 DAX40-05 「知の共有化」の実現を目指した構想【2013年】

# 改版履歴

## 2020年3月12日 DAX40-01の詳細を分冊化

# **「知の共有化」とは**

## アーカイブ機関に限らず、 あらゆる組織、 個人がデジタル化した情報を共通的に保存し利活用できるようにする基盤を、 「デジタル知識基盤」といい、 「デジタル知識基盤」により、 新たな知識を創造し還流することを「知の共有化」として記述する。

# **「未来の図書館を作るとは」と「知の共有化」**

## Web2.0時代と言われた頃、 図書館界ではOpac2.0と称するサービスで、 いわゆるAIとまでは言えないが、 AIを指向した組織化、 検索サービス等の実現を目指してきた。

## 既にインターネット上に広範な情報が公開されているが、 ここ数年で、 文化機関、 公的機関が保有している情報もオープンデータ化が進みつつあり、 ビッグデータとして利活用できる方向に向かっている。

## このような時代に、 従来の業務やサービスは、 AIが組み込まれたシステムやロボットに支援されてサービスが省力化され、 さらにビッグデータとして網羅性が確保されることにより、 「知の共有化」が可能な新たなサービスが生まれようとしている。

## 「未来の図書館を作るとは」（長尾真）より抜粋

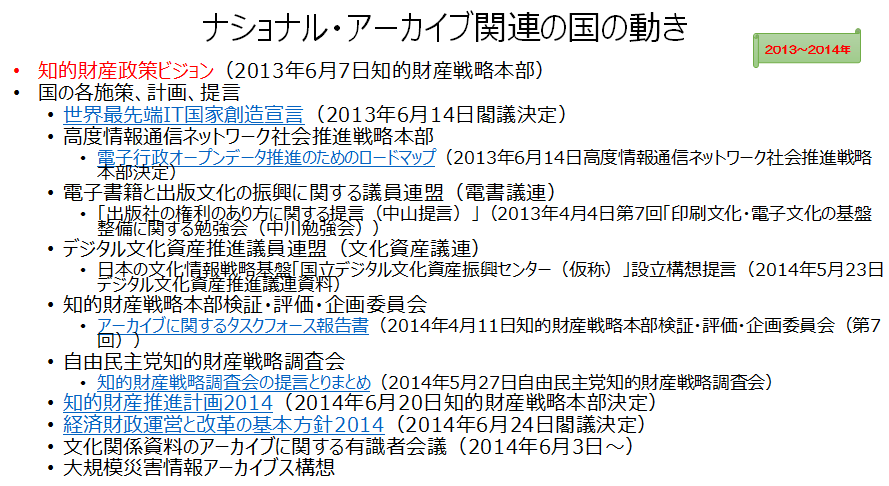
* + 1. 知識インフラ構築における最も大切な概念は、 情報を集め、 これを知識化し活用することによって新しい情報・知識を創出し、 知識インフラに加えるという形で循環的にこのシステムを強化・拡大してゆくことによって、 社会・経済に貢献することである。
    2. 知識が種々の観点から組織化、 構造化されて利用しやすい形で蓄積されて利用に供される

## 図書館の視点で見れば、 専門家によって作成されてきた、 書誌、 辞書、 典拠（シソーラス）類に加えて、 著作物のイメージ画像、 全文フルテキスト、 目次、 索引、 あらすじ、 書評等、 出版界や図書館界で従来から形式知化してきた情報を、 信頼性の高いビッグデータとして利用可能にし、 AIの教師用データとして投入・学習することにより、

## 長尾先生が1994年に提唱した「電子図書館」が、 信頼性の高い知識を豊富に持った 「AI化されたデジタル知識基盤」として実現することになる。

# **アーカイブに関連した国の活動の方向性**

## Subtopic



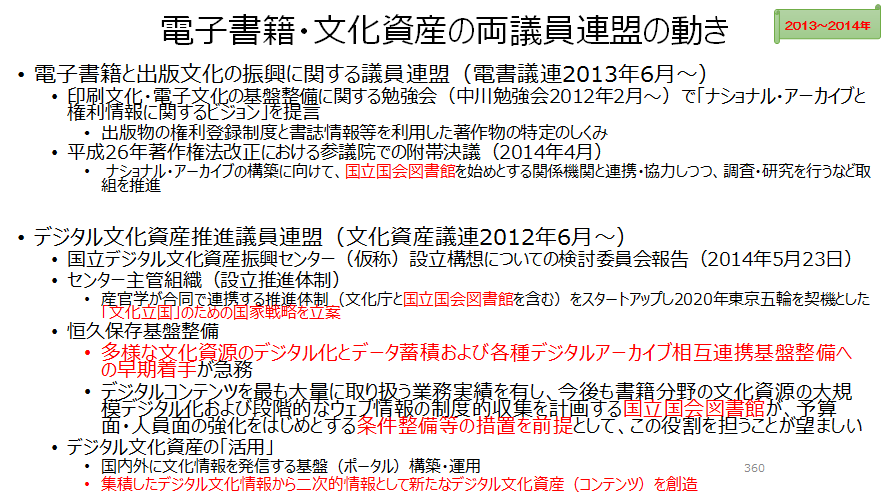
## 国は、 「知的財産政策ビジョン」（2013年6月7日知的財産戦略本部決定）に基づく国の施策の中で、 図書館に直接関連する計画として、 電子書籍化と利活用の促進に関する構想（電子書籍と出版文化の振興に関する議員連盟）、 デジタル文化資産の保存・活用の基盤の整備に関する構想（デジタル文化資産推進議員連盟）、 2020年までの6年間の文化芸術分野での振興に関する基本的な方針（文化庁）、 学術情報の公開と共有の拡充に関する計画（文部科学省）、 大規模災害の記録と記憶の保存などのアーカイブの構築等を検討している。

## 個々の施策の目的は異なっていても、 対象とする文化的資産は相互に関連するものであり、 知識インフラを目指した仕組みとして、 国全体で文化関係資料の価値を高め、 新たな文化や情報を生み出す社会基盤として、 「恒久保存・継承・公開・活用」が可能なナショナルアーカイブとなることが望まれる。 「資料・情報を文化資産として収集・保存する」ということは、 従来からの出版物に相当する情報の範囲ではなく、 美術館、 博物館、 文書館等が保有する無形・有形の文化財をデジタル化した情報を含め、 インターネット上で流通している著作物全てを文化資産としてアーカイブすることであると筆者は考える。

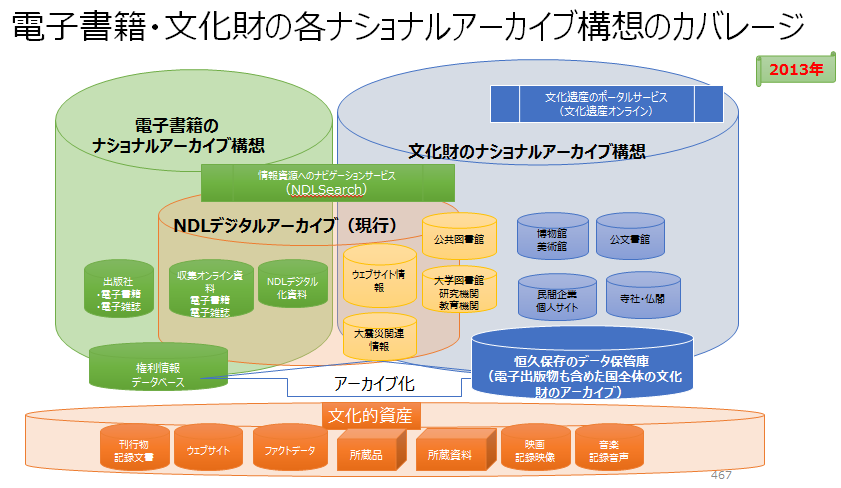
## 文化財のナショナルアーカイブの構築と運用に関しては、 様々な機関において、 必要性の認識の共有やアーカイブ立国宣言の提案17)、 制度的な課題解決のために、 アーカイブ基本法の法制化、 推進体制作りなどが議論されている。 また、 具体的なサービスシステムの仕組みとして、 技術的課題解決のための研究開発も進んでいる。

## 恒久的保存基盤の部分は、 文献も含めて、 すべての文化財、 災害情報が相互に関連付けられて蓄積されている。恒久的保存基盤を活用して新たな知識創造活動が行われ、 知識創造活動の成果を、 様々な目的に応じて利活用されていくことを想定する。

## Subtopic



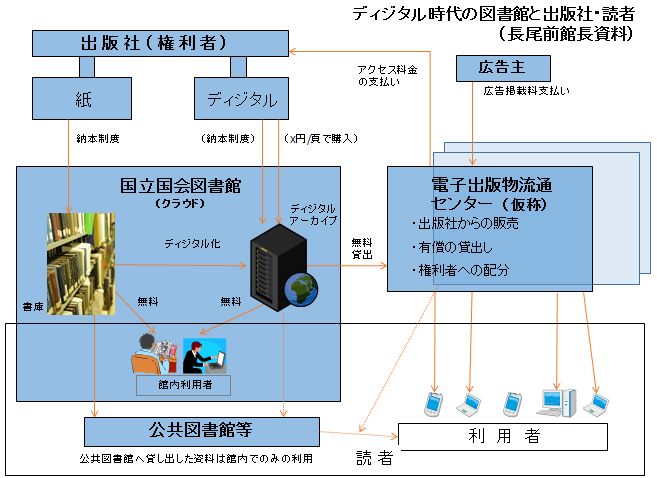
## Subtopic



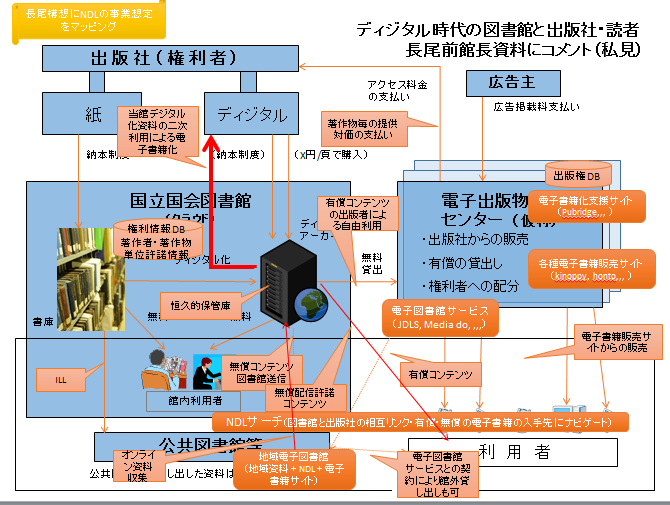
# **電子書籍分野のナショナルアーカイブ**

## **電子書籍分野のアーカイブの役割**

* + 1. 長尾前館長により、 2008年に提案され、 改めて2015年3月に示された電子出版物流通に関する 「長尾モデル」として提案されたモデルと類似したプラットフォームである。
       - Subtopic



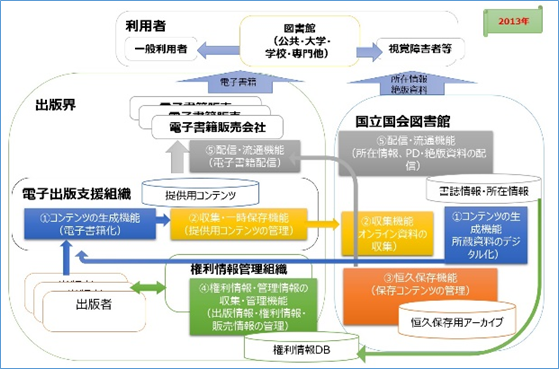
* + - * Subtopic



* + 1. 提供用の商用コンテンツは、 電子書籍サイトが保持し提供する。
    2. NDLは、 DRMによるアクセス制限が掛かっていないコンテンツを収集し、 ダークアーカイブとして恒久的保存の役割を担い、 電子書籍サイトがコンテンツを保持できない、 もしくはサービスを継続できない場合、 NDLの恒久保存用アーカイブを利用して電子書籍サイトから提供する。
    3. また、 NDLが原資料保存のために画像イメージでデジタル化したコンテンツを、 出版界が二次利用して電子書籍化・プリントオンデマンド（POD）化し、 電子書籍サイトから提供することにより、 いわゆる絶版本の電子書籍化を促進する。
    4. さらに出版物・著作権者の情報は、 出版界・図書館界の共通のデータベースとして維持・更新し、 相互に利活用することも想定したモデルである。

## **電子書籍分野のアーカイブの機能モデル**

* + 1. Subtopic



* + 1. 電子書籍に関するナショナルアーカイブは、 コンテンツの生成機能、 収集・一時保管機能、 保存機能、 権利情報・管理情報の収集・管理機能、 配信・流通機能の5つの機能を想定する
    2. コンテンツの創出から収集・保存、 配信・流通に至る全体の流れを一元的に集約するため、 恒久保存のアーカイブと共に、 ナショナルアーカイブ全体のメタデータを集約するデータベースにおいては、 多様な主体が多様なデータを扱えるよう、 柔軟かつ多層的なデータ構造を実現する。
    3. 具体的には、 著作物・著作者・出版者等の書誌情報、 販売データ・販売者・所蔵機関等の所蔵情報、 目次・索引・シソーラス等の情報探索情報等の多様なデータを、 一元的に管理可能な仕組みとすることを想定した。

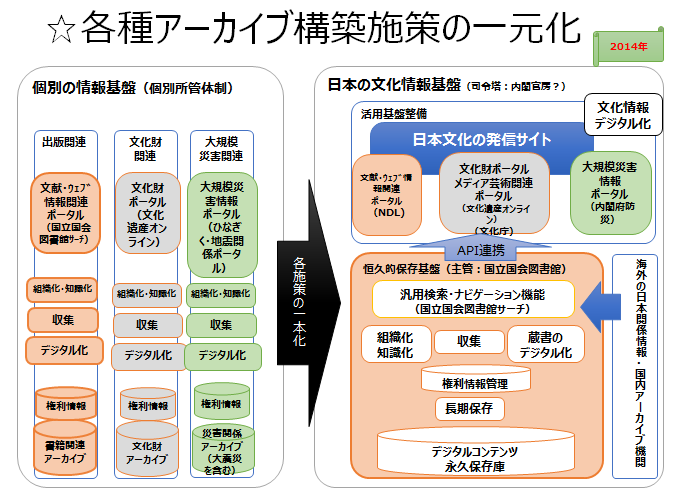
# **文化情報資源のナショナルアーカイブ**

## **「未来の図書館を作るとは」（長尾真）より抜粋**

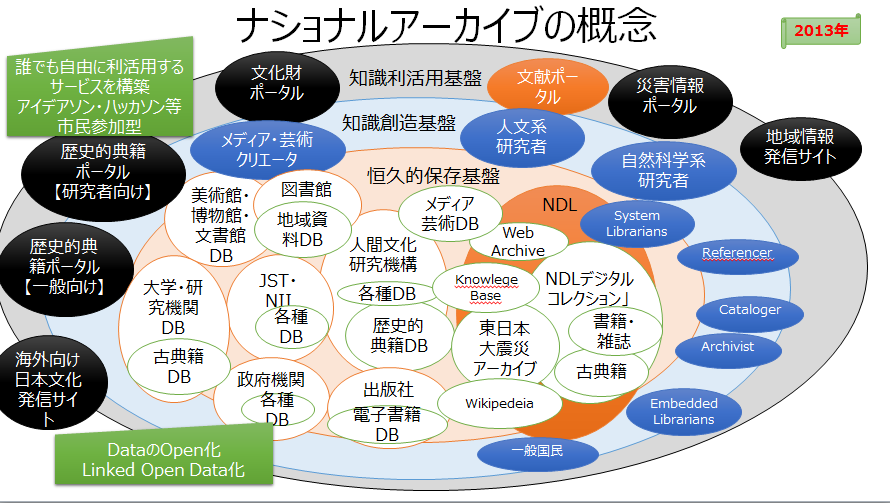
* + 1. 【長尾先生】図書館は、 書物等の情報資源を収集・保存し提供する場であるとともに、 それらを使って関心のある人が集まって議論し新しい知識を創造する場
    2. 【長尾先生】日本としてのナショナルアーカイブは、 日本中に存在する知識情報が有機的に結合され、 日本中の人が自由に使える日本の「知識インフラ」

## **はじめに**

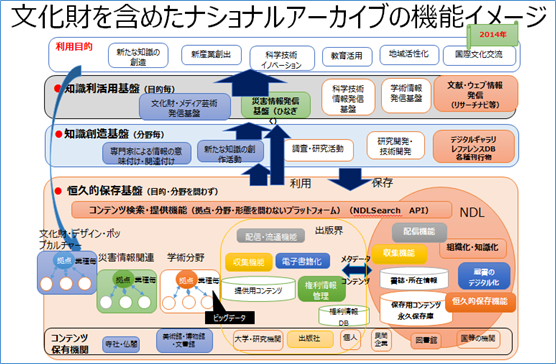
* + 1. 各種デジタルアーカイブ構築施策の一元化
       - Subtopic



* + 1. あらゆる分野、 業種業態を越えた情報資源を一元的に利用できるアーカイブ
       - Subtopic



* + 1. 文化財を含めたナショナルアーカイブの機能イメージ
       - 「文化財を含めたナショナルアーカイブの機能イメージ」は、 2012～2014年に描いた図である。
       - Subtopic



* + 1. 我が国の多様な文化を知識として保存・継承する役割、 様々な分野の専門家が参加し、 新たな文化を知識として創造していくための社会的な基盤としての役割、 それらの知識を利用目的に応じて発信する役割、 そして、 これらの仕組みを統括し運用していく役割が考えられ、 文化的資産を館種毎に集約している拠点と、 新たな知識を創造し発信している拠点等が分担して構築・運用していくことを想定した。

## **恒久的保存基盤（あらゆる情報の恒久的保存）**

* + 1. 恒久保存と利活用のための共通プラットフォームで、 １つの機関にすべてを集約するのではなく、 各分野のアーカイブを集約する拠点が中核となって分散アーカイブを構築する。
    2. あたかも１つのアーカイブとして見えるように、 個々の情報同士を意味的に関連付け、 検索で情報を取り出すだけでなく、 取り出された情報から芋づる式に関連する情報を取り出せるようにする。
    3. 分野での情報の網羅性・完全性が重要であり、 また異分野の情報との関連付けにより、 AIとしての知識として活用できる。

## **知識創造基盤（新たな知識創造活動の場）**

* + 1. キュレーター、 ライブラリアン等の支援のもとで、 それぞれの分野の専門家のみならず、 広く国民も含めて、 情報に付加価値を付けたり、 他の分野の情報と関連付けて、 二次的著作物を創造する場として想定した。
    2.  知識創造支援機能
       - 情報全体の基本情報としてのメタデータを付与する活動、 記事、 章節項、 文節等の単位で組織化・構造化する活動、 情報間を意味的に関連付けるための基本情報として、 用語辞書、 典拠、 シソーラス辞書等を作成する活動である。
       -  辞書類作成
         * 辞書類の作成において、 コーパスは、 AIの支援により省力化が進むと思われる。
         * 件名典拠、 シソーラスは、

本文フルテキストもAIへの入力データとして特徴量を抽出することにより、 書誌データのみならず、 本文テキストから標目を抽出することが可能になり、

また、 コーパスを利用することにより、 同義語、 上位語、 下位語、 関連語、 分類記号の関係づけをすることも可能になると思われる。

* + - * + 固有名典拠は、

既に刊行されている書籍の著者に関して、 AIを利用して、 著作名、 著者、 出版年、 NDC分類等の特徴量を抽出・分析することにより、 著者の自動同定の精度は相当向上すると思われる。

AIにより、 省力化、 自動化ができることにより、 書籍だけでなく雑誌記事等も含めて、 固有名典拠の自動作成が可能になる。

* + - * + NDC分類は、

件名典拠をベースにした分類の番号体系は、 人間の判断によるところが大きいので、 容易にAI化することは困難と思われるが、 分類を番号体系でなく意味的なシソーラス体系として維持し、 改訂していく作業は、 AIにより体系を整理することは可能になる。 また番号体系の版毎の対比表の維持も、 人手を介さずに行えるようになると思われる。

ただ、 配架のための分類という考え方、 資料検索のための分類体系は、 全文検索を含めて様々なレベルでの検索が容易になってくる状況において、 役割を問い直す必要があるのではないか。

* + - *  情報を関連付けるための情報
        + 共通語彙

分野を超えた情報交換を行うには、 個々の単語について、 表記・意味・データ構造を統一し、 互いに意味が通じるようにすることが必要であり、 AIの知識として蓄積する場合、 共通語彙になっていることが学習の効率を高められる。

* + - * + コーパス、 オントロジー、 分野別シソーラス、 各種用語辞書、 DBPedia等

それらが、 永続的識別子となって、 Linked Data化されていると、 より精度の高い関連付けが可能になる。

* + - * + IMI：情報共有基盤への参画

データに用いる文字や用語を共通化し、 情報の共有や活用を円滑に行うための基盤であるIMI(Infrastructure for Multilayer Interoperability:情報共有基盤)の構築への参画、 連携が重要である。

* + - *  情報に関する情報の組織化
        + 情報の組織化は、 二次情報としての書誌、 メタデータの作成があるが、 新しい著作物に対するNDC分類、 件名などは、 既存の書誌、 典拠を教師データとして蓄積することにより、 AIによりかなりの精度の自動付与が可能になる。
        + 専門家は、 AIによる精度に応じてチェック、 補正することが役割になるのではないか。また、 雑誌記事に関しても、 記事を著作単位とした書誌作成が、 AIにより省力化して付与が可能になる。
        + 「未来の図書館を作るとは」（長尾真）より抜粋

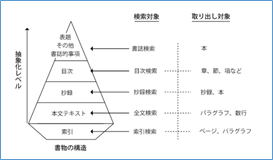
図書・資料は部品に解体され、 それぞれが種々の観点からリンク付けされた巨大なネットワーク構造が作られるようにする

目次情報を付けたり、 本の表紙の画像や数行の簡単な要旨を付ける

動画・静止画・音声は、 画像認識機能により自動タグ付け

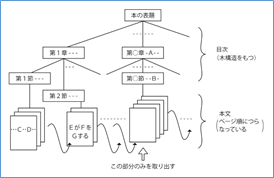
検索システム

Subtopic



書籍の構造化

Subtopic



* + 1.  知識創造活動
       - 新たな知識を創造する活動は、 図書館等のアーカイブ機関の主たる目的の一つであり、 デジタル知識基盤の中で、 恒久的保存基盤に格納された網羅的な情報を活用して、 新たな知識を創作する活動である。
       - 歴史的な文化財や現代文化を映像化、 画像化、 テキスト化する活動、 構造化された情報に解題情報等を付与する活動、 情報間を意味的に関連付ける活動、 テーマを設定してデジタルギャラリを構築する活動等も、 創造活動の一つと位置づけられる。
       - 創造活動が、 効率的に行えることにより、 質、 量ともに多くの知識が再生産され、 創造された知識は、 恒久的保存基盤でアーカイブされることを想定した。
       - 「未来の図書館を作るとは」（長尾真）より抜粋
         * 従来の書誌情報の考え方を、 マルチメディア情報に対応できるように拡張かつ詳細化するとともに、 対象資料に関係する様々な種類の情報にリンクをはり、 それらの情報をたどってゆけるようにする
         * 既存の知識から自分の必要とする部分を切り出して、 新たな発想の著作を創造し、 新たな知識とする
    2.  新たな知識創造の場としての図書館
       - 図書館等の機関は、 自館が保有している資料に加えて、 「デジタル知識基盤」で共有されている知識を組合わせて提供し、 知識創造の場として、 施設としての実空間と、 インターネットを利用した仮想空間が、 相互にシームレスに、 「場」として利用できるものを想定する。
       - 情報を媒介して専門家と専門家を繋ぐこと、 様々な組織が持つ情報が関連付けられて、 それぞれの組織を中心に活動していた人が繋がっていくこと、 専門家を媒介して辞書と辞書を繋ぐことを可能にして、 様々な分野の人たちが保有する知識が有機的に関連付けられ、 それを活用して、 更に利用者同士がつながっていく「場」としての機能を期待する。
       - 「未来の図書館を作るとは」（長尾真）より抜粋
         * 議論の場の提供という図書館機能
         * 解決したいという人達と研究者、 そして図書館司書のグループが種々の角度から資料をもとに議論
         * 考え方の違う人達が知識を共有し、 その違いを議論を通じて明らかにすると共に、 新しい知識・思想を作り出してゆく場
         * 出版社、 著者と読者をつなぐ場としての図書館
         * 図書館が司書による相談サービス、 あるいは自動的な案内サービスをする場合でも、 自分の電子書棚を作りたいという人の多くは図書館から借りるのではなく、 出版社のデータベースの方に行き、 書物を購入することになるわけで、 図書館は出版社と読者を結合する接続業者のようになってゆく
       -  レファレンスサービス
         * 利用者からの問い合わせに対して、 AIを利用して、 必要とする内容に最もマッチングする資料名、 所在場所へナビゲーションする。資料がオープンデータであれば、 参考となる内容の記載部分を列挙し、 さらに答えとしての事実まで提示が可能になる。
         * 自然言語処理機能を利用することにより、 音声での自然言語の問いに対して、 「デジタル知識基盤」を知識としたAIが、 解の候補を抽出し、 それを音声での自然言語で回答することが可能になる。
         * 「未来の図書館を作るとは」（長尾真）より抜粋

司書の世話にならなくてもある程度のレファレンスサービスが受けられるように

自然言語による質問要求を受け付けて、 取り出したものがその要求に対応するものであるかどうかを、 自然言語処理技術によって調べ、 できるだけ質問要求に近いものだけを選択する

* + - *  事実に基づいた報告書
        + 依頼または予測調査に関しては、 AIを利用した文献の意味的内容検索により、 最もマッチングする資料名、 資材場所を探し出し、 その資料がオープンデータ化されていた場合は、 記述された内容を洗い出し、 内容を要約して報告書の原案を作成することも数年で実用可能と思われる。

## **知識利活用基盤（情報発信）**

* + 1. 恒久的保存基盤に格納された一次情報、 コンテンツ創造基盤で創出された二次的情報が、 デジタル知識基盤の中でAIの機能を活用して有機的にネットワーク化され、 あらゆる分野で利活用できるようにする。
    2. 見るだけのアーカイブ」から「使い、 創り、 繋がり、 伝えるアーカイブ」として、 広く国民による新たな知識の創造、 新産業の創出、 地域活性化、 防災・減災、 教育活用、 教養・娯楽、 観光、 国際文化交流等、 様々な利用者毎の目的に応じて、 恒久的保存基盤に格納された一次情報、 コンテンツ創造基盤で創出された二次的情報を有機的に組み合わせて、 利用できるようにする。
    3. 知識検索機能として、 網羅的な情報から、 利用目的に応じてあらかじめ適切に絞り込み、 利用者の属性、 スキル、 利用場所に応じて、 様々な画面インターフェースを用意して、 利用者が必要とする情報、 参考となる関連する情報を容易に得られるようにする。
    4. 「未来の図書館を作るとは」（長尾真）より抜粋
       - 電子図書館になって取り出す単位が書籍の単位ではなく、 書籍の中の章や節、 パラグラフ、 あるいはこんな内容が書かれている部分のみ
       - 自動的な形で適切な知識の所在にまでナビゲートしてゆくシステム
       - その本のどこに書かれているかを探すというのではなく、 自分の欲しい情報そのものが出てくることになる。
       - 書誌検索のような単純、 単一の検索でなく、 種々の検索のモードを提供する
       - 種々のあいまいさを許すあいまい検索の工夫
    5.  自分の知識の外部記憶装置
       - 「未来の図書館を作るとは」（長尾真）より抜粋
         * 個人によって違った知識の構造の部分については、 その人の力によって種々の検索方式を試み、 自分の必要とする情報をとり出して、 中立的な知識の構造に付加してゆくことが出来ねばならないし、 またそれによって自分に合った知識の構造を作りあげてゆく
         * 自分の書棚を電子的に作れば、 自分の連想に基づいて自分の本や本の部分部分をリンク付けして、 自分の知識の構造に合った自分図書館を作って楽しむことができる
       - 現在のAIの技術で、 デジタル知識基盤にある共通的な知識から、 個人のスキル・嗜好に合わせてパーソナライズするとともに、 個人が保有する知識を組み合わせて、 自分専用のインテリジェンスは外部記憶装置を持つことが可能になる。
       - スマートフォン等からこの外部記憶装置にアクセスして、
         * ①可能な限り「事実情報」を提供する、 出来ない場合は、
         * ②参考になりそうな文献候補を提示する、
         * ③その文献候補の所蔵先へ案内する。
         * ④デジタル化された情報があれば、 閲覧サービスまでナビゲート、
         * ⑤デジタル化された情報がなければ、 入手可能なサービスへナビゲートという仕組みは、
         * もう実現可能な段階になっている。

## **運用基盤**

* + 1. データ保存課題、 人材不足問題、 資金不足問題、 権利処理コスト問題など、 デジタルアーカイブの活性化を阻害する課題を制度的に克服する方策 (人材育成策・予算措置・権利処理に関する法改正など)を導き出し、 解決を図っていく中核的な役割を担う推進母体(司令塔)としての 組織・体制も必要である。
    2. この基盤では、 ナショナルアーカイブ全体の戦略企画、 デジタル情報の保存や利活用のための調査研究、 研究開発、 デジタル化支援、 アーカイブに所蔵された資料に関する知識、 読解力とIT技術の知識等も備えた高度な専門的人材の育成、 孤児著作物の権利処理や、 絶版作品のデジタルアーカイブ化における所有権、 肖像権問題も含めた権利情報DBの構築を促進する等の 役割を持つことを想定する。