

国立国会図書館における 情報探索サービスの新たな展開

平成21年10月30日
国立国会図書館
中山 正樹

今日のお話し

- 背景
 - 「知識はわれらを豊かにする」
- 現状の課題解決のために
 - 現状の問題の再認識
 - 情報の利活用のために実施すべきこと
 - 多様なニーズの認識
 - 情報の利活用のための仕組みの概念
 - クラウドの世界でのサービスの連携
 - 情報探索サービスの将来像
(クラウドの世界でのサービスの連携)
 - 利用の視点でのコンテンツの体系的整理
- これまでの経緯
 - 電子図書館中期計画2004
 - PORTAの現状
 - 経緯
 - 統合検索の概念
 - メタデータ収集・横断検索の概念
 - 連携のための共通仕様
 - メタデータの仕様の関係
- 当面の情報探索サービスがめざすもの
 - 情報探索サービスのカバレッジ
 - 情報探索サービスが当面目指す方向性
 - 24年度正式公開に向けて
 - 情報探索システムの機能モデル
(21FYプロトタイプ)
 - 情報探索システムの開発想定
 - サービス要件・システム化要件定義
 - 各種調査
 - 技術標準適用ガイドライン
 - 今年度構築システム
 - プロトタイピングの基本方針
 - プロトタイプシステム機能の構成
 - 構築・提供スケジュール (想定)
- 関係機関との連携
 - 大学図書館等との連携
 - 公共図書館等との連携
- 終わりに

「知識はわれらを豊かにする」を デジタルアーカイブ・情報探索サービスの観点からみて

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

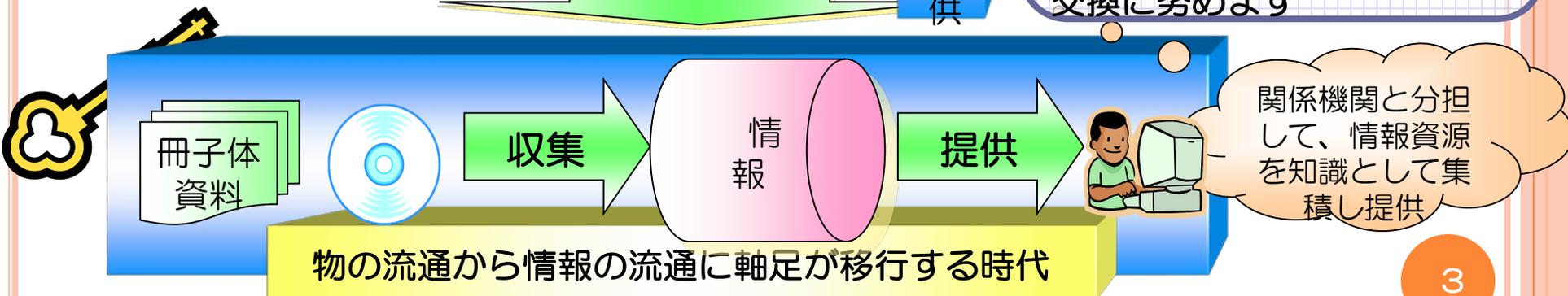
関係機関と協力して収集・提供

(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



現状の問題点の再認識

指数的に増大するデジタル情報

有用な情報は、民間、個人のデータベース、ブログ等に多く存在

情報探索行動の目的は、まさに問題・課題の解決

インターネットで見られないものはないも同然

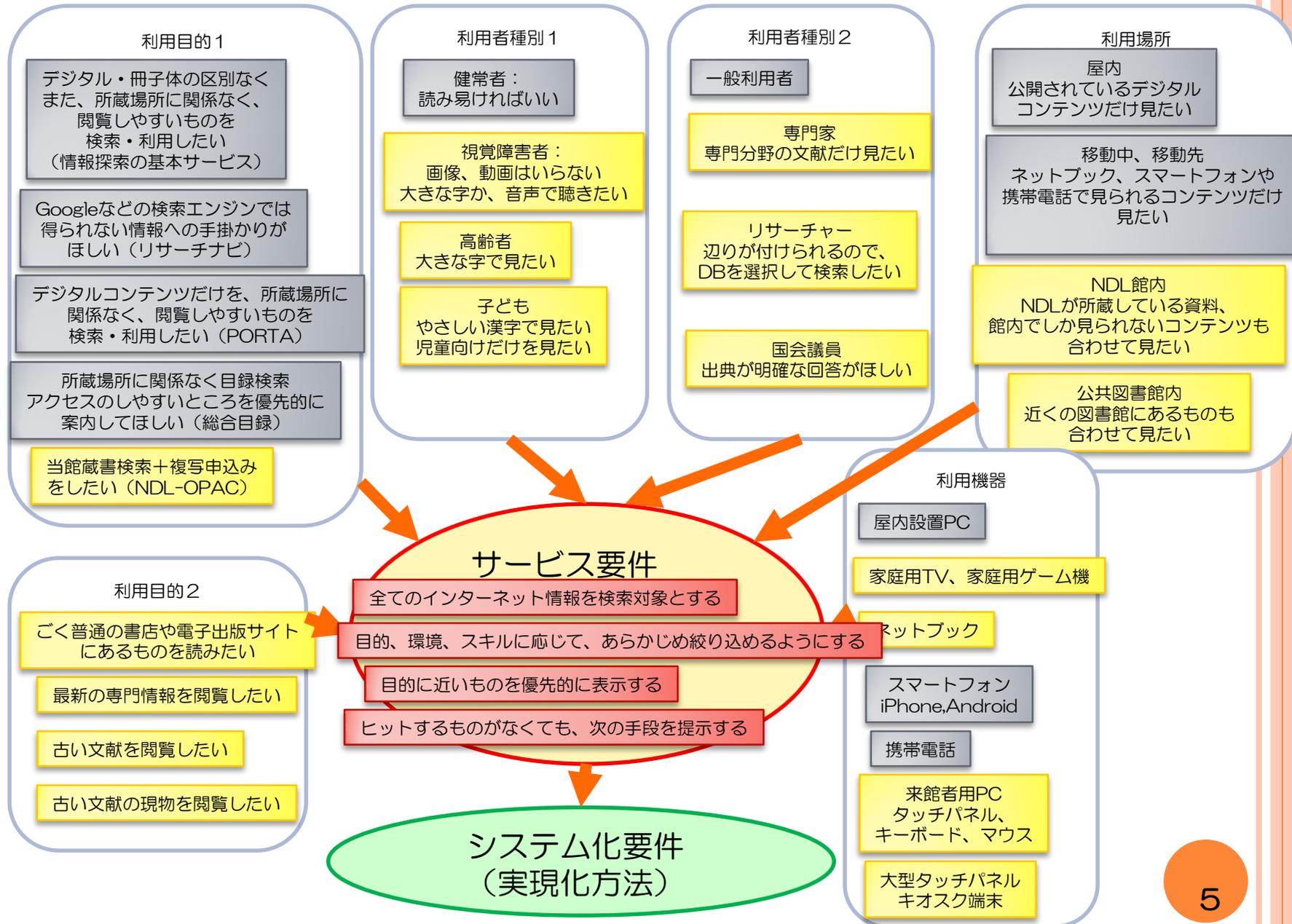
インターネット上の情報のすべてを収集し、提供することは不可能

人の知の集合は専門家をしのご

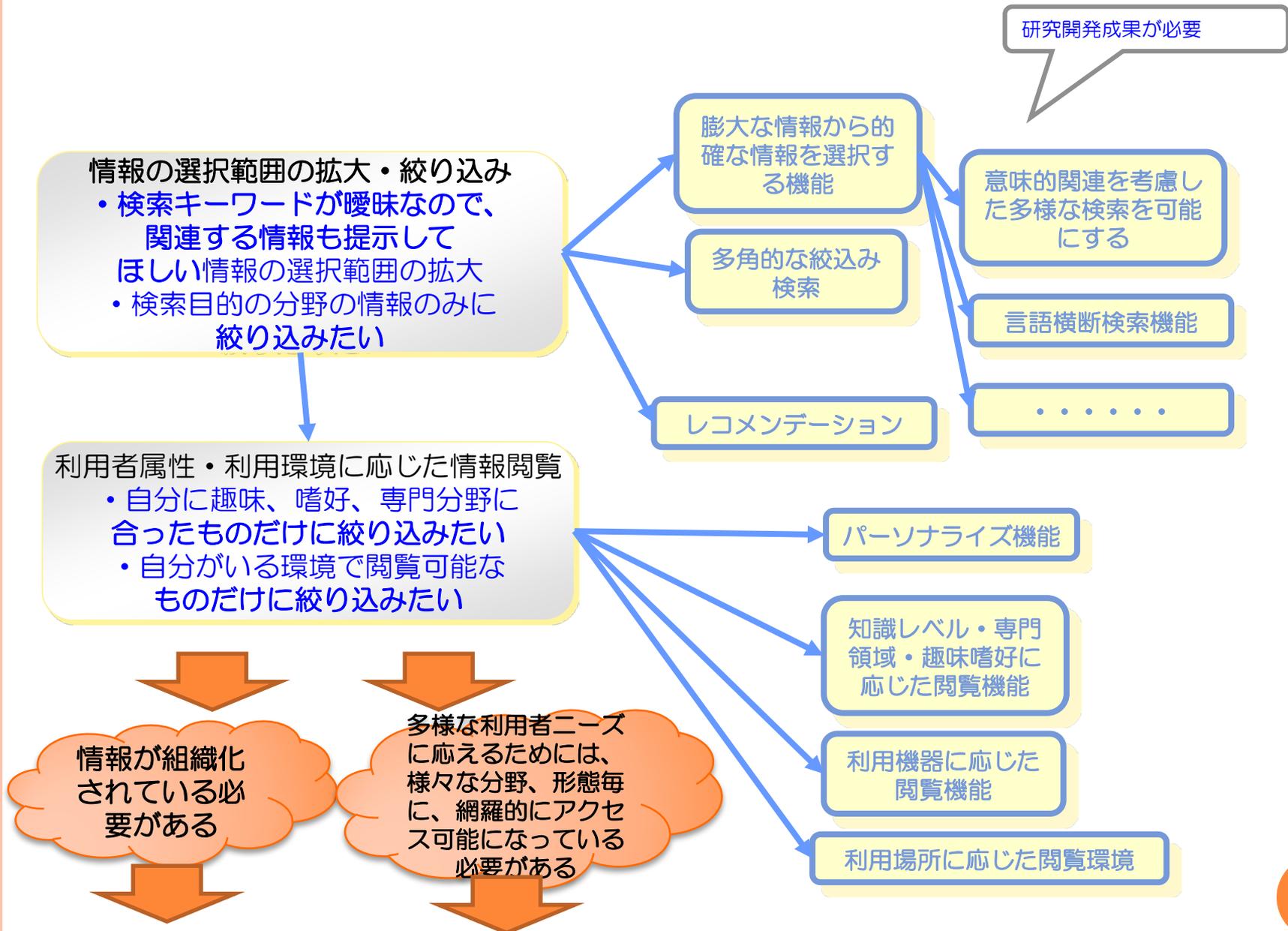
インターネット情報を専門家の人海戦術で組織化は無理

クラウドコンピュータの普及でサービス連携が進む

多様なユーザーニーズの認識 (現在、調査実施中)

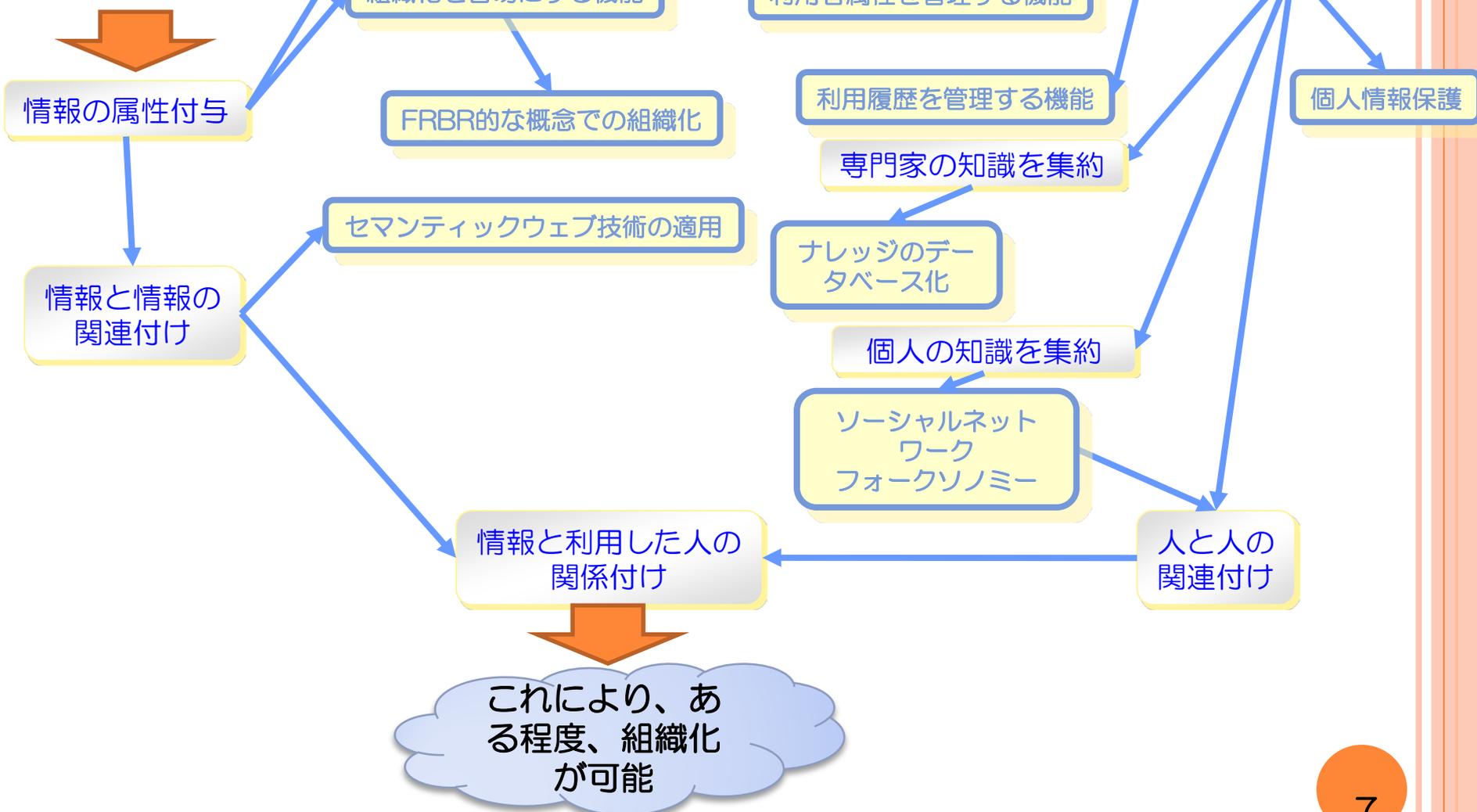


検索に関するユーザーニーズ



情報の組織化

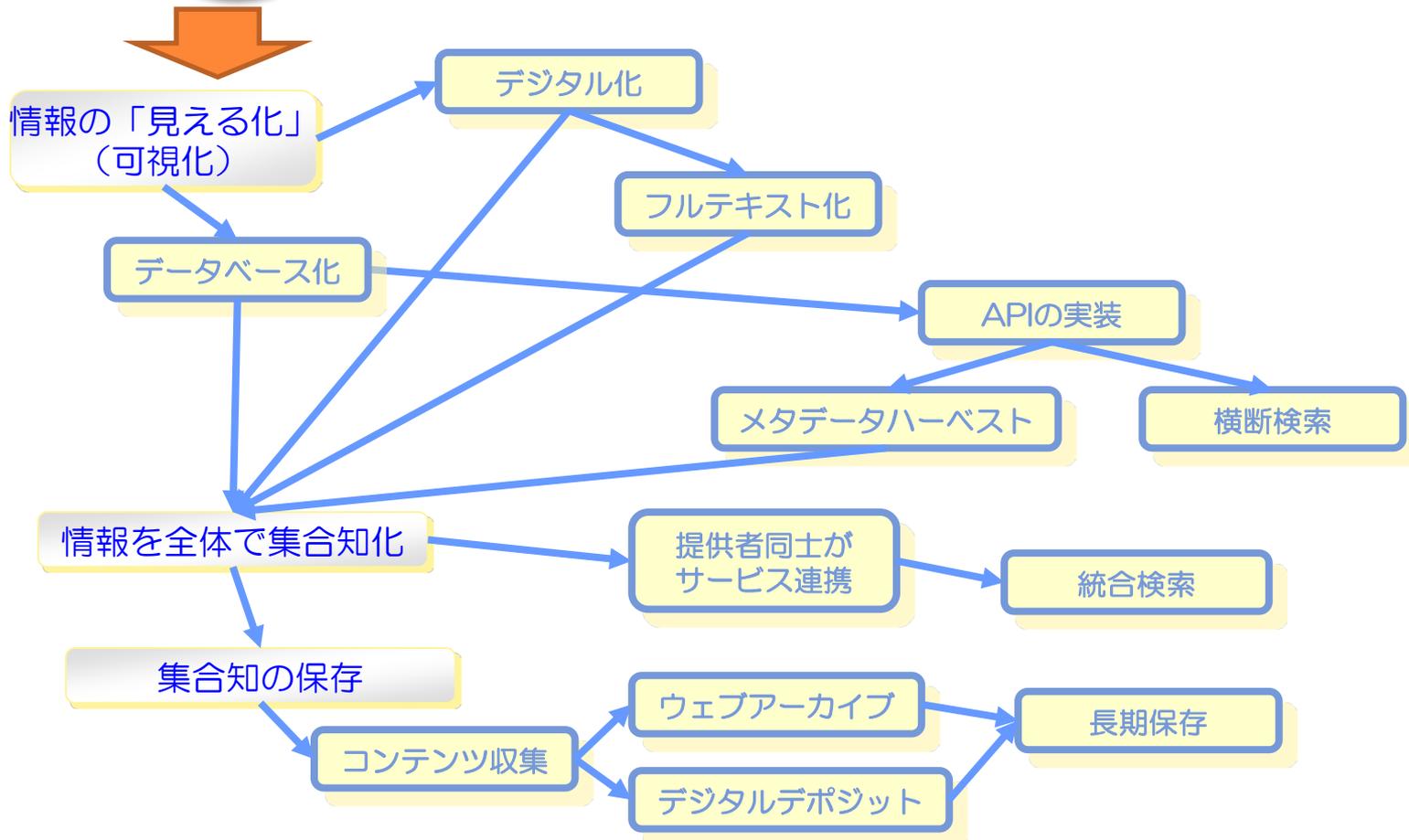
情報が組織化されている必要がある



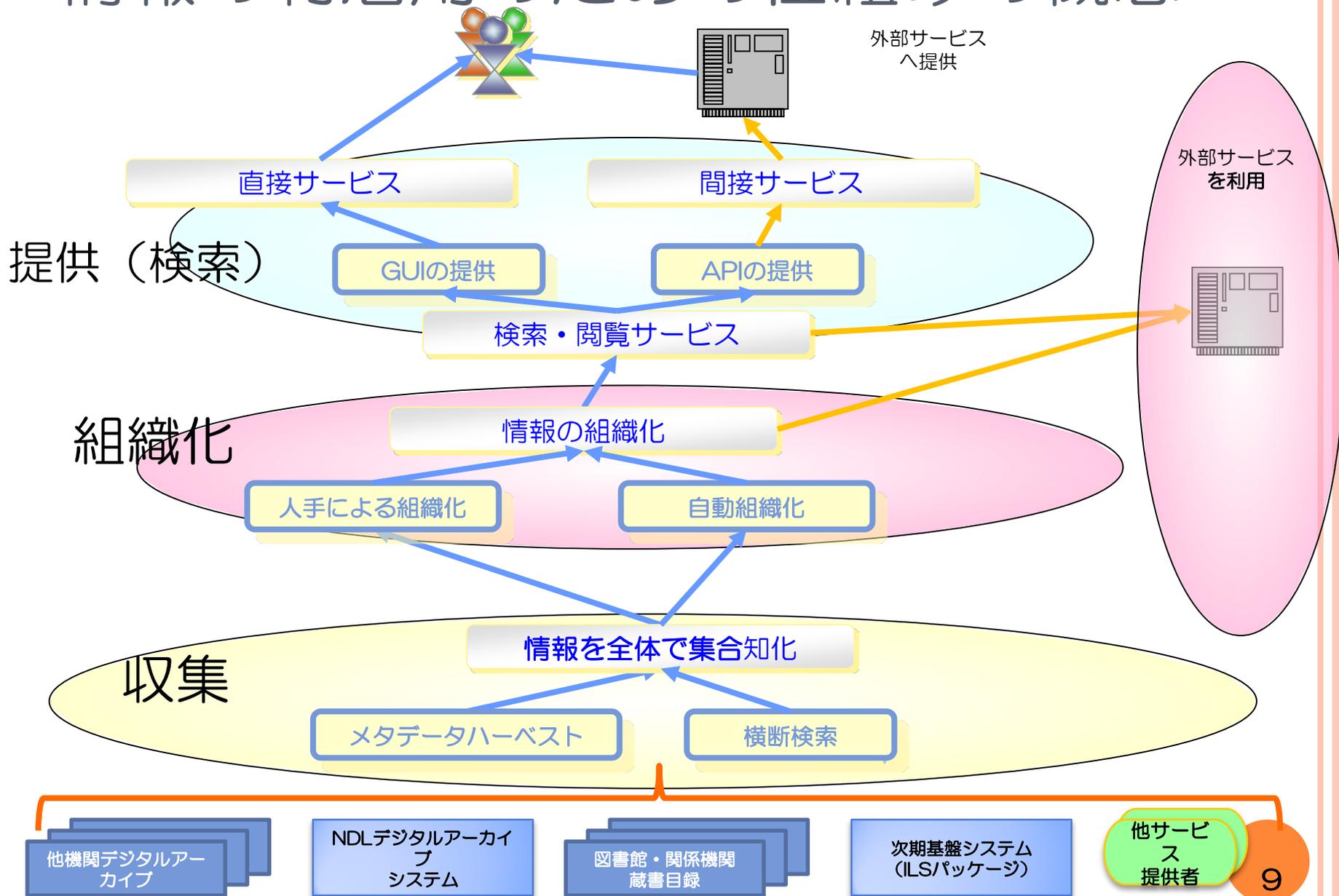
情報の「見える化」、集合知化、保存

多様な利用者ニーズ
に応えるためには、
様々な分野、形態毎
に、網羅的にアクセ
ス可能になっている
必要がある

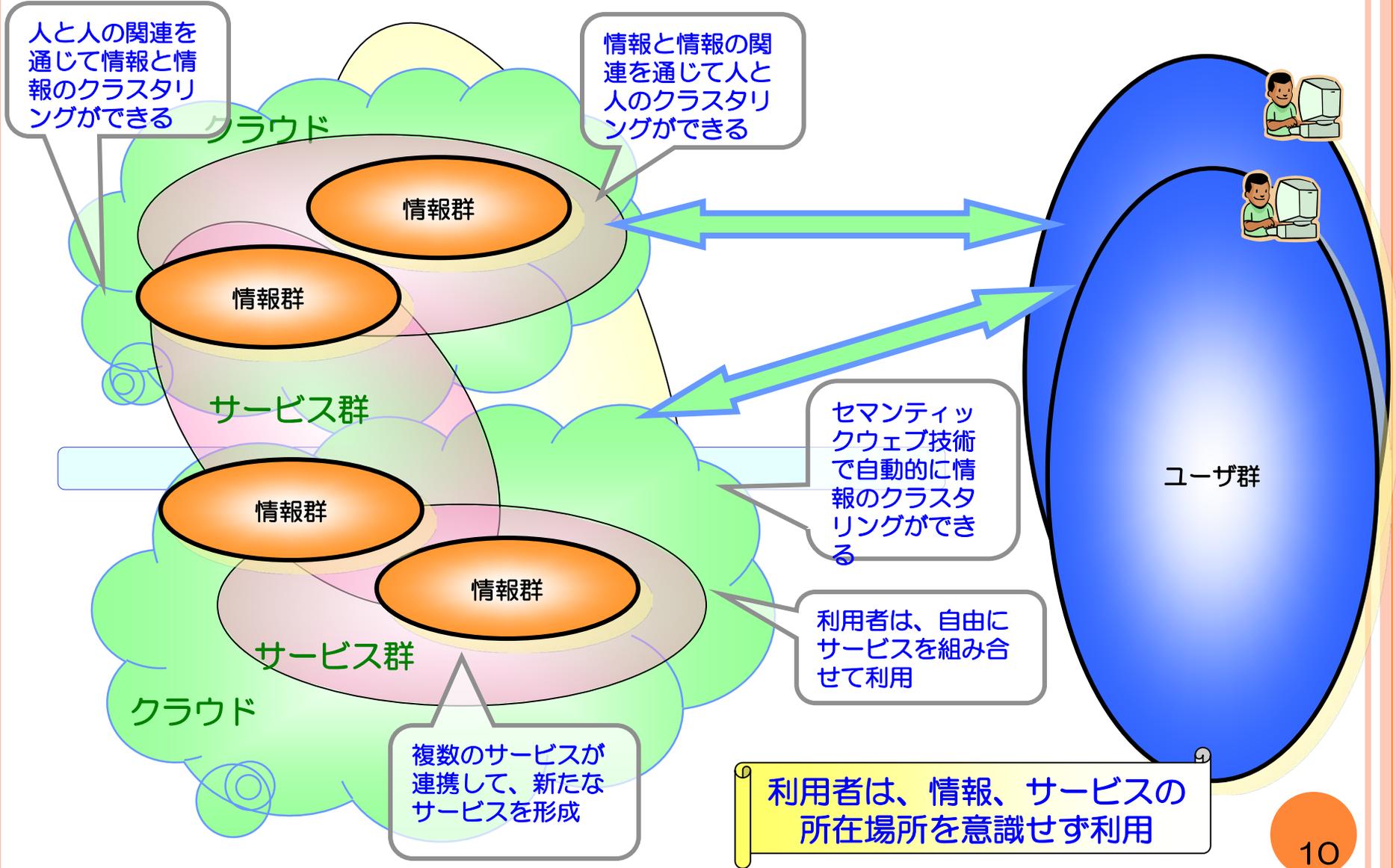
現在、公開されている情報
(表層、深層)
過去、公開されていた情報



情報の利活用のための仕組みの概念

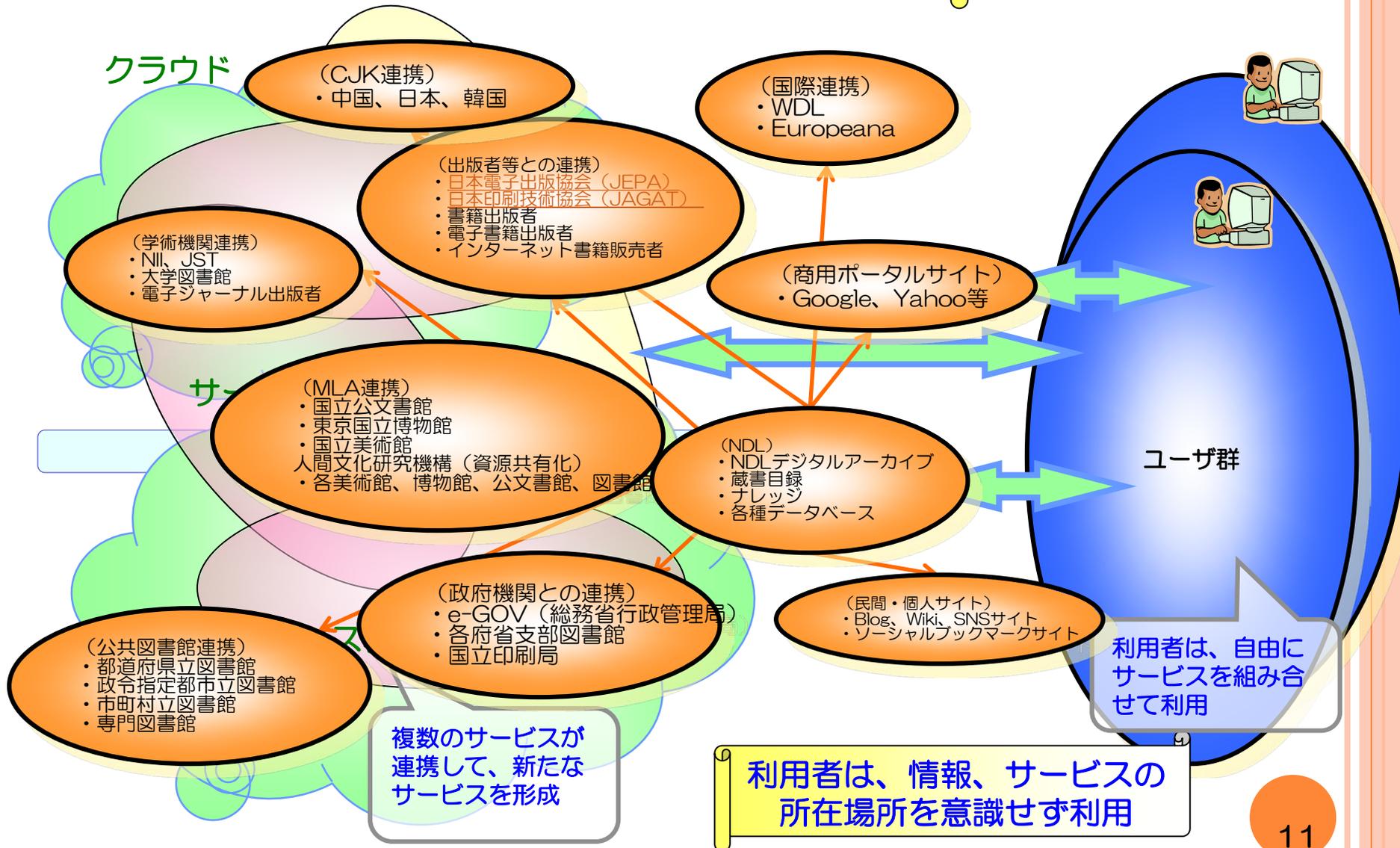


人と情報が関係付けられたサービス (クラウドの世界でのサービスの連携)

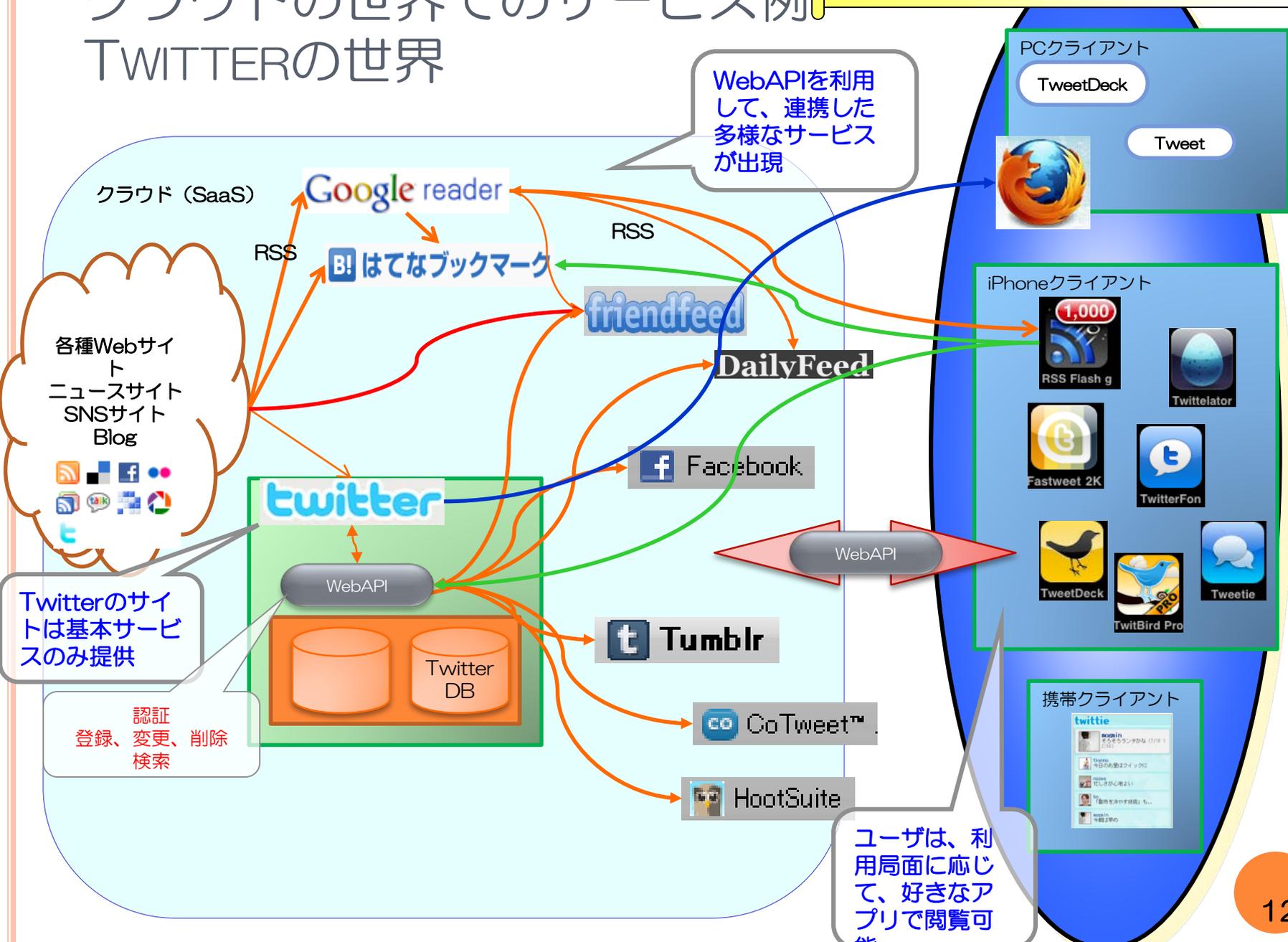


情報探索サービスの将来像 (クラウドの世界でのサービスの連携)

・当館は巨大なデータプロバイダ
 ・巨大なデータプロバイダとして、
 中核的なサービスプロバイダとなる



クラウドの世界でのサービス例 TWITTERの世界



TWITTERの普及とそのインパクト

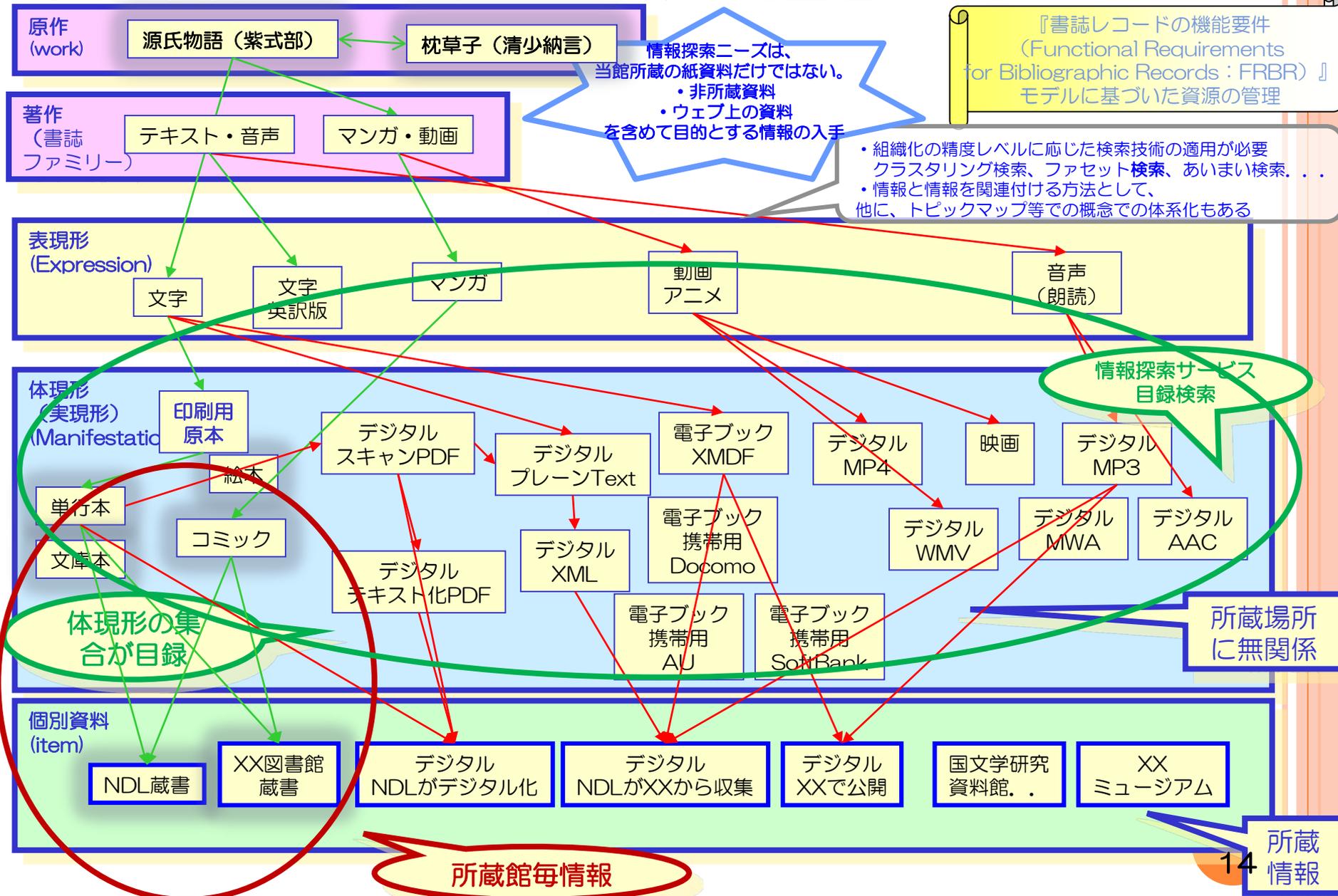
- サービス面
 - 140字でメッセージを不特定多数向けに書き込める
 - Replyでコメントを付けられる
 - 引用したコメントも付けられる
 - 気に入った個人をフォローして、その人の発言を追うことができる
- 仕組み面
 - 何よりも仕組みが簡単。→全文検索エンジンが高速化されたことにより、平易なシステムで実現された。
 - 発言が自分のTLに残る→BlogのTrackbackのようなものだが、Trackbackのように複雑な仕組みではない。(Blogのコメントとは根本的に異なる)
 - ハッシュタグとかで、ソーシャルタギングが容易→本文にタグを含むMicroFormatのようなもの。→セマンティックウェブへの道も開かれる
- 使いやすいAPIの公開
 - 各種端末用に多数の閲覧・投稿アプリが生まれている→まさにいつでもどこでもを実現
 - ウェブサイト(ウェブサービス)、PC用アプリ、スマートフォン用アプリ、携帯(ウェブ、アプリ)
- シンプルなDB構造、サービス機能
 - 全文検索エンジンの高速化をフルに活用したサービスの提供
 - 全文検索、ハッシュタグ検索、ユーザ検索



- ナレッジのもとになる知識の断片が流れている→引用をもとに、集めて整理すれば、有効な知識となる
- ハッシュタグにより、組織化された情報としても利用できる
- Twitter小説で、ごく短編の小説、連載小説も出現→ケータイ小説よりも容易に普及する可能性もある

【情報と情報の関連付けの概念の一つ】

利用の視点でのコンテンツの体系的整理の概念



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＞

『書誌的記録の機能要件

(Functional Requirements for Bibliographic Records : FRBR)』
モデルに基づいた資源の管理

ワンストップ
サービス

ユーザ



メタデータ検索

著作の発見
(Find)

著作メタデータ
(work)

表現形の識別
(Identify)

表現形メタデータ
(Expression)

体現形の選択
(Select)

体現形メタデータ
(Manifestation)

・書誌情報（記述メタデータ）

・言語、利用対象者
・表現形式（冊子体、テキスト、
動画、静止画、音声）

・デジタル化形式（text、html、
PDF、XMDF、eBook形式、MP3、
AAC、WMA形式）、アクセス条件
・所在場所

個別資料の入手
(Obtain)

個別資料 (Item)
単行本・文庫本

個別資料 (Item)
パッケージ
(DVD、CD等)

個別資料 (Item)

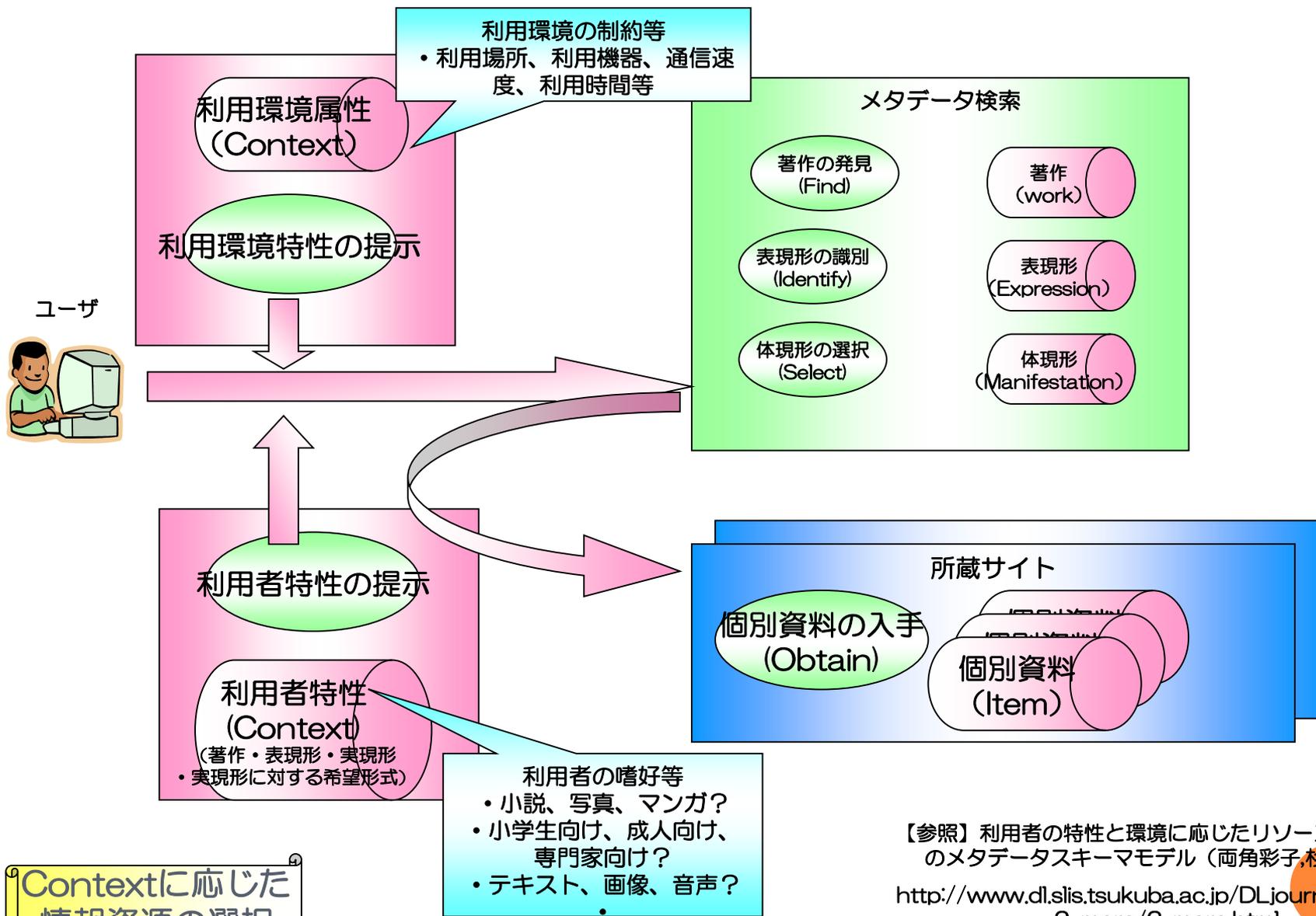
ウェブ情報
・html,xhtml
・pdf
・動画(mpg2,...)
・画像(jpeg,ipeg2000,...)
・音声(mp3,aac,...)
・各種電子書籍
フォーマット(text)

個体情報

【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

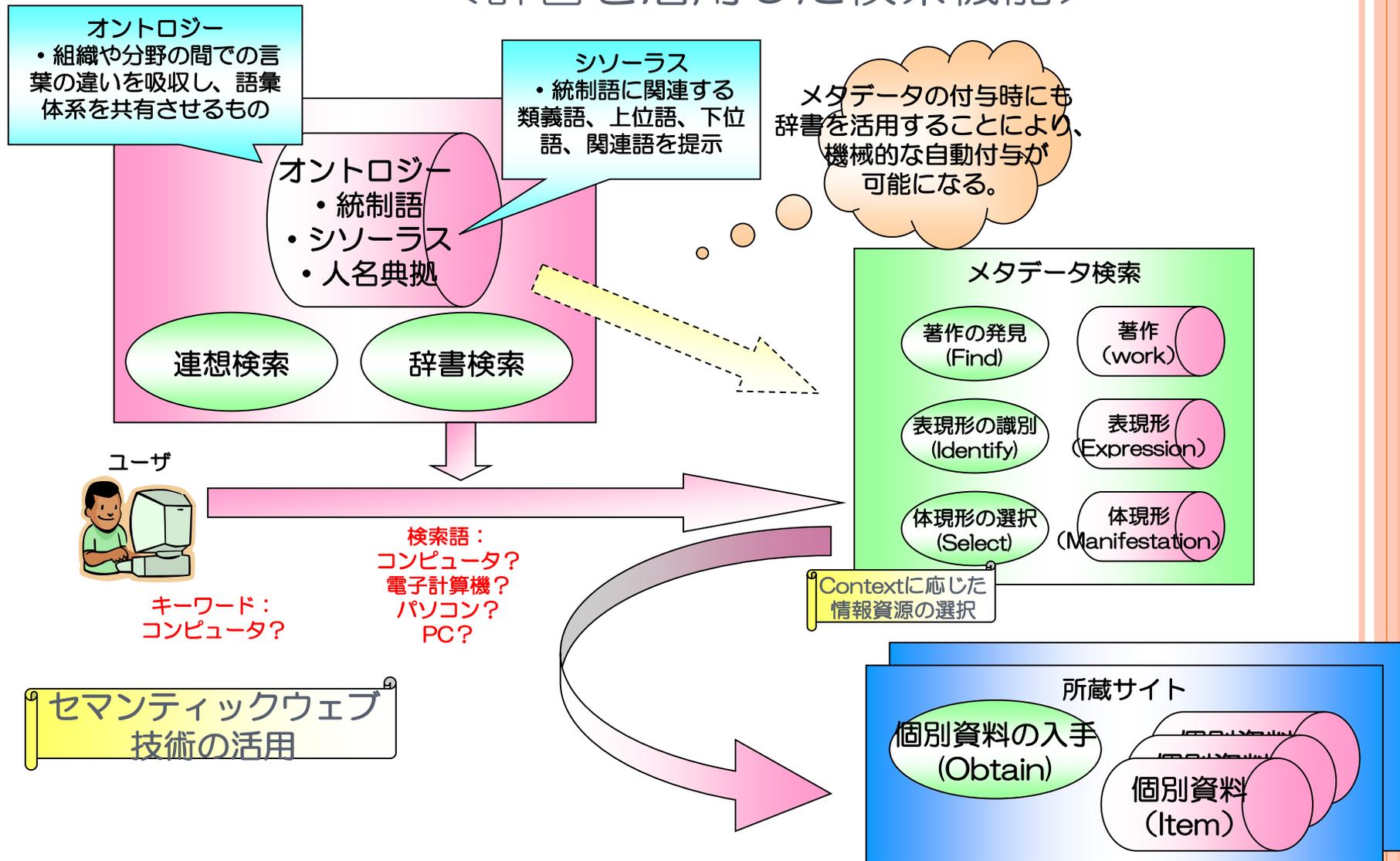
利用者の特性や環境に応じた情報の選択 ＜マイポータル機能＞



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）

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＞



電子図書館中期計画2004

- 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

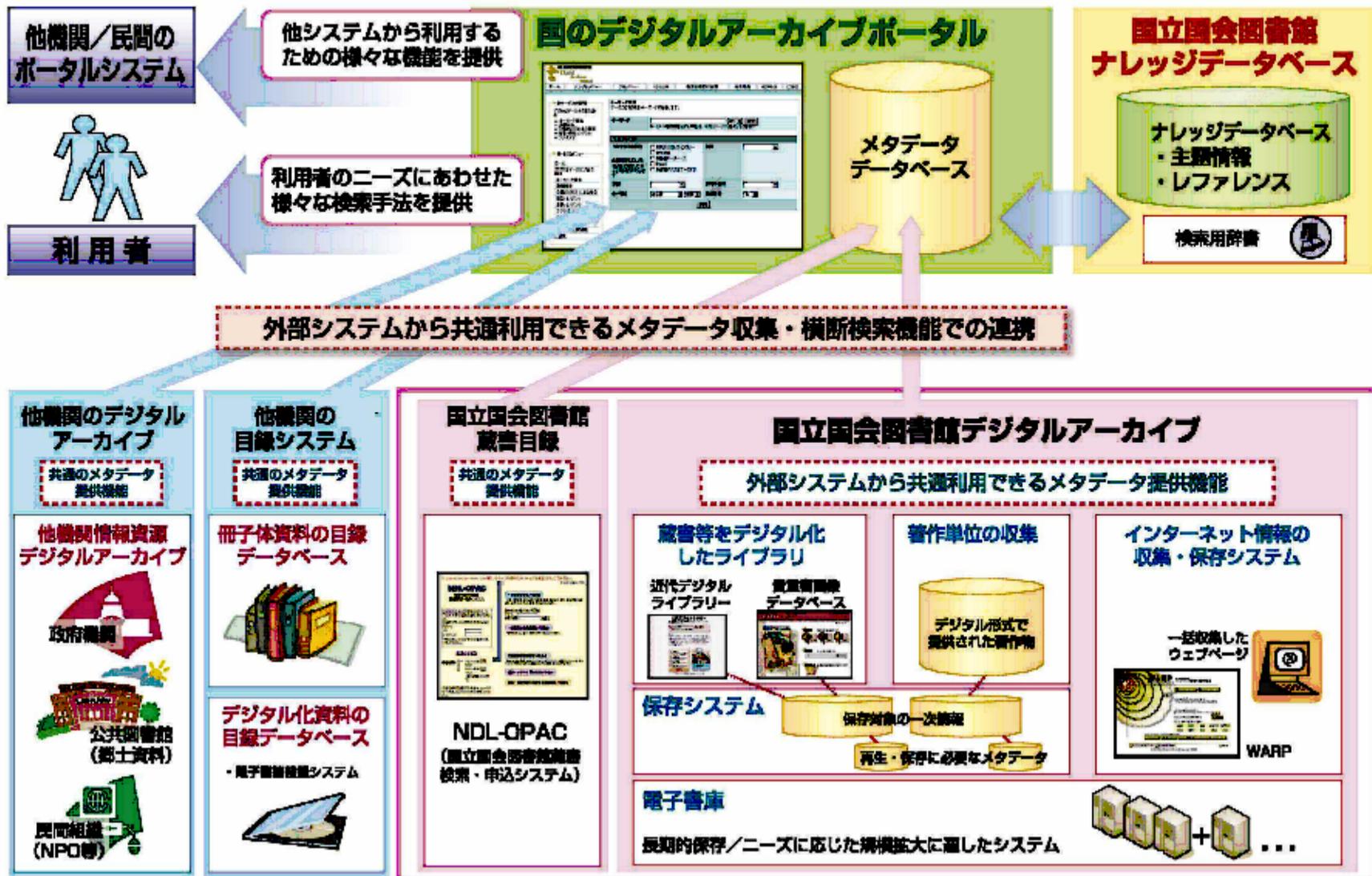
② 情報資源に関する情報の充実

- レファレンス情報、解題情報、統制語辞

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

NDL電子図書館サービスの全体像

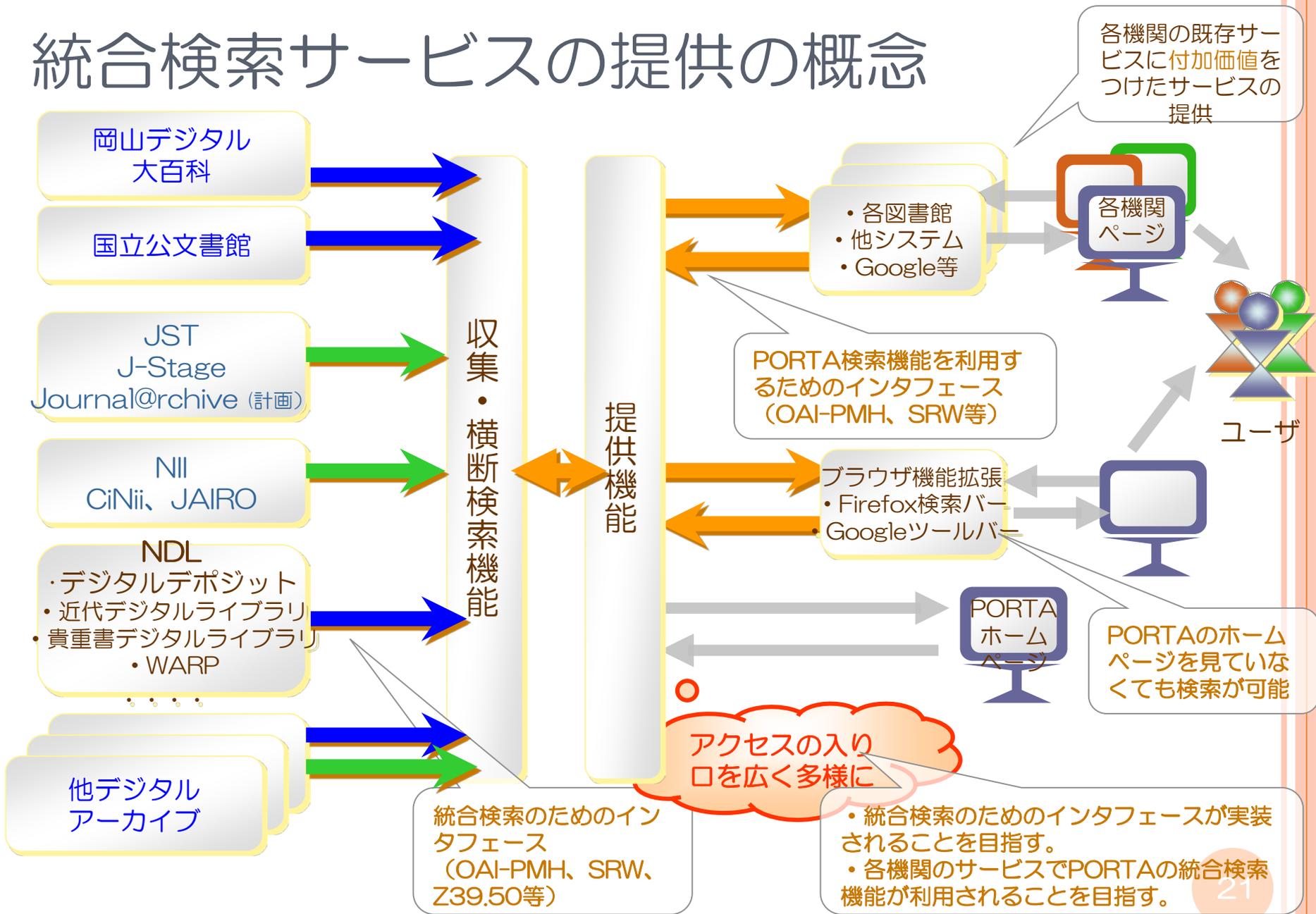


PORTAの現状

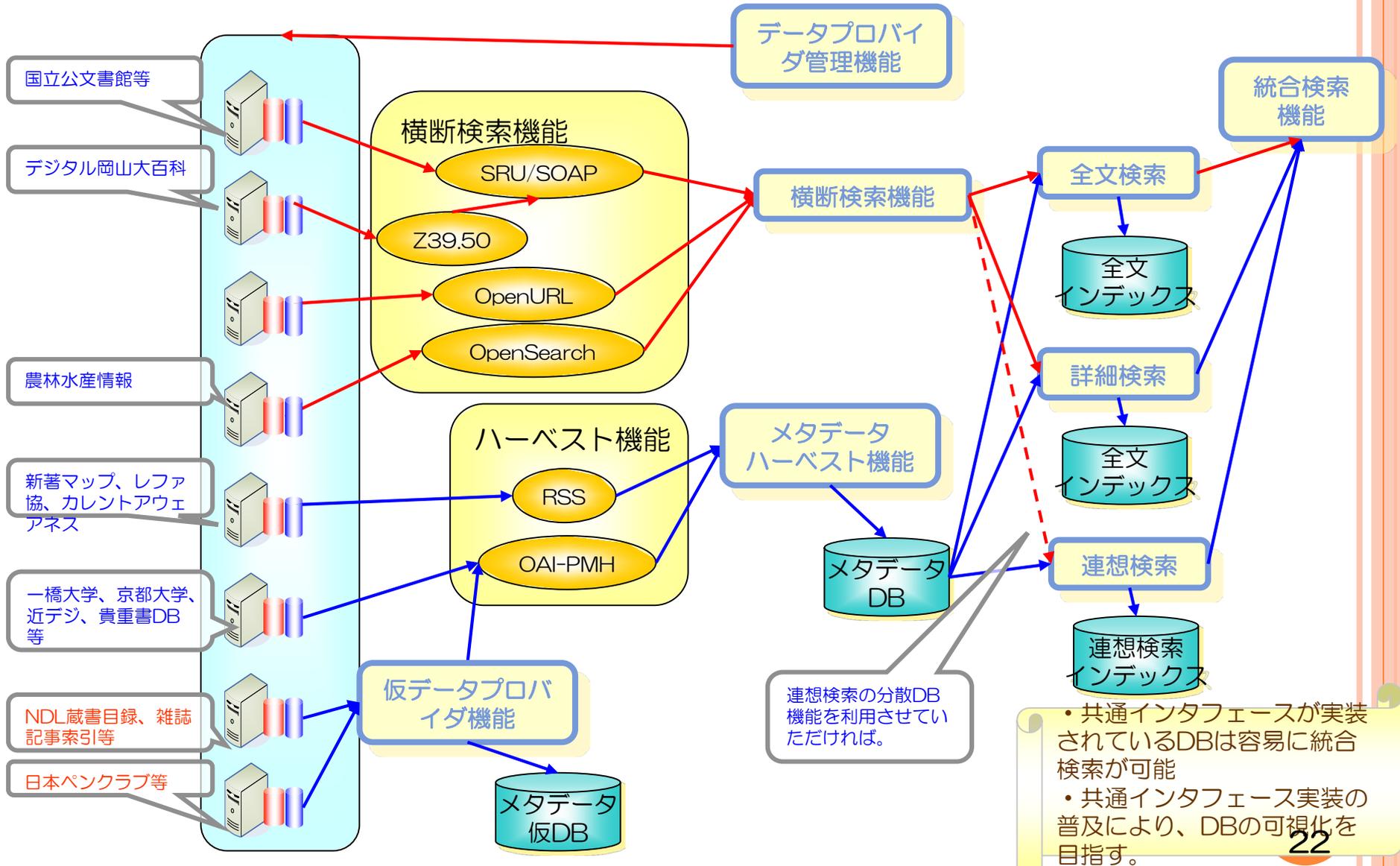
構想からプロトタイプ、正式提供までの経緯

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7月7日	PORTAリニューアル	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

統合検索サービスの提供の概念



メタデータ収集・横断検索の概念



連携のための共通仕様

		メタデータ形式
コンテンツ	DC系 (Dublin Core)	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	SRU/SOAP(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

メタデータ記述要素

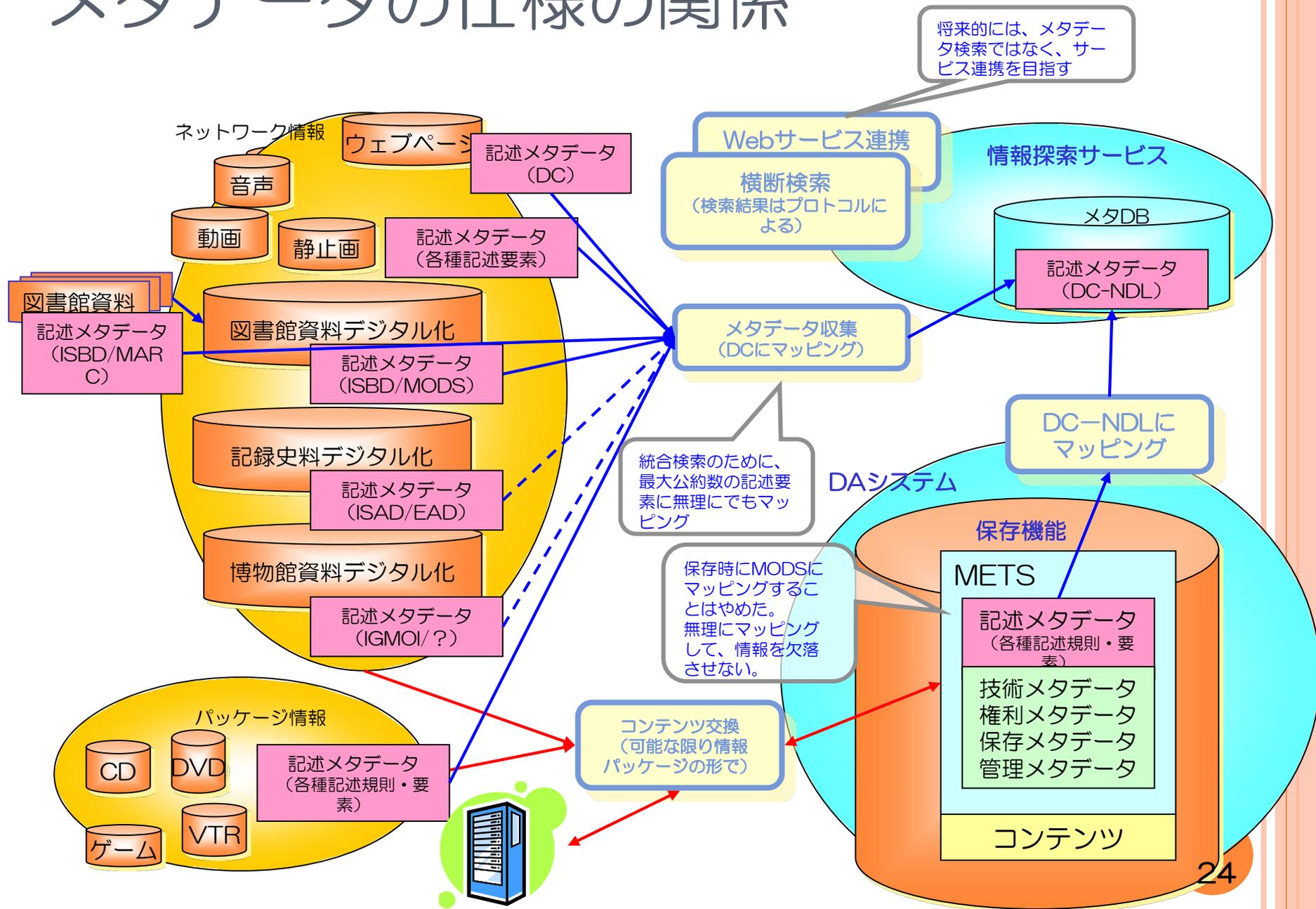
メタデータ記述規則

通信プロトコル	
ハーベスト	OAI-PMH
	RSS
横断検索	SRU/SOAP(SRW)
	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)

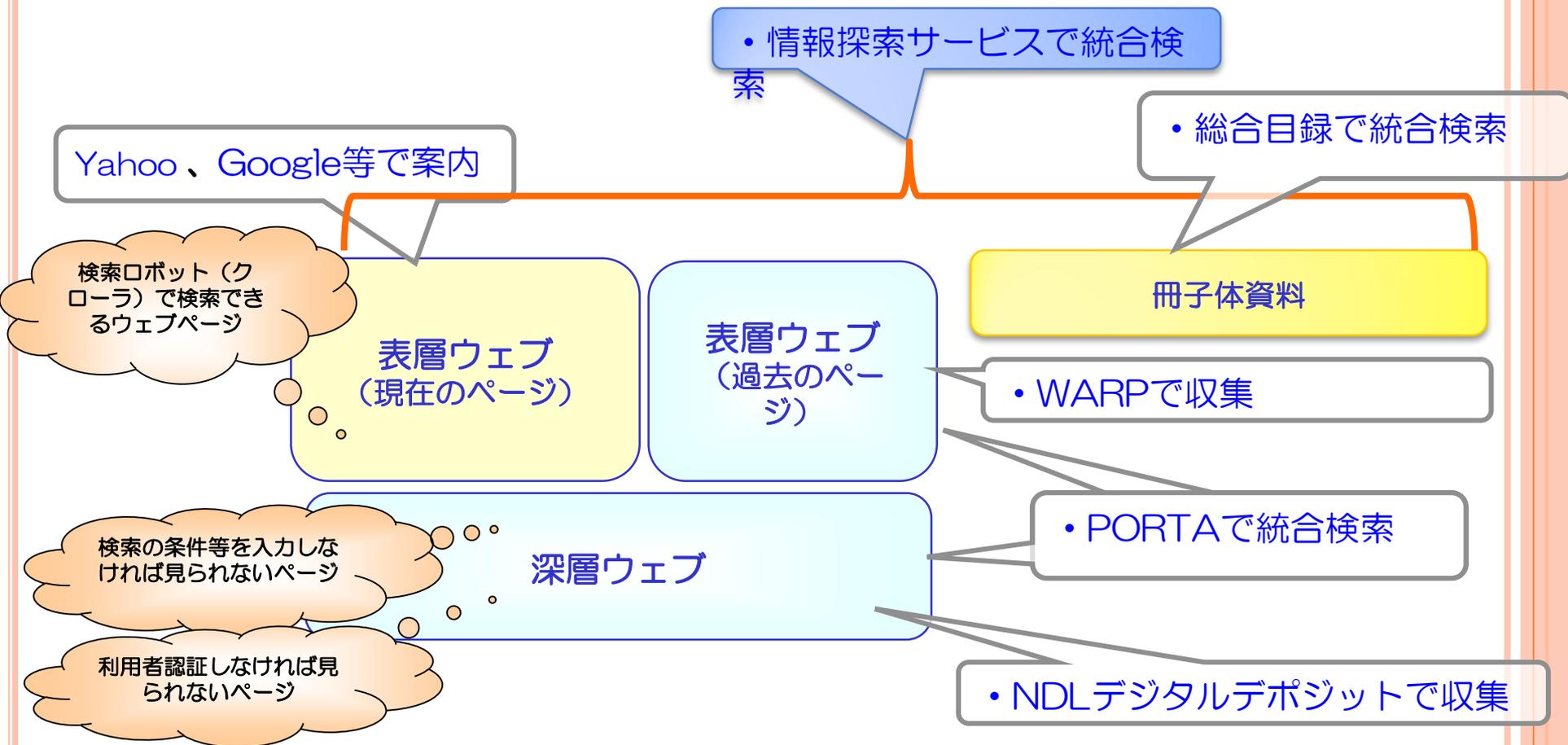
メタデータの仕様の関係



当面の情報探索サービスが目指すもの

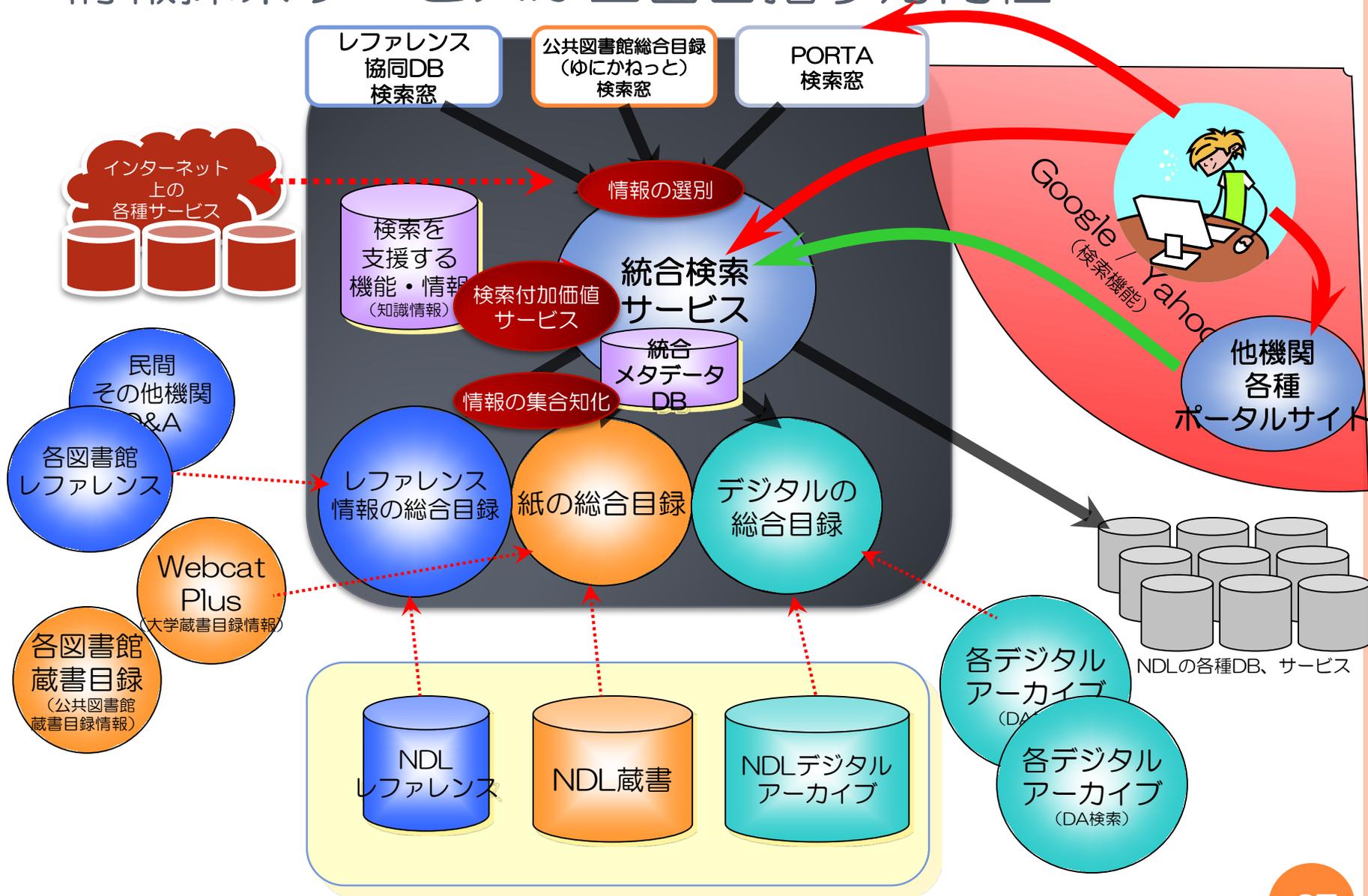
- 情報探索サービスのカバレッジ
- 情報探索サービスが当面目指す方向性
- 24年度正式公開に向けて
 - 情報探索システムの機能モデル（21FYプロトタイプ）
 - 情報探索システムの開発想定
 - サービス要件定義・システム化要件定義
 - 利用者ニーズ調査の作業内容
 - 外部サービス動向調査の作業内容
 - 技術・製品開発及び適用動向調査の作業内容
 - 技術標準適用ガイドライン策定
- 今年度構築システム
 - プロトタイピングの基本方針
 - プロトタイプシステム機能の構成
- 情報探索サービスの構築・提供スケジュール（想定）

情報探索サービスのカバレッジ

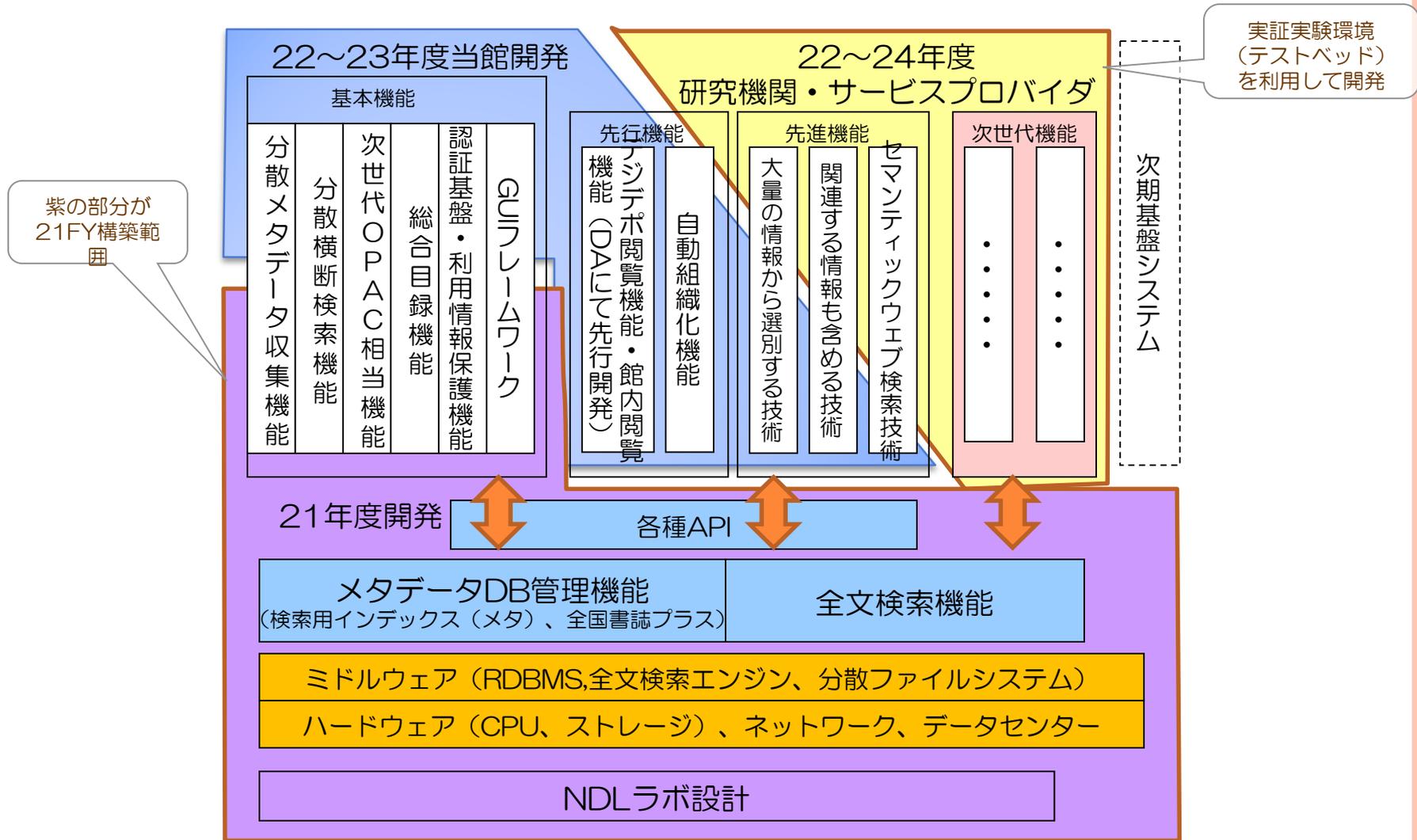


【現在目指している形】

情報探索サービスが当面目指す方向性



情報探索システムの開発想定



サービス要件定義書・システム化要件定義書作成までの流れ

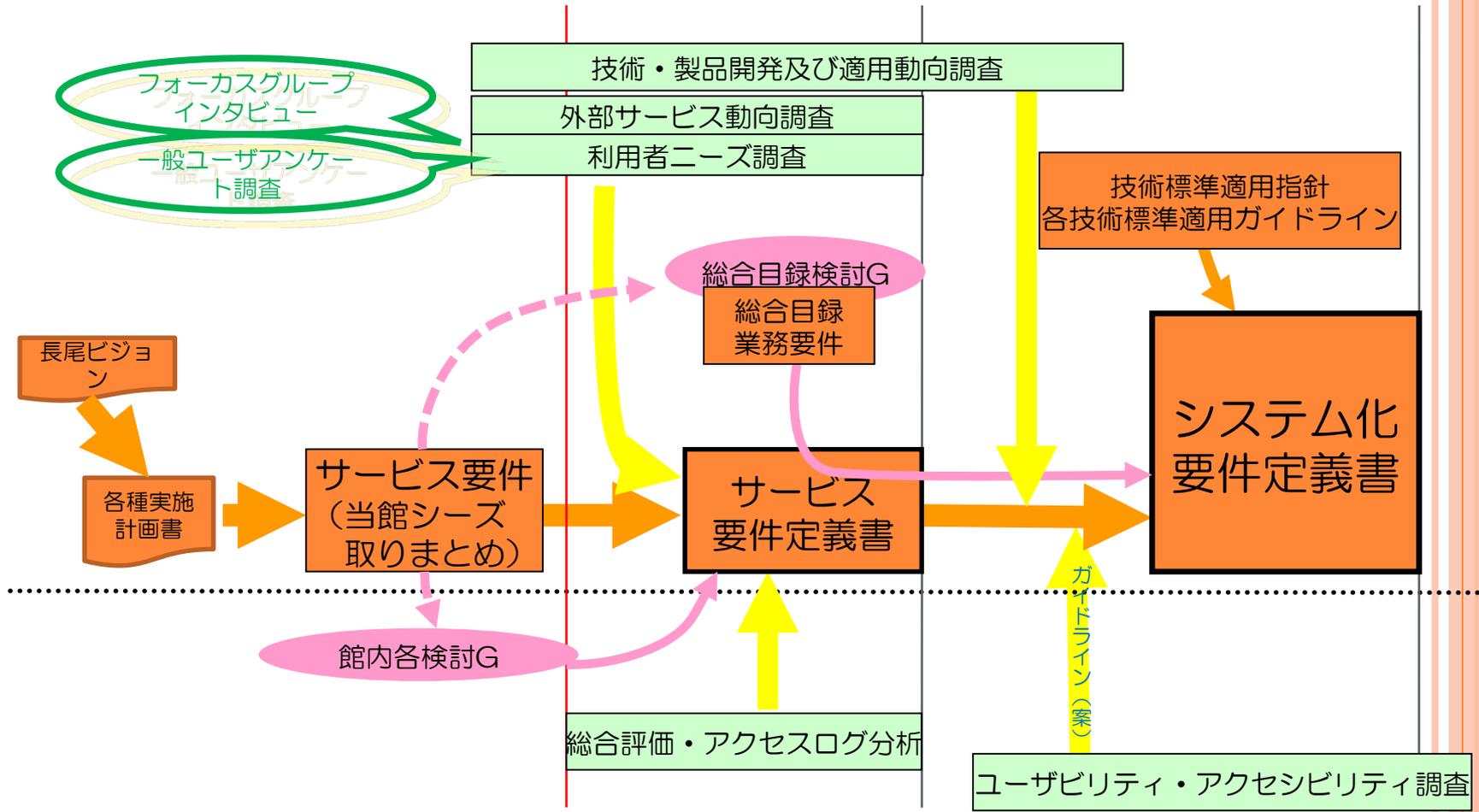
正式システム

プロトタイプ

H21.10

H21.12

H22.2



利用者ニーズ調査

○ フォーカスグループ・インタビュー調査

- ねらい
 - 主に先進的なユーザー/有識者を対象に次期情報探索システムの必要機能を明らかにする。
- 対象
 - 既存の図書館の概念に縛られない広い意見を収集できるようにする。
 - IT技術者、情報サービスの専門家、図書館・情報学の専門家、図書館員（大学、公共、専門）、Webアクティブユーザ、出版、編集、図書販売等、図書館利用ネットアクティブユーザ等

○ 一般ユーザアンケート調査

- ねらい
 - 先進的なユーザを対象としたフォーカスグループインタビューでは得られない一般利用者のニーズを得る
- 方法
 - ウェブアンケート
- 対象
 - 大学研究者、学生、民間企業の調査部門・企画部門・研究開発部門
 - 普通にインターネットで情報検索を利用しているユーザ。
 - いずれの場合も、国立国会図書館の利用経験は前提としない。

外部サービス動向調査

○ 調査対象

- NDLサービスが補完されるサービスと、NDLによって補完されるサービス、技術的な連携と人的・組織的連携と広く考える。
- 商用DB事業者（書籍以外も含む）、書籍出版関連サービス事業者、図書館等

○ 調査項目

- サービス構築の背景・経緯・目的
- サービス内容
- サービスの利用者・利用状況
- 情報システム構成
- 保有データの種類、及び利用法（書誌、全文、レファレンス、機関リポジトリ、デジタルアーカイブ、等）
- 採用技術
 - 情報探索技術（ファセット検索、ラダリング型検索等）
 - 外部連携技術（Webサービス連携等）
 - 採用ソフトウェア（パッケージ、オープンソース等）
- Webサービス等による外部への検索サービスAPIの整備状況
- サービス構築時の課題と解決方法
- NDL情報探索サービスとの連携意向、連携の場合のインターフェース要件
- 今後の構想・予定・課題

技術・製品開発及び適用動向調査

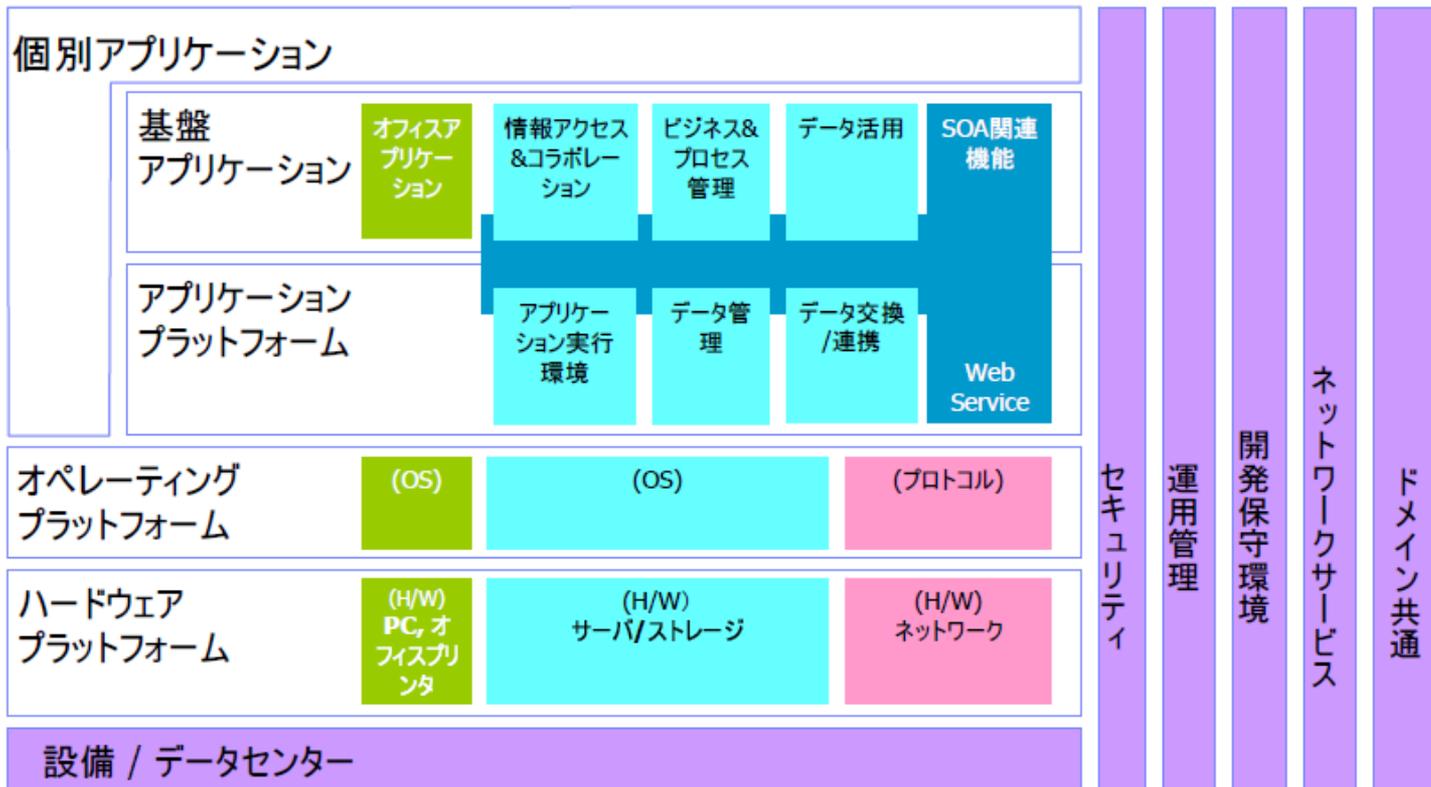
○ 方針

- 情報探索サービスのシステム化に当たって、実装可能又は将来的に実装可能となることが予想される情報技術を調査する。

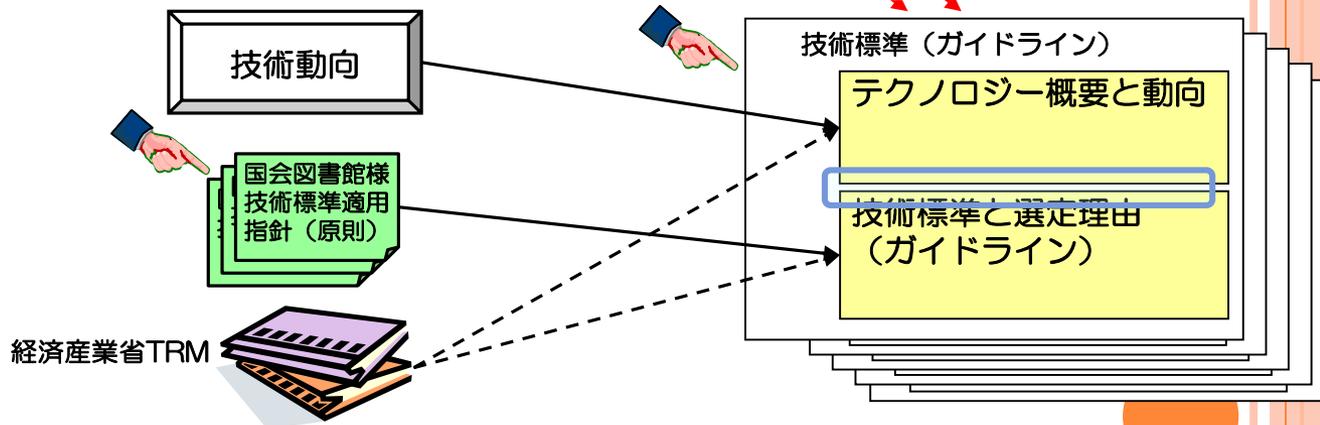
○ 調査内容

No	領域	想定調査対象	調査項目
1	・情報・知識の集合知化技術 ・情報の組織化、情報と人の関連づけをする技術 ・膨大な情報からの的確な情報を選択する技術 ・利用者属性・利用環境に応じた閲覧提供技術 ・次世代検索技術 等	ファセット検索 検索ナビゲーションサービス ソーシャルタギング リッテルナビゲータ ラダリング型検索 など	1) 要素技術名 2) 機能 3) ソフトウェア製品化の状況 4) 情報システムにおける採用状況
2	デジタルコンテンツの配信管理システム調査		1) システム名、提供者名 2) 商用ソフトウェア／オープンソースの区分 3) 機能の概要 （要素技術のサポート状況、APIサポートの状況等を含む） 4) カスタマイジングの可否 5) [オープンソースの場合] 維持管理の状況、運用サポートの状況、コミュニティの状況 6) 提供価格
3	米国議会図書館のWorld Digital Library (WDL)のシステム調査		「4. 外部サービス動向調査の作業内容及び方法」と同様の調査項目とします。

技術標準適用ガイドライン策定



フレームワーク



技術標準適用指針（一覧）

適用指針番号	適用指針の名称	
TD-01	利用者の利便性向上に資する技術の積極的な採用	
TD-02	オープンな標準に基づいた技術・仕様の採用	
TD-03	技術・仕様の共通化	
TD-04	システムの特性に応じた成熟度を持つ技術の採用	
TD-05	パッケージ・ソフトウェアやオープンソース・ソフトウェアの活用	
TD-06	資源の共同利用および柔軟な配分・拡張に資する技術の採用	
TD-07	システムの重要度に応じた障害対策技術の選択	
TD-08	情報セキュリティを考慮した技術の選択	
TD-09	運用・保守業務の集約化・共通化に資する技術の採用	

プロトタイピングの基本方針

- NDLの新しい利用者サービスの方向性を打ち出す
 - 網羅性が保証された情報資源へ利用者をナビゲート
- 利用者オリエンテッドでユーザビリティを追求する
 - 利用者の検索プロセスを考慮したユーザインターフェースの提供
 - 旧来のOPACのような表示にはこだわらない。
- 利用者をターゲティングする
 - あくまでも「一般ユーザ」。未利用者層を開拓
 - 確証をもった形での利用
 - 試行錯誤によりたどり着いた利用
- 検索エンジン経由で訪れるユーザを重視する
 - NDLのサイトという認識なしに訪問したユーザを適切にナビゲートすることを重視
- デザイン・操作性を磨き上げる
 - 他の優れたアイデアや工夫は積極的に取り込み、無理なオリジナリティは追及しない。
- 「いつでも、どこでも」を実現する
 - 携帯端末利用者にも、高い操作性とデザインのGUIを提供
- 新しい付加価値を生み出す
 - 民間企業や非営利団体、個人が提供しているサービスとの連携や複数の異質なサービスの組み合わせ
 - 従来の図書館の枠にとらわれない自由な発想による付加価値創造の仕掛け

プロトタイプシステム機能の構成

○ 情報の収集

- 当館及び他機関のデータベースに格納された書籍、ジャーナル、雑誌、地図、画像、映像、音楽等のコンテンツのメタデータ
- ハーベスティングや横断検索

○ 情報の組織化

- 収集したメタデータについて、インデックスを作成して組織化
- DC-NDL形式により体系化された形でデータベースに保管
- 関連資料をグルーピングするなど、情報の構造的な見せ方も可能に

○ データ管理

- 情報探索サービス用に収集・組織化したデータを最新の状態で管理

○ 情報の検索

- 情報探索サービスシステム内のデータベース及び全国の公共図書館等のサイトから、簡易検索、詳細検索その他様々な検索方法を提供
- その際、サジェスト機能やレファレンス情報、外部機関が提供する連想検索機能等のナビゲーションサービスを活用

○ 付加価値サービスの提供

- 検索機能のほか、RSS配信やブックマーク機能など情報探索に役立つ検索以外のサービスも提供
- また、公共図書館等に対し、情報提供・収集用のAPIも提供
- 民間企業や非営利団体、個人等と連携しながら、保有する情報資源を活用した様々なサービスを提供することを目指す

情報探索サービスの構築・提供スケジュール(想定)

項目	平成21年度				平成22年度	平成23年度	平成24年度
	第1四半期	第2四半期	第3四半期	第4四半期			
インターネットサービス	Web2.0 インターネット上のサービスが連携				クラウドコンピューティングの世界 インターネット上の資源を使ってサービスを展開		
世界規模のデジタル アーカイブ、書誌 データベースの構築					WorldDigitalLibrary 世界規模のデジタルアーカイブ WorldCat 世界規模の書誌データベースへの書誌情報提供		
関係機関と連携によるサービス提供	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		
情報探索サービス	公共図書館等のDA構築支援・デジタル化支援				新PORTA運用		
	サービス要件定義				順次機能強化		
	システム化要件定義 プロトタイプ構築				新情報検索サービス 順次サービス公開		
					多角的な利活用に関する研究開発成果を順次適用		
デジタルアーカイブ システム 次期図書館システム	構築・運用準備				DAシステム運用		
	調達仕様書作成				DAシステム機能強化		
	RFC・入札				次期基盤構築		
					移行作業		
					次期図書館システム		
インフラ	共通インフラの検討(ハード・ネットワーク・認証基盤等)				インフラ調達		

大学図書館等との連携

○ 現状

○ システム連携

- NIIのCiNii（論文データベース・サービス）等との連携
 - NDL-OPAC→(リンクリゾルバ)→CiNii（21年9月）
 - CiNii→NDL-OPAC雑誌記事索引（21年中）
 - PORTA→CiNii（21年7月）
- GeNii、Webcat-plusとの連携
 - NDL-OPAC、PORTA →(リンクリゾルバ)→ GeNii、Webcat-plus（21年9月）
- JAIRO（大学機関リポジトリポータル）との連携
 - PORTA→(統合検索)→JAIRO
 - 各大学の機関リポジトリ内の情報へナビゲート（21年7月）

● 今後（想定）

- 情報探索サービス（24年1月正式公開予定）
 - 画像、テキスト、音声と冊子体資料の統合検索
 - NII、JST、当館（総合目録、PORTA）に加えて、電子ジャーナル、電子書籍、商用DB等を統合検索し、一次情報へナビゲート
 - 情報探索サービステストベッド(NDLラボ(仮称))
 - 保有している情報資源（メタデータ、デジタルコンテンツ）を研究開発等へAPIを通じて提供。
 - 成果を当館システムに適用できることを期待
- 次期図書館システム（24年1月稼働）
 - MARC21で、世界に日本語書誌を提供
 - 件名典拠をウェブサービスで提供
 - 書誌をウェブサービスで、リアルタイムに提供
- デジタル化された学術情報流通の協調、効率的な棲み分け
 - 学位論文のデジタル化、収集、保存
 - メタデータの標準化と相互運用、永続的識別子付与
 - 学位論文のポータルサイトの構築
 - 著作権処理
- 機関リポジトリのアクセスの保証
 - 長期保存、障害時代替提供(ダークアーカイブ)

公共図書館等との連携

- 全国公共図書館総合目録ネットワーク
 - 情報探索サービスの中で、デジタルコンテンツと統合したサービスとして強化
 - 新規参加館は、国際標準(OAI-PMH、DC等)でのメタデータハーベストを検討
 - 横断検索による参加館の拡大も視野に
- レファレンス協同データベース
 - 参加図書館の拡大
 - 検索・更新APIの実装を想定
- 地域デジタルアーカイブの構築支援
 - 地域資料のデジタル化→デジタル化等ガイドライン、メタデータ記述要素、記述規則ガイドラインの提示
 - 地域デジタルアーカイブ構築支援
 - NDLデジタルデポジットシステムのOSSでの提供、ASP、SaaSでの実現も視野に
 - 地域デジタルアーカイブの長期保存、バックアップも想定
 - 地域デジタルアーカイブの統合検索サービスの構築支援
 - 情報探索サービスのOSSでの提供も検討
 - 各公共図書館の検索サービスに、情報探索サービスを組み込むことを支援

⇒ 図書館の郷土資料がインターネットで発見され利活用が進む
各図書館のサイトや来館での利用者が増える

終わりに

- 膨大な情報を知識として活用して、新たな知識の創造をめざす
→「知識はわれらを豊かにする」への実現手段の一つ
- 現状で確立した技術では実現が困難
- 研究開発成果の実用化の実証実験環境の構築
- 実証実験環境の資源を活用して、より先進的なサービスの実用化実験の実施

国立国会図書館: <http://www.ndl.go.jp>

ご静聴ありがとうございました