 DAX40-03-02 知識インフラの構築へ

参照: <https://bluemoon55.github.io/Sharing_Knowledge/MindManager/DAX40-03-02>;

# 改版履歴

## 2020年3月12日 DAX40-01の詳細を分冊化

# **「知の共有化」を試行した活動【2010～2013年】**

## 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ある。

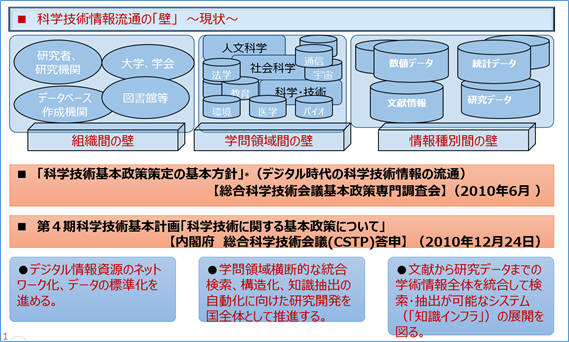
## そこでは、 単に情報を集めたものではなく、 関連するものが有機的に結合され、 ネットワーク的に統合化されたものであり、 日本中にある芸術を含んだあらゆる学問・研究のコンテンツ、 研究ツール、 社会状況データ等が知識の形に組織化され、 これらの知識・情報が公開され、 全ての人が共有できる。

## **知識インフラとは**

* + 1. 「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ある。

## **知識インフラの必要性**

* + 1. Subtopic



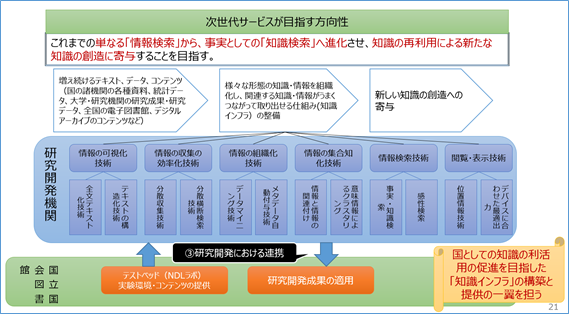
* + 1. 第４期科学技術基本計画「科学技術に関する基本政策について」（内閣府　総合科学技術会議2010年12月24日）が答申され、 文献等研究情報のデジタル化、 オープンアクセスの推進等とともに、 「文献から研究データまでの学術情報全体を統合して 検索・抽出が可能なシステム（「知識インフラ」）の展開を図る」とされた。
    2. これを踏まえて、 NDLは、 2011年に、 第三期科学技術情報整備基本計画－知識インフラ構築に向けて」を策定した。

## **知識インフラの構築の目的**

* + 1. 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、 関連する情報が組織化され、 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。
    2. 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ものであり、 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。

## **知識インフラの機能**

* + 1. 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。
    2. 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。
    3. 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が期待される。
    4. 次世代技術の研究開発成果の活用
       - Subtopic



## **知識情報基盤の構築モデル**

* + 1. Subtopic



* + 1. NDLは、 NDLサーチの前身のPORTAと、 NDLデジタルデポジットシステムが進めてきた概念をベースとして、 知識情報基盤の構築モデルを描き、 後の「東日本大震災アーカイブ（ひなぎく）」により、 大震災の分野に絞った「知識インフラ」の実現を目指した。
    2. 「ひなぎく」の実現に当たっては、 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・表示技術等の次世代技術の研究開発成果の活用なしには実現できないと考える。

# **東日本大震災アーカイブ「ひなぎく」（知識インフラ構築の先行事例）**

## 2011年3月には、 東日本大震災が発災し、 甚大な被害がもたらされた。国は、 この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こととした。

## NDLは、 それを実現するために、 大震災関連にフォーカスした知識インフラの構築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）」を構築した。

## 大震災アーカイブの構築は、 従来の図書館の枠を越え「あらゆる情報、 記録を収集、 永久保存して、 提供する」というNDLの使命を達成するために、 今後数年間で取り組むべき事項の多くが含まれていると考える。

## **収集の基本的な考え方**

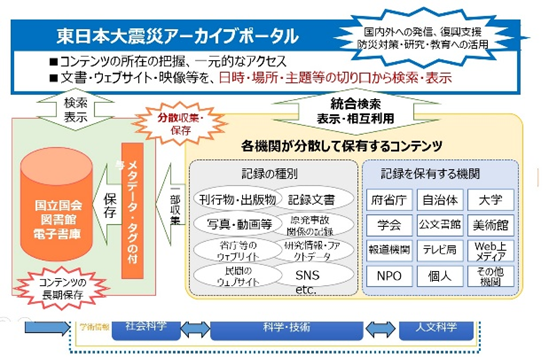
* + 1. 大震災に関連する情報のすべてを、 NDLだけで収集し保存することは不可能であり、 国全体で分担して収集し、 分担して保存する形を目指すべきである。
    2. (1)収集範囲
       - 国全体で．様々な機関が保有している大震災前の記録、 大震災後の事象・被害、 状況の記録、 今後の復旧・復興の記録等、 過去から、 現在、 未来に亘って、 可能な限り収集・保存する。
       - Subtopic



* + 1. (2)収集方法
       - NDLによる直接収集
         * NDLは、 制度的に収集可能な記録のほか、 他の機関が保存の対象としていない記録等については、 積極的に受け入れる。
       - 他機関による記録等の保存の推進
         * 記録の保有機関、 アーカイブ機関、 NDLのいずれかで記録等を保存し、 NDLは、 その所在情報を把握できるようにする。
       - メタデータの収集又は検索の機械的連携
         * NDLは、 現在のNDLサーチと同様に、 可能な限り、 メタデータの収集、 もしくは横断検索による機械的連携を行う。
       - 働き掛け・協議
         * NDLは、 国等の各機関で、 国全体としての記録等の保存の必要性の認識を共有し、 関係機関に協力を要請する。

## **「ひなぎく」のシステムイメージ**

* + 1. Subtopic



* + 1. これまでの成果であるNDLサーチ、 ウェブアーカイブシステム、 デジタルデポジットシステム等の既存のシステムをベースに、 機能拡張する形でシステムを構築した。
    2. 構築に当たっては、 効率的、 効果的に進められるように、 また、 利便性の高いサービスにするために、 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が図れるように調査研究を進めた。
       - 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(Hadoop)、 分散ファイルシステム(GlusterFS)を導入した。
       - 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した。
       - 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の導入を検討した。
       - 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。
       - 次世代の図書館システムのモデルとなることを指向した。

## **「ひなぎく」での課題**

* + 1. 構築・運用においては様々な課題が顕在化した。
    2. (1)技術面での課題として、
       - 大規模分散ファイルシステム、 分散処理システムを適用したが、 実際には、 想定ほどコンテンツが集まっておらず、 将来的な実用化レベルでの検証ができていない。
       - 計測機等から得られた生データ、 分析の中間データなどのファクトデータの永久保存も目指しているが、 情報に関するメタデータまでしか収集できていないため、 ビッグデータへの対応の検証ができていない。
       - ビッグデータとしてのウェブアーカイブから大震災関連の情報を自動的に切出し、 コンテンツ単位で精緻な検索をする仕組みが構築できていない。
       - 各アーカイブにおいて、 検索およびハーベスト用のWebAPIが未実装なものが依然として多い。
       - 収集前データに永続的識別子がない。
       - 各アーカイブのメタデータとの記述規則の差異、 付与されたメタデータに記述要素の不足が多い。
       - 不完全なメタデータへの自動付与機能はない。
       - 利活用のための検索・閲覧機能として、 本文テキスト、 イメージ認識技術等、 内容を関連付けた検索技術が未成熟である。
    3. (2)人材面での課題としては、
       - 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）。
    4. (3)協力体制の課題として
       - 縦割り行政のため、 立法府であるNDLが進めることに警戒感。
       - 「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。
    5. (4)制度面・運用面での課題として
       - 情報公開法では、 各文書に対して最低保存期間が設定されるが、 運用では行政文書管理簿に掲載されていない軽微な資料の保存期限は1年未満とされ、 1年を経過すると廃棄される。
       - 公文書管理法では、 軽微な資料を含めて、 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、 人格権等の権利処理がされなければ提供できないが、 NDLは、 提供を前提としない、 いわゆるダークアーカイブとしての収集は行わないため、 継続して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、 財政法の解釈では、 アーカイブシステム機器等の運用維持が困難になったアーカイブ機関のために、 Webサーバやストレージを無償で貸出すことができない。
       - 現在のインターネット情報の制度的収集は国等のサイトに限定されているため、 民間および個人のサイトは個別許諾により収集しなければならないが、 個別に許諾手続きが必要なため、 悉皆的な収集が困難。民間サイト、 個人サイトは、 ハーバード大学と連携し、 インターネットアーカイブ社に収集を依頼している。
    6. このように課題は多いが、 協力体制、 制度面・運用面の課題は、 大震災に限らない共通の課題であり、 今後、 国全体のナショナルアーカイブ構想の議論の中で、 具体的な解決策の検討が進められることを期待している。