

2016年5月4日

国立国会図書館サーチから
東日本大震災アーカイブへ
そして知識インフラの構築へ【詳細】

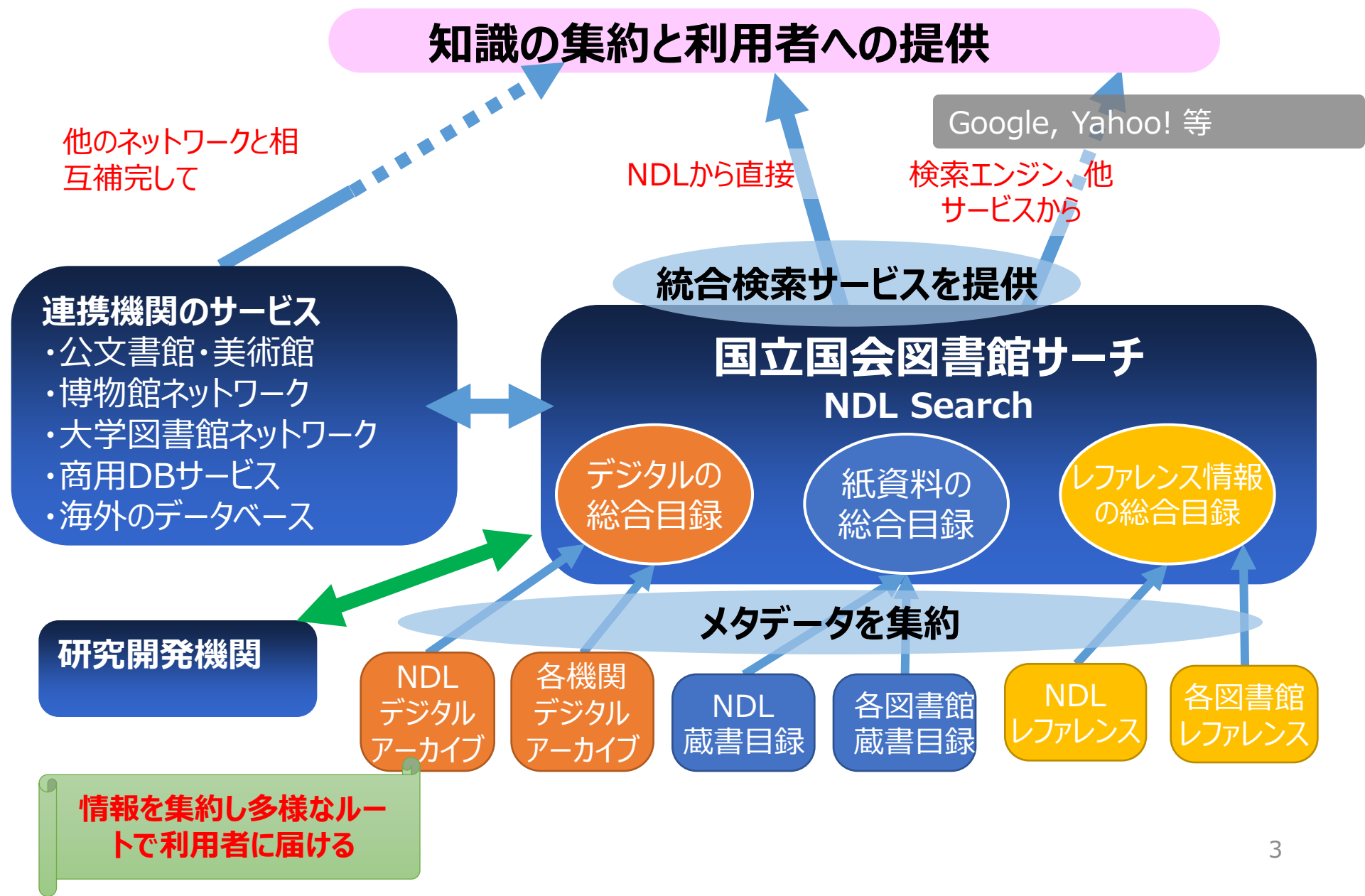
同志社大学大学院総合政策科学研究科
嘱託講師 中山正樹



☆知識インフラ

- ・東日本大震災に限らず、あらゆる分野の記録・記憶を対象に
- ・「ひなぎく」の発展形

☆ 知の提供に向けたNDLサーチのイメージ



近い将来に取り組むべき事項 1/3

国内学術出版物のデジタル化と電子情報資源の収集

- 印刷物のデジタル化を推進し、国民各層の求めに応じた提供
- 電子情報資源に関する収集と保存
- 電子情報資源へ永続的にアクセスしやすい状況を早期に実現
- 出版社、学会、大学等の関連機関の動向に十分留意した上で、持続可能な体制を整備

デジタル化のための環境整備

- 電子情報資源の書誌データ基準の普及、電子情報資源に関する識別子の付与・登録機関の設立推進等を行う
- 国レベルの協議の場の形成に向け、関係機関と協議を進める

電子情報資源の管理・保存

- 電子情報資源についても、「最後の拠り所」として広く国民からのアクセスに応えるために、長期に管理、保存するシステムを構築
- 研究開発や応用、実践に関して、常に最新状況を把握
- 海外の国立図書館等長期保存に取り組む機関との情報共有や、共同研究や調査の連携を実施

参考 近い将来に取り組むべき事項 2/3

電子情報資源の利活用の促進

- 個別システムやデータベースへのリンクや一括して検索を行うシステム
- 科学技術振興機構、国立情報学研究所等関連機関との棲み分けと連携
- 政府が保有する各種統計データを中心とした電子情報資源へのナビゲーションに優先的に取り組む
- 保有する各種電子情報資源のAPI 提供等を推進

従来の所蔵資料・サービスと電子情報資源との有機的連携

- 従来の印刷物を中心とする所蔵資料と電子情報資源との有機的な連携を図り、利用者が資源の種別に関係なく一括して検索でき、その違いを意識することなくシームレスに情報本体へと案内することを目指す。
- 情報検索に習熟していない利用者への支援を進める
- より複雑な情報ニーズを持つ利用者に対しては、そのニーズを的確に判断し、国立国会図書館の所蔵にとどまらず、また、媒体を問わず、適切な資料・情報へと導くレファレンスサービス等の整備を行う。

参考 近い将来に取り組むべき事項 3/3

利用情報の解析と利活用

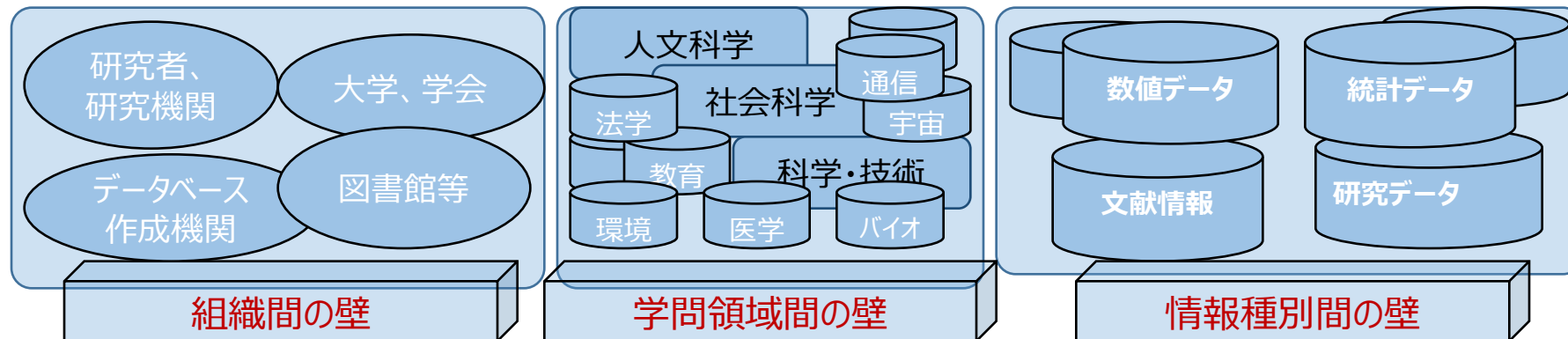
- 検索語、利用された資料名とその頻度、利用した資料・サービスの経路といった利用動向の把握が容易
- 特定の主題分野で利用頻度の多い資料や類似の資料に関する情報を提示
- 個人情報を除く等の統計上の加工を行い利用情報として活用及び提供が可能か、検討

知識インフラのノードとしての社会的な機能の展開

- 国会議員と研究者コミュニティをつなぎ、最新の科学技術の動向を政策立案に組み込むなどの立法府の活動を補佐することを通じて、最終的には科学技術の成果の国民への還元に寄与する。
- 国民が学術情報へ容易にアクセスできることを保障することで、サイエンスコミュニケーションの基盤形成に寄与する。

☆知識インフラの必要性

■ 科学技術情報流通の「壁」 ～現状～



■ 「科学技術基本政策策定の基本方針」*（デジタル時代の科学技術情報の流通） 【総合科学技術会議基本政策専門調査会】（2010年6月）

■ 第4期科学技術基本計画「科学技術に関する基本政策について」 【内閣府 総合科学技術会議(CSTP)答申】（2010年12月24日）

● デジタル情報資源のネットワーク化、データの標準化を進める。

● 学問領域横断的な統合検索、構造化、知識抽出の自動化に向けた研究開発を国全体として推進する。

● 文献から研究データまでの学術情報全体を統合して検索・抽出が可能なシステム（「知識インフラ」）の展開を図る。

☆次世代技術の研究開発成果の活用

次世代サービスが目指す方向性

これまでの単なる「情報検索」から、事実としての「知識検索」へ進化させ、知識の再利用による新たな知識の創造に寄与することを目指す。

増え続けるテキスト、データ、コンテンツ
(国の諸機関の各種資料、統計データ、大学・研究機関の研究成果・研究データ、全国の電子図書館、デジタルアーカイブのコンテンツなど)

様々な形態の知識・情報を組織化し、関連する知識・情報がうまくつながって取り出せる仕組み(知識インフラ)の整備

新しい知識の創造への寄与

研究開発機関

情報の可視化技術

全文テキスト化技術
テキストの構造化技術

情報の収集の効率化技術

分散収集技術
分散横断検索技術

情報の組織化技術

データマイニング技術
メタデータ自動付与技術

情報の集合知化技術

情報と情報の関連付け
意味情報によるクラスタリング

情報検索技術

事実・知識検索
感性検索

閲覧・表示技術

位置情報技術
デバイスに合わせた最適出力

③研究開発における連携

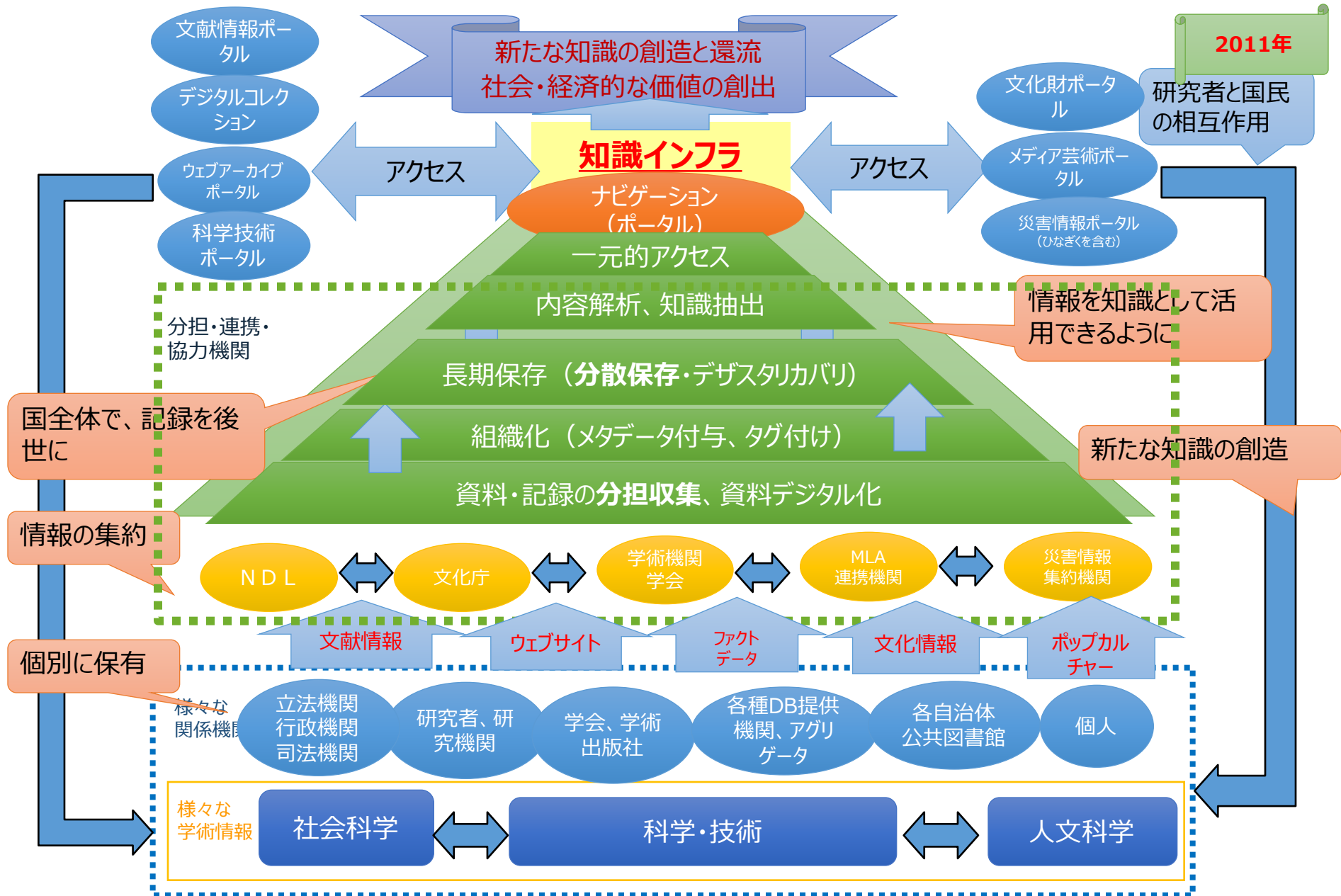
国会図書館

テストベッド (NDLラボ)
実験環境・コンテンツの提供

研究開発成果の適用

国としての知識の利活用の促進を目指した「知識インフラ」の構築と提供の一翼を担う

☆ 知識情報基盤の構築モデル



☆ 知識情報基盤の構築に向けた展開

大震災アー
カイブの構築
を目指して

知識情報基
盤の構築を
目指して

- 大震災アーカイブの延長線上の知識情報基盤を見据えて構築

- 様々な分野の情報資産を知識として利活用し、知識の再生産を支援するサービスの提供を目指す

2008～2011FY

- 業務システム最適化計画に沿ってシステムの再構築

2011FY

- NDLSearchのオープン
- NDL-OPACのリニューアル
- NDLデジタルアーカイブのリニューアル

2012～2013FY

- 大震災アーカイブポータル構築
- NDLSerachを拡張
- NDLデジタルアーカイブを拡張
- 次世代技術を積極的に活用して

2014FY～

- 知識情報基盤構築
- 大震災アーカイブで実証された技術を活用
- 次世代図書館サービスシステムの構築
- NDLSearch、デジタルアーカイブ、ナレッジが融合したシステム
- クラウドコンピューティング環境への移行

東日本大震災アーカイブ

東日本大震災アーカイブの基本理念

2012年

- 東日本大震災の教訓を踏まえた国づくりへの貢献
- 記録保存の重要性についての意識の定着化
- 記録の活用を通じた保存活動の拡大

「防災基本計画」（中央防災会議 2012年9月6日決定）等

国全体としての収
集・保存

官民協力による分
散収集・分散保存

国内外への発信・
永続的伝達

被災地の復興事業、今後の防災・減災
対策や学術研究、教育等への活用

東日本大震災の教訓をふまえた国づくりへの貢献

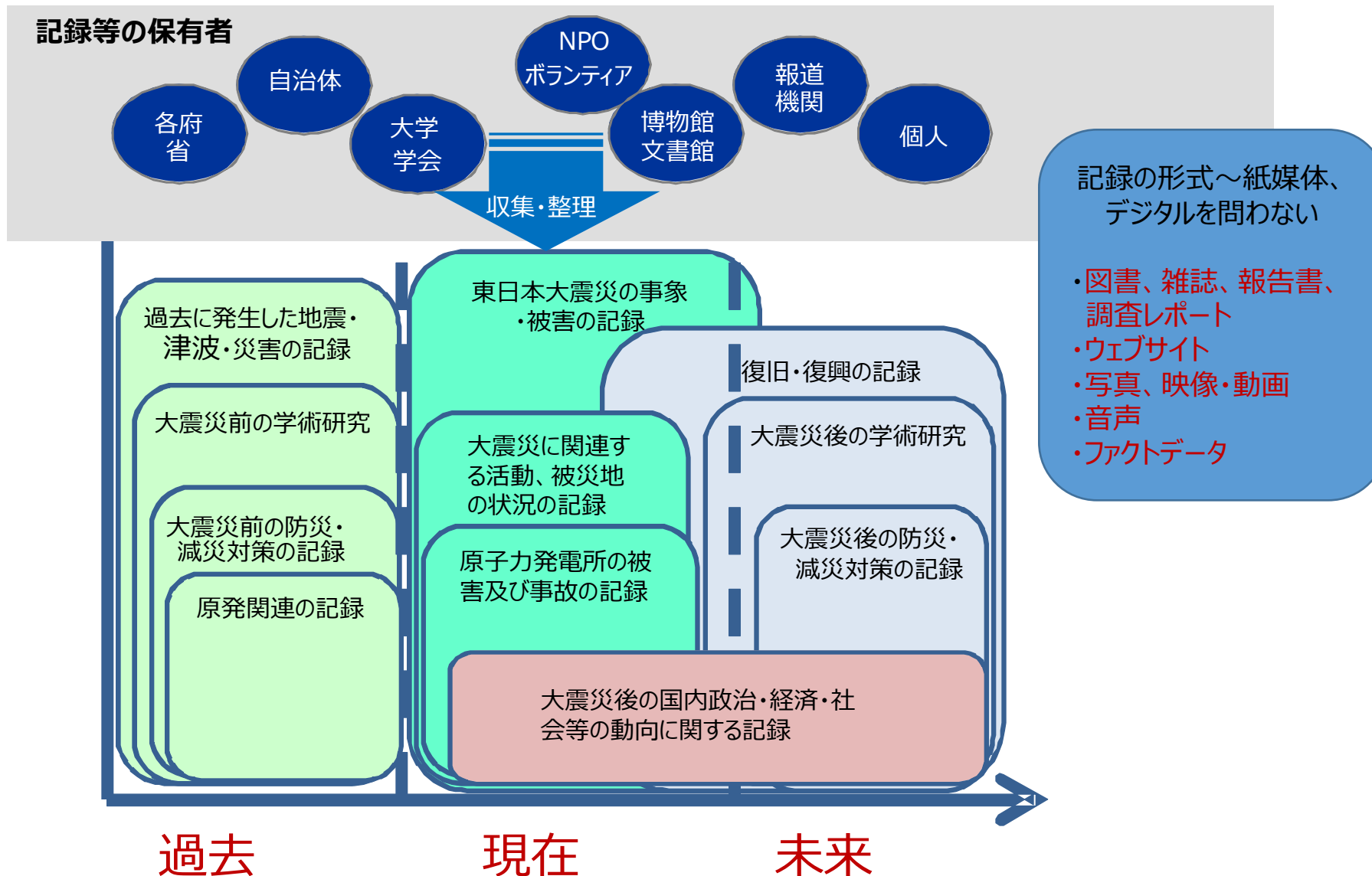
国立国会図書館は**連携の結節点**として機能

国立国会図書館と総務
省が協力して構築

～デジタルアーカイブの連携～ NDL東日本大震災アーカイブの概念



東日本大震災コンテンツ収集対象



コンテンツ収集等の取組

国全体としての収集・保存・提供に向けて

国立国会図書館による記録等の収集

他機関による記録等の保存の推進・支援

他機関が保存する記録等についてのメタデータ収集または機械的連携

これらを実現するための関連機関への働き掛け



国の機関

- 関係各省庁

図書館

- 岩手・宮城・福島各
県立図書館

大学・学会

- 東北大学
- 東京大学
- 日本学術会議

メディア・民間企業

- NHK、Yahoo、
Google
- 日経連

NPO・ボランティア団 体

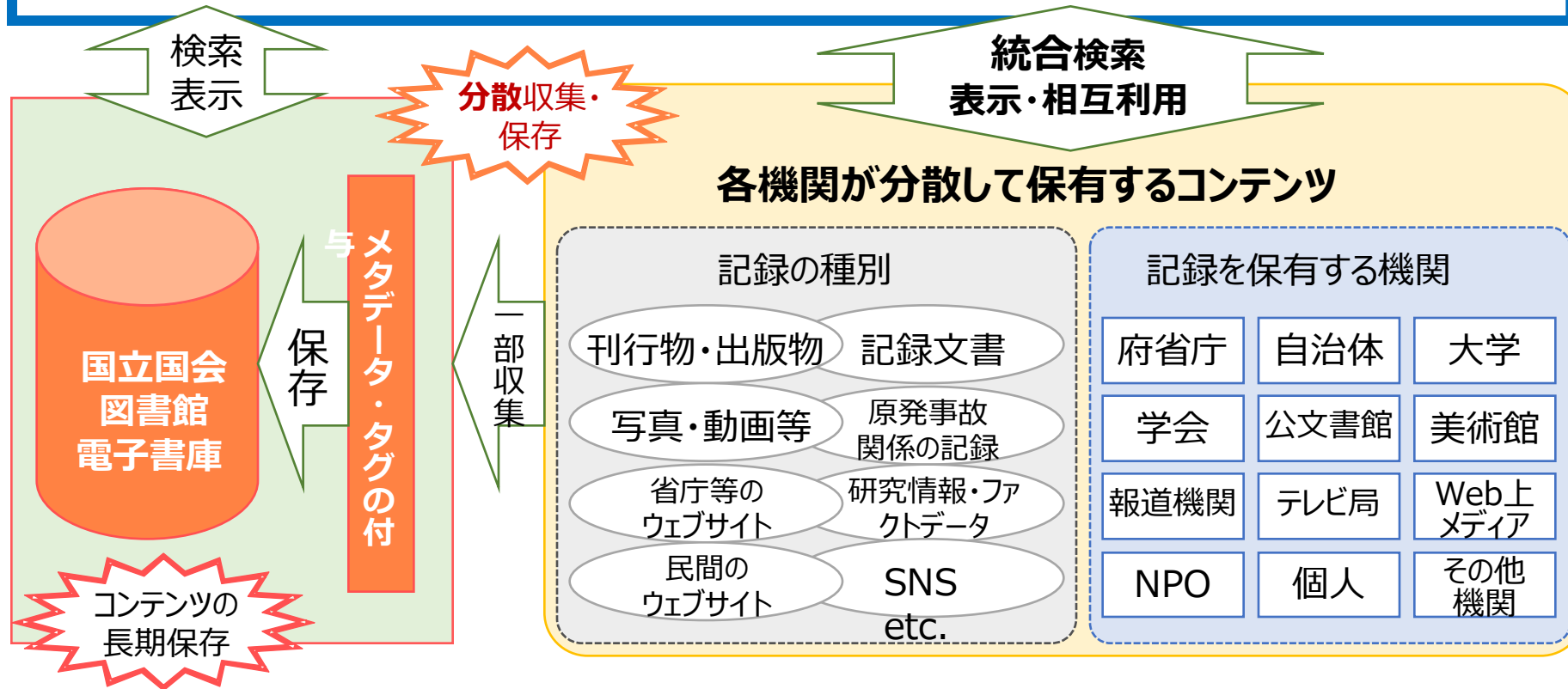
- 日本財団
- 中間支援組織

東日本大震災アーカイブのシステムイメージ

東日本大震災アーカイブポータル

国内外への発信、復興支援
防災対策・研究・教育への活用

- コンテンツの所在の把握、一元的なアクセス
- 文書・ウェブサイト・映像等を、日時・場所・主題等の切り口から検索・表示



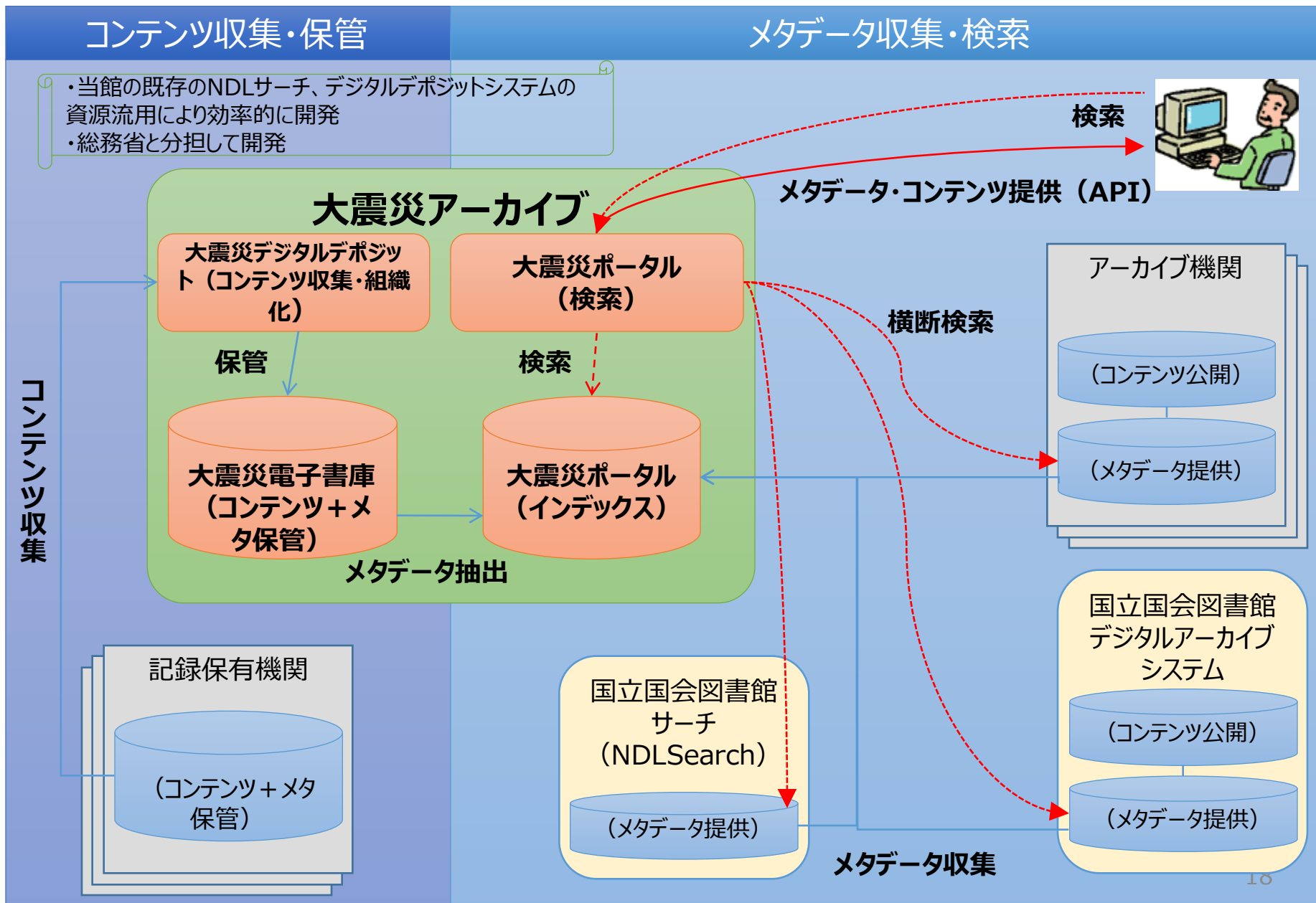
～デジタルアーカイブの連携～ NDL東日本大震災アーカイブの取組



- NDL東日本大震災アーカイブ（ひなぎく）
<http://kn.ndl.go.jp/>
 - 東日本大震災の記録等の国全体としての収集・保存・提供機能
 - 地震・津波災害、原子力災害に関するあらゆる記録を検索・活用できるポータルサイト
 - 文献・資料に限らず、写真、音声、動画などを含む多種多様な記録を対象
 - 関係する官民の機関による分担、連携、協力（分散収集・分散保存）により実施
 - 連携機関39、連携データベース44、検索対象メタデータ数約317万件(2015年9月末現在)
 - 存続できなくなった地方の震災アーカイブをNDLが収集・保存

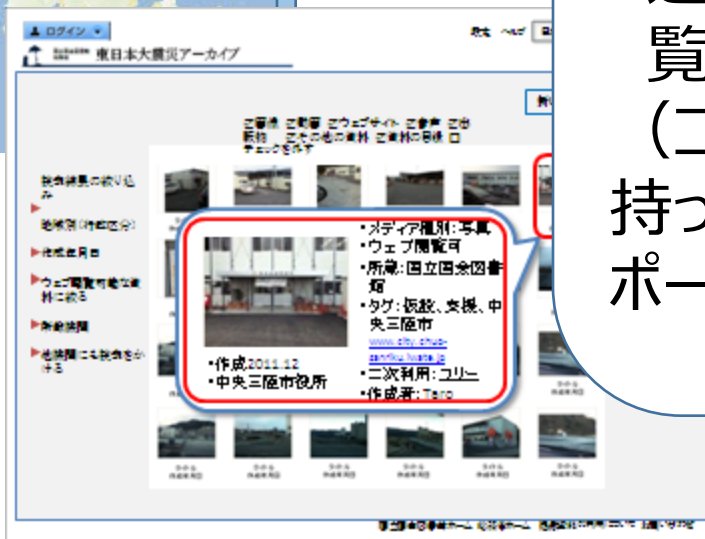


システム機能イメージ



検索・閲覧機能イメージ

- ①地図検索・結果表示、時系列別検索・結果表示及びその組み合わせ
- ②関連キーワード、連想キーワードによる検索や推奨される関連コンテンツの表示
- ③全文テキスト検索
- ④英語・中国語・韓国語への翻訳機能の提供



- ⑤検索結果表示
サムネイル画像やスニペットの表示
(地図・時系列表示)
- ⑥画面遷移を行わない
連携機関コンテンツの閲覧
(コンテンツ提供APIを
持っている場合、大震災
ポータル上で閲覧)

各画面のイメージ ③一覧検索結果画面（一覧表示）

【ファセット】
検索結果の絞込みが可能（11/5時点）

- ①タイトル
- ②作成者/出版者
- ③地点情報
- ④作成日
- ⑤出版年月日
- ⑥種類
- ⑦データプロバイダ
- ⑧Web閲覧可否

【一覧検索結果】
一覧形式で検索結果を表示



【典拠情報検索】

- ・関連キーワード

Web NDL Authoritiesを利用した

た

典拠情報検索が可能

- ・著者名キーワード

インデックスに格納された典拠ID

をもとにした典拠情報検索が可能

【連想検索】

GETAssocを利用した連想検索が可能

【関連リサーチ・ナビ】

リサーチ・ナビを利用した関連ページの検索が可能

【外部サービス検索】

外部サービスを利用したさらなる検索が可能

【検索結果出力】

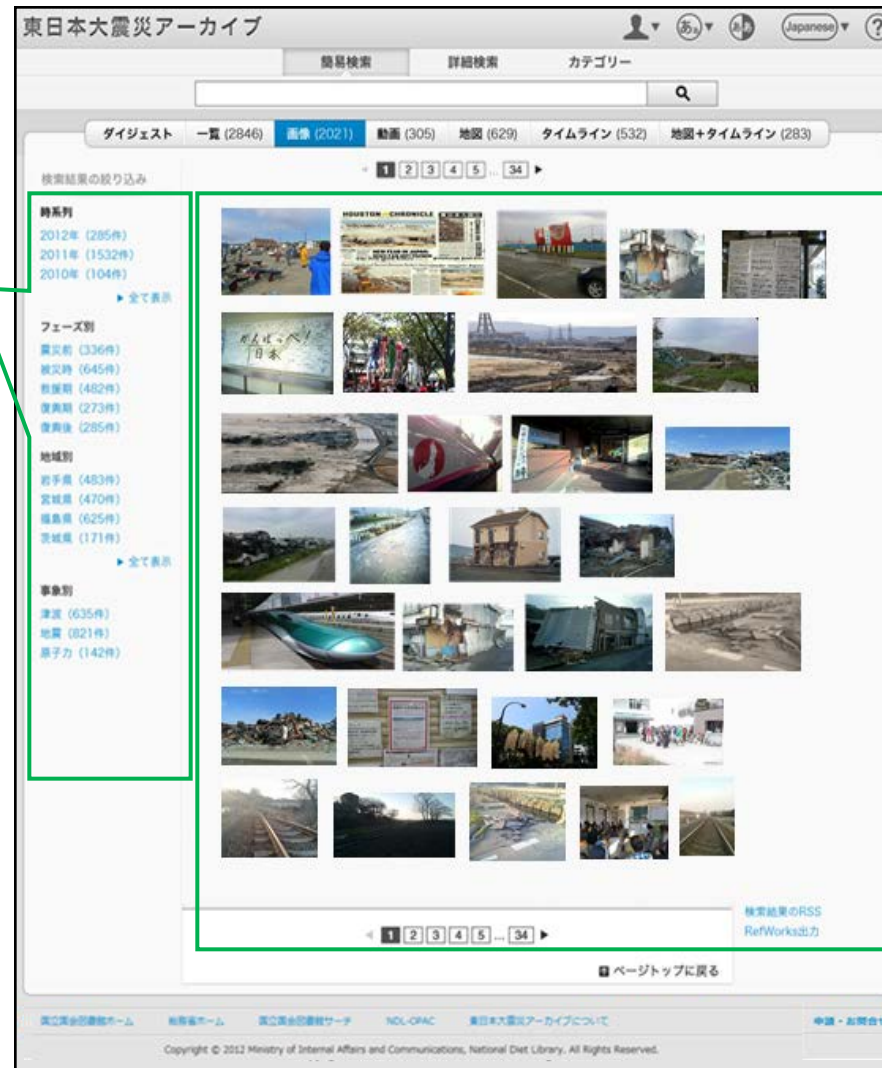
RSS、RefWorksの出力が可能

【凡例】

すべての利用者向けエリア

各画面のイメージ ④画像検索結果画面

【ファセット】
検索結果の
絞り込みが可能



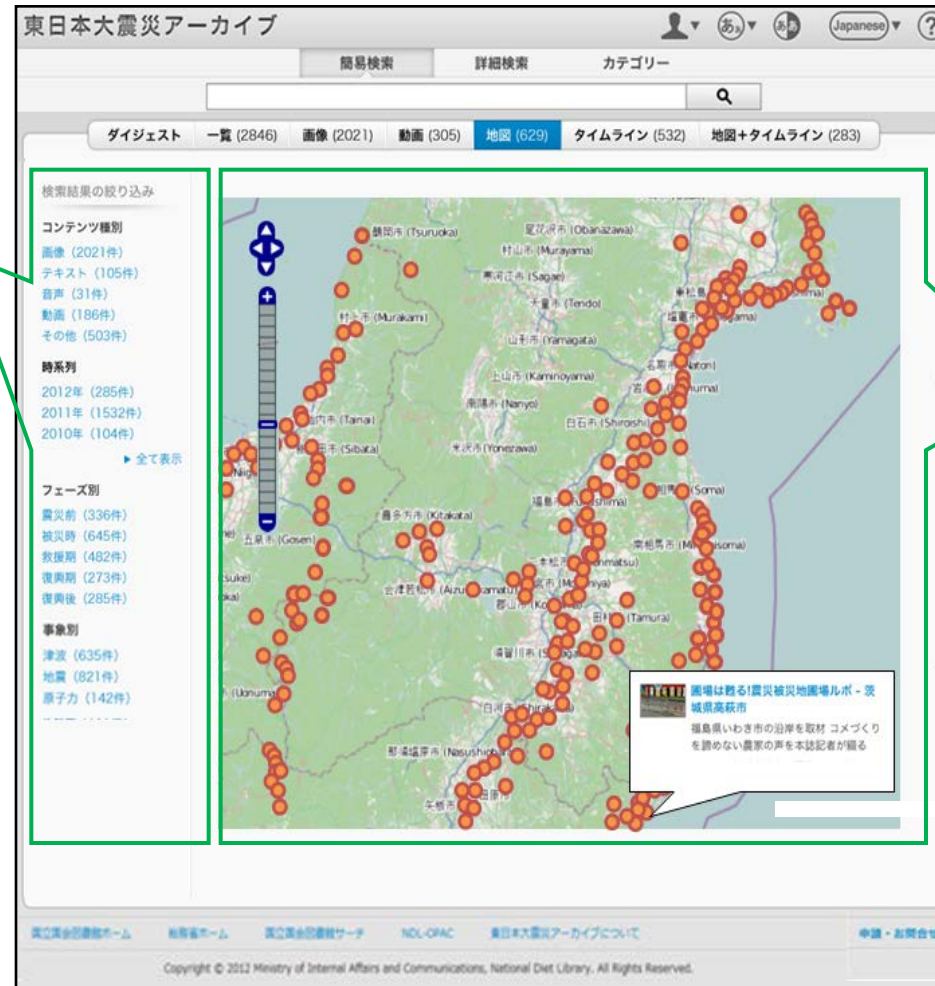
【画像検索結果】
画像に特化した
検索結果を表示

【凡例】

 すべての利用者向けエリア

各画面のイメージ ⑥地図検索結果画面

【ファセット】
検索結果の
絞込みが可能



【地図検索結果】
地域に特化した
検索結果を表示。
データをクリックし
た際、ポップアップ
にてサムネイル及
びメタ情報を表示。

各画面のイメージ ⑦時系列検索結果画面

【ファセット】
検索結果の
絞込みが可能



【時系列検索結果】

日時に特化した検索結果を表示。データをクリックした際、ポップアップにてサムネイル及びメタ情報を表示。

【凡例】

すべての利用者向けエリア

各画面のイメージ ⑧地図時系列検索結果画面

【ファセット】
検索結果の
絞込みが可能



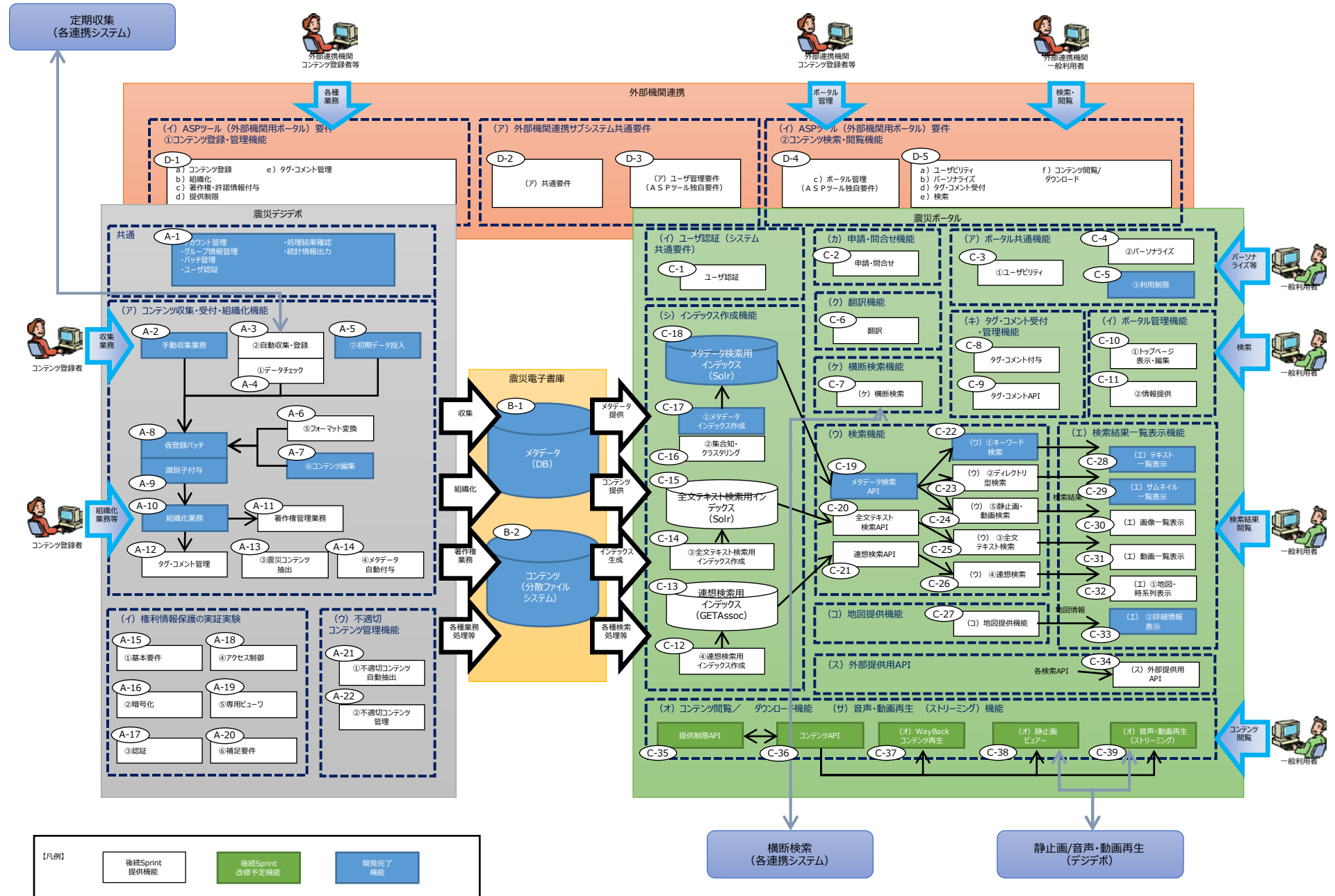
【地図時系列検索
結果】

地域と日時の両
面から検索が可能。
データをクリックした
際、ポップアップにて
サムネイル及びメタ
情報を表示。

【凡例】

すべての利用者向けエリア

機能全体図 (Sprint1リリース時点)



大震災アーカイブの課題と対策

• 課題

- 未だ知られていない記録の把握
- ひなぎくで検索可能な記録が増えない
 - WARPで収集した国、被災地のサイト内の大震災関連の情報が検索できない
 - 各種団体への働きかけを進めているが、実収集へたどり着けない
- 復興支援としての利活用
 - 復興状況のタイムリーな情報提供
- 近い将来での利活用（防災、減災）
 - 関係機関の研究開発成果の技術移転
- 将来的な利活用（記録としての永久保存）
 - 知識インフラとして利用

• 対策

- **WARPのコンテンツの可視化**
 - 「ひなぎく」から横断検索もしくはWARPへのナビゲーション
- インターネット上の大震災関連記録の発見
 - 利用を促進することにより、不足している記録の発見
 - 利用者がインターネット上で発見した記録をソーシャルブックマーク
 - ソーシャルブックマークされたページを「ひなぎく」に登録し検索対象にし、ナビゲート先の拡大
- 発見した記録へのナビゲーションの拡大により、更なる利用の促進
 - ウェブ上からページが消滅する前に、可能な限り許諾を得て収集

「ひなぎく」での課題（１）

ナショナルアーカイブの構築の課題として認識

• 技術面での課題

- 大規模分散ファイルシステム、分散処理システムの適用⇒実際には、想定ほど集まっていない
- NDLサーチおよびNDLデジタルコレクションの発展形として開発⇒Factデータに関してはメタデータまで
- ウェブコンテンツの可視化⇒ウェブアーカイブからの大震災関連の情報の自動切出しが困難
- 検索およびハーベスト用のAPIの未実装、不整合
- 収集前データに永続的識別子がない
- メタデータの記述規則の差異、記述要素の不足
 - 組織、アーカイブごとのメタデータの差異、画像、写真等、メタデータが不完全
- 不完全なメタデータへの自動付与
- 利活用のための検索・閲覧機能⇒本文テキスト、イメージ認識技術等により、内容を関連付けた検索技術が必要

• 人材面での課題

- 専門分野に関する知見、文化資産の収集・保存・修復・公開の技能の不足
- 文化資産を取り扱うための知識・技能の不足（プリザベーションエンジニア、コーディネータ、エンベデッドライブラリアン）
- デジタル技術を活用したアーカイブ化のための知見の不足（アーキビスト、法規担当）
- システム開発・運用管理の一般的な知識・技能の不足（システムライブラリアン、ITエンジニア）

「ひなぎく」での課題（２）

ナショナルアーカイブの構築の課題として認識

- 協力体制の課題
 - 縦割り行政
 - ⇒立法府であるNDLが進めることに警戒感
 - 表現・報道の自由
 - ⇒「放送アーカイブ」とする構想に対する報道圧力・事後検閲の可能性に対する警戒感
- 制度面・運用面での課題
 - 情報公開法
 - 行政文書管理簿に掲載されていない軽微な資料は、保存期限は1年未満
 - 公文書管理法
 - 軽微な資料も歴史的公文書と指定されなければ公文書館に移管されない
 - 肖像権、プライバシー権、人格権
 - 権利処理がされなければ、収集できない（ダークアーカイブ）
 - 維持できなくなったアーカイブのコンテンツを預かることもできない
 - 国有財産法、財政法
 - 維持できなくなったアーカイブのために、Webサーバやストレージを貸し出すことができない
 - 民間および個人のサイトは個別許諾に基づく収集
 - 個別に許諾手続きが必要なため、網羅的な収集が困難

次世代サービスの実現に向けた過 渡期の計画

背景：検索エンジンの高度化と情報資源の多様化により、 図書館の情報提供の役割が大きく変化

資料

- 所蔵資料のデジタル化の推進と提供
- 制度的に収集した紙資料とオンライン資料の一体的な提供
- ライセンス契約電子資料
- オープンアクセス資料と当館の情報資源を統合的に検索・参照できる環境の整備

遠隔

- 複写物の電子的形態での送信
- 情報発信型サービスの改善
(調べ方案内)
- 個別回答型サービスの改善
(オンラインレファレンス)

来館

- 情報提供機能の充実
- 利用者用端末
- 閲覧体制
- ホール及び専門室の機能の整理
- 18歳未満者への対応
- 三施設の役割分担と一体化の促進

図書館協力

- 図書館間貸出し
- 図書館送信サービス
- 遠隔複写サービス
- 研修の充実
- 障害者サービスの展開

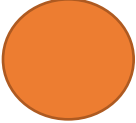
統合的オンラインサービス

登録利用者制度



☆統合的オンラインサービス：概要

- 利用者サービス部から提案したNDLの次期検索サービスの構想
- H29から段階的に実施する業務基盤システムのリニューアルが完了するH32時点での実現レベルを想定
- 一次情報とそれを利用するための各種機能（申込機能、閲覧機能 等）が個別システムに散在している状況を改善し、**求める一次情報を利用者自身が効率良く選択・入手**することを可能にするサービス
- 特に実現したいのは以下の3点
 - ① **NDLサーチとNDL-OPACの2系統の検索サービスの統合**
⇒NDLの一次情報資源へのアクセス向上を図る
 - ② **契約電子ジャーナルの記事情報、目次情報等の統合的な検索**
⇒NDL内外の一次情報資源のネットワーク利用を促進する
 - ③ **調べ方案内等の主題情報と書誌情報のシームレスな結合**
⇒他のウェブサービスとの差別化を図る
- ナショナルアーカイブ構想のうち、特に以下の箇所を具体化したもの
 - **文献・ウェブ情報の検索サイト**（＝情報発信基盤の一部）
 - **新たな知識創造の場としてのリサーチ・ナビ等の活用**（＝知識創造基盤の一部）



☆ 統合的オンラインサービス：実現したい機能

① NDLサーチとNDL-OPACの2系統の検索サービスの統合

- 利用者種別（来館、遠隔 等）と資料形態（紙資料、デジタル化資料、電子ジャーナル 等）に応じて、**可能な一次情報入手手段**（閲覧申込み、遠隔複写申込み、デジタル資料の閲覧 等）**ヘナビゲート**する
- 遠隔利用者用インタフェースに加え、来館すると利用できる資料・申込みに特化した**来館利用者用インタフェース**を設ける
- 統合的オンラインサービス内で閲覧・複写申込みまで実行可能とする

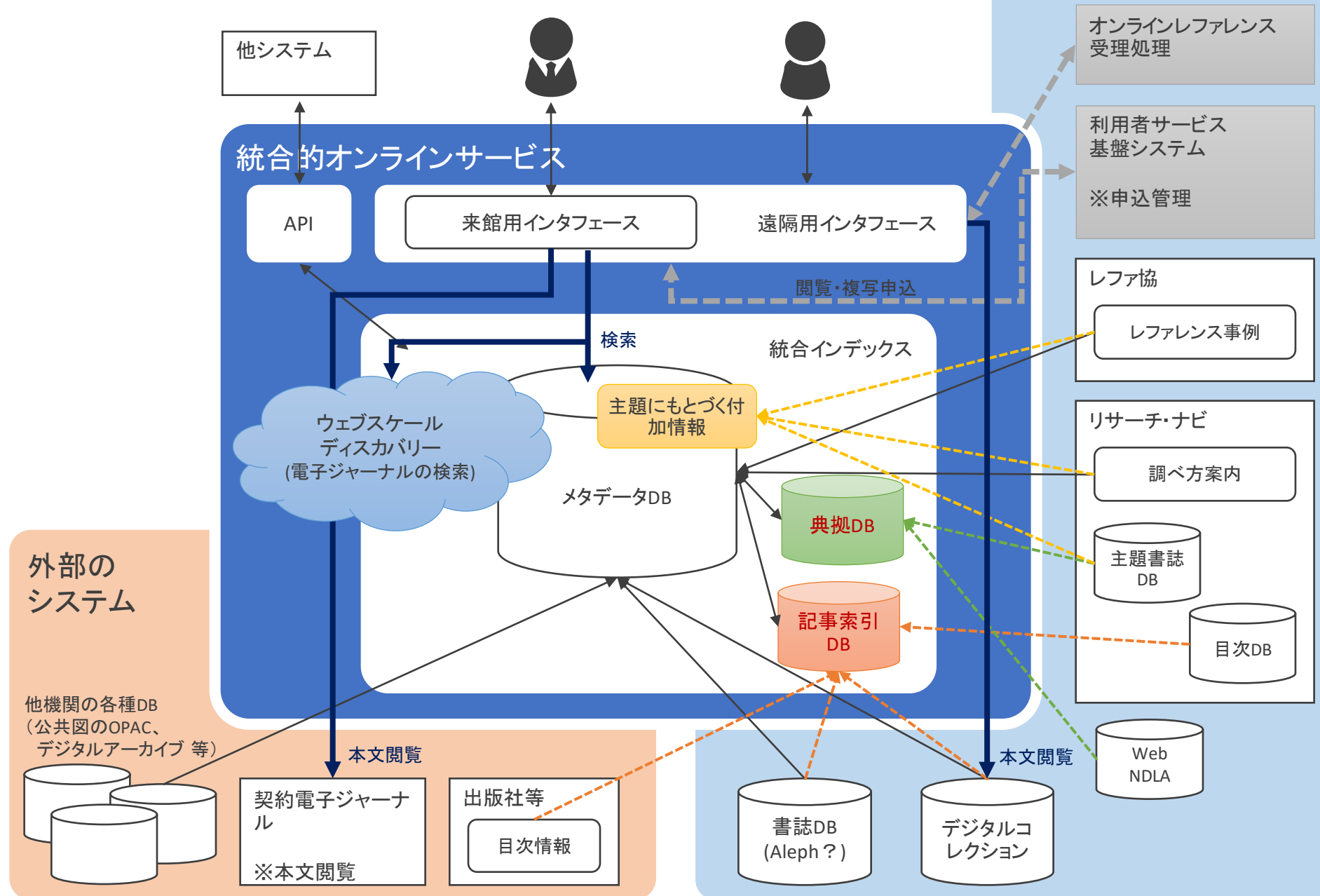
② 契約電子ジャーナルの記事情報、目次情報等の統合的な検索

- ウェブスケールディスカバリーサービス（例：Summon、WorldCat Local、Primo Central、EBSCO Discovery Service）の導入等により、**契約電子ジャーナルの記事情報の統合検索**を実現する
- リサーチ・ナビの目次DB、デジタルコレクション、雑誌記事索引、出版社等外部から収集した**目次情報を統合・組織化**し、書誌情報と目次情報の一体的な検索を実現する

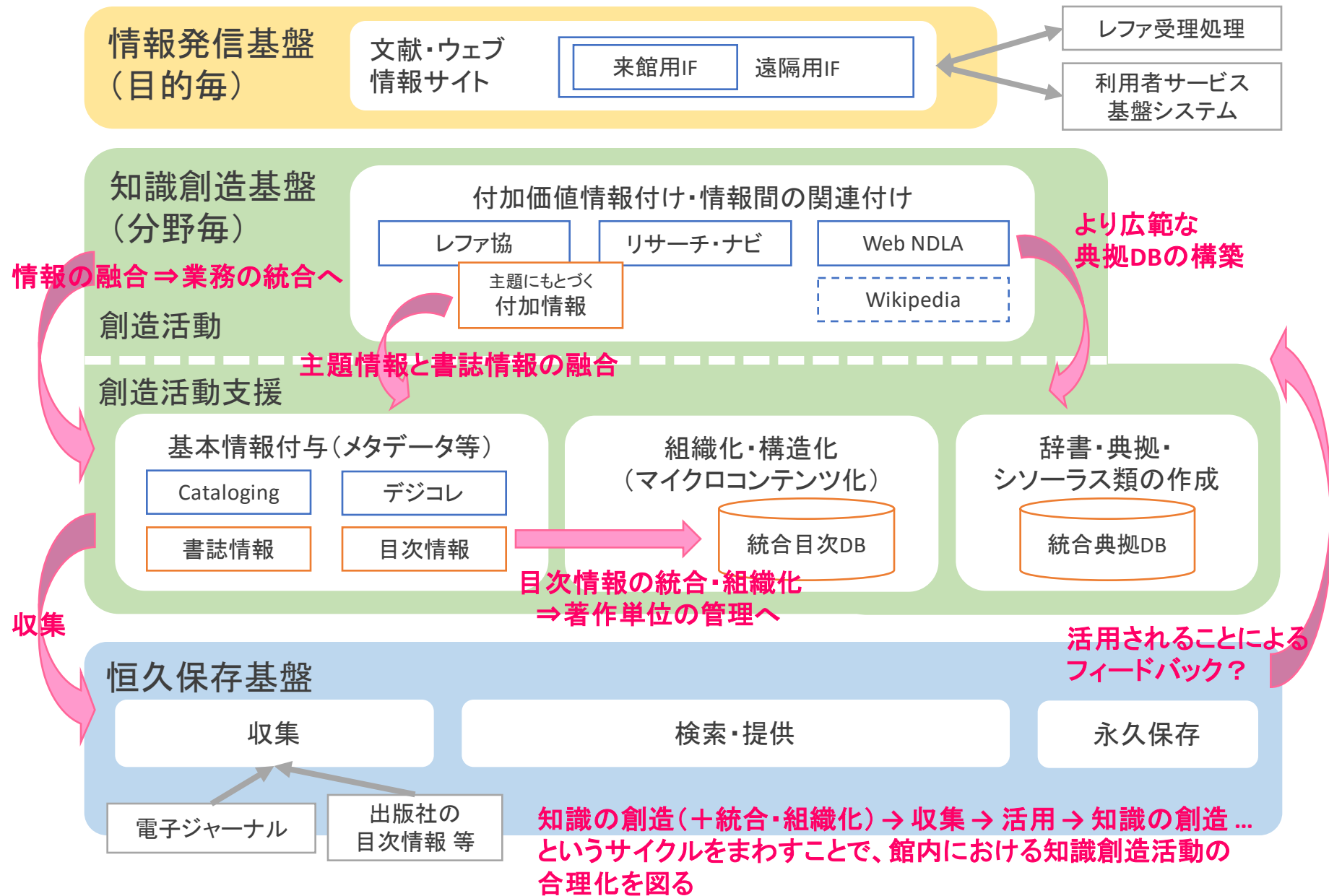
③ 調べ方案内等の主題情報と書誌情報のシームレスな結合

- 調べ方案内やレファレンス事例で紹介されている資料について、**書誌情報と調べ方案内・レファレンス事例を相互にリンク**し、利用者が書誌情報と主題情報をシームレスに行き来しながら情報探索することを可能にする
- リサーチ・ナビの主題書誌DBに登録された人名や分類、Web NDLA、Wikipedia 等を融合した**典拠DBを構築し、典拠コントロールされた検索機能**を実現する

☆ 統合的オンラインサービス：実現イメージ



ナショナルアーカイブ構想へのマッピング



5. NDLが保有する主題情報まとめ（主なもののみ）

【表1】NDLが保有する主な主題情報（サービス・システム別）

サービス・システム名	含まれる情報
リサーチ・ナビ	
調べ方案内	各主題の調べ物に役立つ参考資料・ウェブサイトとその解説
主題書誌DB	参考資料の内容情報（解題、目次、独自分類、掲載人名 等）
レファレンス協同データベース	
レファレンス事例	レファレンスのQ&A、参照した資料・ウェブサイト
調べ方マニュアル	レファレンス協同データベース参加館が作成した調べ方案内
その他	
Books on Japan	日本に関する欧文の資料のリスト

【表2】NDLが保有する主な主題情報（情報の性質別）

情報の種類	内容
① 書誌情報を補強する情報	<ul style="list-style-type: none">・ 主題書誌DBの内容情報（解題、目次、独自分類、掲載人名 等）・ 調べ方案内やレファレンス事例で各資料に付された説明文
② 資料の有用性に関する情報	<ul style="list-style-type: none">・ 調べ方案内、主題書誌DB、Books on Japanに掲載されている資料・ レファレンスの回答に使用した資料 ＝当該資料は参考資料である（＝調べ物に有用である）という情報
③ コンテンツそのもの	<ul style="list-style-type: none">・ 調べ方案内やレファレンス事例の各記事
④ メタ情報	<ul style="list-style-type: none">・ 調べ方案内で紹介されているウェブサイトの情報・ レファレンスの際に参照したウェブサイトの情報 ＝調べ物に有用なウェブサイトのメタ情報

6. 主題情報の分解・利活用案

① 書誌情報を補強する情報

主題書誌DBの解題、目次／調べ方案内やレファレンス事例で各資料に付された説明文

- 書誌ID等をキーに同一資料についての情報を同定、書誌情報とマッチングし、書誌情報と統合的に検索・表示

主題書誌DBの独自分類、掲載人名

- Web NDLA、Wikipedia 等と融合した典拠DBを構築し、典拠コントロールされた検索機能を実装

② 資料の有用性に関する情報

調べ方案内、主題書誌DB、Books on Japanに掲載されている資料／レファレンスの回答に使用した資料 = 当該資料は参考資料である（＝調べ物に有用である）という情報

- 書誌IDをキーに当該資料の書誌情報にフラグを付与し、検索・絞り込み条件、GUIでの参考資料マーク（？）表示等に活用

③ コンテンツそのもの

調べ方案内やレファレンス事例の各記事

- 書誌情報と同粒度の情報として検索対象とし、資料が紹介されている箇所は書誌情報と相互リンク

④ メタ情報

調べ方案内で紹介されているウェブサイトの情報／レファレンスの際に参照したウェブサイトの情報 = 調べ物に有用なウェブサイトのメタ情報

- 書誌情報と同粒度の情報として検索対象とし、出典となる調べ方案内等と相互リンク

背景: 検索エンジンの高度化と情報資源の多様化により、図書館の情報提供の役割が大きく変化

資料

- 所蔵資料のデジタル化の推進と提供
- 制度的に収集した紙資料とオンライン資料の一体的な提供
- ライセンス契約電子資料
- オープンアクセス資料と当館の情報資源を統合的に検索・参照できる環境の整備

遠隔

- 複写物の電子的形態での送信
- 情報発信型サービスの改善（調べ方案内）
- 個別回答型サービスの改善（オンラインレファレンス）

来館

- 情報提供機能の充実
- 利用者用端末
- 閲覧体制
- ホール及び専門室の機能の整理
- 18歳未満者への対応
- 三施設の役割分担と一体化の促進

図書館協力

- 図書館間貸出し
- 図書館送信サービス
- 遠隔複写サービス
- 研修の充実
- 障害者サービスの展開

統合的オンラインサービス

登録利用者制度