

国立国会図書館の新たな動き

—デジタル情報も含めた統合検索サービスを目指して—



平成20年2月21日

国立国会図書館関西館

中山 正樹

「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

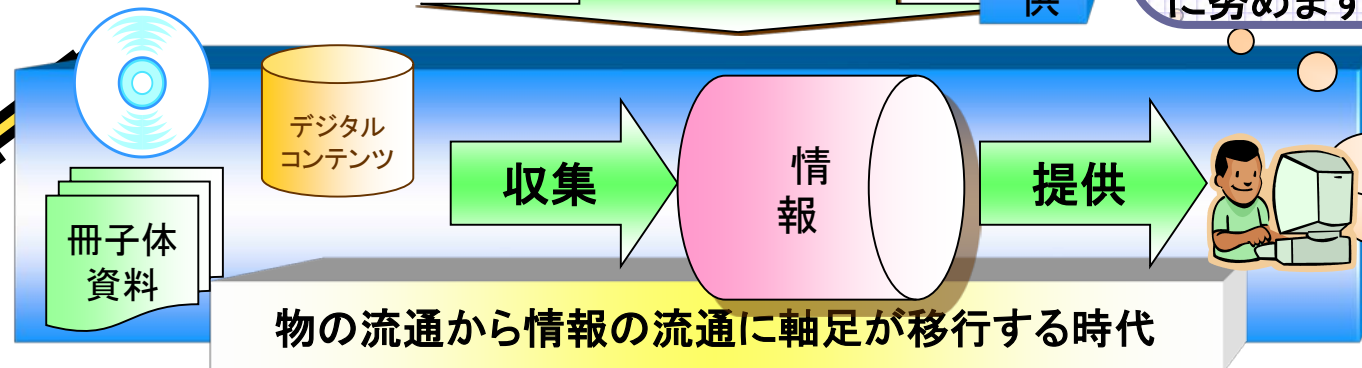
関係機関と協力して収集・提供

(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



関係機関と分担して、情報資源を知識として集積し提供

現状のデジタルアーカイブサービス

◆史料に見る日本の近代

- ◆描かれた動物・植物
- ◆近代日本人の肖像
- ◆日本国憲法の誕生
- ◆世界の中のニッポン
- ◆デジタル貴重書展

- ◆PORTA
- ◆Dnavi

ホームページ

編集コンテンツ
(電子展示会)

二次資料

一次資料

- ◆国会会議録
- ◆貴重書画像データベース
- ◆近代デジタルライブラリー
- ◆WARP

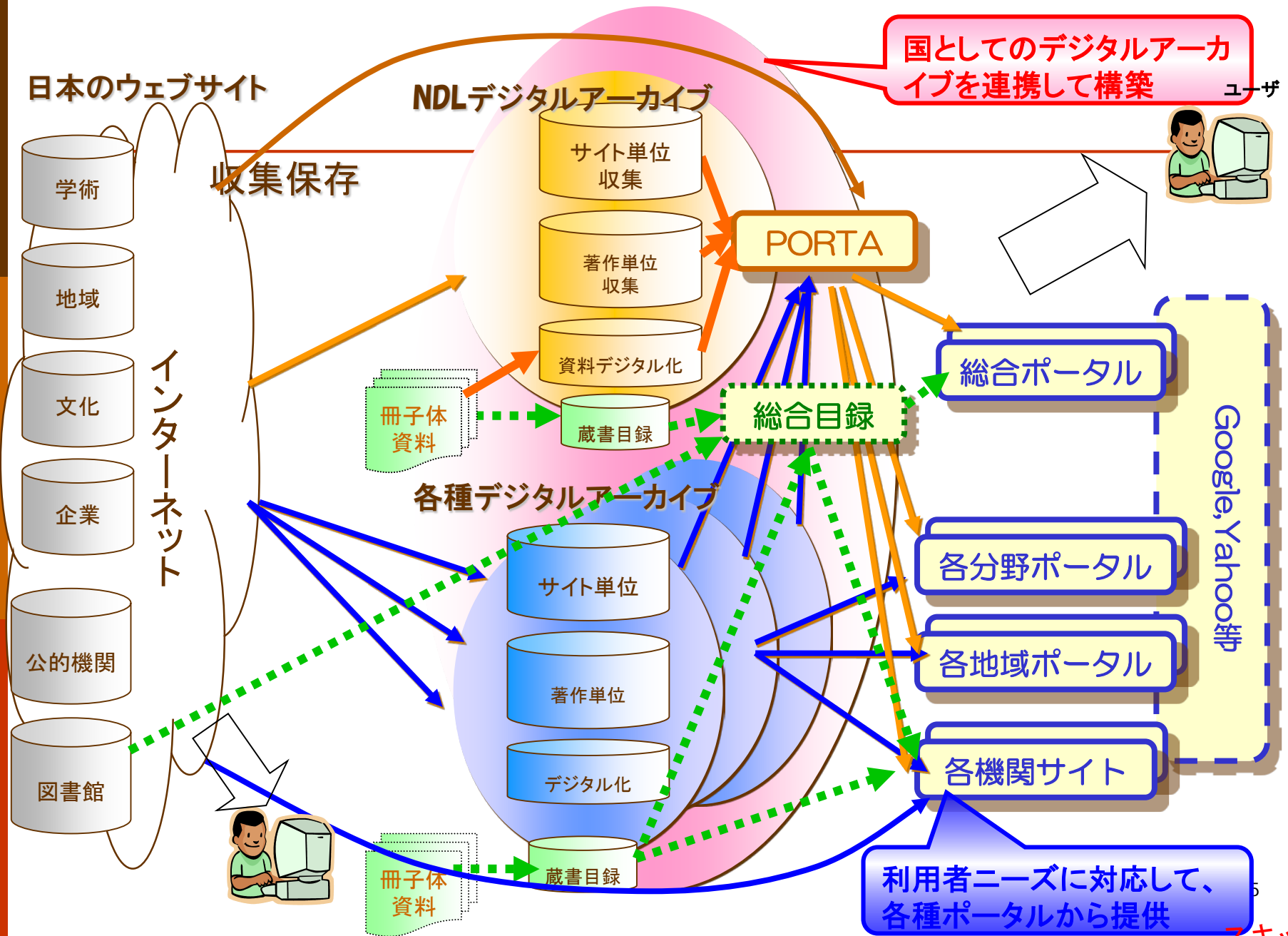
サービスを支えるシステム

紙ベース
資料の目
録

NDLのデジタルアーカイブ事業

- 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
 - デジタル化した資料、デジタルで生まれた情報をウェブで公開
- 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
 - 各アーカイブ機関をあわせて国のデジタルアーカイブ
 - NDLだけでなく、各機関協力して、分野、地域、業種毎でも
- 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

国のデジタルコンテンツの分散保存と提供



(デジタルアーカイブ構築のご説明の前提)

紙資料とデジタルコンテンツの違い

□ 紙資料

- ものを見れば何かが分かる
- 紙資料の組織化→書誌情報付与


□ デジタルコンテンツ

- ただのファイル→何のファイルであるかの情報が必要
- デジタルコンテンツの組織化→メタデータ情報付与

□ メタデータと書誌

- 書誌情報＝紙ベースの資料につけた情報
- メタデータ＝デジタルコンテンツにつけた情報
- **メタデータ≒書誌情報**

- 厳密には、メタデータの種別のうちの「記述メタデータ」



タイトル、著者名、出版社名、件名、主題。。。。

(デジタルアーカイブ構築のご説明の前提)

表層と深層の違い

□ 表層ウェブ

- 検索ロボット(クローラ)で検索できるウェブページ
 - GoogleやYahooなどで検索できるページ

□ 深層ウェブ

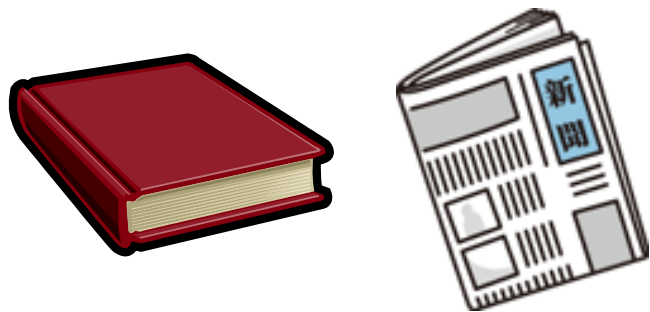
- 検索の条件等を入力しなければ見られないページ
 - データベース内のページ
- 利用者認証しなければ見られないページ

検索ロボットでの
アクセスが困難

有用な情報は
深層ウェブに
多い

表層ウェブの量
≪ 深層ウェブの
量

デジタル資料を探す



紙資料はNDL-OPACや
総合目録等で検索可能



デジタルアーカイブは様々な
機関が作っていて、どこに何
があるか分からない！

デジタルアーカイブのポータルが必要


Google、WARP、PORTAの違い

□ Google

- 現在見られるウェブサイトの表層ページの検索
- 深層ウェブは見られない
 - (深層でもメタデータを提供しているものは見られる。Cf. CiNii等)

□ WARP

- もう消えてしまった過去のウェブサイト
- 定期的な収集により時系列変化がわかる
 - (収集周期の間の変化はわからない。。。)



平均余命
は44日

□ Dnavi

- 各機関が提供しているデータベース(深層ウェブ)の入り口まで案内

□ PORTA

- 基本は、各機関が提供している深層ウェブ内のコンテンツへ案内
 - 表層ウェブでもコンテンツ単位にメタデータが収集できるものも
- デジタルコンテンツだけでなく、紙ベースの資料の目録も検索対象にできる

デジタルアーカイブポータル(PORTA)のご紹介

- 特徴
 - 国全体のデジタルアーカイブの構築と提供を目指して
- 統合検索対象
 - ハーベスト先、横断検索先
- 提供機能
 - 検索の流れ
 - 辞書による検索のサポート
 - パーソナライズ機能
- 連携機能
 - PORTAで検索可能となるためには
 - PORTAの検索機能を活用するためには



- 国の電子情報資源や情報提供サービスにナビゲートする総合的なポータルサイト
- デジタル資源の保管庫である「デジタルアーカイブ」から情報を探す
- 国立国会図書館のデジタルアーカイブに加え、協力機関の複数のデジタルアーカイブを対象に一元的な検索
- 利用可能なコンテンツやサービスへ案内
- PORTA(ポルタ)
ラテン語で「門」や「入り口」を意味する、当ポータルの通称です。

	NDL	各機関
デジタル 一次情報	近代デジタルライブラリー 貴重書画像データベース 児童書デジタルライブラリー 貴重書サンプル	デジタル岡山大百科(岡山県立図書館) 秋田県立図書館デジタルライブラリ 青空文庫 国立公文書館デジタルアーカイブ アジア歴史資料データベース
二次情報 参考情報 etc.	蔵書目録 雑誌記事索引(2003年以降) プランゲ文庫雑誌・新聞目録 児童書総合目録 WARP、Dnavi レファレンス協同DB カレントアウェアネス	秋田県立図書館記事索引 府省所管デジタルアーカイブサイト情報 (内閣官房) 新書マップ(NPO法人連想出版)

データプロバイダ数全20種 (館内12、館外8)

データ件数合計約800万件 (平成19年10月公開時点)

□ 検索機能

- 簡易検索／検索項目を特定しないキーワード検索
- 詳細検索／タイトル、作成者、分類等の検索項目を特定したキーワード検索
- 連想検索／フレーズや自然文から連想される語を利用した検索
- 分類検索／NDC(日本十進分類法)、NDLC(国立国会図書館分類表)のカテゴリをたどりながら検索

□ 補助機能

- 辞書検索／辞書データ(国立国会図書館の典拠データ)を検索し、グラフィカル表示
- 辞書による検索サポート／キーワード検索時に、辞書データの検索結果と「Wikipedia」へのリンクを表示
- 関連情報リンク／検索結果からさらに関連サイト等を検索

●検索からコンテンツ閲覧まで



簡易検索

検索結果一覧



結果詳細表示



関連情報リンク



資料の所在検索へ

書籍販売サービスへ

ウェブサイト検索へ

コンテンツ閲覧へ



●辞書検索



統制語、上位語、下位語、関連語、類義語をグラフィカル表示。

●辞書による検索サポート

検索結果一覧 (簡易 / 詳細検索)



キーワードで辞書データも同時に検索。

Wikipediaの検索結果へ遷移。

選択した語を用いた再検索 (OR検索) が可能。

[デモへ](#)
[Wikipediaへ](#)
[ツールバーへ](#)

簡易検索：可能な限り情報入手先までナビゲーション

- 「終戦 詔書」で検索してみると...

簡易検索

キーワード

検索

すべてく 選択しない場合はすべてになります)

本文、デジタル画像等(一次情報)

近代デジタルライブラリー

貴重書画像データベース

国立公文書館デジタルアーカイブ

デジタル岡山大百科

青空文庫

貴重書サンプル

秋田県立図書館(デジタルライブラリー)

児童書デジタル・ライブラリー

アジア歴史資料データベース

目録、索引等

サイト情報

調べ物に便利な情報、参考情報

検索結果一覧

簡易／連想検索

キーワード

検索先毎の件数

横断検索は200件、その他は1000件まで一覧表示出来ます。

並び替え 1ページ最大表示件数

印刷用表示

1ページ目(19件中)

選択した語で検索(OR検索) >>

- 「終戦の詔勅」を見たい。**

『国史大辞典』に「しゅうせんのかうしやく 終戦の詔書」の項があった。写真が掲載されていた。また、アジア歴史資料センター。

[レファレンス協同データベース](#) [提供元詳細画面](#) [関連情報リンク](#)
- 昭和史【録音資料】完全版 第18巻 むすび 昭和史二十年の教訓 半藤一利【述】**

録音ディスク1枚(24分):CD 発売番号: ANOC-9041 収録:2003年4月-12月 内容:
(三百万の死者が語りかけてくれ...)

エニー

[NDL蔵書目録\(和図書・和雑誌\)](#) [提供元詳細画面](#) [関連情報リンク](#)
- 天皇の玉音放送 小森陽一著**

293p 20cm CD1枚(12cm):終戦の詔書

小森 陽一(1953-)

五月書房

[NDL蔵書目録\(和図書・和雑誌\)](#) [提供元詳細画面](#) [関連情報リンク](#)
- 大東亜戦争陸海軍の部隊長 藤沼清輔編**

119p 図 21cm 部隊長会議の写真・陸大卒業記念写真・開終戦の詔書・戦争史年表・部隊名・配置状況・部隊長艦長名・位階

藤沼 清輔
日本勲章章協会

[NDL蔵書目録\(和図書・和雑誌\)](#) [提供元詳細画面](#) [関連情報リンク](#)
- 天皇の玉音放送 小森陽一 著**

293p 20cm CD1枚(12cm):終戦の詔書

小森 陽一(1953-)

五月書房

[NDL蔵書目録\(和図書・和雑誌\)](#) [提供元詳細画面](#) [関連情報リンク](#)

詔書

Wikipediaで検索

<<統制語>>

詔勅

<<読み等>>

ショウチョク

Syoutyoku

<<類義語>>

詔書

勅書

勅語

<<下位語>>

宣命

教育勅語

<<関連語>>

天皇

公文書

- 各アーカイブの検索結果を一覧で表示
- 入力キーワードの関連語で再検索も
 - 上位、下位、関連語
 - Wikipediaへのリンク
- 検索結果がなかった場合、Google, Yahoo!へのリンクも

- 入力キーワードの追加により検索範囲を拡大
- 情報の種別により絞込みも可
- 検索結果がなくても次の手がかりへ案内

詳細検索：きめ細かな検索オプション

詳細検索(キーワード検索)
各項目に入力されたキーワードで検索します。

すべて 部分一致 ▼

AND ▼ タイトル 部分一致 ▼

AND ▼ 作成者 部分一致 ▼

正式発行日 ~ 「年月日(yyyymmdd)」の形式で入力して下さい

ISBNコード

分類(NDC) 選択

分類(NDLC) 選択

1ページ最大表示件数 並び替え

▲検索オプション設定を開じる。

以下の初期設定はユーザの所属するグループの設定となっております。

■検索先
(選択しない場合はすべて)

すべて

本文、デジタル画像等(一次情報)

目録、索引等

サイト情報

調べ物に便利な情報、参考情報

■作成者種別
(選択しない場合はすべて)

民間の機関等 美術館・博物館 教育・研究機関 図書館

地方公共団体 国の機関等 個人

■表記揺れの吸収

ひらがな/カタカナ表記揺れも吸収

異体字表記揺れも吸収

単語レベルの表記揺れも吸収

■推定語の利用 機械的に推定したメタデータ(よみ、NDCなど)の検索対象としての利用可否

利用する 利用しない

- 詳細項目指定
- 種別選択
- 検索先選択

- 表記ゆれ、推定の選択

連想検索：文章との関連度で検索

- 入力された文章と関連性の高いと思われるコンテンツを検索

- Wikipediaの「天文学」の解説ページから文章を入力してみると

The screenshot shows the PORTA search engine interface. At the top, there's a search bar and navigation tabs. Below, there are search results for the keyword '天文学'. The results are listed in a table-like format with columns for item number, title, author, and publisher. On the right side, there's a sidebar with a '関連キーワード' (Related Keywords) section containing a list of related terms with checkboxes.

Item No.	Title	Author	Publisher
1	岩波講座物理学及び化学 宇宙物理学 1 A-2 C 岩波書店編	14冊(合本2冊) 22cm 内容:宇宙物理学 1 A 大気物理学(藤原暁) B 空中電気学(執山大三) 気象光学(藤原...)	岩波書店 岩波書店
2	天文宇宙物理学総論 第1部 第3編 上 荒木俊馬編	282p 22cm 古典天文学 天体力学 荒木俊馬, 清水高一共著 コチン テンモンガク テンタイ リキガク 荒木俊馬(1897-1978) 宇宙物理学研究会	NDL蔵書目録(和図書・和雑誌) 提供元詳細画面 関連情報リンク
3	天文宇宙物理学総論 第1部 第3編 上 荒木俊馬 再版	282p 22cm 古典天文学 天体力学 荒木俊馬, 清水高一共著 コチン テンモンガク テンタイ リキガク 荒木俊馬(1897-1978) 宇宙物理学研究会	NDL蔵書目録(和図書・和雑誌) 提供元詳細画面 関連情報リンク
4	日本の物理学100年とこれから 観測で探る天体・宇宙の物理--見えない宇宙の観測	雑誌名: 日本物理学会誌 巻号・年次: 60(5) (8月号 649) [2005] 掲載ページ: 337~343 松岡 勇 日本物理学会	NDL雑誌記事索引 提供元詳細画面 関連情報リンク
5	新天文学講座 第13巻 新版	260p(図版共) 22cm 天体の位置観測 清水高編 テンタイ ノ イチ カノンク 内容:天体位置の子午線観測(中野三郎)...	恒星社厚生閣 提供元詳細画面 関連情報リンク
6	新天文学講座 第15巻 新版	262p(図版共) 22cm 内容:天体写真測光法(高藤国治) 太陽分光(末元善三郎) 天体光電測光(大沢清輝)...	提供元詳細画面 関連情報リンク

汎用連想計算エンジン(GETA)

関連情報リンク(リンクリゾルバ)

★関連情報リンク

■他のサイトで検索する

選択されたコンテンツの書誌情報で以下のサイトより検索します。

項目名

- NDL総合目録ネットワーク
- Amazon.co.jp
- Kinokuniya BookWeb
- オンライン書店 ビーケーワン(bk1)
- Google ブック検索
- Google Scholar
- Google
- Google イメージ検索

- 検索結果から、関連ウェブサイトや資料の入手場所へナビゲート
- 選択されたデータのタイトルやISBN等を用いて、任意のサービスへ

分類検索：カテゴリをたどって検索

分類検索 (NDC)

日本十進分類法(NDC)のカテゴリをたどりながらデータを検索します。

機械推定した分類を

キーワード検索を行うことによって、現在選択されているカテゴリに属するデータを絞り込むことができます。

0 総記	98438
00 総記 (26051)	
01 図書館, 図書館学 (9269)	
02 図書, 書誌学 (20874)	
03 百科事典, 用語索引 (3291)	
04 一般論文集, 一般講演集, 雑著 (25009)	
05 逐次刊行物, 一般年鑑 (1288)	
06 団体, 博物館 (2321)	
07 ジャーナリズム, 新聞 (3168)	
08 叢書, 全集, 選集 (7076)	
09 貴重書, 郷土資料, その他の特別コレクション (91)	
1 哲学	177829
2 歴史	445730
3 社会科学	675198
4 自然科学	242552
5 技術	281422
6 産業	210893
7 芸術	356452
8 言語	65327
9 文学	534466

- 日本十進分類法 (NDC) 第三階層
- 国立国会図書館分類表 (NDLC) 第二階層
- 最上位の階層を選択した時点でキーワードを用いて絞り込み

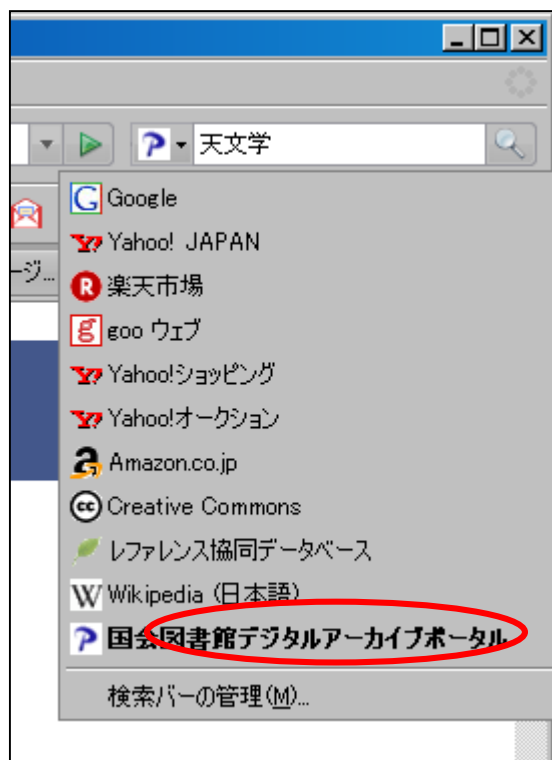
辞書検索：相関を視覚的に表示

上位語
 下位語
 関連語

統制語: 情報処理
 よみ: ジョウホウシヨリ
 Zyouhouisyori
 上位語: 情報活動 ジョウホウカンドウ Zyouhoukatudou
 下位語: データ処理 データシヨリ Detasyori
 画像処理 ガンシヨリ Gazousyori
 音声処理 オンセイシヨリ Onseisyori
 パターン認識 パターンニンシキ Patanninsiki
 分散処理(コンピュータ) ブンサンシヨリ(コンピュータ) Bunsansyori(konpyuta)
 並列処理(コンピュータ) ヘイレツシヨリ(コンピュータ) Heiretusyori(konpyuta)
 情報検索 ジョウホウケンサク Zyouhoukensaku
 数式処理 スウシキシヨリ Suusikisyori
 関連語: コンピュータ教育 コンピュータキョウイク Konpyutakyouiku
 アルゴリズム アルゴリズム Arugorizumu
 コンピュータ コンピュータ Konpyuta
 デジタルアーカイブ デジタルアーカイブ Dezitaruakaibu
 分類(NDC): 007.6
 分類(NDLC): M151
 典拠ID: 00575015
 チェックした辞書語で書誌の検索をします。

- 上位・下位・関連語でキーワード検索が可能
- 今後2階層まで表示して自由に動かせるようにする予定
- 例えば「情報処理」で検索してみると...

ブラウザのツールバーからの検索



Firefoxの検索バー



Googleのツールバー

- PORTAのページへ行かなくても検索が可能
- Googleパーソナライズのカテゴリとして実装予定

検索結果を利用者が加工して再利用

- 検索結果一覧のXMLファイルの提供
 - 現在は印刷イメージのみ
 - XML文書の形での提供も準備中
- 新着、更新情報のRSSの提供
 - PORTAのサイトに行かなくても、RSSリーダーで、登録されたメタデータを受け取ることができる

キーワードランキング、アクセスランキング

キーワードランキング

検索に多く使用されたキーワードのベスト20位を表示します。

折れ線(赤色)は日単位の使用回数、縦棒(水色)は週単位の使用回数累計を表示します。

集計期間: 2007/12/28 - 2008/01/28

順位	キーワード	推移
第1位 (88回)	歴史	
第2位 (60回)	社会科学	
第3位 (57回)	哲学	
第4位 (48回)	文学	

よく使われた

- 検索キーワード
- 検索されたコンテンツ

ユーザ登録
すると

- ユーザグループの選択
 - 一般、図書館員、理系研究者、文系研究者、子どもの5種類から所属するユーザグループを選択
 - 各グループに応じた検索条件(検索対象、分類等)があらかじめ設定
- デザインの選択
 - 6種類の画面デザインから、好みのデザインを選択
- 検索オプションの初期設定
 - 詳細検索等の検索オプションのうち、検索対象、分類等の絞込み条件設定を保存
- ダイレクトメニューの編集
 - 「ダイレクトメニュー」に表示するメニュー項目の選択・追加編集

お気に入り機能(ブックマーク)

私の **みんなの** ブックマーク

●キーワードで検索:

●タグ一覧から参照

●分類(NDC)から参照

すべて	総記	哲学	歴史	社会科学	自然科学	技術	産業
29 件	5 件	0 件	0 件	0 件	0 件	2 件	0 件

タイトル: [RSS\(RDF Site Summary\)を活用した新たな図書館サービスの展開—OPAC2.0へ向けて](#) 2users
 NDLL雑誌記事索引
 コメント: [図書館]
 タグ: 図書館
 登録日付: 2007/11/27
[ブックマーク編集](#)

タイトル: [E708 \(No.116\) - 学術機関リポジトリの課題と次なる戦略は? <報告>](#) 1user
 カレントアウェアネス
 コメント: [図書館]
 タグ: 図書館
 登録日付: 2007/11/26
[ブックマーク編集](#)

タイトル: [Apache Lucene入門 Java・オープンソース・全文検索システムの構築 関口宏司 著](#) 2users
 NDLL蔵書目録(和図書・和雑誌)
 コメント: [情報技術]
 分類(NDC): 総記
 タグ: 情報技術
 登録日付: 2007/11/13
[ブックマーク編集](#)

タイトル: [CA1624 \(No.291\) - 次世代の図書館サービス? —Library 2.0とは何か / 村上浩介](#) 1user
 カレントアウェアネス
 コメント: [情報技術] [図書館]
 タグ: 図書館
 登録日付: 2007/11/06
[ブックマーク編集](#)

- お気に入りのデータを、コメントをつけて保存
- みんなで共有(ソーシャルブックマーク機能)

PORTAで検索可能となるためには・・・

- PORTA等の他のシステムから、検索できるようにすること
- 共通仕様及び連携に関するガイドラインの実装
 - メタデータをPORTAで収集(ハーベスト)または横断検索するため
 - メタデータ要素通信プロトコル等を共通仕様及びガイドラインで規定
 - 各システム側で仕様に沿った実装を行う ⇒ PORTAの検索対象としてシステムの連携を行うことが可能となる
- データプロバイダディレクトリへの情報登録(登録機能準備中)
 - PORTAの「データプロバイダディレクトリ」にシステム連携に必要な情報を登録
 - (データプロバイダ名称、機関名、連携方法、リクエスト先URL等)

※ データプロバイダとは:コンテンツもしくはメタデータを提供するシステム。何らかのインタフェースを通じて、保有データ等を外部システム等へ提供可能なシステムのこと。

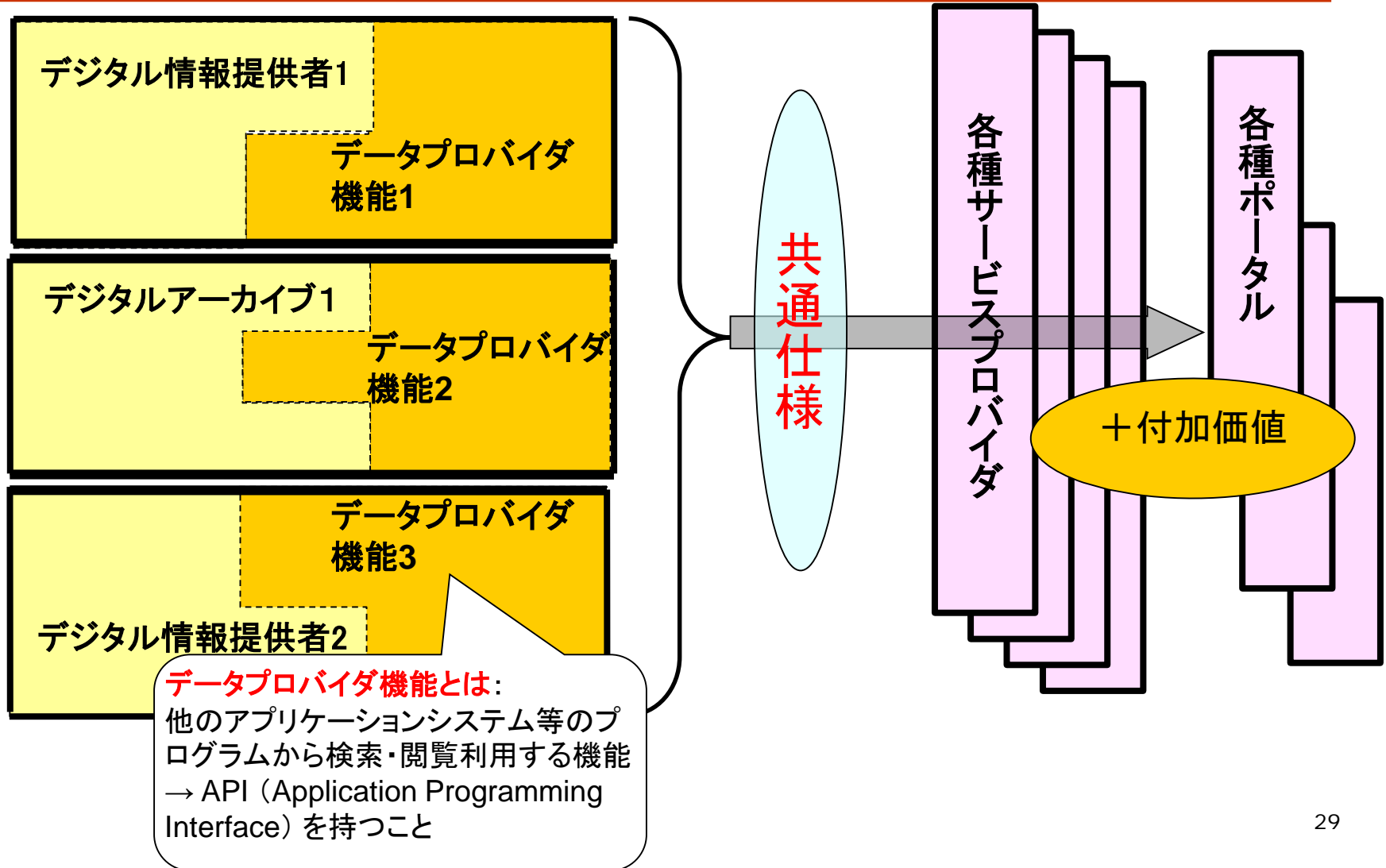
PORTAの検索機能を活用するためには・・・

- NDLは、外部システム等がPORTAの検索機能をシステムの的に活用できるようにしたインタフェースを提供

- PORTAが提供しているインタフェース（提供準備中）
 - 検索用としてSRW、OpenSearch、OpenURL、（Z39.50は準備中）
 - 将来的にデータ収集用としてOAI-PMHなど

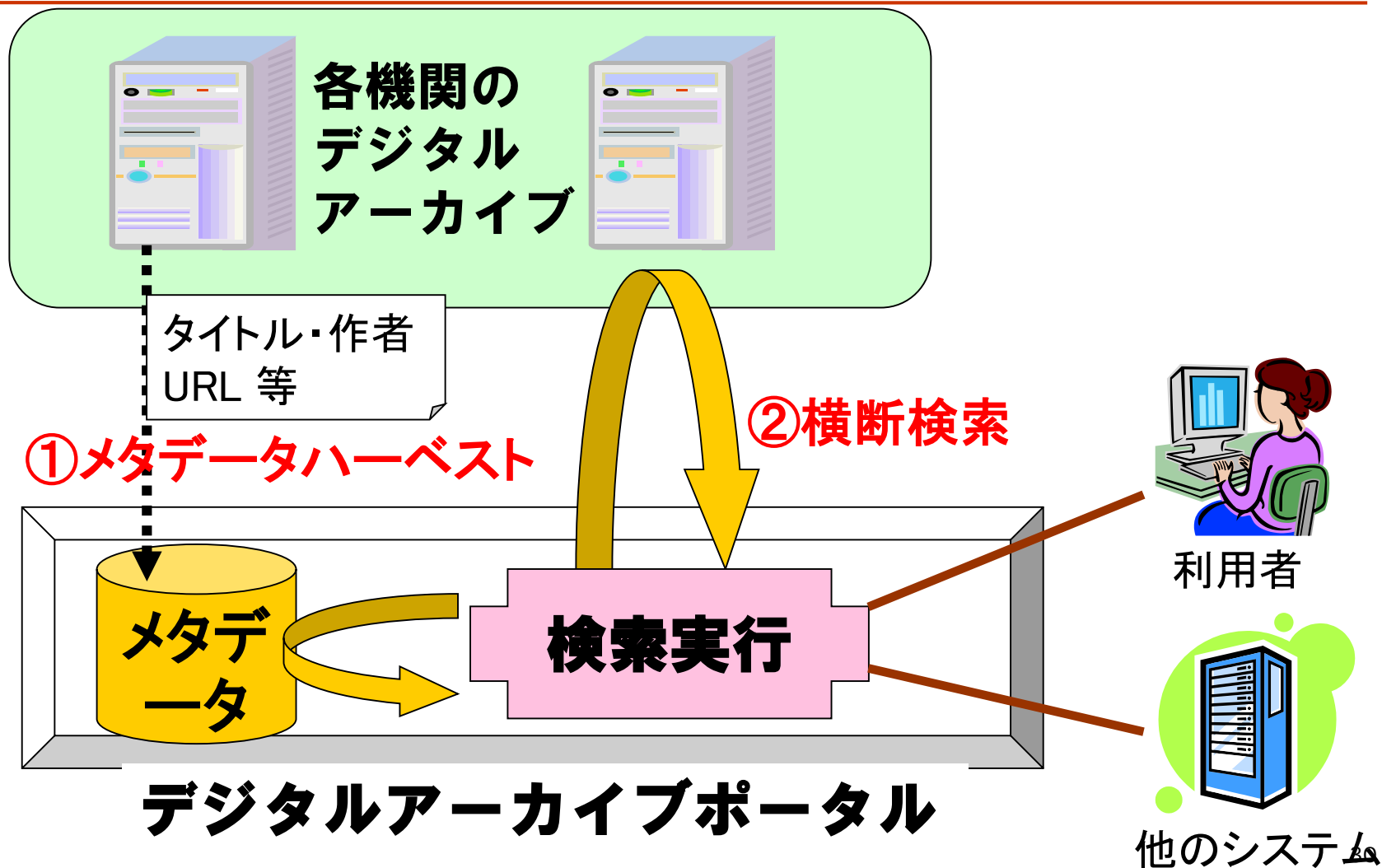
統合検索の仕組み1

データ提供側で共通インターフェースを !!



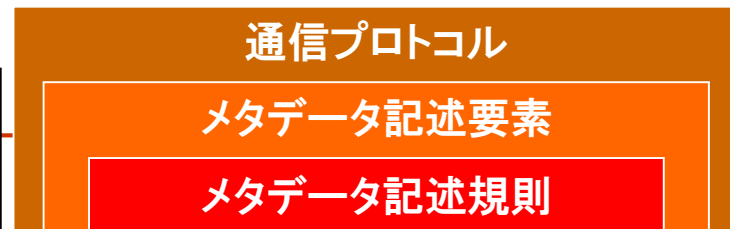
統合検索の仕組み2

メタデータ収集と横断検索



具体的な連携のための共通仕様は

		メタデータ形式
コンテンツ	MODS系	NDL-DAメタデータスキーマ(METS,MODS準拠)
	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
	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
サイト情報もしくはサイト内基点情報	OPML形式(RSSリスト)	
	Google SiteMap形式(適用検討中)	
	UDDIディレクトリ形式(拡張仕様)	

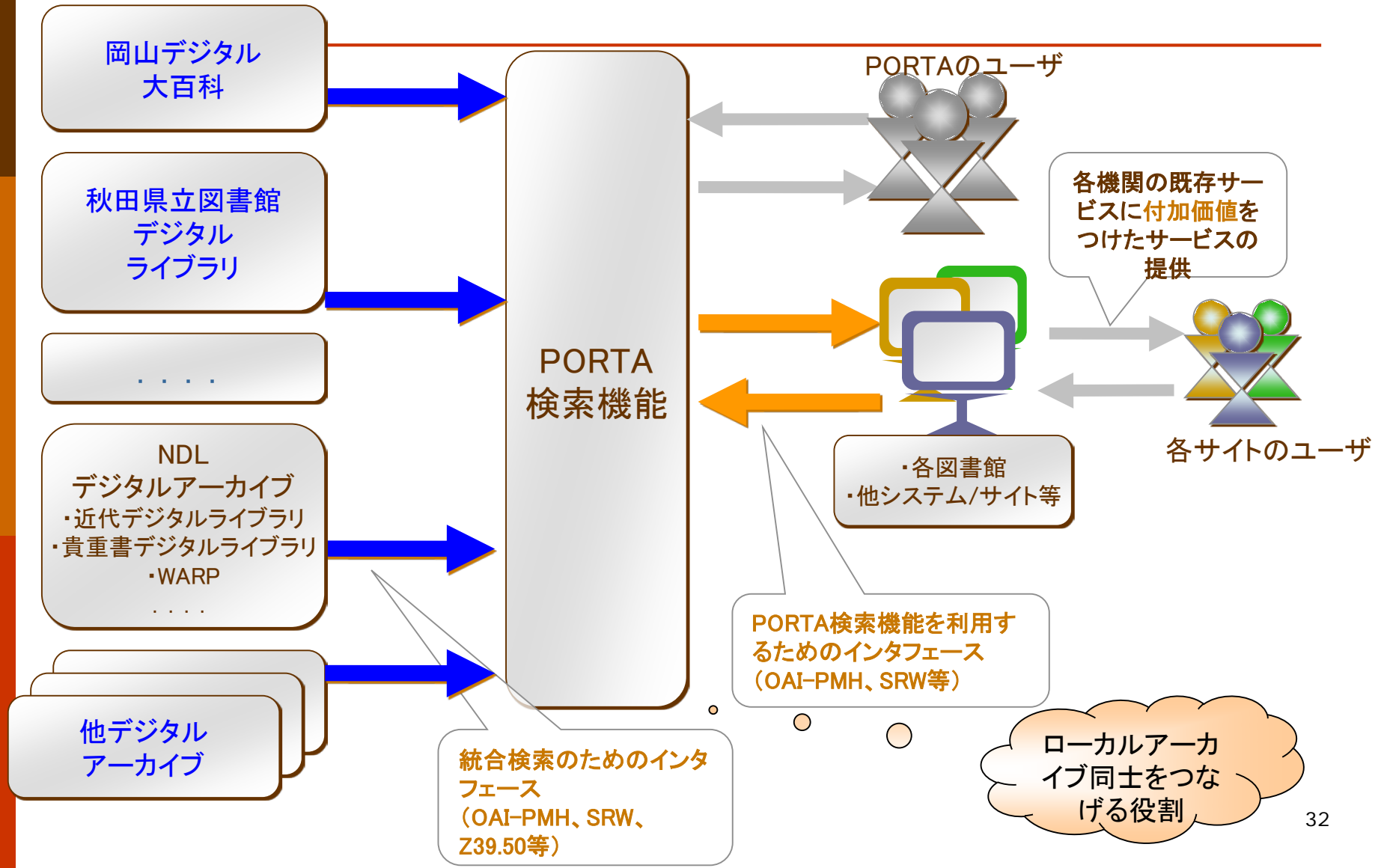


通信プロトコル	
ハーベスト	OAI-PMH
	RSS
横断検索	SRW(SRU/SOAP)
	Z39.50
	OpenSearch
リンクリゾルバ	OpenURL
サービス連携	WebサービスAPIによるシステム連携
	UDDI

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)

PORTAでの統合検索と 各機関への統合検索機能の提供



PORTA公開時の反応・課題

□ ブログ、レポート

- 国立国会図書館がソーシャルブックマークを始めると誰が予想したのか。(図書館退屈男)
 - http://toshokan.weblogs.jp/blog/2007/10/post_4226.html
- 5年の歳月が育んだもの ―荒野の日々を超えて(岡本真)
 - http://www.maruzen.co.jp/business/edu/lib_news/pdf/library_news150_04-05.pdf
- 各Blogで. . . .

□ 意見等

- 操作性が悪い、不安定、レスポンスが悪い、
- NDLのサービスらしくない. . .

現状の最大の
の問題

褒められて
いる??

統合検索のために苦労した点

検索先がまだ少ない理由

□ 統合検索できることのメリットの理解

- 各機関のコンテンツの利活用が促進されることの理解
- 利用者の利便性が向上することの理解
 - サービスが横取りされると誤解される
 - アクセスが増えることを不安視する

□ 標準プロトコルの実装

- ウェブページに貼り付けて、データベースとして検索できる形になっていない
 - コンテンツに対応したメタデータがない
- データベース化されていても、外部提供インタフェースを持っていない
- 外部提供インタフェースを実装する費用がない

□ メタデータのマッピングの調整

- 同じ要素を使っている、記述規則が異なり、同じ内容として認識できない
 - ダブリンコアをベースにした記述要素を使っている、使い方がそれぞれまちまち
- 検索機能で表記のゆれはある程度カバーできるが、横断検索の場合は、システムに依存

調整には、膨大な時間を費やしている⇒汎用的なマッピング、記述規則等のルールが必要

各図書館のデジタルアーカイブの構築を支援

今までの
経験を教
訓に

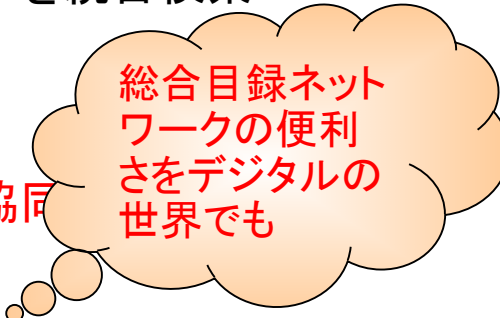
- 各図書館のデジタルアーカイブ構築を支援
 - 「デジタル化の手引き」(公開中、改定準備中)
 - 各図書館での郷土資料等のデジタルコレクションの構築
 - 「デジタルアーカイブ構築の手引き」(計画中)
各図書館でのデジタルアーカイブの構築のための手引き
 - 「共通仕様及び連携に関するガイドライン」(公開準備中)
 - 各図書館のDAに標準的な連携機能の実装
 - 「メタデータスキーマガイドライン」(策定中)
 - 標準的なメタデータ記述要素、記述規則
 - 「各図書館でのデジタルアーカイブの汎用パッケージ」の提供も検討中
 - たとえば、Dspace、XooNipsをベースに
 - NDLデジタルアーカイブからの提供も可能
 - 各図書館でデジタルアーカイブを構築することが困難な場合
 - NDLでの長期保存だけでなく、提供も
- 各図書館での検索サービスの構築を支援
 - PORTAの「外部インタフェース仕様書」(公開準備中)
 - 各図書館システムでPORTAの検索機能を利用するための機能の実装

ベンダーの協
力が必要

組織の壁を越えて

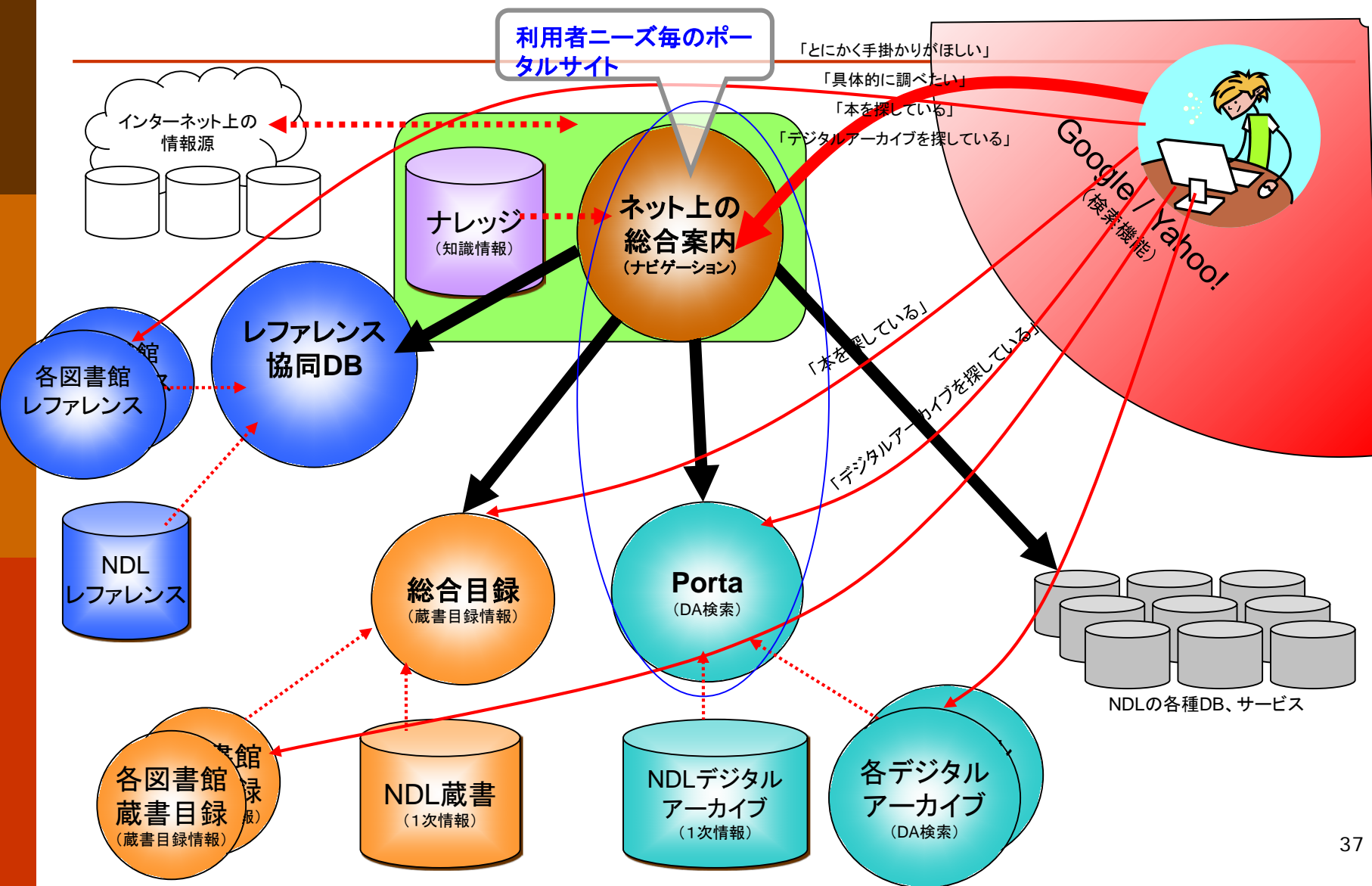
網羅的なデジタルコレクションの提供を目指して

- 館種、官民、分野、時代の枠を越えて、利用者へコレクションを提供
- コレクションの種別ごとのポータル提供
 - 学術著作物、自然科学系、人文科学系、、、
 - 政府および公的機関情報
 - 明治期刊行物、江戸期以前の貴重書、各地域の郷土資料
 - 京阪奈地域の図書館、資料館等のコレクション構築も想定
 - 電子書籍、電子ジャーナル、音楽・映像関連アーカイブ
- MLA連携
 - 美術館、図書館、公文書館、博物館で、同種のコレクションを統合検索
- 地域情報ハブとしての図書館が連携
 - 全国規模での図書館情報の共有利用
 - 総合目録ネットワークシステム(蔵書目録)、レファレンス協同(レファレンス事例)に加えて、デジタルコンテンツも
 - デジタル岡山大百科のような形で連携を



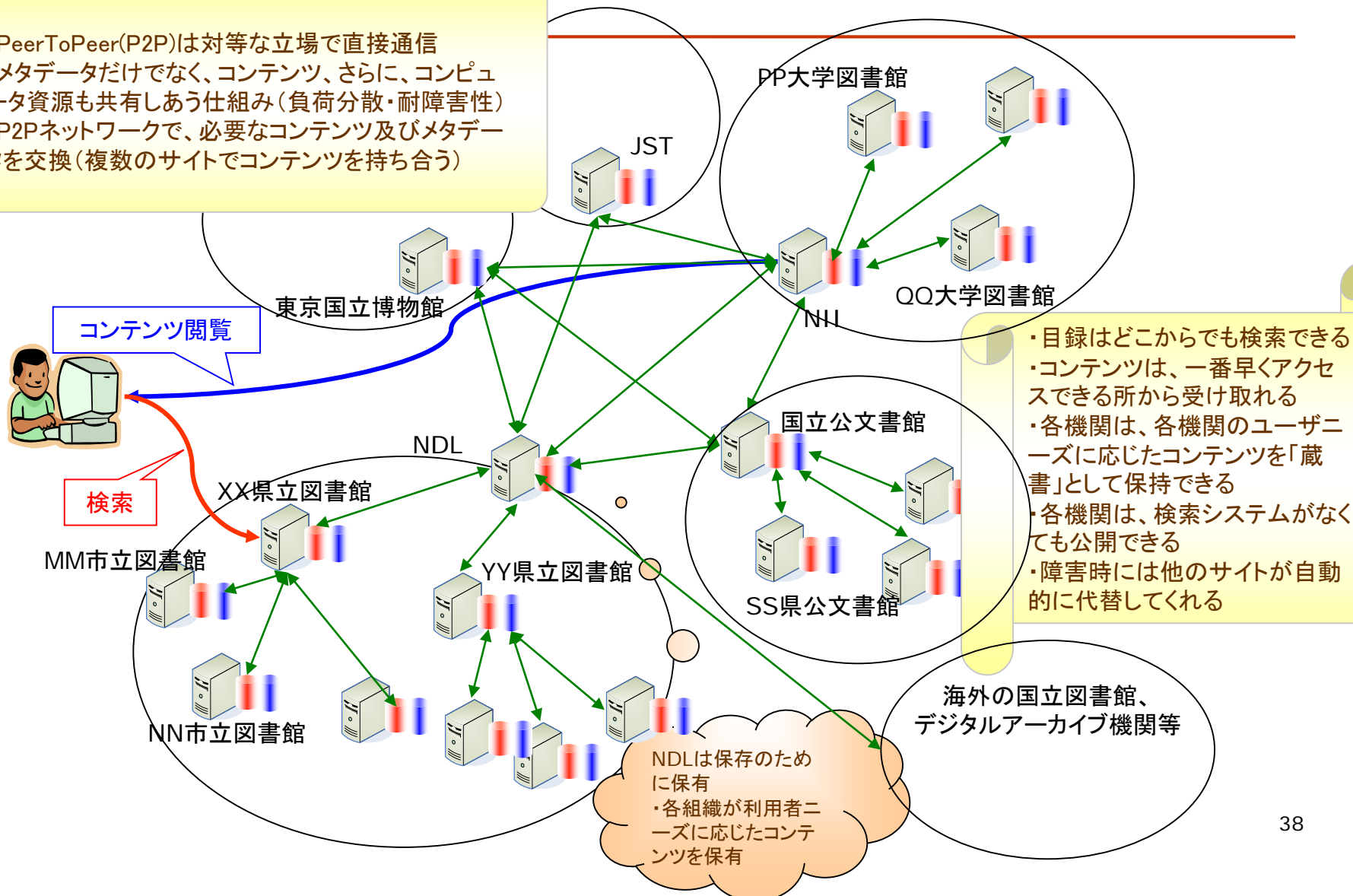
総合目録ネットワークの便利さをデジタルの世界でも

紙資料とデジタルコンテンツを含めた統合検索



5~10年後にはこんなイメージも (P2P型ネットワークによる分散アーカイブ)

- PeerToPeer(P2P)は対等な立場で直接通信
- メタデータだけでなく、コンテンツ、さらに、コンピュータ資源も共有しあう仕組み(負荷分散・耐障害性)
- P2Pネットワークで、必要なコンテンツ及びメタデータを交換(複数のサイトでコンテンツを持ち合う)



終わりに

- 情報を現在及び将来に亘って活用できるように
 - 各機関でデジタルコレクションを構築
 - 統合検索できるように公開を...
 - NDLは、コンテンツの長期保存と、提供を支援します。

国立国会図書館: <http://www.ndl.go.jp>

デジタルアーカイブポータル: <http://porta.ndl.go.jp>

国立国会図書館デジタルアーカイブポータルの現状と今後
(情報管理, Vol.49 (2006) No.6 pp.313-323)

ご静聴ありがとうございました

参考資料

ーデジタル情報も含めた統合検索サービスを目指してー



(現状のサービス)

近代デジタルライブラリー



Digital Library from the Meiji Era

◆当館所蔵の

明治期(1868-1912)及び
大正期(1912-1926)刊行
図書を提供する画像データ
ベース

約14万3千冊

【約9万7千タイトル】

<http://kindai.ndl.go.jp/>



『浮草』

『京都遊覧案内』

(現状のサービス)

WARP (インターネット情報選択的収集事業)

WARP 国立国会図書館 インターネット情報選択的収集事業
- National Diet Library, Japan - Web Archiving Project

主事業の紹介 ▶ 推薦のお願い ▶ ヘルプ ▶▶ 国立国会図書館

お知らせ

- 2006年8月 パンフレットを掲載しました。詳しくは[こちら](#)をご覧ください。
- 2006年7月 本格事業化にともない、デザインを一新しました。

簡易検索 書誌・本文検索 テレクトリ検索

簡易検索
(書誌情報のタイトル、編者、出版者・公開者、起点URLと本文情報の本文、ページタイトル、URLが検索できます。)

検索 クリア

キーワード

コレクション種別

- 国の機関 都道府県 政令指定都市 市町村合併
- 法人・機構 大学 イベント 電子雑誌
- その他

NDL資源タイプ

- 電子雑誌 サイト

収集日付範囲

年 月 年 月

検索 クリア

All Rights Reserved (C) National Diet Library, Japan 2002-2007

<http://warp.ndl.go.jp/>

- ➔ 『[2002 FIFA WORLD CUP](#)』
([2002/10/28](#))
- ➔ 『[Dinosaurs](#)』 ([2006/06/19](#))

- ◆ WWWの情報資源を選択的に収集・蓄積(2002年～)
- ◆ 2006年に本格事業化
- ◆ 政府、地方自治体、市町村合併、イベント、電子雑誌など、約3,600タイトル【約16,000件】を公開

(現状のサービス)

WARP 収集コンテンツ

Collection of WARP

コレクション		タイトル数	個体数	容量(GB)
サイト	国の機関	51	373	1,153
	都道府県	32	140	872
	政令指定都市	12	55	425
	市町村合併(合併協議会を含む)	1,705	6,191	1,299
	法人・機構	151	947	976
	大学	118	171	1,808
	イベント(2002FIFAワールドカップetc.)	31	116	12
電子雑誌		1,505	8,017	701
合 計		3,605	16,010	7,246

(平成19年12月末) ⁴³

(現状のサービス)

Dnavi(データベース・ナビゲーション・サービス)

国立国会図書館 データベースナビゲーションサービス
National Diet Library, Japan Database Navigation service

■ 本事業の紹介 ■ ヘルプ >> 国立国会図書館

お知らせ

簡単検索 キーワードを入力してください。 検索 クリア

詳しく検索 色々な項目から検索できます 詳細検索へ→

テーマ検索 探しているテーマを選んでください。

総記	哲学	歴史	社会科学	自然科学
技術	産業	芸術	言語	文学

電子化資料 館蔵品・所蔵品の一覧を表示します。 表示

All Rights Reserved (C) National Diet Library, Japan 2002-2006

<http://dnavi.ndl.go.jp/>

- ▶ 『上方浮世絵館』
- ▶ 『スラブ研究リンク集』

- ◆ WARPでは収集できない深層ウェブ(Deep Web)の入り口へ案内
- ◆ 日本国内のウェブ上のデータベースを幅広く収録
- ◆ タイトル、作成者、分類、内容等の様々な情報を付与
- ◆ 約10,000件を収録

貴重書画像データベース

- 和漢書、錦絵、絵
図
- 重要文化財
 - 平成12年3月から公開
 - 861タイトル
 - 約 37,000コマ



(現状のサービス)

電子展示会

「日本国憲法の誕生」

「描かれた動物・植物

—江戸時代の博物誌—」

「近代日本人の肖像」

「絵本ギャラリー」

等々10種



ブラウザ画面のスクリーンショット。URL: http://www.diet.go.jp/const/110n/110n003shohshi.html

日本国憲法の誕生

資料と解説

[資料と解説](#) > 第1章 戦争終結と憲法改正の始動

1-4 トルーマンの日本国民に対する声明 1945.5.8

資料名	日本国民諸氏 アメリカ合衆国大統領ハリイ・エズトルーマンより一書を呈す
年月日	
資料番号	憲政資料室収集文書 1313
所蔵	国立国会図書館
原所蔵	
注記	

ドイツが降伏した1945(昭和20)年5月8日にハリイ・トルーマン米大統領が発表した対日声明文は、トルーマンの写真付きで日本語に訳されどらとして投下された。トルーマンは、日本国民と軍部とを明確に区別しながら、日本軍が無条件降伏するまで攻撃を続けると警告した。同時に、日本軍の無条件降伏は、日本国民の「抹殺」や「奴隷化」を意味するのではなく、むしろ日本を「破壊の淵に誘引」している軍部の消滅、前線で戦う兵士たちの「愛する家族」のもとへの復帰、そして「現在の難難苦痛」の終わりを意味すると説いた。

世界の中のニッポン

Nippon in the world

日本の風景記憶

ウィーン万国博覧会

憲政資料 (日本近現代政治史料)

デジタルコンテンツの収集・蓄積・提供の まとめ

- 収集・蓄積(情報の利用を将来に亘って保証するために)
 - 消えてしまう前に、デジタルコンテンツを収集
 - 収集したデジタルコンテンツを将来に亘って利用できる形で保存
 - ロングテールの情報を含めたナショナルデジタルコレクションの構築
- 提供(Web2.0世代の情報提供を目指して)
 - デジタルコンテンツの統合検索は、メタデータの収集、横断検索もしくは、ウェブシステム連携で
 - 直感的な操作による検索
 - シングルソース・マルチユースの統合検索
 - 利用者の特性や、利用環境に応じた情報の検索→マイポータル機能
 - 利用者も含めた集合知の活用
 - CGM(Q&Aサイト、ユーザレビュー、ソーシャルブックマーク、Wiki等)
 - 辞書を活用したセマンティックウェブ技術の適用
 - 自動メタデータ付与、あいまい検索...

構築中のデジタルアーカイブシステム (DA)の機能構成

構築中のデジタルアーカイブシステム(DA)の機能構成

- デジタル情報の収集・組織化・保存・提供（開発中 平成22年3月運用開始予定）
 - 収集先選定・許諾機能
 - 著作物単位収集・組織化機能
 - 当館資料デジタル化コンテンツ
 - デジタルデポジット
 - （著作物単位、パッケージ系電子出版物、深層ウェブも）
 - サイト単位収集・組織化機能（サイト全体、特に表層ウェブ）
 - 長期保存システム及び電子書庫
- デジタルアーカイブポータル（平成19年10月公開）
 - メタデータハーベスト機能
 - 横断検索機能
 - 検索API提供機能（公開準備中）

統合検索に必要な技術要素

- 深層ウェブの表層化(検索エンジンがインデキシングしやすいように)
 - GoogleのSitemapsのような仕組み
 - 検索機能API
- 共通仕様の適用
 - 統合検索のための通信規約
 - OAI-PMH
 - SRW(SRU/SOAP)
 - RSS
 - Z39.50
 - 書誌情報及びメタデータの記述要素、記述規則
 - DC、MODS、.....
 - J/M⇔MODS、DC⇔MODS.
 - デジタル化仕様
 - テキスト、イメージ、音声、動画...

デジタルアーカイブ構築における 研究成果の活用の必要性

指数的に増大するデジタル情報

図書館側での問題

従来の人手に頼った収集確認や
組織化(書誌情報付与)では破綻し、
組織化が十分に行えなくなる

利用者側での問題

全文検索だけでは的確に絞り込めない。
また、検索語など適切な検索条件を与えるのは
一般的には難しい

ツールや手法の適用性の検討

- ・自然言語処理、機械学習
- ・Webコンテンツ解析
- ・セマンティックウェブ技術

「収集・組織化支援機能」

収集確認や組織化に要する人手の工数を削減し、
あわせて件名標目や典拠DB等の辞書類の増強

「利用者向け検索機能」

書誌情報を使った検索絞り込みや、
類似文書提示などによる検索拡大といった
様々な検索機能の提供や、
地図や時間軸での検索インタフェースの提供

デジタルアーカイブの構築において 今後適用すべき技術と進め方

□ 技術

- 収集を容易にする技術
 - 収集効率を高める技術(差分収集と再現技術を含む)
 - 収集品質を高める技術
 - 収集したサイトイメージから著作物を切り出す技術
- メタデータの付与を省力化する機能
 - 可能な限り自動化
 - セマンティックウェブ技術の活用
 - 将来的には、FRBRの概念を適用した組織化も
 - 自動化の精度が低いものに関しては、人手による確認作業を省力化できる機能
- 検索を容易にする技術
 - キーワードマッチングだけでなく、推定する技術も活用した検索技術
 - 検索、クラスタリング検索等

□ 進め方

- 国際標準、業界標準の積極的な適用
- 政府機関、民間を問わず、国内外の研究機関、研究者等と連携して調査研究
- 各機関の研究成果を組み合わせることで機能を実現
- まずは、海外ではIIPC、国内では研究開発を行っているJST、NII、NICT、AIST、IPA等の政府機関、大学の研究室との連携
- 有用な検索サービスを実施している商用ポータル機関(Google、Yahoo等)との連携

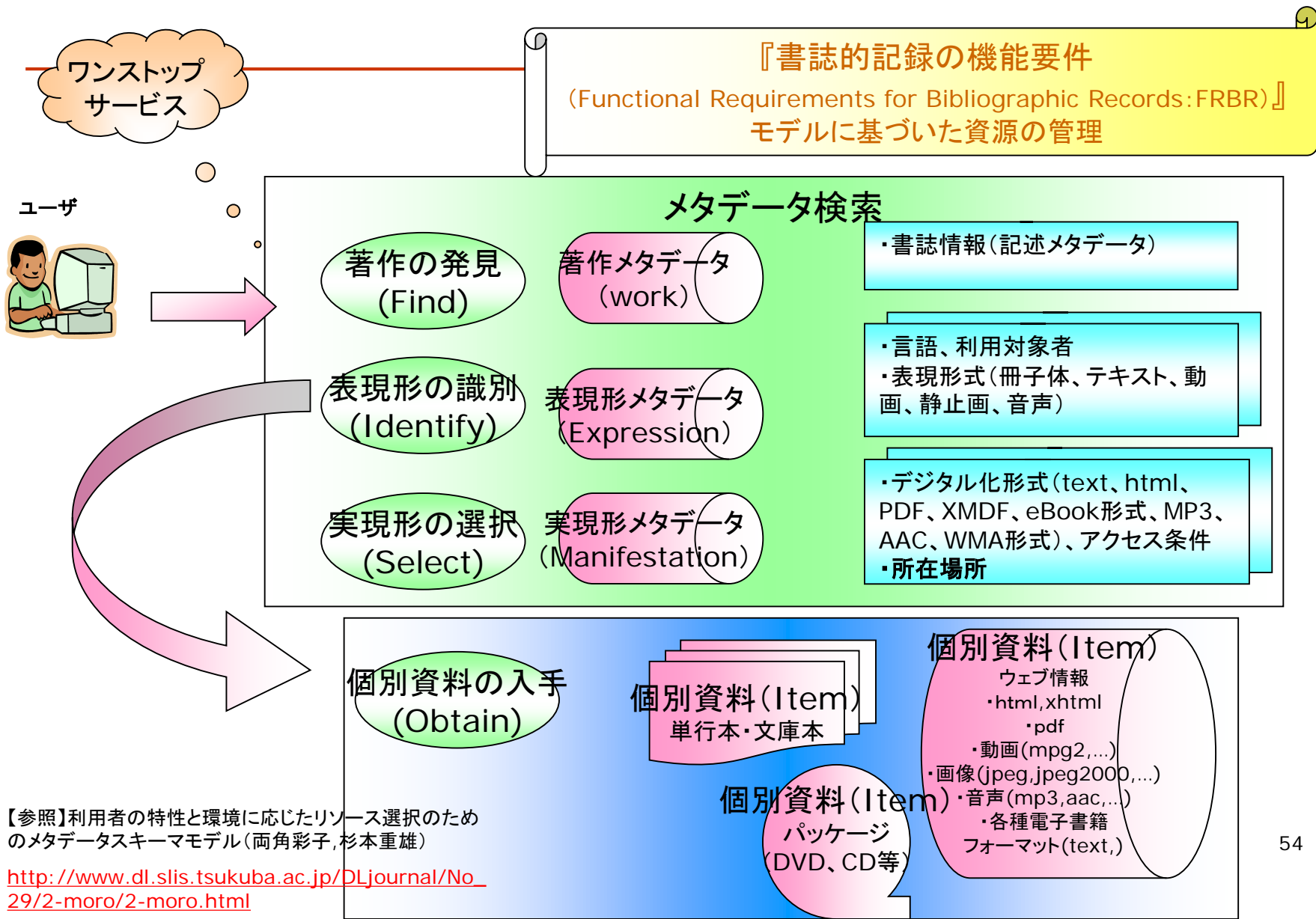
今後の図書館サービスの方向性(概念)

- 蔵書管理指向からサービス指向へ
 - 図書館システム内で利用者を囲い込まない
 - 他のシステムに任せられる部分は任せる、
 - 必ずしも図書館システムはトータルシステムでなくてもよい
- 利用者、関係機関との関係の緊密化
 - ユーザの情報利用行動の客観的把握
 - 利用者満足の向上のために、マーケティング手法の適用
 - 利用者 と 図書館蔵書のマッチングを支援
 - Google 等から図書館資料へのアクセスを可能に
 - 利用者個人への直接サービスを志向
 - サービスをパーソナライズ
 - 関係機関が連携・分担してサービスを提供

メタデータ、書誌データの統合検索から、意味的情報の相互利用による統合検索へ

- 集合知を利用するWeb2.0から、意味的情報の相互利用のWeb3.0へ
 - Web2.0: Blog、Forksonomy、Ajax、Mashup、Long Tail
 - Web3.0: RDF、SKOS、オントロジー、GRDDL、SPARQL
- 関係機関による共同構築、オープンソースの利用、マッシュアップによるサービス構築
- ボーンデジタルを含めた蔵書構築、総合目録
 - 統合から、ウェブサービスの協調によるサービス提供へ
 - 図書・雑誌における所蔵のほかに、電子情報資源のライセンス(利用権)等の管理も
 - メタデータを発生源に近い所で作成。さらにメタデータの自動付与技術の適用
 - 将来的には、目録を利用者側の視点から見直すモデル(FRBR)に基づくメタデータの再構築の検討も

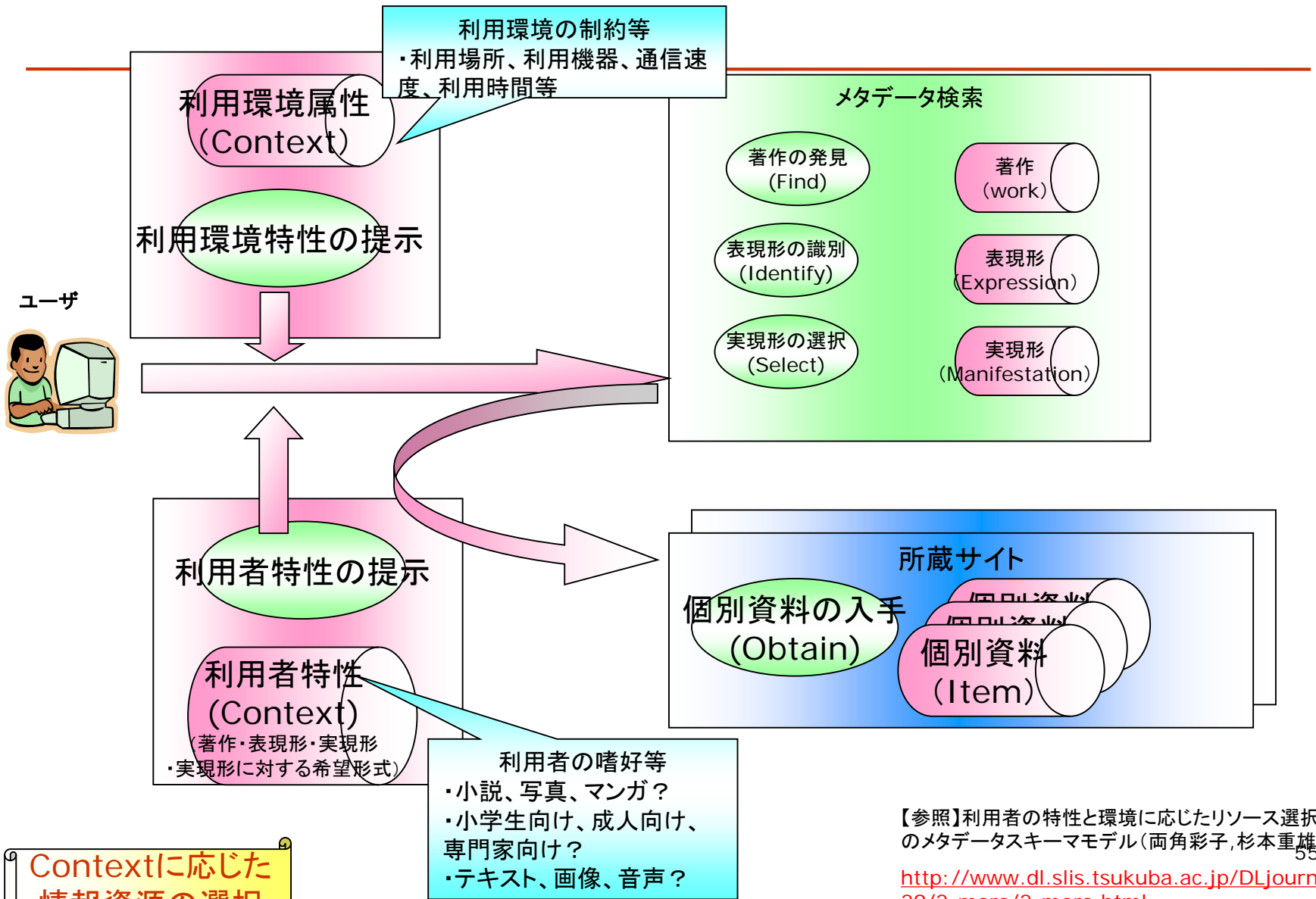
所在や資料の形式を問わないワンストップサービス ＜シングルソースマルチユースの統合検索＞



【参照】利用者の特性と環境に応じたリソース選択のためのメタデータスキーマモデル(両角彩子,杉本重雄)

http://www.dl.slis.tsukuba.ac.jp/DLjournal/No_29/2-moro/2-moro.html

利用者の特性や環境に応じた情報の選択 ＜マイポータル機能＞

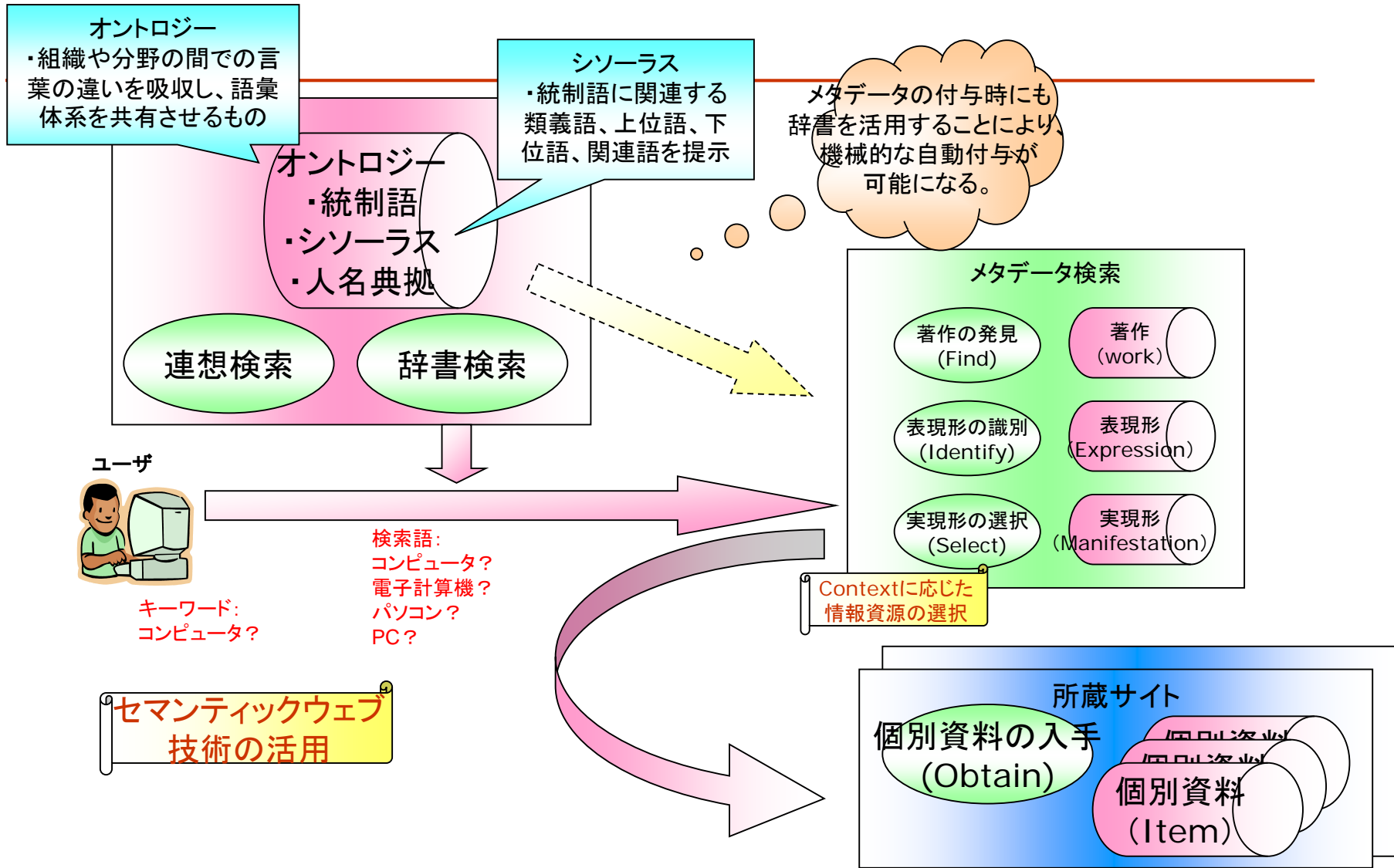


Contextに応じた
情報資源の選択

【参照】利用者の特性と環境に応じたリソース選択のためのメタデータスキーマモデル(両角彩子, 杉本重雄)

http://www.dl.slis.tsukuba.ac.jp/DLjournal/No_29/2-moro/2-moro.html

提供元の語彙の違いや関連語も含めた情報の選択 ＜辞書を活用した検索機能＞



地域情報ハブとしての図書館の利用イメージ

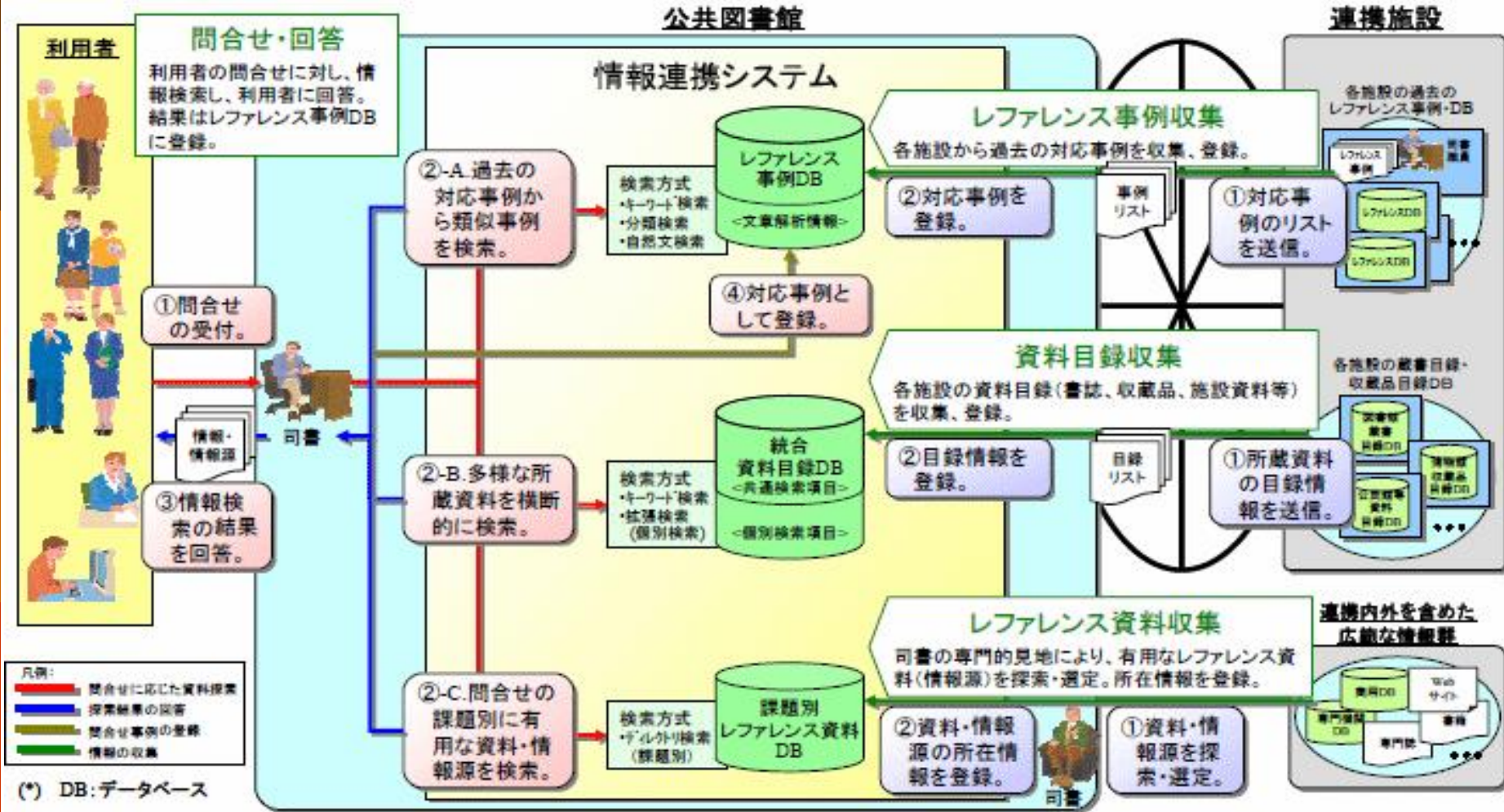


図 7 他施設・他機関と連携した情報連携システムの利用イメージ

❖ 引用【地域の情報ハブとしての図書館－課題解決型の図書館を目指して－ 第4章²-2:】

❖ http://www.mext.go.jp/a_menu/shougai/tosho/houkoku/05091401/017.htm

地域情報ハブとしての図書館に必要な機能

引用【地域の情報ハブとしての図書館－課題解決型の図書館を目指して－ 第4章－2:】

- 資料目録を総合的にデータベース化し、高度な情報検索を支援するための仕組み
 - 公共図書館及び他施設・他機関保有の資料を課題別に体系化する取組を進め、
 - その整理に従いメタデータを付与することによって、
 - 資料目録を総合的にデータベース化し、高度な情報検索を支援するための仕組み
- 司書の課題解決能力の向上と地域課題解決のノウハウの蓄積に資する仕組み
 - 司書のレファレンスに関する経験・ノウハウが集めたレファレンス事例をデータベース化し
 - 共有するための環境整備(課題別レファレンス機能等)を通して、
 - 司書の課題解決能力の向上と地域課題解決のノウハウの蓄積に資する仕組み
- 地域資料(郷土資料)の電子化と、地域のウェブ情報を含む電子資料のアーカイブ化
 - 将来にわたり公共図書館及び他施設・他機関の共有・活用に資するため
- 公共図書館における情報基盤の整備
 - 利用者の公共図書館利用環境の向上や、
 - ウェブ上からの公共図書館サービスの利用等へのアクセスを容易にするため

公共図書館としての 電子図書館サービスの構築コンセプト

- 資料提供のみならず、図書館としての付加情報の提供
 - インターネット時代にふさわしい、新たな付加価値を有するインタフェースを作り上げることが必要
 - (インターネット情報の活用、専門司書の暗黙知の形式知化、形式知化された知識の関連付け)
- 図書館サービスのインテグレーション
 - 他機関等と連携した図書館ナレッジの構築と統合検索機能
 - 利用者を情報そのものまでナビゲーション(情報の所在だけでなく、答えまで)
 - 利用者個人の要求に応じたサービスを提供するためのパーソナライズ機能
 - デジタルレファレンス、オンラインチュートリアル、サブジェクトゲートウェイ
- 媒体を問わず保有している資料のメタデータ・データベースの構築と提供
 - 電子コレクション構築とアクセス支援機能の橋渡し
 - (調べ方案内から、その一次情報へのワンストップ連携)
 - インタフェースの標準規格と要素技術の適用
- 電子保存のための技術及び企画、統合検索のための技術の適用
 - デジタル化技術、保存システム構築技術、日本語構文解析技術の適用等

連携に当たっての基本的な考え方 ナショナルデジタルコレクションの構築を目指す

□ 現状認識

- NDLは冊子体資料と同様にデジタルコンテンツに関しても網羅的に収集し保存する使命を持つ
- しかしながら、NDLのデジタルアーカイブは、国のデジタルアーカイブのOneOfThemにすぎない
 - NDLが全てのデジタルコンテンツを収集・保存できるとは思っていない。
 - NDLは冊子体資料の全てを保有しているわけではない。NDLが全てをデジタルコンテンツを収集することはできない
- 分散デジタルアーカイブを統合検索するサービスとして、PORTAもOneOfThemにすぎない

□ コンテンツを保有している関係機関と共同で、分散デジタルアーカイブの形での実現を目指す

- 分野、種別、地域毎に、分散デジタルアーカイブが構築されることが望ましい
- 分野・種別毎に中核となる組織が連携・分担して、それぞれの関係機関のデジタルアーカイブの構築を支援する形が望ましい
- 提供者にとって利活用が促進されることが重要
 - 利用者にとっても、様々な手段によりコンテンツに辿り着けることが、好ましいのではないか。

連携に当たっての基本的な考え方

NDLは何をすべきか

- 国のデジタルアーカイブ構築の一翼を担う
 - 国の有用なデジタルコンテンツの保存と利活用が促進されるための施策を推進する。
 - コレクションの利用を保証するために
 - 将来に亘って利用を保証するために、提供機関の要請に従って、必要に応じて収集し、長期的に保存する
 - 各機関のデジタルアーカイブの障害時に、必要に応じて、バックアップサーバとして機能できるようにする
 - 各機関、各ポータルが、利用者ニーズに応じて、様々な検索・閲覧サービスを提供できる機能を提供する
- 図書館のコンテンツの利活用を促進する
 - 自ら保有する資料のデジタル化を推進する
 - 公共図書館等のデジタルコレクション構築を支援する
 - 公共図書館等が保有する資料のデジタル化およびデータベース化を支援する
 - 公共図書館等が自らのコレクションに加えてNDLコンテンツも含めて電子展示会等のサービスを構築することを支援する
 - 同種のコンテンツを保有する機関との相互運用性を確保した形で推進する。

連携に当たっての基本的な考え方

PORTAが果たすべき役割

□ PORTAの位置づけ

- PORTAの大きな役割は、各機関のデジタルアーカイブに外部インターフェースが実装され、統合検索できるようにすることのインキュベーション。
- GUI機能
 - API機能を利用して、コンテンツの種別・特性を考慮せずに、全てのコンテンツを一元的に閲覧できるようにした、最もベーシックなサービス
- API機能
 - 各機関が、PORTAの統合検索APIを利用して、多くのDBを一度に検索できるようにして、新たなサービスを展開できることを期待。

□ PORTAが今後果たすべき役割は？

- PORTAはデータベースアクセスのゲートウェイの一つ。
 - 各機関のDBが外部提供APIを持つようになれば、誰でも統合検索サービス機能を構築できる。そのときには、役割は終わる？
- コンテンツ種別ごとに見せ方の工夫が必要
 - 最適化された検索・閲覧のGUIサービスは、センスのいいサービス組織に委ねたほうがいい？