

国の「デジタルアーカイブ」構築に向けて — 今後実施すべきことと連携協働 —



平成20年11月26日

於・図書館総合展フォーラム(第11会場10:30～12:00)

国立国会図書館関西館

主任司書 中山正樹

今日のお話し

- 現状の再認識
- 何をすべきか
 - データベースの見える化
 - 情報を集合知化
 - 情報の組織化
 - 利用者の組織化
 - 情報と利用者の関連付け
 - 情報探索方法の変化
- どのようにシステムを構築するか
 - 当館システム構築の考え方
 - デジタルアーカイブシステム構築に向けた連携協力
- どのようにしてデジタルコレクションを構築するか
 - 政府の施策(2008年)
 - 公共図書館との連携
 - 学術機関との連携
 - 商用サイトとの連携
 - 出版社、著作権者との連携
 - MLA連携
 - 日中韓連携
 - WDLとの連携
- 5年程度の将来に向けた方向性

現状の課題の再認識

指数的に増大するデジタル情報

.....

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

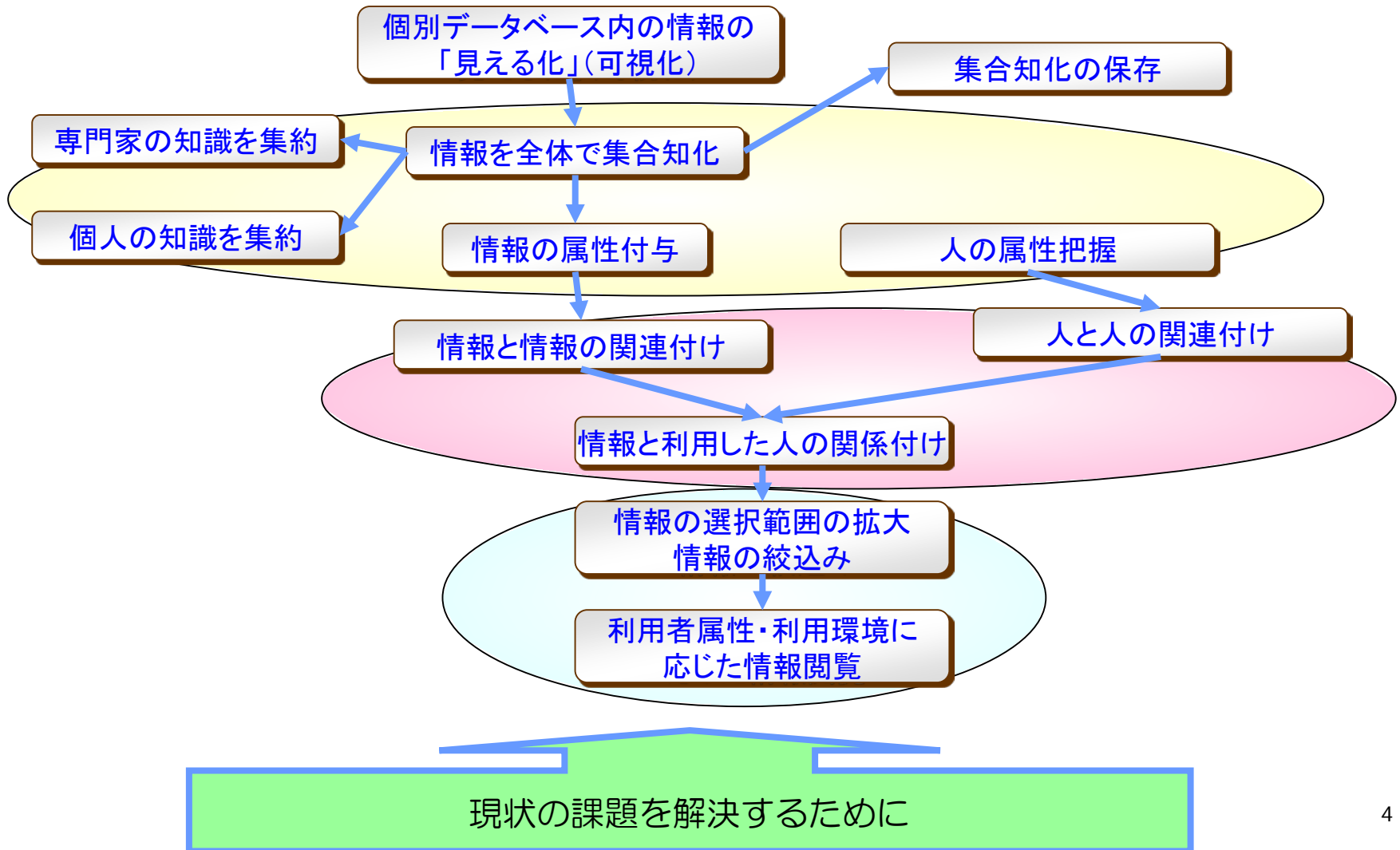
有用な情報は、民間、個人の
データベース、ブログ等に多く存在

インターネット上の情報は
すべてを収集することは不可能

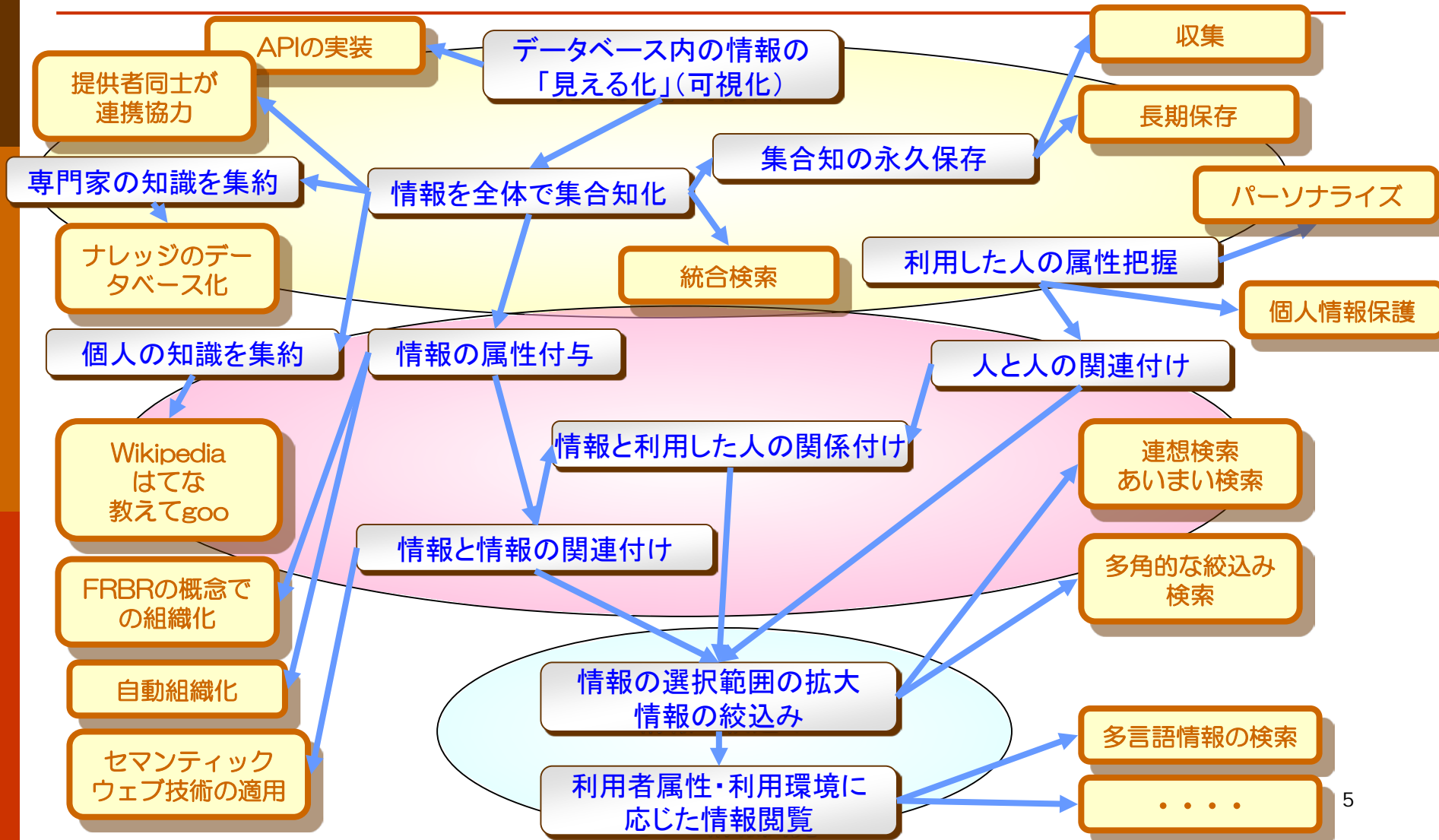
専門家の人海戦術では無理

人の知の集合は
専門家をしのぐ

実施すべきこと

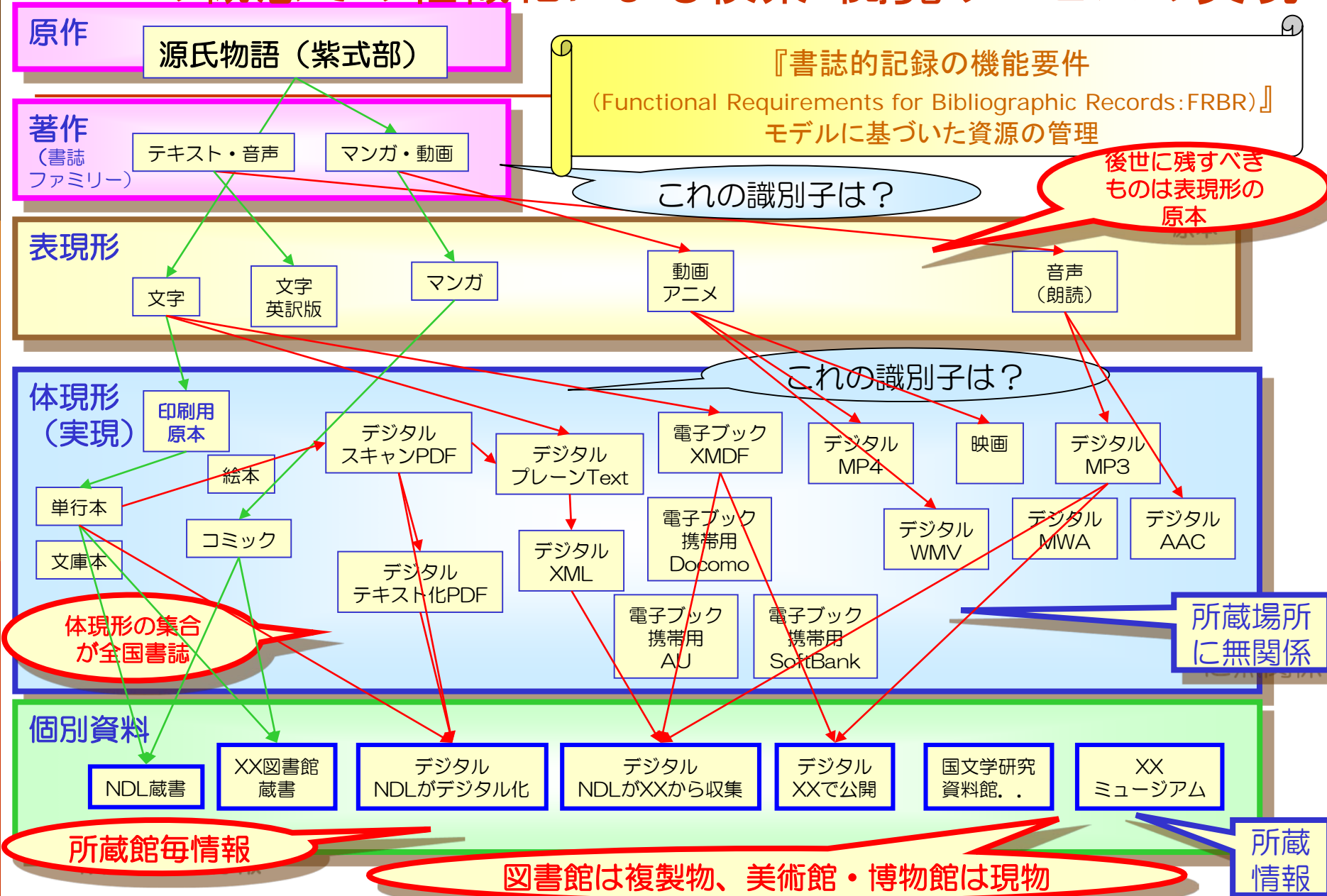


実施すべきことと、必要な機能の関連



【参考】情報の組織化の1つ

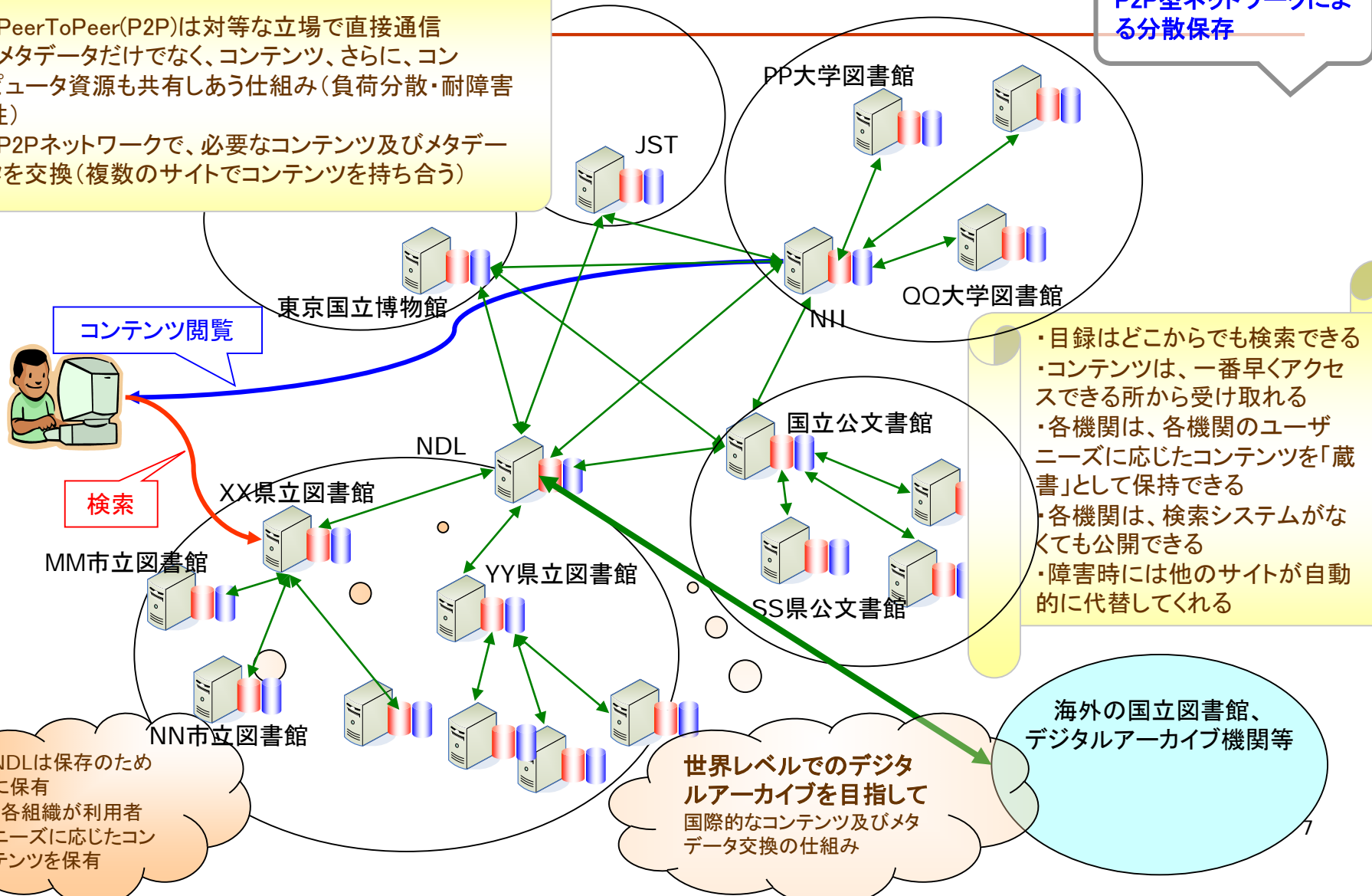
FRBRの概念での組織化による検索・閲覧サービスの実現



【参考】コンテンツ交換の方法の1つ (PeerToPeer) 対等な関係でのデジタルアーカイブの構築イメージ

- ・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)

P2P型ネットワークによる分散保存



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ーク（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

産学官の研究機関との連携

技術の例

- ・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
- ・長期保存を実現するストレージ技術（次世代ストレージ）

- ・文明開化プロジェクト（総務省）
- ・情報大航海プロジェクト（経済産業省）
- ・情報爆発IT基盤（文部科学省）

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、大学の研究室等との連携
- ・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

⇒ 当館コンテンツを利用して、研究開発が進むように。
研究開発成果が、当館システム構築に活用できるように。

【連携協力】

国際インターネット保存コンソーシアムとの連携

国際インターネット保存コンソーシアムへの加盟

(International Internet Preservation Consortium : IIPC)

- ・IIPCは、Webアーカイビングに資する相互運用可能なツールや技術の開発・標準化を推進し、国際的な利用を促進することを目的として結成された機関
- ・現在、各国の国立図書館や公文書館など38の機関が参加
- ・IIPCを通じて国際的に貢献することはもちろん、相互運用可能なツールや技術を国内に還元することで、日本のWebアーカイブ進展に資する

スタンダード

短期的にはWARCの標準化、将来的にはメタデータ、評価基準の標準化等の標準化

ハーベスティング

Heritrixの開発、WARCフォーマットの開発とサポート、クローラに特化した要望、深層ウェブの収集、ビデオやストリーミングの収集

アクセス

アクセスツール（Wayback machine,WERA）改善、NutchWAXを使った全文検索のテスト、ユーザ認証、ユーザ権限、アクセスコントロール、アーカイブのコンテンツ解析のためのツール、アーカイブの構造解析のためのツール

プリザベーション

デジタル情報の保存をめぐる技術的課題への取り組み、ストレージの拡張性に関する研究、WARCフォーマットの保存機能の向上、ウェブをめぐる技術環境の変化の記録、再現できなくなりつつあるファイルフォーマットの特定

デジタルコレクションの構築に向けた 主な連携協力

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

【参考】NDLのデジタルアーカイブ事業に関連する政府の施策(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2008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>

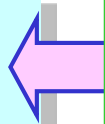
重点計画2008(8月20日)IT戦略本部

「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>

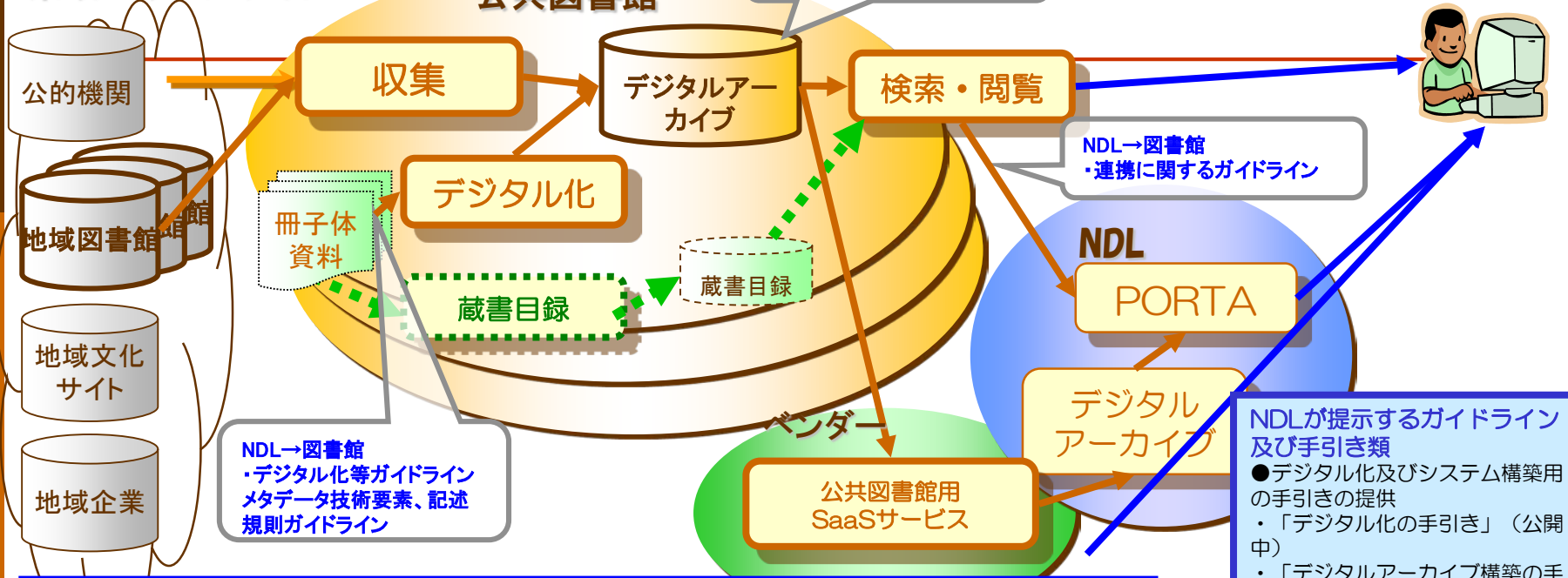
経済財政改革の
基本方針



【連携協力】公共図書館支援のイメージ

県域でのデジタルアーカイブを構築

県域のウェブサイト



公共図書館に対して、地域情報ハブとしての機能の実現を支援

- ・公共図書館が、県内有用サイトの収集もしくは横断検索することを支援
- ・公共図書館が、郷土資料をデジタル化することを支援
- ・公共図書館が、収集コンテンツ及びデジタル化コンテンツのデジタルアーカイブを構築し、提供することを支援（必要に応じて当館が利用しているデジタルデポジットシステム等を貸与する）
- ・デジタルアーカイブを構築し運用することが困難な図書館に対しては、ベンダーのSaaSサービスの利用を支援。
- ・NDLは、各公共図書館のデジタルアーカイブを長期保存もしくはバックアップのために収集保存する。
- ・NDLは、各公共図書館のデジタルアーカイブのメタデータをハーベストもしくは横断検索して、利用者に対して、全国公共図書館が持つ郷土資料を統合的に検索し、各図書館サイトへナビゲートするポータル機能を提供する。

NDLが提示するガイドライン及び手引き類

- デジタル化及びシステム構築用の手引きの提供
 - ・「デジタル化の手引き」（公開中）
 - ・「デジタルアーカイブ構築の手引き」（計画中）
 - ・「共通仕様及び連携に関するガイドライン」（公開中）
 - ・「外部提供インタフェース仕様書」（公開中）
 - ・「メタデータスキーマガイドライン」、「メタデータ記述規則」（策定中）
- 各機関での地域情報収集用及びアーカイブ構築用の汎用ソフトウェアの提供

概念的には、岡山県立図書館が実施している「デジタル岡山大百科」のような事業が全国展開されることを目指す

【連携協力】関係機関拡大策の例 公共図書館支援のスキーム

公共図書館所蔵資料の可視化で苦労したこと

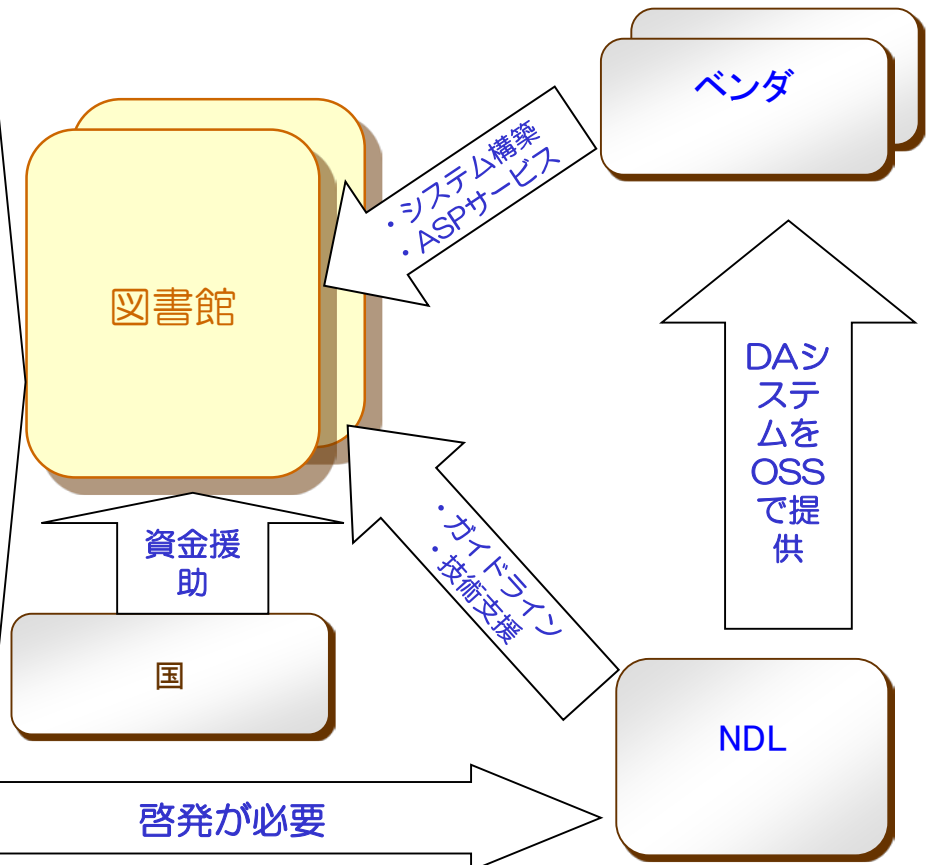
- 資料を、デジタル化する費用がない
- 自前で、サーバを立ち上げられない。運用できない
- 意義は理解したが、標準プロトコルを実装できない
 - ウェブページに貼り付けて、データベースとして検索できる形になっていない
 - データベース化されていても、外部提供インターフェースを持っていない
 - 外部提供インターフェースを実装する費用がない
- メタデータのマッピング調整に膨大な時間がかかる
 - DCをベースにした記述要素を使っているが、使い方がそれぞれまちまち
 - 同じ要素を使っているが、記述規則が異なり、同じ内容として認識できない

□ 統合検索できることのメリットの理解

- アクセスが増えることが不安
- サービスが横取りされるのでは？

これを解決することが、関係機関の拡大につながる

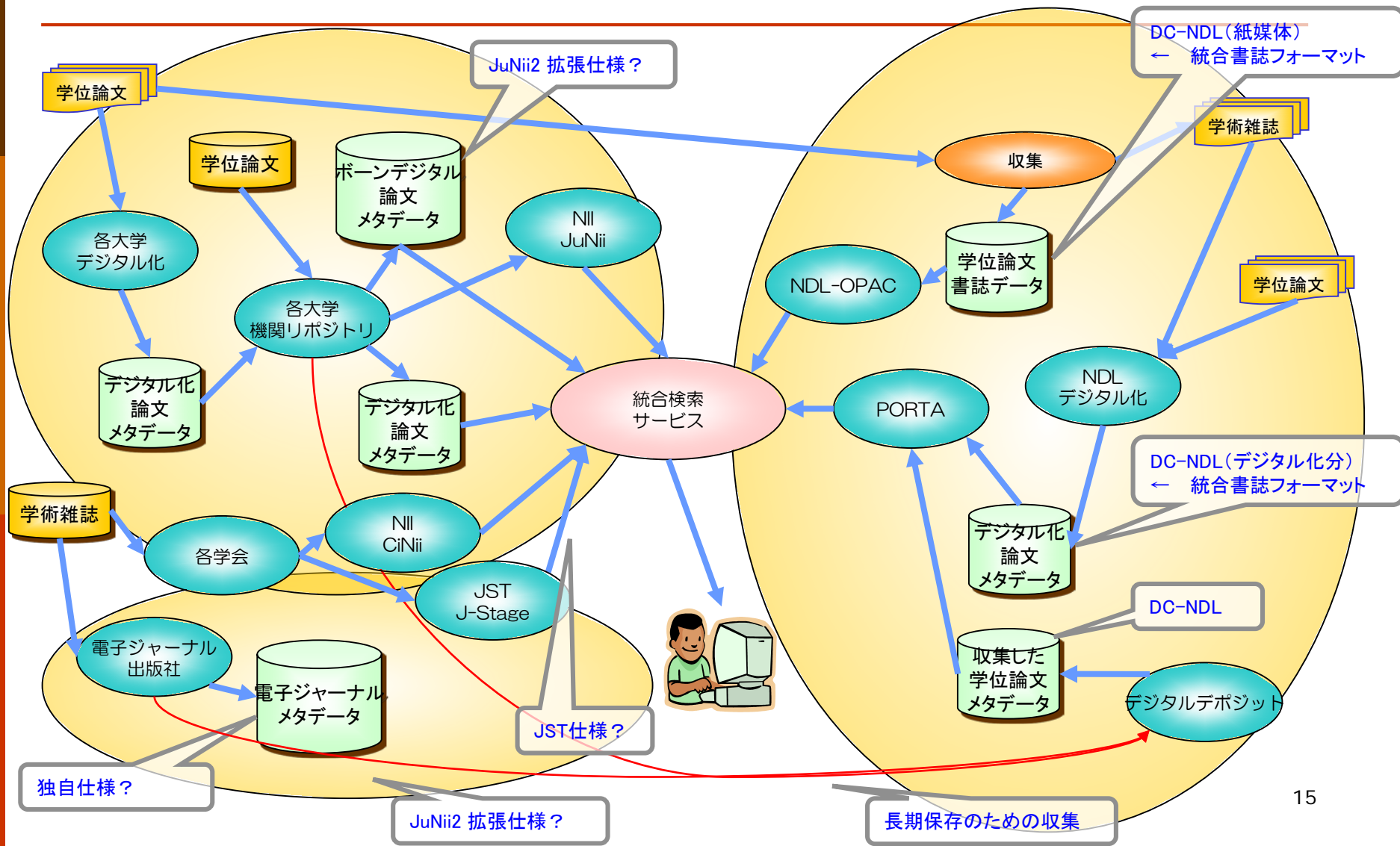
支援が必要



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

【連携協力】学術機関との連携

大学機関リポジトリと遡及デジタル化コンテンツの統合利用環境



【連携協力】民間サイトとの連携協力

商用デジタルアーカイブ

• hon.jp

約9万タイトルの電子書籍を検索可能とし、それぞれのタイトルのPC用、各社携帯電話用、PDA用の電子書籍（約30万件）へナビゲートできるデータベース

• 統合検索の準備中（12月初旬公開予定）

• 日経BP

日経BPbizボードにおいて、雑誌記事、Webニュースを公開中。PORTAとの連携に関して意見交換中

民間団体デジタルアーカイブ

• 歴史的音盤アーカイブ協議会（HiRAC）

検索エンジンサービス

• Google、Yahoo等

民間・個人サイト

• Blog、Wiki、SNSサイト

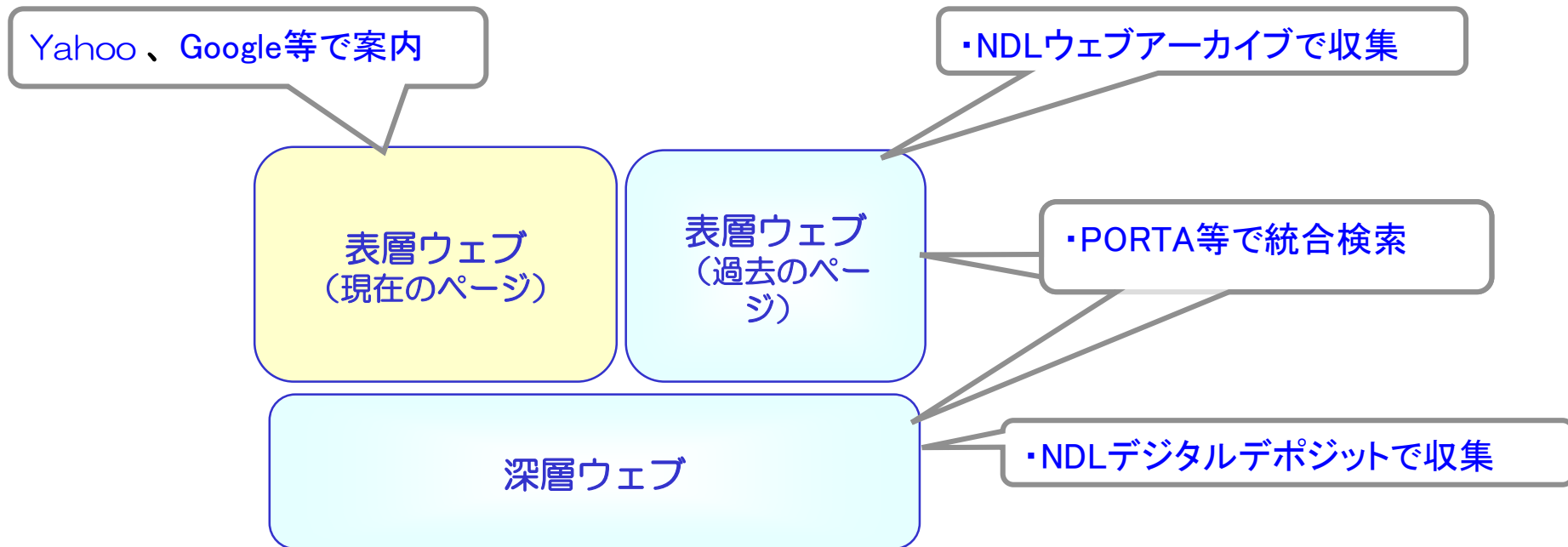
• Wikipedia等

ソーシャルブックマークサイト

• はてな等

【連携協力】

Yahoo、Google等との連携の可能性



Yahoo、Googleと連携することにより、
網羅的な検索の実現を目指す

【連携協力】

出版者、著作者等との連携協力

- **資料の長期保存と利用の促進のために**
 - **当館はデジタルコンテンツを長期保存**
 - デジタルコンテンツの収集
 - マルチユースコンテンツの原本、印刷用原本等、DRMのないもの
 - 可能な限り詳細なメタデータの提供
 - 資料のデジタル化
 - 出版物を当館がデジタル化することの許諾
- **出版者のビジネス支援**
 - **各出版者の電子書籍作成支援**
 - 当館がデジタル化したコンテンツを電子出版用に提供
 - **各出版者の電子書籍データベースの立上げ支援**
 - 当館デジタルアーカイブを利用して、出版者が販売することも検討
 - **各出版者サイトへのナビゲーション**
 - 出版者の電子書籍データベースを、PORTAで統合検索
 - 電子出版物の購読が促進されるように
 - PORTAでの統合検索結果に、入手先として出版者へのリンクを表示

⇒ 出版者、著作権者と利用者の橋渡し。出版者等のビジネスが拡大できる形で連携・協力

【連携協力】MLA連携

デジタル情報資源ラウンドテーブル(仮称)の発足

□ 背景

- 文書館、博物館、図書館においては、これまで大量に蓄積されてきた過去の文化資源のデジタル化は、未だ進んでいない。
- また、デジタル化形式の標準化や、各組織・機関の作成するデータベース間の連携についても、その利活用において、多くの課題を抱えている。

□ 目的

- MLAの各機関が保有するコレクション全体を、インターネットで、一元的に利用できるように
- 検索結果から各機関のDBへ案内することにより、各機関のDBの利活用が進むように。
- さらに来館して実物を見る利用者が増えるように。

□ 状況

- 平成20年3月準備会開催
 - 国立公文書館、東京国立博物館、国立情報学研究所、科学技術振興機構、慶応大学、奈良女子大、国立国会図書館
 - 横断的アーカイブズ論研究会
- 平成20年12月に、ラウンドテーブルを発足させる
 - 日本国内の博物館、図書館、文書館が、館種を超えたデジタル化による仮想的なコレクションの構築を目指す。
 - デジタル化の進展に係る共通の課題について、その解決に向けた協議・検討を行う。

⇒ 関係機関のコンテンツへのアクセス、来館利用が促進される形での連携・協力

【連携協力】

歴史的音盤アーカイブへの協力

- 歴史的音盤アーカイブ推進協議会(HiRAC)
 - 趣意
 - 文化的遺産であるSPLレコード原盤が散逸・劣化し始めている
 - これをデジタル化して保存し、広く国民に公開することによって、日本の文化と音楽産業界の発展に寄与する
 - 対象
 - SPLレコード:3万タイトル(6万曲)、その他、商業用レコード以外の記録物
 - HiRACメンバー
 - 日本放送協会、日本音楽著作権協会(JASRAC)、日本芸能実演家団体協議会、日本伝統文化振興財団、映像産業振興機構、日本レコード協会
- 考えられる協力
 - 方針
 - HiRACへのオブザーバ参加
 - HiRACによる音盤デジタル化の支援
 - デジタル化音源の当館への納本
 - NDLデジタルアーカイブからの公開（館内、公共図書館、インターネット(?)）
 - オーディオアーカイブの標準規格の策定作業(JEITA)への参画
 - アーカイブ及び公開用システムの構築
 - デジタル化、統合検索等に関する研究開発成果の実装
 - 著作権管理機能の研究開発成果の実装

⇒ 歴史的音盤を将来に亘って聴くことができるように

【連携協力】

日中韓の国立図書館での連携

- 中国国家図書館、韓国国立中央図書館、国立国会図書館(日本)
- 3国が協力して、東アジア圏のデジタルアーカイブを構築して、コンテンツの長期保存と提供が行えるようにする

日中韓連携合意事項(10月24日)

□ メタデータ基準

- 保存のためのメタデータの形式は、各国の基準を尊重する。
- メタデータ交換は、DCベースとする。
- 語彙等の違いは、オントロジー等による吸収を検討する

□ 統合的な情報サービス

- コンテンツは3館で分散して持ち、統合検索できるようにする
- 機械翻訳は研究開発の状況をウオッチ
- 統合検索のプロトコルは、ハーベスト系(OAI-PMH、RSS)、横断検索系(SRU/SOAP、OpenSearch、OpenURL、Z39.50)のどの方法も3国とも可能
- 3国共通のポータルを立ち上げるのではなく、各国のポータルで見せていく
- まずは、歴史的資料から統合検索

□ 長期保存

- DRセンターは、まず各国内で立ち上げ、その後、3国でのDRセンターの立上げを検討する。その際は、国際的なDRの方向性で検討する。
- OAISはフレームワークであり、実装は各国の開発状況により異なる
- 情報パッケージの仕様は、3国とも手探り状態。まだ共通化する段階にない。

□ 情報交換

- 収集・組織化・保存及び提供機能に関する研究開発成果に関する情報
- デジタルアーカイブについての制度化に関する各国状況

【連携協力】

World Digital Library 構想と連携

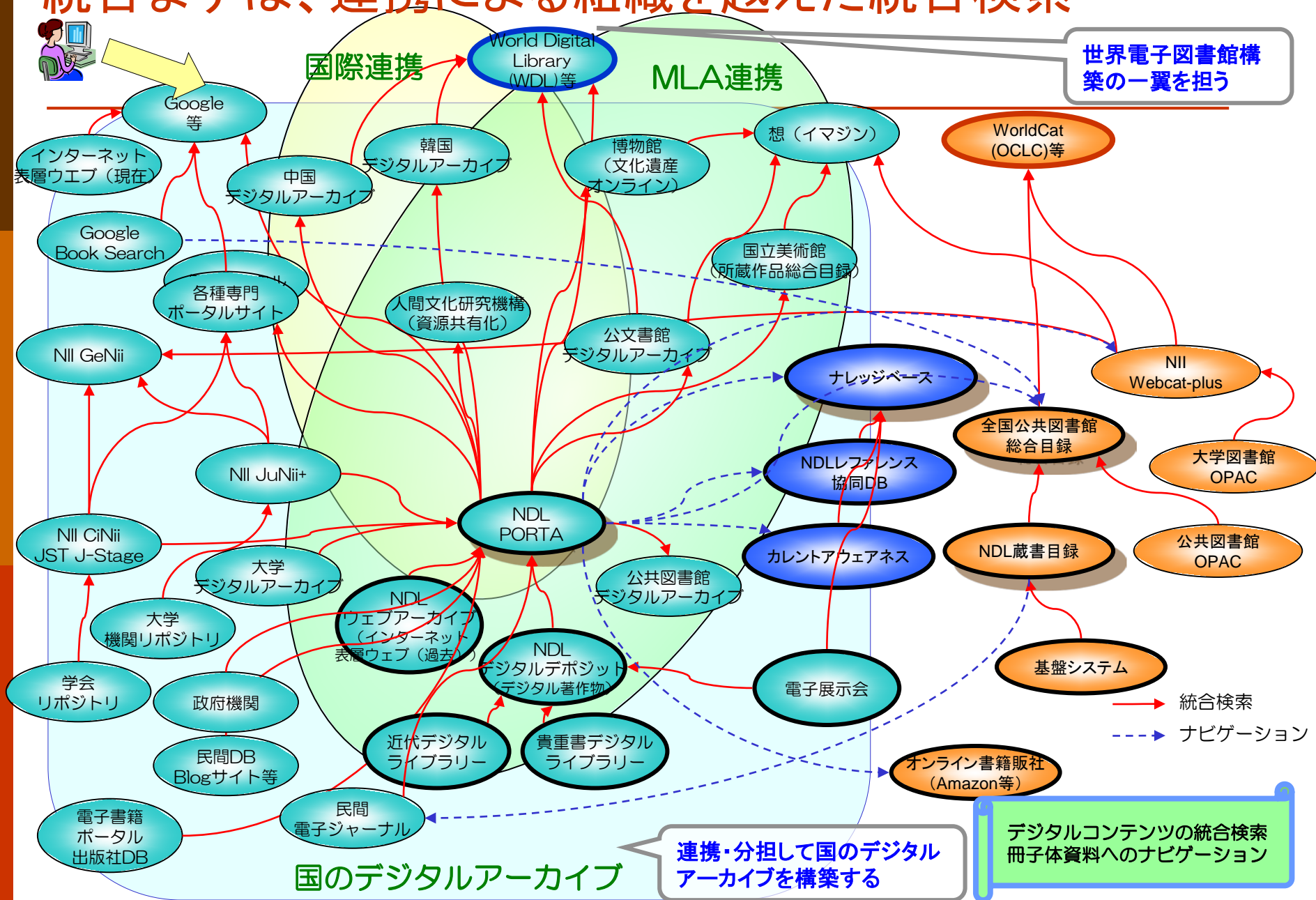
- 米国議会図書館(LC)ビントン館長が打ち出した構想
 - 2005年6月、ユネスコ米国国内委員会のユネスコ米国国内委員会で
- 貴重書など一つしかない歴史的文化遺産を、世界各地の国立図書館等と協同でデジタル化し、無料公開しようとするもの
- 当面のアクション
 - WDLサイトに、テーマを特定したコンテンツを提供
 - 画像データ、メタデータ、解題

⇒ 日本として、東アジアとして参画し、世界規模のデジタルライブラリが構築されることを目指す

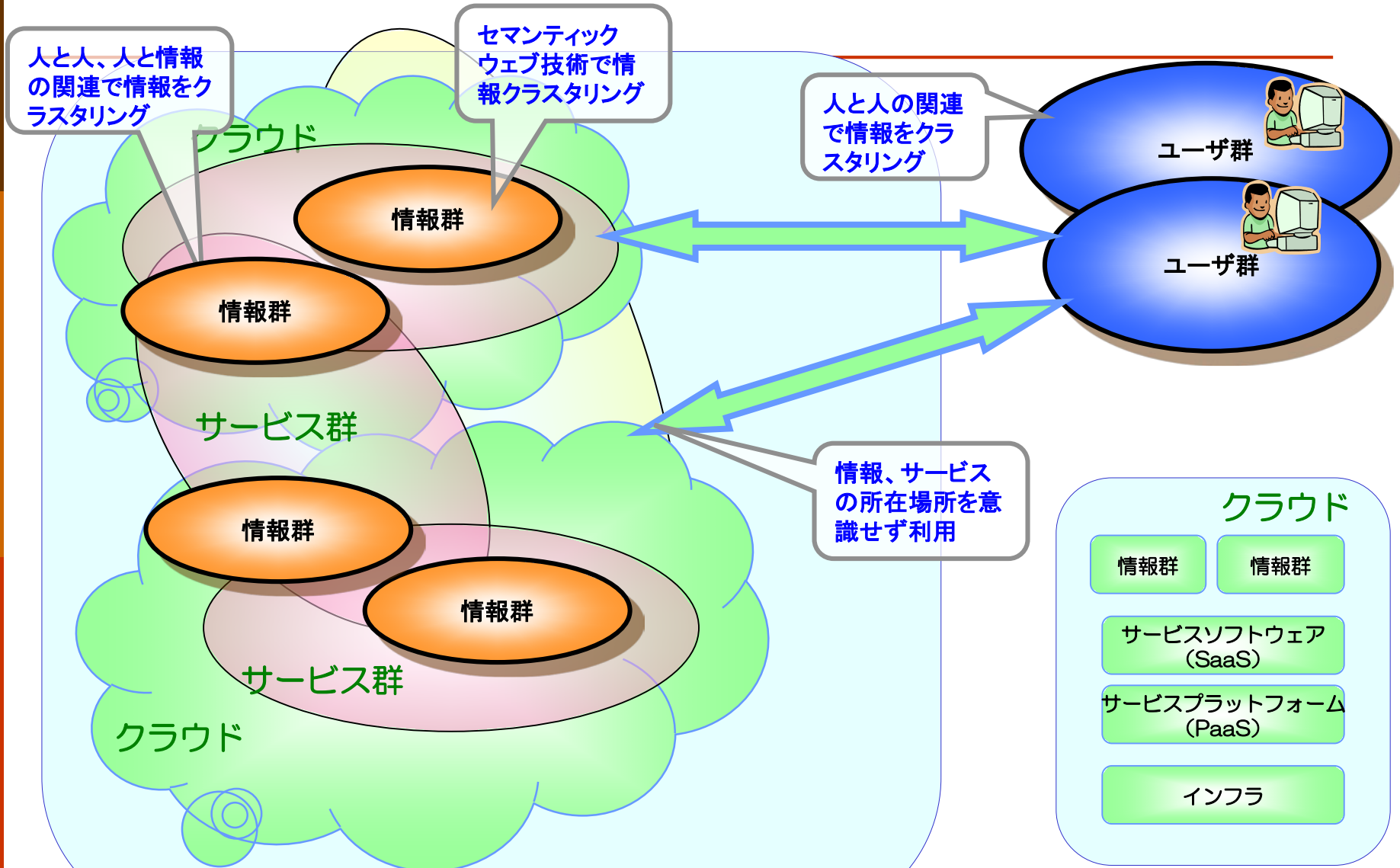
終わりへ

【まとめ】

統合まずは、連携による組織を越えた統合検索



【まとめ】次世代は、 人と情報が関係付けられたサービスの連携



【まとめ】

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		
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				
	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				
PORTA・DAシステムの環境の構築	磁気ストレージ増強		OAIS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		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
				半導体ストレージ導入			

終わりに

関係機関と連携して、国のデジタルコレクションの構築を目指す

- 連携・分担して、人と情報が有機的に関係付けられた利用環境の提供していく
- 分担して、国のコレクションを後世に残していく

□世界レベルでのデジタルアーカイブの構築の一翼を担う

国立国会図書館: <http://www.ndl.go.jp>

PORTA: <http://porta.ndl.go.jp>

ご静聴ありがとうございました