルのプロトタイプシステム実装イメージ】





# ポータル成功の鍵は？

- いかにかデジタルアーカイブ内のコンテンツのメタデータを収集できるか
- 過渡期は
  - 表層にあるコンテンツはRSSにより収集
    - ごく一般のサイトで普及。むずかしいことは言わずに、実現可能？
    - RSS等の提供ができない場合は、メタデータの半自動生成ツールで生成
  - 深層にあるコンテンツはOAI-PMH、SRW (Webサービス化したZ39.50) のようなインターフェースで収集
    - OAI-PMH、SRWは業種業態を越えて普及することは困難？
    - 共通項目として、最低限シンプルダブリンコア、RSSの情報でXML Webサービス化を要請
  - ウェブでのコンテンツ管理システム (CMS) の利用が進めば
    - セマンティックWeb化が自然に進む
      - 一次情報、二次情報にメタタグが付与され、機械的に意味解析が容易に
      - メタデータの自動収集が可能に
  - メタデータが収集できることにより
    - メタデータのURLを、ウェブ上の著作物収集の起点として利用できる
- 将来は
  - Webサービス機能によるシステム連携
    - Amazon.comのようなインターフェースの実装
- 結果
  - **ポータルができて利活用推進！**
  - **アーカイブができて保存も安心！**

# ポータル提供に向けて実施すべきこと

- **一次情報の網羅的な検索サービスの提供を目指して**
  - 対象コンテンツの拡大
  - データプロバイダ機能の実装支援
- **個別利用者の必要に応じたサービスの提供を目指して**
  - ナレッジデータベースの提供
  - 検索機能の充実
  - マイポータル機能
- **各種ポータルサービスの発展を目指して**
  - 共通仕様の公開と普及啓発
    - データプロバイダの共通仕様の公開と普及活動
    - デジタルアーカイブ登録システムの構築と提供
- **本格サービスのためのシステム基盤整備**
  - 対象データプロバイダの拡大、高アクセス、大量ユーザ管理に対応
  - 高アクセスに対応できるコンテンツマネジメントシステム(CMS)
  - システム間連携仕様及び連携のためのモジュール群

# 書誌データでの利用例

## □ 蓄積されたデータの開放とサービス連携

- メタデータを記述し配信するRSSは、サービス、データベース間のスムーズな連携をサポートし、有機的なサービスが行われる基盤を生成するもの

## □ 現状の書誌(メタ)検索サービス

- 現在のサービスは、GUIでキーワードを入力し検索結果として、書誌データの一部を返すもの
- 他のデータと統合し再利用することが困難

## □ 検索結果のRSSデータ開放では

- RSS書誌データのRSSは、一般の利用者がURLで簡易に指定し自由にダウンロードできる
- Amazonでは、通常のインターフェースを介さずに外部プログラムから自由に検索できるWebサービスを公開

## □ 図書館としての新たな公共サービス

- 目録規則などによりコントロールおよびメンテナンスされた信頼ある書誌データを公開することは、ネットワーク上のデータサービスの基盤をなす公共財として利用されうる



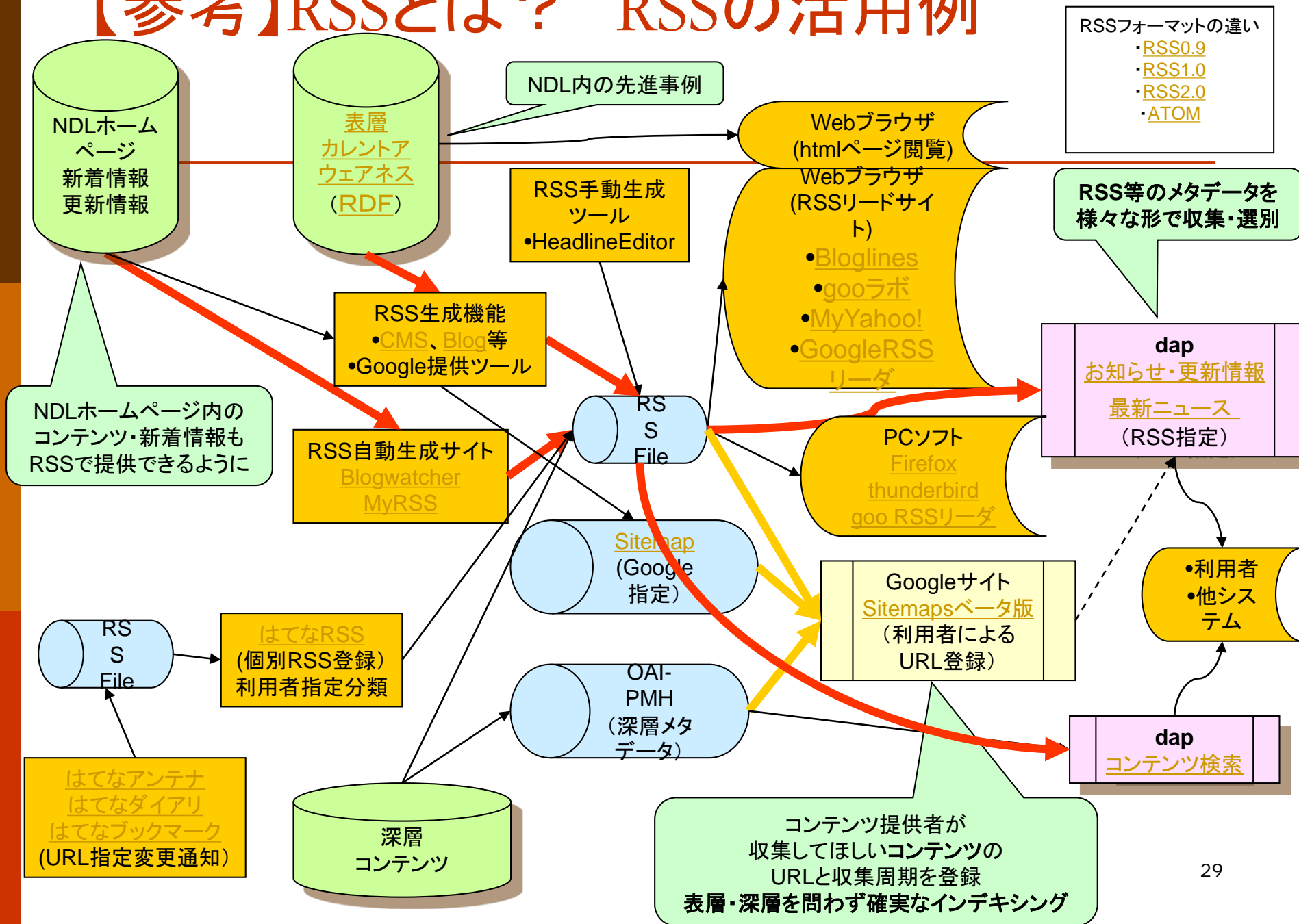
# RSSの利用の現状

- RSSとは、
  - ウェブ情報の見出しや要約などのメタデータを構造化して記述するXML仕様の書式。
  - 主要要素
    - Channel要素でサイト情報、item要素で、見出し、概要、リンク先、日付
- 利用者にとって
  - コンテンツをWebブラウザで直接見る方法に加えて、RSSファイルを機械的に収集して、その内容を分類して、必要な情報を選択しなら、様々な方法で閲覧できる
- RSSの普及の現状
  - 現在、サイトの新着情報等で利用されている
    - ニュースサイト、各組織お知らせ、Blog
  - コンテンツマネジメントシステム(CMS)にはRSSを出力する機能が実装されている
  - メタデータの項目が最低限になっており、サイト運営側での導入が容易
  - メーラと同じようなGUIでRSSで提供されている新着情報を読めるツール、Webアプリが広く普及してきている
  - 深層ウェブの情報のメタデータ提供にも適用が可能
  - 適用範囲は拡大傾向
    - 情報提供に積極的なサイトは、サイト内にRSSファイルを設置
    - 政府機関でも適用が始まりつつある
      - コンテンツの所在情報としても適用が進む
    - PodCasting等による音楽・映像情報のメタデータの収集

# RSSの将来的な普及の可能性

- Web2.0やセマンティックウェブでの中核的な要素
  - Web2.0とは、
    - Webをプラットフォームとした次世代的なWebのあり方
      - 専用アプリを利用せず、Webブラウザ上で動作し利活用できるサービス
    - 利用者が自分のデータベースなどとXMLで連携させて自由に利用できる
  - セマンティックウェブとは、
    - ウェブ上のページ等のリソースに、構造化されたメタデータを埋め込むことで、コンピュータ自身が情報の意味を解釈し、推論することを可能にする技術
    - 例えば、htmlページにmetaタグをつける。画像ファイルのプロパティに記述メタデータを埋め込む等...
    - 埋め込むメタデータの記述様式として、RSSがある
- 市場では、
  - 図書館を含めて、多くの業種業態の組織で付与可能なメタデータ仕様として普及が見込まれる。
- 図書館での対応
  - 今後はRSS関連技術を最大限に活用した仕組み作りに注力する必要がある

# 【参考】RSSとは？ RSSの活用例



# 関係機関と協力することが必要性

---

- 他機関の資源の活用(機能、コンテンツ)
- 共通仕様の普及
  - 共通仕様の策定
    - メタデータ記述要素
      - 例えばNIIとNDLメタデータ記述要素のスキーマの整合性を！
    - インターフェース
      - メタデータハーベスト、横断検索から、Webサービスとしてのシステム連携へ
  - 共通仕様の実装
    - データベースに機械的アクセスを可能とする共通仕様の実装
    - 表層ウェブ上の著作物のメタデータ提供機能の実装