DAW19-20知識インフラの構築

2016年5月29日

中山正樹

内容

[図書館情報学研究(図書館システム・オープンデータ) 1](#_Toc452311445)

[1. 東日本大震災アーカイブ（ひなぎく） 2](#_Toc452311446)

[1.1. 基本理念 2](#_Toc452311447)

[1.2. 東日本大震災アーカイブの全体イメージ 3](#_Toc452311448)

[1.3. NDL東日本大震災アーカイブの取組 4](#_Toc452311449)

[1.4. 「ひなぎく」での課題 4](#_Toc452311450)

[2. 知識インフラ 5](#_Toc452311451)

[2.1. 知の提供に向けたNDLサーチのイメージ 5](#_Toc452311452)

[2.2. 知識インフラの必要性 7](#_Toc452311453)

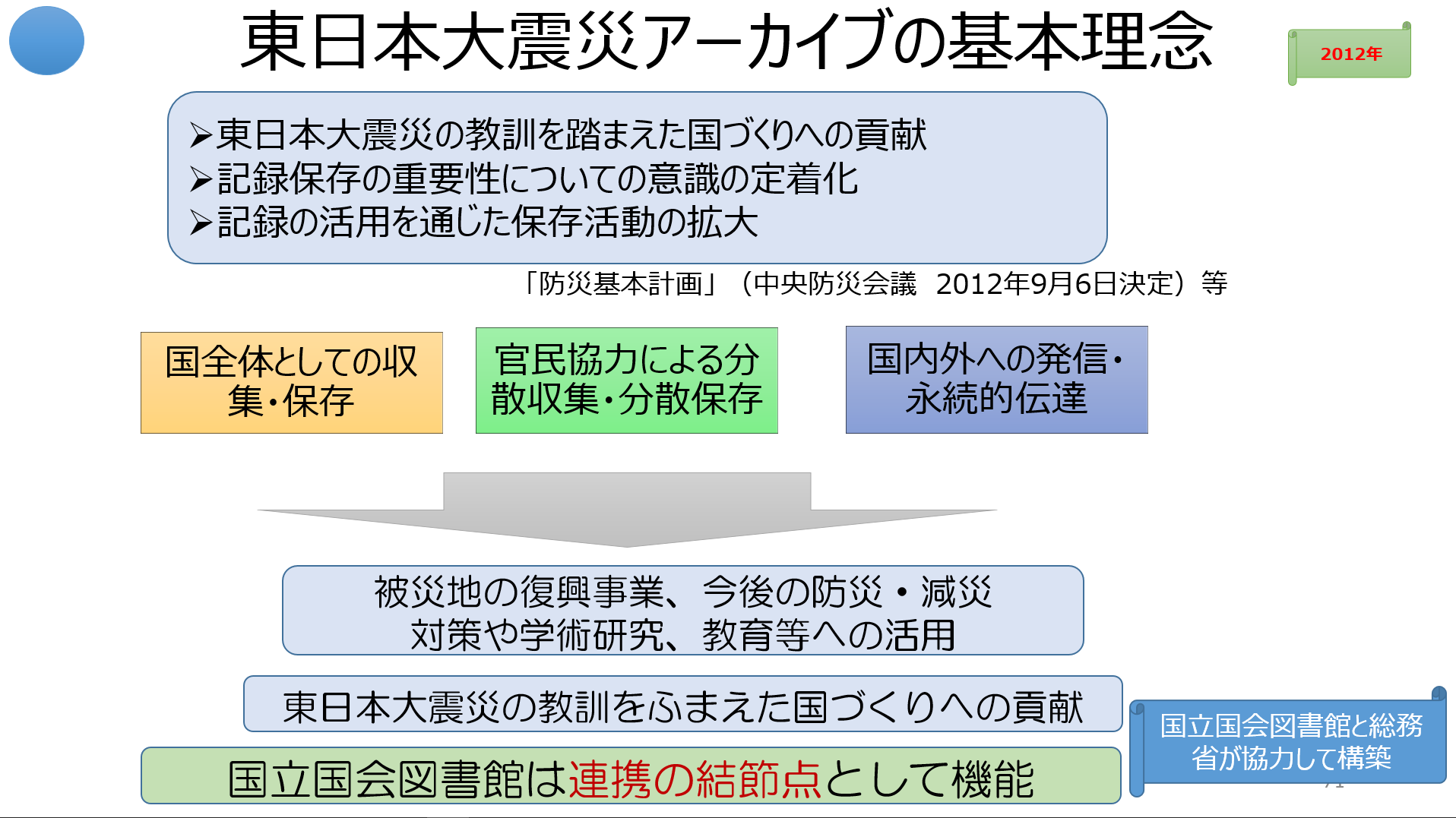
[2.3. 次世代技術の研究開発成果の活用 8](#_Toc452311454)

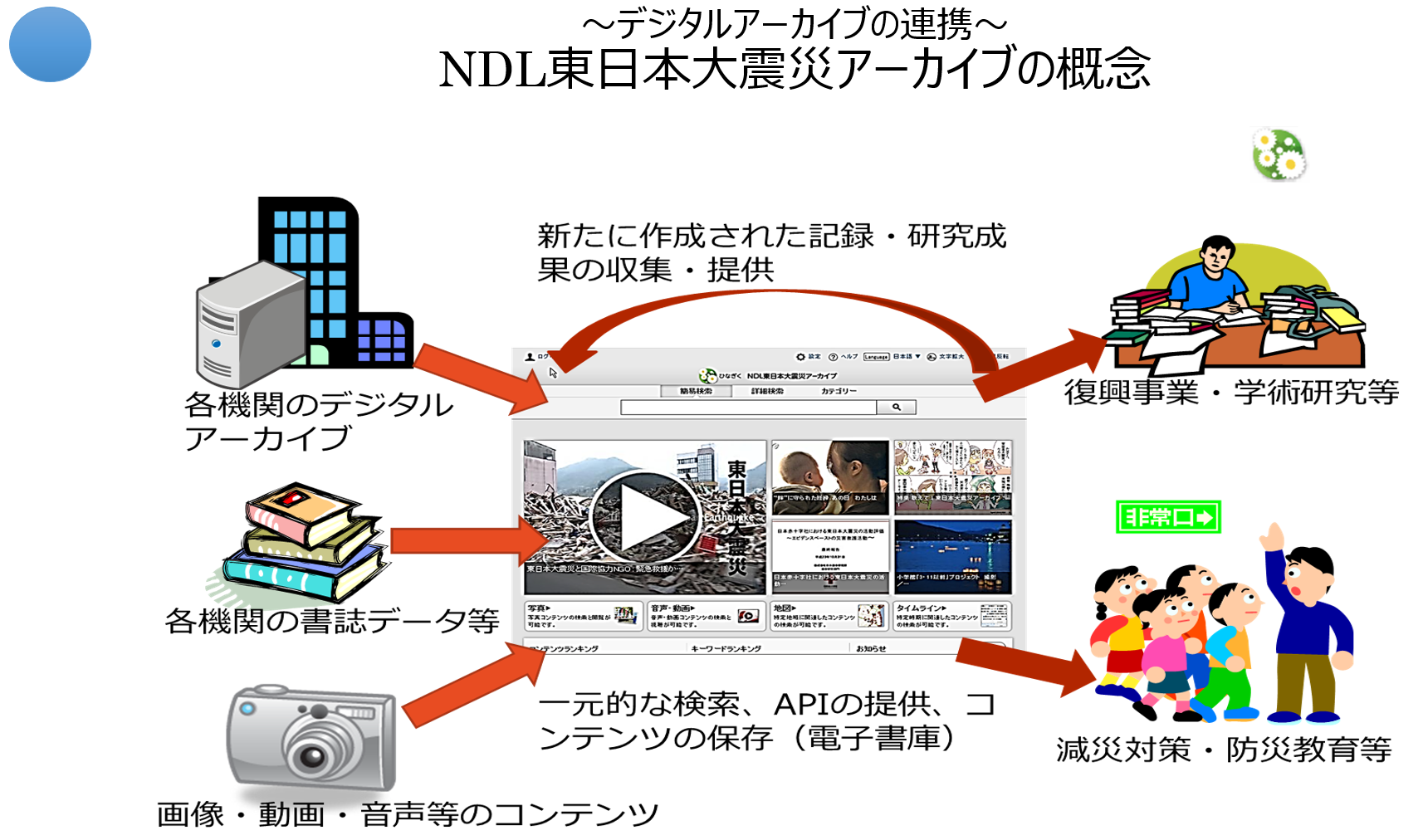
[2.4. 知識情報基盤の構築モデル 9](#_Toc452311455)

[2.5. 知識情報基盤の構築に向けた展開 10](#_Toc452311456)

# 東日本大震災アーカイブ（ひなぎく）

## 基本理念





・文献に限らず、大震災に関するあらゆる記録・記憶を分担収集・保存し、将来にわたって利用を保証する「知識インフラ」構築の特定分野の実現形の１つ

・国立国会図書館サーチの発展形

＝＝＝＝

●大震災に関連する災害現象そのもの、災害前・災害直後・復興の過程、災害時の対応、他地域・次世代への教訓等のあらゆる記録を後世に残す

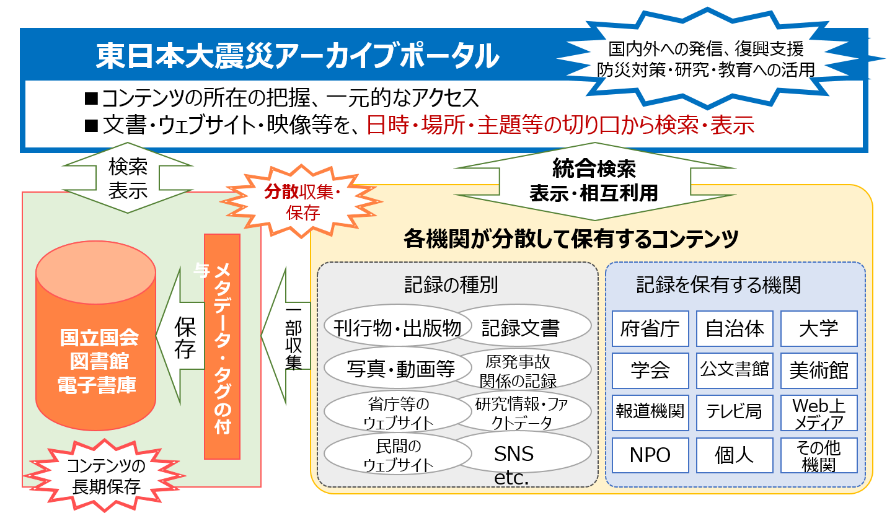
●今後の防災に生かすため、関係府省、各種震災関連情報の保有機関と協力して分担収集・保存し、一元的に検索・閲覧できるように

●媒体、形式を問わず

・

●分散処理システム(Hadoop, GlusterFS)、分散ファイシステム

## 東日本大震災アーカイブの全体イメージ



*2011年3月には、東日本大震災が発災し、甚大な被害をもたらしました。*

この被災・復興の記録を後世に残すとともに、今後の防災・減災に役立てるように、**知識インフラの構築の一環で、分野を特定した知識インフラの実現形の先行事例**として、2013年3月には、大震災に関するあらゆる記録、記憶を保存し、一元的に検索できるようにする「東日本大震災アーカイブ（ひなぎく）」を構築しました。

～～～

構築に当たっては、図書館サービスを効率的、効果的に進められるように、次世代技術の実用化実証実験に取り組み、成果の積極的な活用を図りました。

・映像・観測データ等を受け取ることを想定して、処理能力、ストレージ容量を必要に応じて増強できる仕組み、として、分散処理サーバ、分散ファイルシステムを導入

・大震災アーカイブ自身が、災害で消失してしまわないように、ディザスタリカバリも考慮

・画像・映像なども的確に検索ができるように、明確なメタデータが付与されていない情報にも可能な限りメタデータを自動付与する仕組み

・本文も含めたテキストの全文インデキシング等を試行した

*・*最も次世代の図書館システムを指向したシステム

＝＝＝＝＝

分散保存、一元的アクセス

各機関が分散して保有するコンテンツ

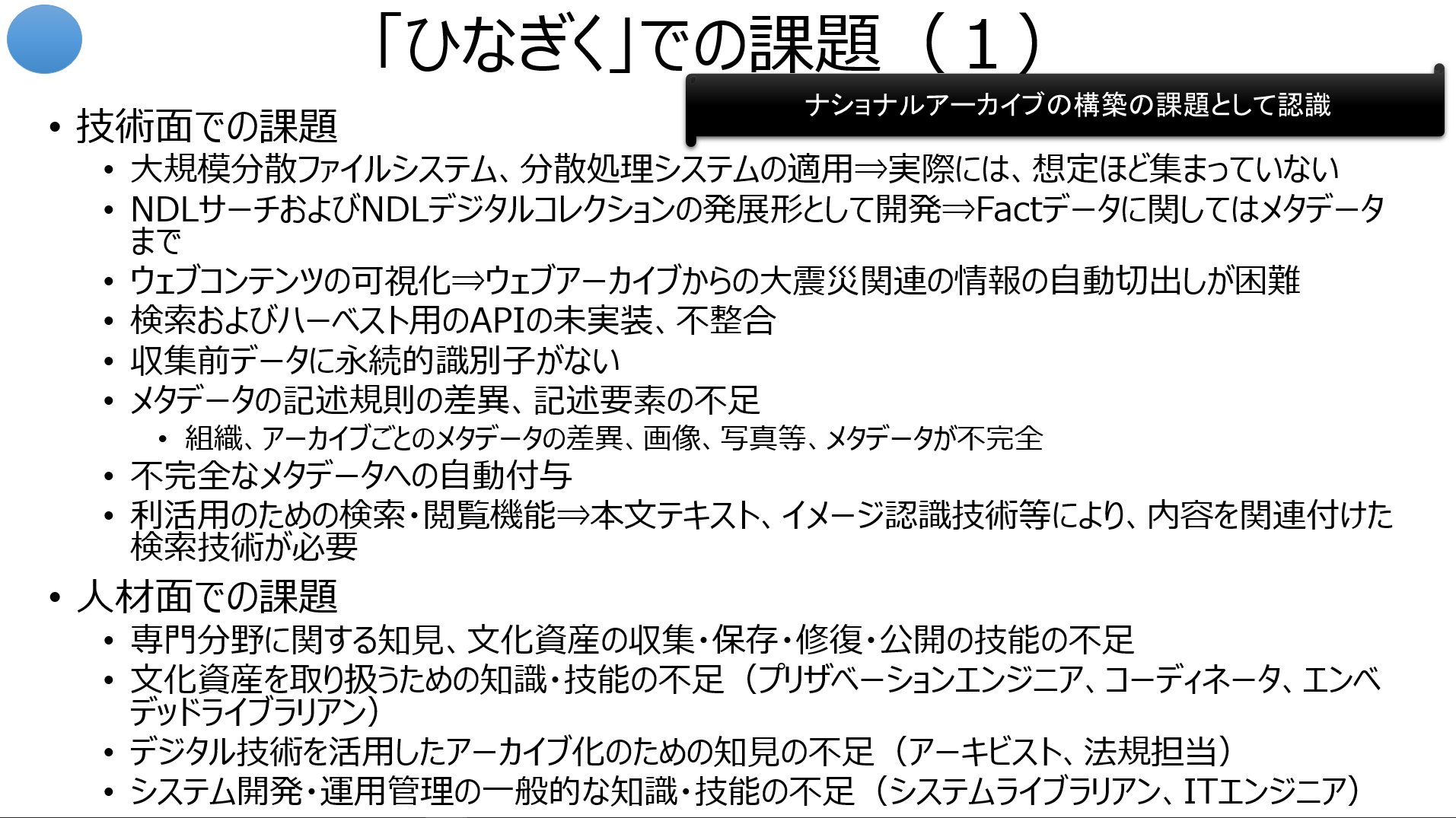
NDLで一部収集し、分散保存

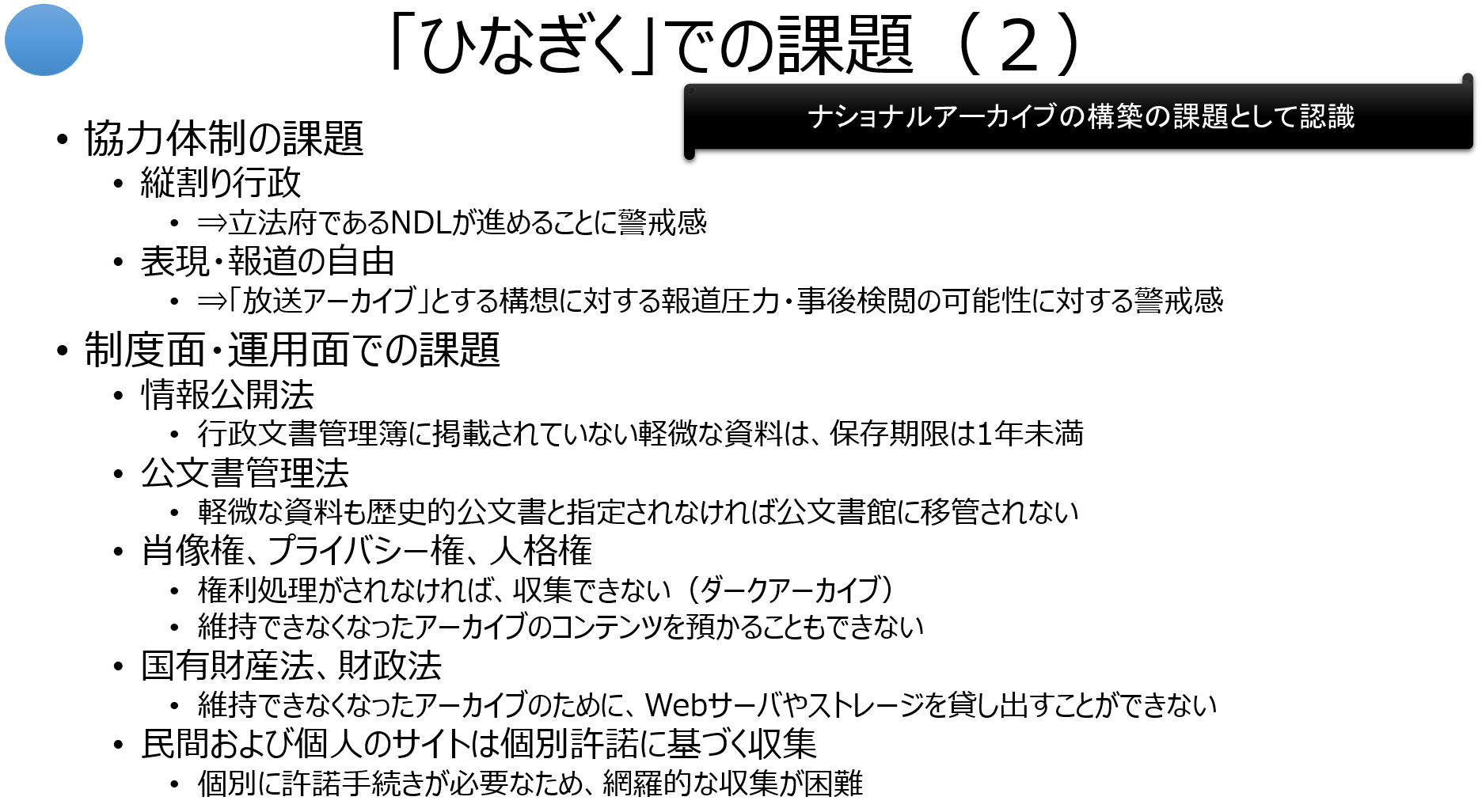
全体をポータルで一元的にアクセス

## NDL東日本大震災アーカイブの取組



## 「ひなぎく」での課題

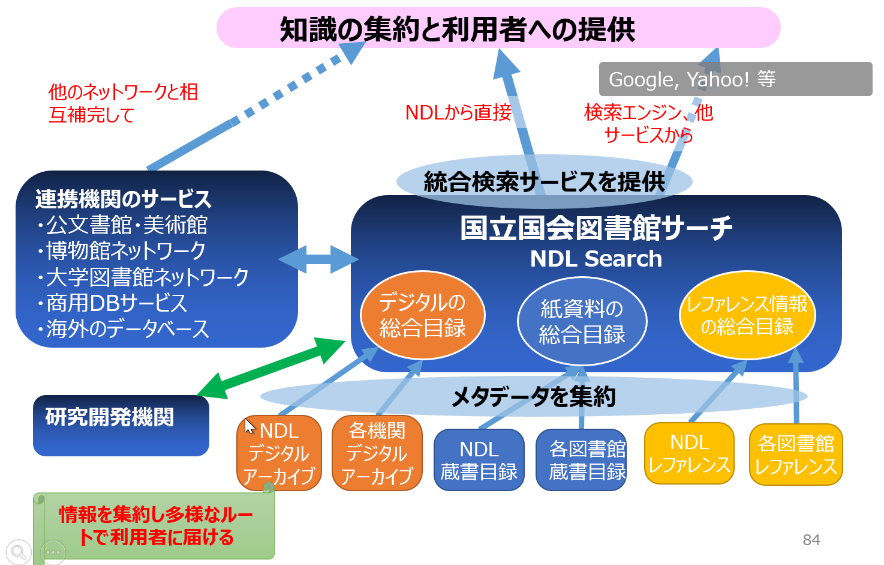




# 知識インフラ

「知の共有化と利活用」を目指したナショナルアーカイブの方向性

## 知の提供に向けたNDLサーチのイメージ



NDLサーチを核とした、現状のサービスのイメージです。

●まず、情報を集約

*国立国会図書館【蔵書目録、デジタルコンテンツ（当館デジタル化、収集コンテンツ）】、公共図書館、大学図書館【デジタルアーカイブ】、専門図書館*

●そして、それを提供する

*提供方法として、NDLが直接利用者に届けるサービス*

*関係機関と相互補完して利用者に届けるサービス*

*サービスプロバイダを通じて利用者に届けるサービス*

●NDLSearchでは、

デジタルネットワーク時代に、**利用者に求められるサービスと機能を持ったシステムを構築し提供するためには、外部の機関との連携協力が必須**であり、当館は積極的に連携協力を行っていきます。その連携の姿勢として、次のような方針を掲げています。

①統合検索サービスの提供

外部機関・サービスが提供するコンテンツのメタデータを当該機関・サービスの許諾を得て収集、もしくは横断検索します。

②外部Webサービスとの連携

外部で提供されている連想検索サービスや機械翻訳サービス等のウェブサービスを有機的に組み合わせて、付加価値の高い検索サービスを実現します。

また、外部の情報サービスへの効果的なナビゲーションを実現することにより、利用者の情報探索を支援します。

③研究開発における連携

利便性の高いシステム構築のためには、現状で確立した技術のみでは実現が困難です。

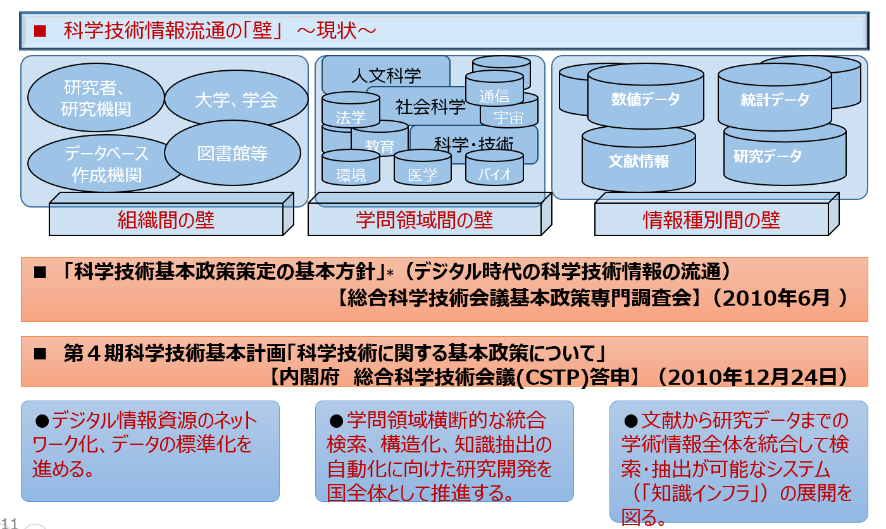
大学の研究室、官民の研究機関、ベンチャー企業等による各種の情報技術に係る研究開発を支援するために、当館の情報資源を利用した実用化・実証実験を行うことができるよう、テストベッドの場を提供します。

④統合利用促進のための環境整備

有用なコンテンツを保有しているにもかかわらず、データベースの構築や検索サービスの提供ができない機関に対して、データベースの構築やAPI実装等を支援します。

このように、**他の機関との連携により、補完しあいながら、利用者が必要とする情報を利用できるようにすることを目指しています。**

## 知識インフラの必要性



●知識インフラとは、

研究情報全体を統合して検索、抽出することが可能な基盤

●知識インフラの必要性

・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】（2010年12月24日）を答申された。

・文献等研究情報のデジタル化、オープンアクセスの推進等とともに、「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」とされている 。

・今後、科学技術研究等を推進していくためには、国全体として新しい科学技術情報基盤として「知識インフラ」の構築及び推進が必要である。

●知識インフラの構築の目的

知識インフラは、科学技術研究活動の実践を根本で支え、科学、技術、学術、文化活動によって生み出される多様なデータ、情報、文献、知識を開放し、それらへの迅速で適切なアクセスを可能にすることで、次の研究、開発、教育、その他の社会的・文化的実践へとつなげる動的サイクルを形成することを目的としている。つまり、情報の生産→流通→アクセス→再生産という**知識の循環を促進するネットワーク、プラットフォーム**となることを目指す。

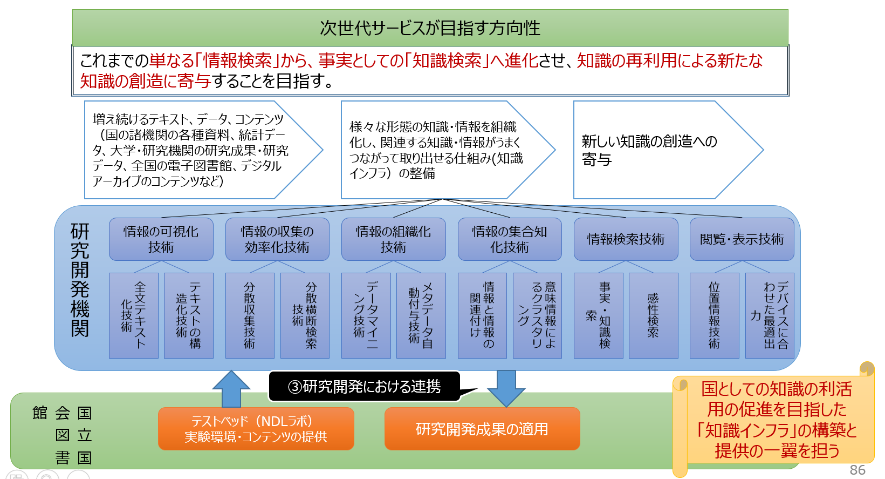
また、**組織や個別学術分野を越えた知識の融合を可能**とし、学際的な新しい知識やイノベーションの創造を容易にするものである。

*●機能*

*・知識インフラにおいては、文字データだけでなく多様な形式で表現されるデータや情報を対象とし、収集、保存、識別、組織化、検索、表示、公開といった機能を実現させる必要がある。利用者は、大量のデータに対して特定条件に適合したデータだけを抽出したり、多様な分野の情報を一括して検索したり、自分の関心に合わせて実体のリンクやネットワークを形成したりといったことが自由にできることが求められる。*

*・また、単語等による検索だけでなく、自動分類や収集された全体を見通した上での体系化や秩序化がなされることも期待される。*

## 次世代技術の研究開発成果の活用



次世代技術の研究開発成果を活用して、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。

図のようなアプローチを想定。

■目的

新しい知識の創造への寄与

■背景

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）

■研究開発の概念

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

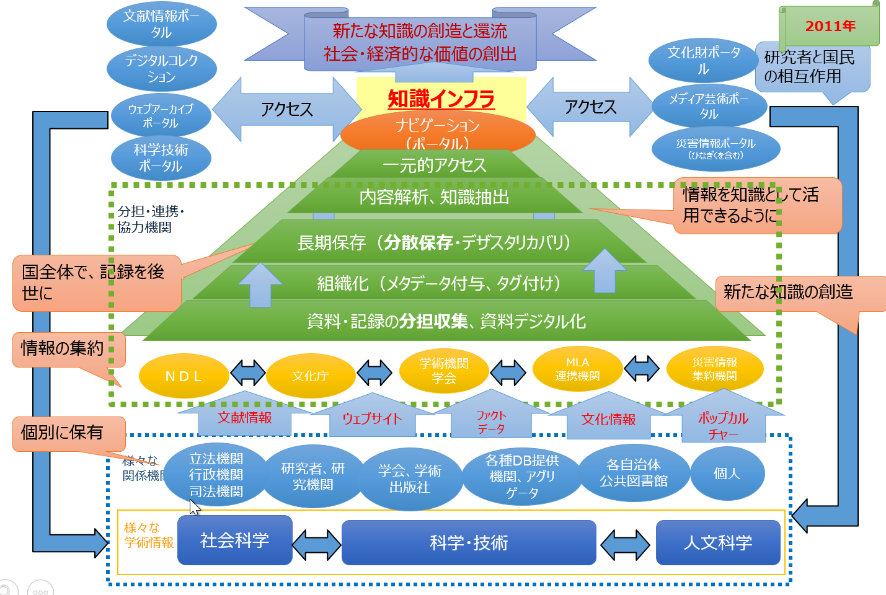
■研究開発機関の研究成果

情報の可視化技術、情報の収集の効率化技術、情報の組織化技術、情報の集合知化技術、情報検索技術、閲覧・表示技術

■これを実現するために

NDLはテストベッド（NDLラボ）において、情報資源（実験環境・コンテンツ）を提供し、研究開発成果を適用したサービスシステムを構築する。

## 知識情報基盤の構築モデル



NDLサーチの前身のPORTAと、NDLデジタルアーカイブが進めようとしていた概念がベースになって、知識情報基盤の構築モデルが描かれた。

のちの「ひなぎく」により、実現を目指した。

～～～～～

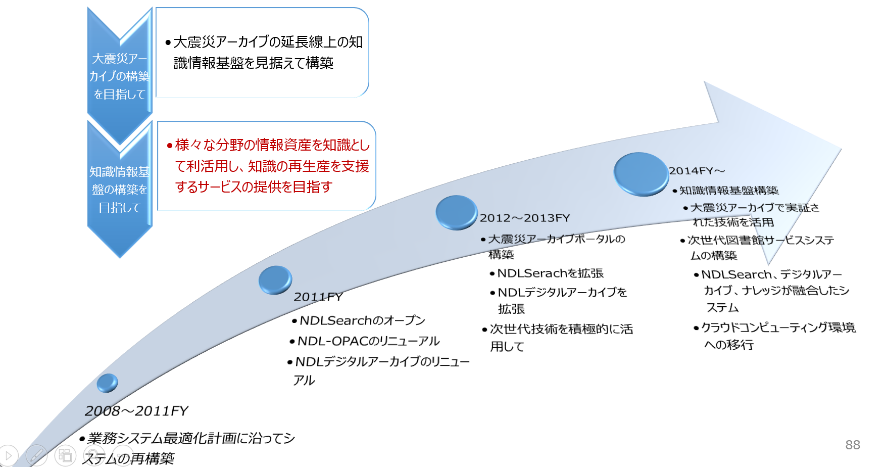
また、2010年 に、我が国の第4期科学技術基本計画の策定に向けて決定された「科学技術基本政策策定の基本方針」*（2010年6月総合科学技術会議基本政策専門調査会決定 ）*で、学術会議での長尾先生のご尽力により「文献から研究データまでの学術情報全体を統合して検索・抽出が可能なシステム（「知識インフラ」）の展開を図る」という方向性が提示されました。

「*これを踏まえて、*NDL*において、2011年に「第三期科学技術情報整備基本計画」を策定しました。*

「知識インフラ」とは、情報資源を統合して検索、抽出することが可能な基盤で、国内の各機関が保有する情報を知識として集約し、新たな知識の創造を促進し、知識の集積・流通・活用と創造するサイクルの構築を目指すものです。

*新たな知識の創造のためには、分野を越えた知識の関連付けが必要であり、日本中に散在するコンテンツの所在を集中管理し、そこに検索をかければ、関連する全ての必要なコンテンツが得られるようにするものです。そこでは、単に情報を集めたものではなく、関連するものが有機的に結合され、ネットワーク的に統合化されたものであり、日本中にある芸術を含んだあらゆる学問・研究のコンテンツ、研究ツール、社会状況データ等が知識の形に組織化され、これらの知識・情報が公開され、全ての人が共有できることを目指すこととされました。*

## 知識情報基盤の構築に向けた展開



目指すところは、「知識インフラの構築」であるが、

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

大震災アーカイブは、2012年１月にリリースした、NDLSearchとNDLデジタルアーカイブをベースに機能拡張して構築

大震災アーカイブで実証された技術・スキームで、2014年度頃から、本格的な知識インフラの構築と提供を目指していきたいと考えていた。