AIを活用した「知の共有化」システムの方向性  
―「未来の図書館を作るとは」の実現に向けてー

2017年10月3日

元国立国会図書館　専門調査員・電子情報部長

中山正樹

# はじめに

1980年代後半に検討が開始された電子図書館事業は、1990年前半に策定された電子図書館関連の構想において「地球規模の知的財産を誰でも容易に利用できるようにする」という目標を掲げると同時に、1990年代中旬に我が国で最初の実用化実証システムとして、「ネットワーク環境における情報と文献の利用のための高度検索システム（Ariadne）」（京都大学）1）、パイロット電子図書館実証実験プロジェクト（国立国会図書館（NDL））での実用化実証実験2）が行われた。国の動きでは、2003年には、e-Japan重点計画2003、e-JAPAN戦略Ⅱ加速化パッケージ、（内閣官房IT戦略本部）において、「国のデジタルアーカイブ構想」、「ジャパンウェブアーカイブ構想」の実現を、また2004年には、e-Japan重点計画2004において「国立デジタルアーカイブポータル構想」を一層推進することが明記された。このような動きと同期して、国立国会図書館では、国立国会電子図書館中期計画20043）を策定し、①「デジタル･アーカイブの構築」として、国立国会図書館デジタルコレクション、国立国会図書館インターネット資料収集保存事業（WARP）」、 ②「情報資源に関する情報の充実」として「リサーチナビ、レファレンス協同データベース」③「デジタル･アーカイブのポータル機能」として「国立国会図書館サーチ」を構築し提供した。2005年7月に試験公開したPORTAのプロトタイプは、ナショナルアーカイブポータルの原点と言え、インキュベータの役割を果たした。4）さらに「東日本大震災アーカイブ（ひなぎく）」により大震災関連の記録・記憶に限定はされているが、従来の図書館の枠を越えたデジタル知識基盤の構築を進めてきた。

現在は、分野を特定せず、図書館を含めて文化情報資源、知的情報資源を保有する機関（以降、「アーカイブ機関」という。）が連携して、日本全体でのデジタル知識基盤を構築し、様々分野の情報を知識として保存し活用できるようにする機運が高まり、活動が活性化してきた。

デジタル知識基盤を構築するに当たっては、ビジネス、制度、組織、技術等の観点から相互に関連し合う課題が多い。情報システムの構築とサービス展開、知的情報のデジタル化においては、いわゆる「第4次産業革命」、「デジタルトランスフォーメーション（デジタル革命）」と言われる技術革新とビジネス変革の時期に来ている。そのような現状と今後の展開を考慮すると、長尾元国立国会図書館長が2012年3月にNDLを退官される際に職員に配布された「未来の図書館を作るとは」5）の中で示された「未来には実現できるだろう」とされたことが、「今、この時代」の図書館で実現できるレベルにあり、これから構築する「デジタル知識基盤」は、第4次産業革命、デジタル革命の方向性に沿うことが必要と考える。

　各アーカイブ機関が、デジタル知識基盤に対応したデジタルアーカイブシステムの構築を、適正な内容と費用で行うためには、有効性が確認されている標準的な調達手順（プロセス）で行うことが重要である。「政府情報システムの整備及び管理に関する標準ガイドライン」（各府省情報化統括責任者（CIO）会議）6）は、政府機関全てでのシステム構築のプロセスと成果物を規定しているガイドラインであり、これを参考にすることにより過不足のない調達要件を受託者に提示して、適正な手順で、適切な技術、パッケージを活用したシステム開発を効率的・効果的に行うことができる。

さらに、「政府情報システムの整備及び管理に関する標準ガイドライン」に沿った開発プロセス（タスク）を遂行できる情報システム関連の担当者の人材育成は、開発工程の個々のタスクに必要なスキルと知識の項目が示されている「iコンピテンシ・ディクショナリ(iCD)」（2016年情報処理振興機構（IPA））7）が参考になる。iCDで網羅的に示されたタスク毎のスキル、知識の中から、担当するタスクに応じて選択的に身に付けることにより、実践的な人材を効率的に育成することができる。

# 「知の共有化」とは

図書館が保有する冊子体資料目録のデジタル化から始まり、冊子体資料のデジタル化、ボーンデジタルの著作物を扱う「電子図書館」、「デジタルライブラリ」、学術情報、研究データも含めあらゆる情報を知識として活用できるようにする「知識インフラ」、美術館、博物館等を含めて文化情報資源全体の「文化資産ナショナルアーカイブ」等、更に同義語的に、インターネット資料、オンライン資料、電子出版物、電子書籍、電子雑誌、知的情報基盤、文化情報資源、文化情報資産、デジタルアーカイブ、ナショナルアーカイブ、インターナショナルアーカイブ等、時の流れの中で、目指すところは同様でありながら全体集合、部分集合の違いにより、様々な呼び方をされてきたが、知識全般を扱う適切な言葉が共有されていないのが現状である。

本稿では、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。

# デジタルトランスフォーメーション(デジタル変革)

今は、IoT、ビッグデータ、ロボット、AI等の技術革新による、いわゆる「第4次産業革命」の入り口にいる。あらゆるものがインターネットに接続するIoTの広がり、あらゆる情報がビッグデータとして活用され、AI技術により、様々な分野で定型的な業務はもとより、人海戦術では不可能だった業務まで、AI技術を適用したサービス、ロボットの適用が始まっている。既存のビジネスや業務に新技術を取り入れるだけでなく、ビジネスモデルを変え、経済活用のみならず、個人の生活や社会構造まで影響が及ぶ。8）

デジタルフォーメーション（デジタル変革）とは、あらゆる情報がデジタル化され、IT技術によって、社会や産業、企業、人のあり方や働き方が変わっていくこと。第4次産業革命が進むにつれて、発展するビジネスと縮小するビジネスが明確になっていく。時代環境が大きく変わる時、それにそぐわないビジネスは淘汰されていく。匠の技的な高度な伝統的技能を要する作業や、旧来の延長線で仕組みの高度化、洗練により、生き残れるビジネスもあるが、現状維持のビジネスの多くは、相対的に意義を失う可能性が高い。

しかし、IoT、ビッグデータ、ロボット、AI等の技術を、クラウドコンピューティングやモバイル環境で容易に活用できるようになったことは、少ない投資で事業や先進サービスを立ち上げることが可能になった。各アーカイブ機関においても同様であり、利用者へのサービスを向上して、機関の存在意義を高める大きなチャンスでもある。

# 「未来の図書館を作るとは」【長尾先生】を読み返して

「未来の図書館」の概念

* 「人間の持っている知識は頭脳の中にあり、種々の知識が何らかの関係性によってつながれていて、連想的に関係する知識が取りだされている」
* 「図書館においてもぼう大な書物の中に存在する知識が関連性をもって書物という単位を超えてつなげられ、それが取り出されることが大切であろう。」
* 「本のある部分に存在する単語や概念を集め、それらに近い単語や概念が存在する部分を他の本について網羅的に調べる」
* 関連する知識を人間頭脳の中のネットワークのようにつないで、利用者の要求に応じて提示できるような形の電子図書館の内容の組織化が望まれているのである。
* 電子図書館における図書・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。これは1つの社会で共有する中立的な知識構造、知識システムである。
* 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ことができるだろう。
* 現実世界の本や情報の大切さ以上にヴァーチュアルな世界における情報処理と表現力の可能性にもっと大きな関心を持つべき時代に来ていると言えるのではないだろうか。

「未来の図書館を作るとは」（長尾真）より抜粋

「電子図書館」（1994年長尾真著）では、「既存の図書や資料をデジタル化すればそれで電子図書館が実現するかといえばそうではない。あるべき姿はデジタル化された情報を縦横に使いこなし、まったく新しい知的空間を創造するための図書館である。」9）とされており、「Ariadne」はその理念に基づいた実用化実証実験システムである。

「未来の図書館を作るには」5）が発行された2012年初めは、まだ第3次人工知能（AI）ブームの前で、AIはまだブレークスルーしていなかった。しかし、2012年以降のAIにおける機械学習はディープラーニング手法等により飛躍的に進展し、また、アーカイブ機関での資料のデジタル化、デジタルコンテンツのオープンデータ化、LOD化の加速化により、AIが扱える質の高いビッグデータが揃いつつある状況で、「未来の図書館を作るには」の中で「未来」と示唆されていた相当な範囲の仕組みが、今後5年程度で実用化を見通せるようになった。そこで、2017年時点での第4次産業革命等の動向を踏まえて、「未来の図書館を作るには」で示唆された提言の実現性について考察したい。

# 知的情報資源のビッグデータとしての活用の可能性

図書館等のアーカイブ機関は、従来から専門員の知識・ノウハウにより統制され形式知化してきた情報を保有している。二次情報としてのメタデータ、書誌情報、情報を見つけ出すための情報としてのレファレンス情報（参考情報）、Q&A、調べ方案内情報等は、既にビッグデータとして活用できる状況であるが、一次情報としてイメージデータ化されたデジタルコンテンツの内容本文は、検索にはほとんど利用されていない。

今後、ビッグデータとして活用が期待される情報として、アーカイブ機関のデジタルアーカイブ内で保有している一次情報があり、オープンデータ化されていない情報も含めて、全文テキストを活用した検索インデックスの作成をはじめ、AIの学習データとして活用するサービスが認められる方向である。

# 「知の共有化」システムの方向性

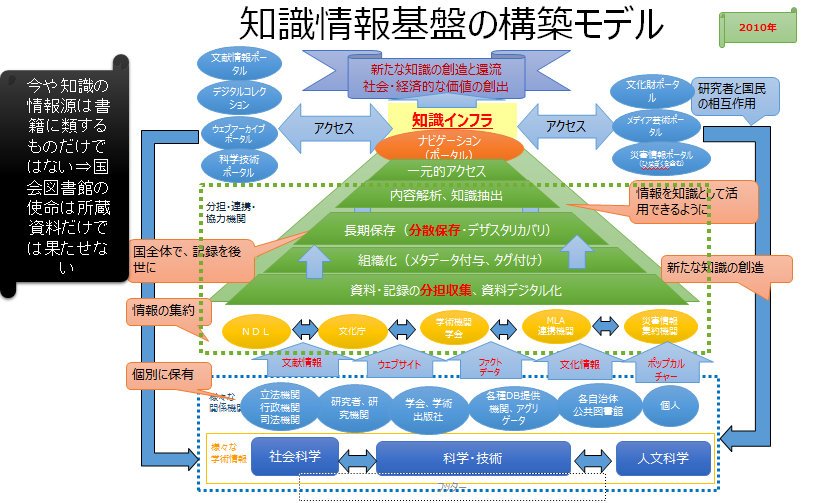
Web2.0時代と言われた頃、図書館界ではOpac2.0と称するサービスで、いわゆるAIとまでは言えないが、AIを指向した組織化、検索サービス等の実現を目指してきた。

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。

## 知識インフラの概念

* 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。
* 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

「未来の図書館を作るとは」（長尾真）より抜粋



既に、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）では、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」とされていた。10)

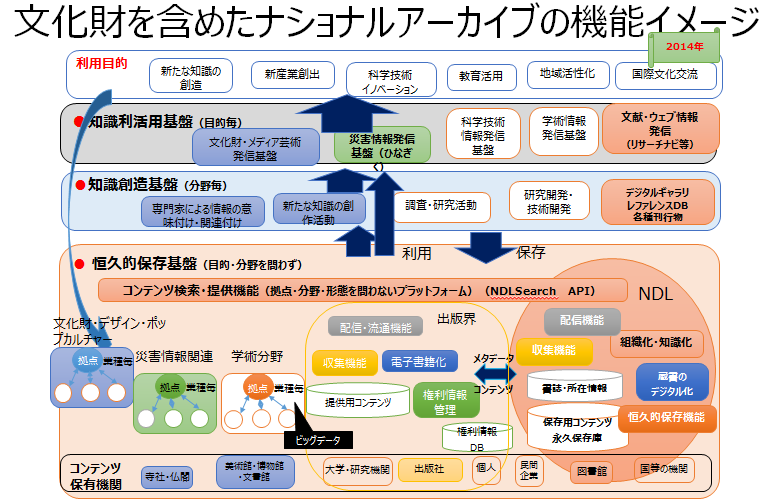
図1 知識情報基盤の構築モデル

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。

## 文化情報資源のナショナルアーカイブ

* 【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
* 【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」

「未来の図書館を作るとは」（長尾真）より抜粋

　「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。11）12）

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。13）

図2 文化財を含めたナショナルアーカイブの機能イメージ

## 恒久的保存基盤（あらゆる情報の恒久的保存）

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。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。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）

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。

### 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**

情報を関連付けるための情報として、まず共通語彙がある。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。

　情報と情報を結びつけるための情報として、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等があり、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。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。

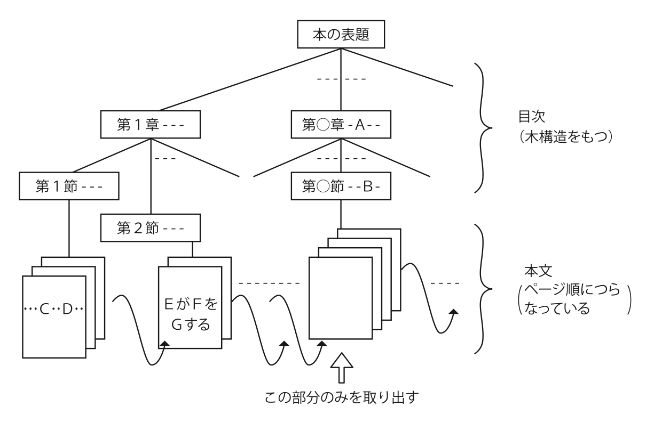
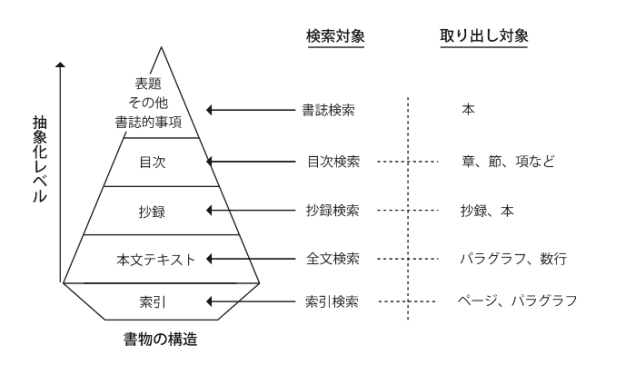


図4 書籍の構造化（「未来の図書館を作るとは」より）

図3 種々の検索システム（「未来の図書館を作るとは」より）

* 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

「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

「未来の図書館を作るとは」（長尾真）より抜粋

### 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。

* 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

「未来の図書館を作るとは」（長尾真）より抜粋

### 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。

* 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

「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

「未来の図書館を作るとは」（長尾真）より抜粋

* 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。

* 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）

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。

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。

* 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

「未来の図書館を作るとは」（長尾真）より抜粋

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。

### 自分の知識の外部記憶装置

* 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

「未来の図書館を作るとは」（長尾真）より抜粋

　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。

# 「知の共有化」システムの構築業務と、人材の資質とスキル

AIと人間の能力と役割の一般論として、AIが実用段階に達した今、今まで人が担ってきた部分の作業も、精密化するとAIを活用したほうが効率的なことが多々ある。それはアーカイブ機関でも同様であり、マニュアル化、ルーチン化されている業務は、一気にAIに置き換わらなくても、専門職員から、外部委託、非常勤職員、等へシフトされていくものも多い。その中でアーカイブ機関の職員としての仕事を見極める必要がある。

アーカイブ機関において、システムの調達・構築・運用のスキル・知識を持った人材が不足しているのが現状であり、適正な調達を行うために、政府機関での調達の標準ガイドラインに沿って開発タスクとドキュメントをひな形に進めることが効果的である。また、開発タスクを遂行する職員のスキル・知識の習得は、iコンピテンシ・ディクショナリ等を活用することが効率的である。特に、デジタルアーカイブ構築は、ビッグデータや人工知能の活用が必須であり、データサイエンス領域のスキル標準を活用すること有効と考える。

## 政府標準ガイドラインに沿ったシステム構築と運用

「政府情報システムの整備及び管理に関する標準ガイドライン」（政府標準ガイドライン）6）は、業務の効率化及び高度化、情報セキュリティを含む情報システムの運用リスクへの適切な対応等、具体的な取組を政府横断的に進めるため、情報システムの標準的な整備及び管理について、その手続・手順に関する基本的な方針及び事項並びに各組織の役割等を定める体系的な共通のルールとして策定し、政府機関全体でこのルールに沿って運用されている。

これは政府機関における調達の共通ルールであるが、知の共有化を進める公的機関でも適正な調達を行う際の手引きとして参考にし、調達手続き、調達仕様の内容を必要に応じて取捨選択して、システム構築。運用を進めることにより、重要事項の考慮漏れをなくすことができ、発注者と受注者の認識の齟齬を減らすことができる。

## 政府標準ガイドラインに沿った開発タスクとドキュメント

この図は、政府標準ガイドラインに沿って、組織としての事業計画に基づいた、業務・サービスの企画段階から、運用・保守、その後のシステム監査まで様々な業務（タスク）があるが、その各工程でのドキュメントを抜き出したもの。工程ごとに様々な種類の仕様書類があるが、テンプレート的にデフォルトとなる記述内容が示されており、計画書・企画書をベースに、必要な手続き、仕様内容を加筆訂正していく形でドキュメント化していくことが可能である。

アーカイブ機関のデジタルアーカイブ構築においては、特に、要求要件を明確にして、個々の機能要件部分を精緻化していくことに注力する必要がある。

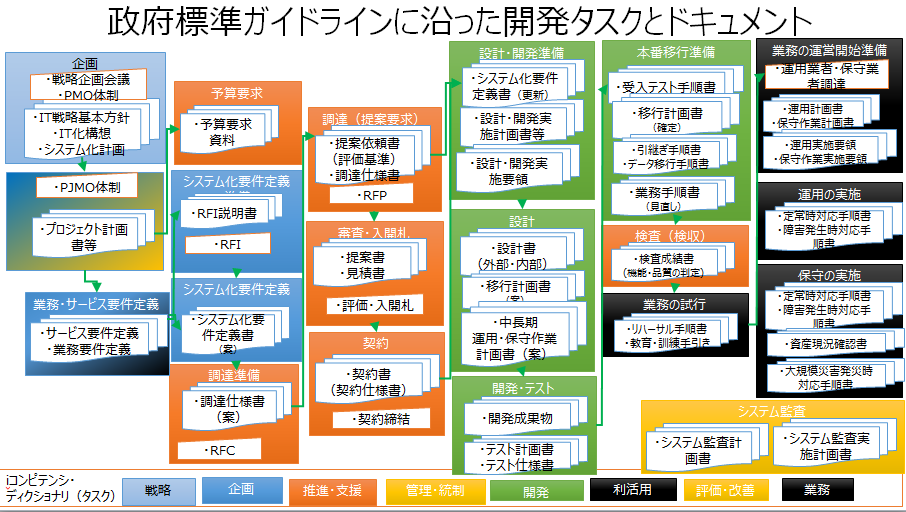


図5 政府標準ガイドラインに沿った開発タスクとドキュメント

### システムで何ができるようにするのか（個別機能要件）

機能に関する事項では、①アーカイブ化機能として、メタデータ、画像データの登録・変更・公開機能、AIを活用したレファレンスに必要な情報の蓄積機能、②サービス提供機能として、利用者に効率的で快適な体験を提供するカスタマーエクスペリエンス（CX）デザイン思考8）でのサービスを実現するユーザインタフェース、内部処理機能、外部サービスとの連携などの要求要件を明確にする必要がある。

　特に、AIを活用できる外部サービスを利用することとして、音声の自然言語による入出力、知識として学習の効率を高めるための教師データの選択、教師データを最適に学習できる機械学習APIを持つAI処理モジュールの選択が重要である。それにより、従来からのレファレンスの回答として、参考になる資料の選択と、その資料の所在情報の提示だけでなく、答えそのものの検索（事実検索）が可能になる。現在もスマートフォン等で利用可能になっているGoogle Assistant、 Apple Siri、Microsoft Cortana、Amazon Alexa、IBM Watson等の知識にデジタルアーカイブ機関が組織化して保有した情報が知識として蓄積されていれば、より的確の事実検索が可能になり、この外部サービスのAPIを利用することにより、アーカイブ機関でのAIを活用したサービスの構築が効率化する。

### どのようなデジタルコンテンツを用意するか（コンテンツ構築要件）

アーカイブ機関がデジタルコンテンツとして構築するコンテンツの要件も明確にして、①二次情報としての蓄積情報・データに関する事項である、永続的識別子、メタデータ、目次・索引データ、関連データ、②一次情報としてのコンテンツそのものである、動画、音声、画像データ、全文テキストデータ、更にIoT等からの収集データ等が他のアーカイブ機関のコンテンツと合わせて、ビッグデータとして利活用しやすいようにするために、より一層の共通的な仕様を適用することにより、AIでの活用を加速させられる。

　著作物の電子書籍化に関しては、原資料からのデジタル化（イメージ化、テキスト化）と、文献の作成段階からデジタル化され電子書籍と印刷書籍が同時並行で進むものがあるが、作成過程及び最終成果物の仕様を標準化することにより、シングルソース・マルチユースが容易になり、利用者の読書環境に応じて様々な形態で提供されることにより、著作物の利活用が促進される。

　現時点でのシングルソースとしてのマスター原稿段階では、章節項、目次、索引、引用等の構造を明示するXSLが適用されたXMLテキストとし、電子書籍化する場合は、様々な閲覧環境に最適な書式とできるように、html5+CSS3をベースとしたEPUB3.1の仕様を適用する。AIシステムの教師データ付き学習データとしては、マスター原稿段階のXMLテキストを利用することにより、構造化された知識として活用が容易になる。例えば、平成28年度情報通信白書（総務省編）は、マスター原稿から、ページレイアウト固定版（PDF版）、スマートフォン・PC・タブレット向けのリフロー版電子書籍（EPUB版）、Webブラウザ版（html版）が用意されている。

## iコンピテンシ・ディクショナリ（iCD）

i コンピテンシディクショナリ（iCD）は、情報処理振興機構（IPA）が作成した、システム開発におけるタスクとスキル・知識を体系的に洗い出したものである。それぞれ、タスクの種類を列挙したものが、「タスクディクショナリ」、スキルの種類を列挙したものが、「スキルディクショナリ」、知識の種類を列挙したものが、「知識ディクショナリ」として辞書化したものである。個々のタスクに必要なスキル、スキルに必要な知識が紐付けられている。また、事業を実施するために様々な業務があり、その業務も複数のタスクにより遂行される。そのタスクの固まりが「タスクプロフィール」として例示的に示されている

各アーカイブ機関が人材育成について検討する際、事業の内容に合わせて、「タスクディクショナリ」からタスクを選択することにより、そのタスクを遂行するために必要なスキル、知識が提示される。タスクを担当する人材は、そのスキル・知識を絞り込んで習得するようにすることにより、短期間に効率的に人材育成ができる。

## 業務遂行のタスクとスキル・知識の蓄積のスキーム

全てのスキルと知識を身に付いていれば、どんなタスクもこなせるが、そんな人材の確保は困難である。

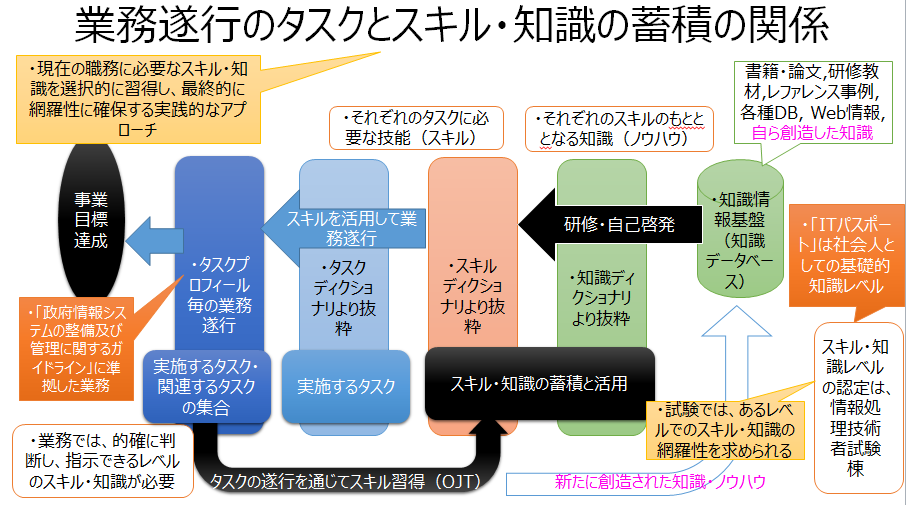


図6 業務遂行のタスクとスキル・知識の蓄積の関係

１つのタスクを実施するためには、様々なスキルが必要であり、１つのスキルは、経験と様々な知識により身に付く。現在の職務に必要なスキル・知識を選択的に習得し、最終的に網羅性に確保する実践的なアプローチが、iコンピテンシ・ディクショナリの考え方である。

業務を遂行するためには、的確に判断し、指示できるレベルのスキル・知識が必要であり、業務に必要なスキル・知識を、事前に選択的に習得して、業務を遂行することが望ましいが、業務を遂行する前に持っていないスキル・知識は、タスクの遂行を通じて習得（OJT）することで、業務を円滑に進められるようになる。

## iコンピテンシ・ディクショナリ（iCD）を活用した業務の遂行とスキル・知識の選択的習得

図7は、政府の標準ガイドラインで示された開発プロセスでの業務をタスクとして、それぞれのタスクの遂行に必要なスキル・知識の選択の概念を示したものである。

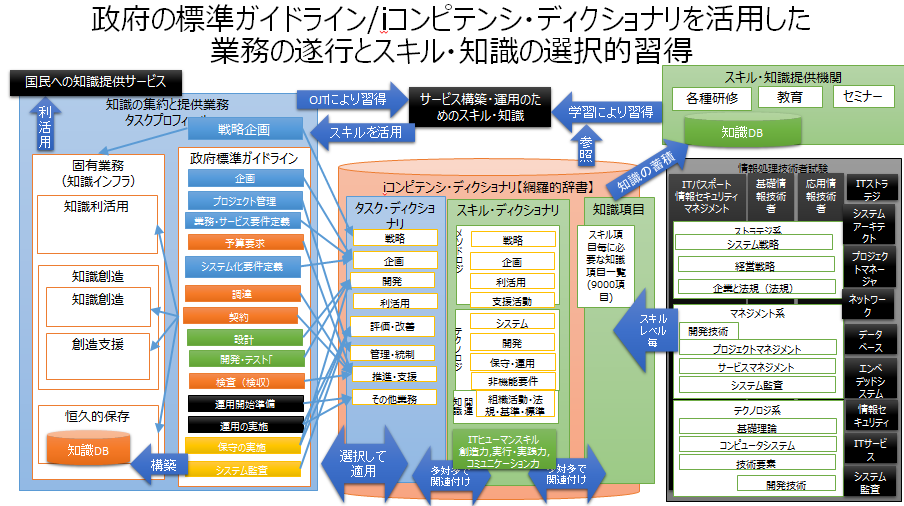


図7 政府標準ガイドライン／iコンピテンシ・ディクショナリを活用した業務の遂行とスキル・知識の選択的習得

更に一定レベルの網羅的な知識の習得状況を評価する情報処理技術者試験との関係も示している。

iCDでは、網羅的なスキルの領域として、①ビジネス活動の様々な手法、方法のスキルとして「メソドロジ」、②IT関連技法などのスキルとして「テクノロジ」、③対象となる業務を進めるための関連知識や業務固有のスキルとして「関連業務知識」、④ IT に関するタスクを実行する際に必要となる実行力・実践力、創造力、コミュニケーション力等のスキルとして「ITヒューマンスキル」とに分類されている。

「④ITヒューマンスキル」は、AIが実用化され、社会の構造が大きく変革しそうな時代において、「価値の創造・問題解決」を着実に遂行する力としての「創造力」は特に重要なスキルである。

また、第4次産業革命に対応した新スキル標準（ITSS+）14）として、「セキュリティ領域」とともに、「データサイエンス領域」に関して、大量データを分析し、その分析結果を活用するための一連のタスクとそのために習得しておくべきスキルカテゴリ、タスク構造が網羅的示されている。業務部門のタスクとして、業務設計、データの作成と収集、構造化データ加工、解析用データ準備、データの準備、データ解析、データ可視化、非構造化データ処理、評価、業務への組み込みと評価の工程が定義されている。

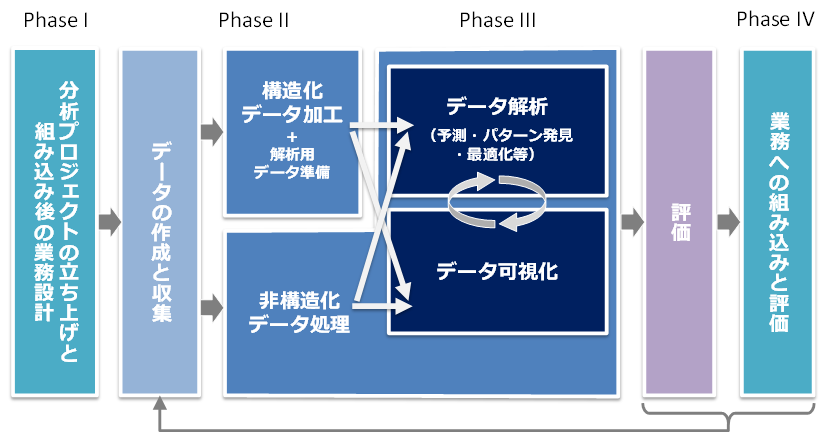


図8 データサイエンス領域のタスク構造

これからのデジタルアーカイブの構築は、ビッグデータやAIを活用が必須であり、各アーカイブ機関の業務部門が中心となって「データサイエンス領域」のタスク工程に沿って確実に進め、業務への適用を評価していくことが重要であり、業務部門での人材育成、人材確保が課題となる。

# まとめ

個々のアーカイブ機関でのデジタル化が進む状況において、文化的資産をあらゆる人々が将来にわたり享受、活用できるようにし、人々の創造的な活用に貢献するためには、個々の知識情報の「見える化」をするだけでなく、全体のそれぞれの施策が同一の方向性を持って、相互に資源を補完し合って、社会全体でより効率的な利活用の保証に取り組む必要があり、デジタル知識基盤は重要な役割を果たすことになる。文化情報資源を中核としたデジタル知識基盤を利活用したサービスの方向性を見極めて、個々のアーカイブ機関は、構成する一機関としてデジタルアーカイブの構築と提供を考えていくことが重要と考える。

今後5～10年のデジタル情報化、AIの実用化レベルの進展を見据えると、アーカイブ機関の業務は根幹から変革が求められると思われる。社会の変革の中でのアーカイブ機関は、従来の延長線上での業務を行っていては、有益なサービスを提供する機関としての存立が危ぶまれる。

AIは人の仕事が奪うよりも、人間がより人間らしい仕事にシフトしていくきっかけでもあり、仕事が奪われるとしたら、むしろ、AIを活用した省力化と新しいサービスに取り組まなかったために事業を継続できなくなる機関・組織であろう。知識創造のための情報の多くがインターネットから入手できる状況において、知識創造の支援の一翼を担う機関としての役割と必要な機能を再検討し、速やかに対応して、「知識インフラ」で目標とした「新たな知識の創造と還流により社会・経済的な価値の創出する仕組みが確立する」が加速され、「未来の図書館」が作られることを期待する。

# 参考文献

1. 兼松芳之，BBCC における電子図書館の研究：けいはんな学研都市フェスティバルより，カレントアウェアネス，CA979，http://current.ndl.go.jp/ca979（accessed 2017-9-1）
2. 田屋裕之．「パイロット電子図書館システム」の現状と課題．出版ニュース．1996，96 1/ 上・中，p.14-17
3. 中山正樹．国立国会図書館電子図書館中期計画2004の実施に向けて，情報の科学と技術，2004，Vol.54，No.9，p.453-460．
4. 久古聡美・吉田曉・中山正樹．Web2.0世代のデジタルアーカイブポータルの提供を目指して．情報管理．2006，Vol.49，No.6，p.313-323．
5. 長尾真，未来の図書館を作るとは．長尾真，2012，27p
6. 政府情報システムの整備及び管理に関する標準ガイドライン．各府省情報化統括責任者（CIO）連絡会議，2014，<http://www.soumu.go.jp/main_sosiki/gyoukan/kanri/infosystem-guide.html>
7. ｉコンピテンシ・ディクショナリ，情報処理振興機構，2015，https://www.ipa.go.jp/jinzai/hrd/i\_competency\_dictionary/icd.html（accessed 2017-9-1）
8. ベイカレント・コンサルティング．デジタルトランスフォーメーション．日経BP 社．2016．
9. 長尾真．「電子図書館」．岩波書店．1994年
10. 内閣府科学技術政策・イノベーション担当．“第４期科学技術基本計画（平成23年８月19日閣議決定）”．内閣府．<http://www8.cao.go.jp/cstp/kihonkeikaku/4honbun.pdf>　(accessed 2017-9-1）．
11. 中山正樹．国立国会図書館におけるデジタルアーカイブ構築―知の共有化を目指して―．情報管理．2012，Vol.54，No.11，p.715-724．
12. 中山正樹．国立国会図書館のサービスシステムの歩みと新たな方向性の模索―電子図書館事業20年の歩み―．国立国会図書館月報．2015，Vol.648，p.18-24
13. 中山正樹，電子図書館サービスからナショナルアーカイブの構築へ― LOD 化によるデジタル文化財の利活用を目指して―，TP&D フォーラム実行委員会，2016．25巻．p.13-33
14. データサイエンス領域のスキル標準「ITSS＋」，情報処理振興機構，2017，https://www.ipa.go.jp/jinzai/itss/itssplus.html（accessed 2017-9-1）

# 概要【削除予定】

## 目的・伝えたいこと

#### 目的

自分の理念で進めてきたことの集大成

今後進めてほしい方向性

現役時代に目標としてきたが到達できなかったこと

実現に当たっては、効率的・効果的に進めてほしいこと

#### 伝えたいこと

# 構成

#### はじめに

電子図書館構想、知識インフラ構想が、夢ではなく、やっと現実にできる時代になった

#### すでにデジタルトランスフォーメーション(デジタル変革)は始まっている

目の前の業務だけにとらわれることなく、広く視野を持って進むべき道を探り、学ぶ。勉強会やコミュニティなど、学びの場は周囲にある。自己研さんによって能力を高めれば高めただけ、社会をリードしていく人材になっていく

#### 本格的なAIの実用化を迎えて

AIの進展

デジタルアーカイブは、文化資源を保有する組織、その資源を利活用する組織のデジタルフォーメーションである

「未来の図書館を作るには」【長尾先生】の再認識

「未来の図書館を作るとは」も、AIが実用レベルになったことで、「未来の図書館」ではなく、「今の図書館を作るとは」として設計構築できるようになった

#### 文化情報資源のナショナルアーカイブの方向性は

ビッグデータとしての知の共有化、AIが適用された知識インフラの構築

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、さらに今まで不可能だった新たなサービスが生まれようとしている

ナショナルアーカイブの各基盤の概念

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・運用していくことが必要

#### AI等の新技術と図書館サービス【未完】

ビッグデータとしての知的情報資源

#### 次世代の図書館サービスの業務と、業務に従事する人材の資質とスキルは？

AIと人間の能力と役割（一般論）

AIの実用段階に達した今、今まで人が担ってきた部分の作業も、精密化すると一部がAIを活用したほうが効率的なことが多々ある

図書館員の役割と資質

一気にAIに置き換わらなくても、少なくとも職員から、アルバイト、非常勤、外部委託へとシフトしていくものが多い。その中で図書館職員としての仕事を見極める必要がある

政府標準ガイドラインに沿った開発タスクとドキュメント

開発は、政府の調達ガイドラインに沿うことが有効

業務遂行のタスクとスキル・知識の蓄積の関係

iコンピテンシ・ディクショナリを活用した業務の遂行とスキル・知識の選択的習得方法

必要なスキル。知識は、網羅性を確保するより、担当するタスクに応じて選択的に順次習得することが効率的

#### データサイエンス領域のスキル標準「ITSS+」の適用【要約】

知識インフラの構築は、ビッグデータやAIの活用が必須であり、典型的なデータサイエンス領域として位置づけられる

#### 「見たことのない図書館を考える（2015年中山）」でお話ししたことの実現性は？

図書館員等に求められるもの

#### 次世代サービス構築及び提供に当たって留意してほしいこと

効率化、人はより創造的な業務へ

今の延長で可能な範囲での実施でなく、今後10年のAI技術の進展、その時点の社会のニーズを見据え、国全体で何をする必要があるかを考える

#### まとめ

文化情報資源を中核とした「知識インフラの構築と利活用」のサービスの方向性を見極めて、提供する機関の1つとしての図書館情報サービスの構築と提供を考えていくことが重要と考える

2040年代には、AIが意志や感情を持って人間を超える「シンギュラリティ―時代」が来ると言われて、また10年後にはどんな新しい技術が確立しているかわからないが、現時点でのAIの実用化レベルでも、社会は大きく変革する。

その社会の中での図書館サービスは、従来の延長線上での業務を行っていては、有益なサービスを提供する機関としての存立が危ぶまれる

AIにより、人の仕事が奪われるのではなく、より人間らしい仕事にシフトしていく。 仕事が奪われるとしたら、むしろ、AIを活用した省力化と新しいサービスに取り組まなかったために事業を継続できなくなる機関・組織であろう。あ

# 【個別詳細編】

# 背景

【経緯】

1980年代から、海外の多くの先進的な図書館同様、日本でも電子図書館の研究開発プロジェクトを発足させ、1994年には、京都大学において、我が国で最初の実用化実証システムとして、「ネットワーク環境における情報と文献の利用のための高度検索システム（Ariadne）」が開発され、NDLでは、大規模な電子図書館の実証実験プロジェクトが実施された。その後、NDLは、2002年から本格的なサービスとして離陸し発展させて、現在に至っている。

NDLは、2004年2月に「電子図書館中期計画2004」（中期計画）を策定した。この中期計画において、デジタルコンテンツを広汎な利用者に提供するために、NDLが国のデジタルアーカイブの重要な拠点となるということ、また国内外の多様な利用者層の需要に応じ、日本のデジタル情報全体へのナビゲーションを行う総合サイトとして「国立国会図書館サーチ」を構築し、利用者がワンストップで利用できるようにすることを目指した。

図書館に限らず、様々な業種・業態の組織において、大量の情報の中から効率的に適切な情報を集約し、必要な情報を取り出す情報提供システムが構築・運用されている。情報システムの観点では、図書館システムは、そのような情報システムの中での１アプリケーション分野であり、文化資源のナショナルアーカイブサービスとしても、利用者にとってOneOfThemのサービスであると考える。

そのような基本的な考え方で、図書館以外の情報資源、技術ノウハウを活用して、利用者の情報取得、知識創造を効率化するサービスを、如何にして効率的に構築・運用できるようにするかを念頭に進めてきた。次世代図書館サービスシステムは、従来型の図書館の延長線で考えるのではなく、世の中の多くの情報提供サービスの中で信頼性の高い情報を提供する機関として、分野横断的な「知の共有化」の観点で、中核的な組織としての責任と義務を果たせるサービスの提供を目指すべき考える。

【今後】

2000年代より、クラウドコンピューティングの普及に伴って、ウェブ上の情報を意味的に理解できる形で公開・共有を目指すセマンティックウェブ技術の中で、構造化された情報同士をリンクさせるための技術や方法論の総称としてのLinked Open Data(LOD)が注目されてきた。この概念は、NDLが1990年代に構想した電子図書館構想の「地球規模の知的財産を誰でも容易に利用できるようにする」という理念の実現のために、電子図書館サービスの構築の仕組みの考え方や、NDLが個別に適用してきた要素技術の集合としてみることができる。電子図書館サービスの発展形としての国の文化財全体の利活用を目指した図書館を中核とした情報保有機関によるデジタル文化財のアーカイブ構築において、LODの考え方がどのように適用されていくべきかの方向性について、技術的な観点にフォーカスして考察する。

さらに、昨今の「第4次産業革命」と言われる産業の変革は、図書館を含めて文化資産が保有している情報のデジタルトランスフォーメーション（デジタル革命）によって集約される情報と、IoTにより産業界、学術機関が収集・蓄積するデータを、ビッグデータとしてAIにより「知識化」されることにより知識が再利用されて、「新たな知識」として創出されることが期待できる。

「未来の図書館を作るとは」（2012年長尾元国立国会図書館長）が示されたことの多くが「未来」ではなく、ごく数年の間に実現できるレベルまで「第4次産業革命」は進んできている。

このような時代の「情報システム構築と運用」は、「政府調達標準ガイドライン」に沿って構成で効率的な開発工程（タスク）で進めることとして、そのタスクに従事するために、iコンピテンシ・ディクショナリを参考にして、必要なスキルと知識を絞り込んで習得していくことが効率的・効果的である

～～～～

図書館サービスの将来方向として、出版物に限らず様々な機関が保有する文化情報資源を一元的に利活用できるようにして、新たな知識を創造する「知識インフラ」としての方向性を検討してきた

その方向性の実現に向けて、メタデータの集約による統合検索が可能な「知識インフラ」を発展させ、情報資源そのもの（画像データ、本文フルテキスト等）をビッグデータとして活用して、人工知能により知識化し、ファクトを取り出せる基盤としての「真の知識インフラ」の構築の実現を見通せる時代になった

「電子図書館」は、「真の知識インフラ」として、「情報の集合体としてのアーカイブ」に留まらず、「人工知能により知識化」され、利用者に対して「電子図書館」を「脳の外部記憶」として利用を可能にする実用化システムの構築も実現可能なところまで到達しつつある。

そのような時代に、図書館を含めた文化情報資源の保有機関は、どんな業務・サービスに人工知能を活用すべきか、そのためにどんな知識とスキルを持つべきかを考察する

# すでにデジタルトランスフォーメーション(デジタル変革)は始まっている

## 第4次産業革命の入り口

**[【参照】IT人材白書2017【2017年4月IPA】](https://www.ipa.go.jp/jinzai/jigyou/about.html)**

今は、IoT、ビッグデータ、ロボット、AI等の技術革新による、第4次産業革命の入り口にいる。あらゆるものがインターネットに接続するIoTの広がり、あらゆる情報がビッグデータとして活用され、AI技術により、様々な分野で定型的な業務はもとより、人海戦術では不可能だった業務まで、AI技術を適用したサービス、ロボットの適用が始まっている。既存のビジネスや業務に新技術を取り入れるだけでなく、ビジネスモデルを変え、経済活用のみならず、個人の生活や社会構造まで影響が及ぶ。

デジタルフォーメーション（デジタル変革）とは、あらゆる情報がデジタル化され、IT技術によって、社会や産業、企業、人のあり方や働き方が変わっていくこと。第4次産業革命が進むにつれて、発展するビジネスと縮小するビジネスが明確になっていく。時代環境が大きく変わる時、それにそぐわないビジネスは淘汰されていく。匠の技的な高度な伝統的技能を要する作業や、旧来の延長線で仕組みの高度化、洗練により、生き残れるビジネスもあるが、現状維持のビジネスの多くは、相対的に意義を失う可能性が高い。

しかし、IoT、ビッグデータ、ロボット、AI等の技術を、クラウドコンピューティングやモバイル環境で活用できるようになったことは、少ない投資で事業を立ち上げることが可能であり、中小企業、ベンチャー企業や個人の活躍のまたとないチャンスである。

## 組織として、

時代の潮流を捉えて、組織が社会の変化の中で、時代に適合して発展できる道を探り、ビジョンをはっきり示すことが重要であり、それは組織管理者者の責務

「デジタルトランスフォーメーション」を実現するには、ビジネスとデジタルのスキルを併せ持った人材の育成と獲得をしていく必要がある

## 個人として、

自らも「デジタルトランスフォーメーション」の流れの中にあることの意識

求められるのは、周囲を巻き込みながら改革を進める能力やビジネスとデジタルを結び付けて全体をデザインする能力を持った人材になること

目の前の業務だけにとらわれることなく、広く視野を持って進むべき道を探り、学ぶ。勉強会やコミュニティなど、学びの場は周囲にある。自己研さんによって能力を高めれば高めただけ、社会をリードしていく人材になっていく

# 本格的な人工知能の実用化を迎えて

電子図書館は、図書館界、出版界におけるデジタルフォーメーションであり、また、ナショナルアーカイブは、文化資源を保有する組織、その資源をデジタルアーカイブとして利活用できるようにするデジタルフォーメーションである。

~~今までは、~~~~「電子図書館サービス構想」の実現形としての「デジタルコレクション」、「NDLサーチ」、「リサーチナビ」を構築し提供してきた。~~

電子図書館サービスの発展形として、20XX年に「知識インフラ構想」が提唱された。

「知識インフラ構想」の実現形を目指して分野を特定した「東日本大震災アーカイブ」が、そして「東日本大震災アーカイブ」の発展形として分野を文化情報資源まで拡大した「文化資源のナショナルアーカイブ」が構想されたが、今後は、「デジタルトランスフォーメーション時代の知識インフラの構築」を意識していく必要がある。

人工知能の進展は、ビッグデータをディープラーニングにより学習できるようになったことにより、第3次人工知能ブームが起き、様々な分野で実用化が進んできた。今の技術の延長線でも、コンピュータが人間の頭脳を上回るシンギュラリティ（技術的特異点）が2044年までに起きると言われている。また、突然変異的に技術が生み出された場合、もっと早く訪れる可能性もある。

## 「未来の図書館を作るには」【長尾先生】を読み返して

「電子図書館」（1994年長尾真著）では、「既存の図書や資料をデジタル化すればそれで電子図書館が実現するかといえばそうではない。あるべき姿はデジタル化された情報を縦横に使いこなし、まったく新しい知的空間を創造するための図書館である。」とされており、「Ariadne」はその理念に基づいた実用化実証実験システムである。

「未来の図書館を作るには」（2012年長尾真）が発行された2012年初めは、まだ第3次人工知能ブームの前で、人工知能がブレークスルーしていなかった。しかし、2012年以降のディープラーニングに代表される人工知能の進展、資料のデジタル化、デジタルコンテンツのオープンデータ化、LOD化の加速化により、「未来」とされてきた相当な範囲の仕組みで、今後5年程度で実用化できるものが多くなった。そこで、現時点での第4次産業革命等の動向を踏まえて、「未来の図書館を作るには」で示唆された提言の実現性について考察したい。

### ~~「電子図書館 」岩波文庫、 1994年、 長尾真~~

~~「未来の図書館を作るには」~~

* ~~第2次人工知能ブームの終盤、第3次人工知能ブームへの助走段階~~
* ~~既存の図書や資料をデジタル化すればそれで電子図書館が実現するかといえばそうではない。あるべき姿はデジタル化された情報を縦横に使いこなし、まったく新しい知的空間を創造するための図書館である。そのために何が必要かを説く本書の構想は挑戦的かつ刺激的な未来の設計図だ。~~

~~「電子図書館」（長尾真著）より抜粋~~

### ~~「未来の図書館を作るとは」が刊行された2012年からの技術の進展~~

~~「未来の図書館を作るには」が発行された2012年初めは、まだ第3次人工知能ブームの前で、人工知能がブレークスルーしていなかった。~~

~~「未来の図書館を作るには」では、「未来」の話しとして、人工知能を活用したシステムにより、図書館サービスや図書館員の仕事の変革が示唆されていたが、人工知能の実用化により、今まさに実現可能な段階になっていると思う~~

~~しかし、2012年以降のディープラーニングに代表される人工知能の進展、資料のデジタル化、デジタルコンテンツのオープンデータ化、LOD化の加速化により、「未来」とされてきた相当な範囲の仕組みで、今後5年程度でできることが多くなってきたように思える~~

### 人工知能と図書館

* 「人間の持っている知識は頭脳の中にあり、種々の知識が何らかの関係性によってつながれていて、連想的に関係する知識が取りだされている」
* 「図書館においてもぼう大な書物の中に存在する知識が関連性をもって書物という単位を超えてつなげられ、それが取り出されることが大切であろう。」
* 「本のある部分に存在する単語や概念を集め、それらに近い単語や概念が存在する部分を他の本について網羅的に調べる」
* 関連する知識を人間頭脳の中のネットワークのようにつないで、利用者の要求に応じて提示できるような形の電子図書館の内容の組織化が望まれているのである。
* 電子図書館における図書・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。これは1つの社会で共有する中立的な知識構造、知識システムである。
* 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ことができるだろう。
* 現実世界の本や情報の大切さ以上にヴァーチュアルな世界における情報処理と表現力の可能性にもっと大きな関心を持つべき時代に来ていると言えるのではないだろうか。

「未来の図書館を作るとは」（長尾真）より抜粋

### 図書館関連が保有しているビッグデータとしての知的情報資源

#### 従来から図書館員の知識・ノウハウにより統制され形式知化してきた情報

図書館等は、従来から図書館員の知識・ノウハウにより統制され形式知化してきた情報を保有している。二次情報としてのメタデータ、書誌情報、情報を見つけ出すための情報としてのナレッジデータ、レファレンス情報（参考情報）、Q&A、調べ方案内情報等があり、これらは既にビッグデータとして活用できる状況であるが、一次情報としてイメージデータ化されたデジタルコンテンツの内容本文は、検索にはほとんど利用されていない。

今後、ビッグデータとして活用が期待される情報として、デジタルアーカイブ内で保有している情報があり、オープンデータ化されていない情報も含めて、バックヤードにおいては、検索・分析のために、全文フルテキスト化し、人工知能の学習データとして活用したサービスが認められる方向である。

~~を活用した新サービスが認められる方向である。に活用できる本文フルテキストがある。これらは、著作権者の許諾なしに書籍の全文をテキスト化し、検索のために人工知能などを活用した新サービスが認められる方向である。図書館等が収集した資料のうち一定期間公開しない有償オンライン資料、東日本大震災アーカイブ等の資料も検索のためには活用できる方向になるだろう。~~

#### *今後活用を期待したい情報*

*今後活用を期待したい情報として、図書館が保有するオープンデータ化された情報のコンテンツ**、図書館の枠を越えたナショナルアーカイブ内のコンテンツ**、図書館員が持つノウハウ等の暗黙知**としてのレファレンス業務過程情報も形式知化することにより活用が期待できる。*

*また、今後、図書館情報システムから収集し、活用が期待される情報として、**匿名化した加工した利用実績情報（統計化されてない情報）がある。*

*その外にも、**図書館内での各種センサー等のIoT機器から収集する情報の移動履歴もサービスの向上につながる*

# 文化情報資源のナショナルアーカイブの方向性は

ビッグデータとしての知の共有化、人工知能が適用された知識インフラの構築

## 今後

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、いわゆる人工知能とまでは言えないかもしれないが、人工知能を指向した組織化、あいまい検索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ように方向に向かっている

~~また、今後の著作権法の改正により、著作者の許諾なく、著作物の本文そのもの活用して、人工知能等により高度な検索、分析に活用できるようになる~~

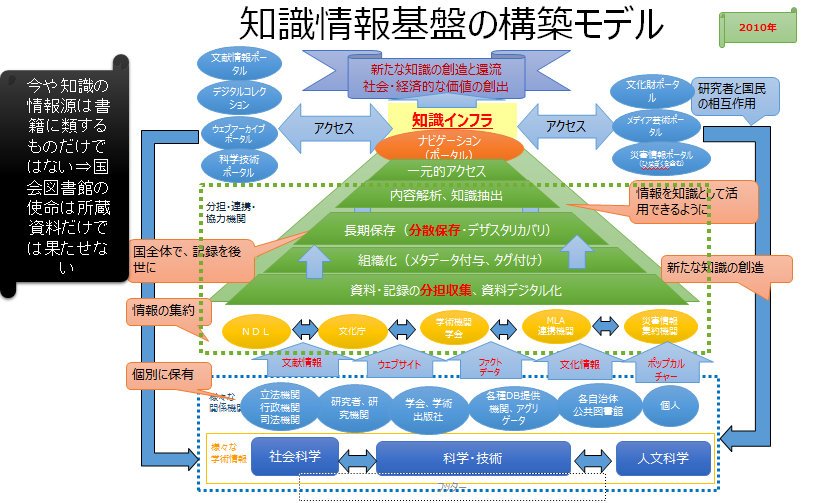
このような時代に、従来の業務やサービスは、人工知能が組み込まれたシステムやロボットに支援されてサービスが省力化され、さらに今まで不可能だった新たなサービスが生まれようとしている

## 文化情報資源のナショナルアーカイブの方向性

### 知識インフラ

* 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。
* 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

「未来の図書館を作るとは」（長尾真）より抜粋



### ナショナルアーカイブは

世界中の知識情報が有機的に結合したものが、「インターナショナルアーカイブ」と言える

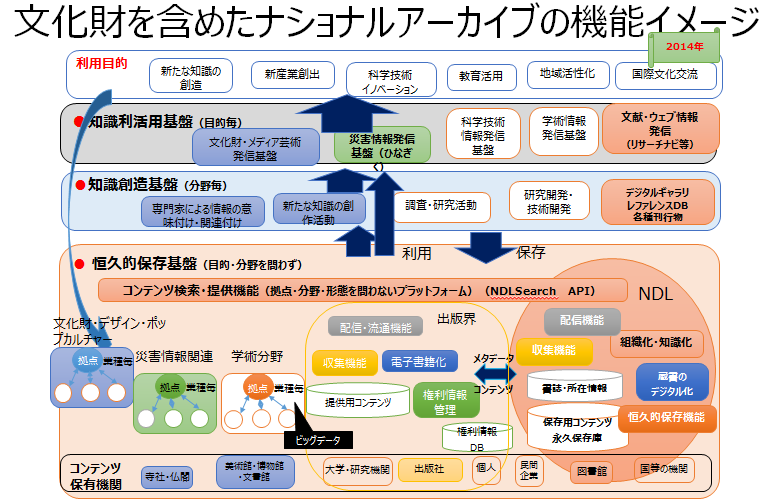
* 図書館は、書物等の情報資源を収集・保存し提供する場であるとともに、それらを使って関心のある人が集まって議論し新しい知識を創造する場
* 日本としてのナショナルアーカイブは、日本中に存在する知識情報が有機的に結合され、日本中の人が自由に使える日本の「知識インフラ」

「未来の図書館を作るとは」（長尾真）より抜粋

### 今、ディープラーニングが実用化になった時代に、ナショナルアーカイブはどのような方向へ進んでいくか

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、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、人工知能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「人工知能を持った知的情報基盤」となる。

## ナショナルアーカイブの各基盤の概念



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・運用していくことが必要

### 恒久的保存基盤（あらゆる情報の恒久的保存）

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。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。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）

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

#### 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、人工知能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人工知能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し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、人工知能を利用して、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。人工知能により、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、容易に人工知能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、人工知能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**として、まず共通語彙がある。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、人工知能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。

情報と情報を結びつけるための情報として、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等があり、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。

~~分野別辞書~~

~~大漢和辞書等~~

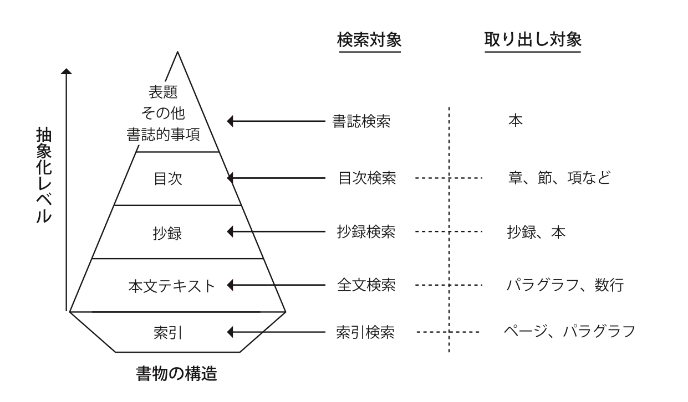
~~分野別シソーラス~~

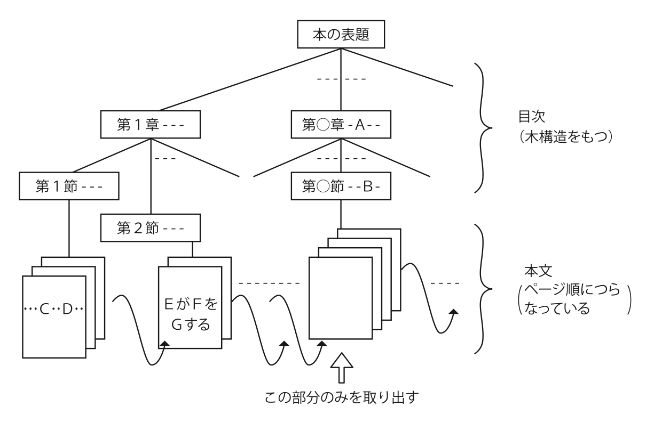
~~JSTシソーラスmap等~~

~~個別情報の永続的識別子（Linked Data化）~~

●情報の組織化

**情報に関する情報の組織化**として、書誌、メタデータ作成があるが、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、人工知能により自動付与が可能になる。専門家は、人工知能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人工知能により省力化して付与が可能になる





* 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

「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

「未来の図書館を作るとは」（長尾真）より抜粋

~~全文テキストを利用することにより、新たに時間軸、空間軸に沿った知識の関連付け等の組織化も可能ではないか~~

~~レファレンス事例、司書の知識・ノウハウを、信頼性の高い情報を、人工知能の教師用データとして特徴量を自動抽出し、情報を知識化して人工知能の知識データベースとして蓄積~~

#### 知識創造活動

新たな知識を創造する活動であり、図書館等文化情報資源を保有する機関の主たる目的の一つである。これは、知識情報基盤の中で、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り、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る。創造された知識は、恒久的保存基盤でアーカイブされる。

* 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

「未来の図書館を作るとは」（長尾真）より抜粋

#### 新たな知識創造の場としての図書館

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あり、相互にシームレスに、「場」として利用できるものにしていく。

* 情報を媒介して専門家と専門家を繋ぐ

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

* 専門家を媒介して辞書と辞書を繋ぐ

様々な分野の人たちが繋がって、情報に関する情報が蓄積され、それによって様々な各種辞書が関連付けられていく

* 読者と読者を繋ぐ場としての図書館
* 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

「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

「未来の図書館を作るとは」（長尾真）より抜粋

* 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、人工知能を利用して、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。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、知識インフラを知識とした人工知能が、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。

* 事実に基づいた報告書

依頼または予測調査に関しては、人工知能を利用した文献内容検索により、最もマッチングする資料名、資材場所を探し出し、オープンデータに関しては、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可能になる

~~知識創出システムの構築~~

~~人と人工知能の協調による創造的な課題解決~~

~~人工知能の助けを借りて新サービスの構想を練り開発する~~

~~人とシステムの得意な作業が連携~~

### 知識利活用基盤（情報発信）

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。

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

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

#### 知識検索機能

* 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**

「未来の図書館を作るとは」（長尾真）より抜粋

#### 自分の知識の外部記憶装置（自分用の電子図書館【長尾先生】）

* 【長尾先生】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【長尾先生】

「未来の図書館を作るとは」（長尾真）より抜粋

# 人工知能等の新技術と図書館サービス【未完】

## 従来型システム構築

要件定義⇒システム化要件定義⇒設計⇒開発⇒運用

手続きをアルゴリズム化⇒設計書⇒プログラミング

## 要素

AI

ロボット

IoT

クラウド

テレワーク

マーケティングオートメーション

Web接客ツール

チャットボット

システム

クラウドMLの利用

自社だけでは開発できないような最新のアルゴリズムでも、低価格で利用可能になる

自力で機械学習の実験が可能になる

今はプログラミング技術を持たなくても、機械学習を活用できる時代になった

## 機械による情報処理の進展

今でいう人工知能でなくとも、情報処理の進展により、人に代わって、サービスを提供してきている

10年先の新技術はわからないが、今確立しつつある技術は、明らかに実用レベルになる

今後10年で広く普及する次世代技術

あらゆる情報のビッグデータ化

図書館としてのデジタルアーカイブ⇒国としてのデジタルアーカイブ⇒業種業態を問わない機関が保有する情報、関連情報を合わせて、ビッグデータ化

第3世代人工知能

loT

## ビッグデータとしての知的情報資源

### 図書館での従来から形式知化してきた情報

情報に関する情報⇒メタデータ

書誌情報

情報を見つけ出すための情報⇒ナレッジデータベース

レファレンス情報（参考情報）

Q&A

調べ方案内情報

情報資源⇒デジタルアーカイブ

### ビッグデータとして活用していく情報

#### 図書館情報システムから収集する情報

匿名加工情報

**利用情報**

#### 図書館が保有する情報資源のオープンデータ化

図書館の世界でのデジタルアーカイブ

#### IoTから収集する情報

利用行動履歴

利用した情報の移動履歴

研究データ

レファレンス業務過程情報⇒教師データ

#### 図書館の枠を越えて⇒ナショナルアーカイブ

#### 知的情報資源の分散保有

ブロックチェーン

図書館外が保有する情報

図書館が保有する情報

### サービスのマッピング（図書館での活用例）

#### デジタル化

スキャナーによるイメージ化

OCRによるテキスト化

画像認識機能

本文構造化

#### 組織化

●人（カタロガー）

教師データ作成

書誌作成

クラス分類

辞書作成

Wikipediaの内容評価

機械学習の精度を上げるための教師データ化([**書誌作成**](#______16)、 [**クラス分類**](#_______8)、 [**辞書作成**](#______17))

●AI

書誌的事項の作成

過去の書誌データを教師データとして

新しい文献の書誌を自動生成

AIによる自動書誌作成([**過去の書誌データを教師データとして**](#___________________12)、 [**新しい文献の書誌を自動生成**](#_______________6))

本文組織化

クラスタリング

サブトピック 3

#### 知識DBを構築

●従来アーカイブ

書誌情報

本文情報

レファレンス情報

●ビッグデータ

図書館保有情報

人が作った情報

書誌情報

本文情報

レファレンス情報

予測調査レポート

参考文献の信頼性評価

機械学習の精度を上げるための教師データとして利用([書誌情報](#______20)、 [本文情報](#______21)、 [レファレンス情報](#__________7)、 [予測調査レポート](#__________8))

システム、IoT等から収集

利用者

利用情報

利用者行動情報

実績値として利用([利用情報](#______22)、 [利用者行動情報](#_________12))

図書館員

書誌作成業務の過程情報

レファレンス業務の過程情報

教師データとして利用([書誌作成業務の過程情報](#_____________9)、 [レファレンス業務の過程情報](#_______________7))

他機関保有情報

#### 情報探索

●従来型システム

書誌的事項検索

キーワード検索（書誌的事項全文検索）

典拠検索

連想検索．．．．

従来型OPAC([書誌的事項検索](#_________15)、 [キーワード検索（書誌的事項全文検索）](#____________________5)、 [典拠検索](#______25))

AI指向のOPAC([連想検索．．．．](#__________10))

●AIシステム

シソーラス、オントロジー検索

協調フィルタリング

アイテムベースレコメンド

「ユーザベースレコメンド」

キーワードサジェスト

曖昧検索

あいまいな検索キーワード

意味的に類似性の高いパラグラフの抽出

本文全文テキストの意味的検索

パターン認識検索

音声認識検索

自然言語音声による探索指示と、解答そのもの（事実）の音声回答

AIを活用した検索([シソーラス、オントロジー検索](#________________9)、 [協調フィルタリング](#___________5)、 [キーワードサジェスト](#____________8)、 [曖昧検索](#______26)、 [本文全文テキストの意味的検索](#________________10)、 [パターン認識検索](#__________11)、 [音声認識検索](#________18)、 [自然言語音声による探索指示と、解答そのもの（事実）の音声回答](#________________________________3))

#### レファレンス業務

利用者の求める資料名と所在情報を提示

●人（レファレンサー）

情報検索支援

教師データの作成

司書の知識・ノウハウに基づいて、複数の資料をピックアップし内容を確認

利用者が必要とする資料に最もマッチングする資料を提示

事実検索支援

●AI

ビッグデータから的確な情報を探し出す

チャットボット

音声認識機能

#### 業務

アナリスト

markerpeople-red

#### 課題

「図書館の自由」を尊重する中での利用情報の高度利用の可能性

# 「知の共有化」システムの構築業務と、業務に従事する人材の育成

markerflag-red

## 人工知能が果たす役割・機能

不足している労働力を補完する

既存の労働力を省力化する

既存の業務効率・生産性を高める

既存の業務の提供する価値（品質や顧客満足度など）を高める

これまでに存在しなかった新しい価値をもった業務を創出する

既存の業務に取組む意欲や満足度を高める

新しい業務に取組む意欲や満足度を高める

その他

## 人工知能（AI）の活用が一般化する時代における重要な能力

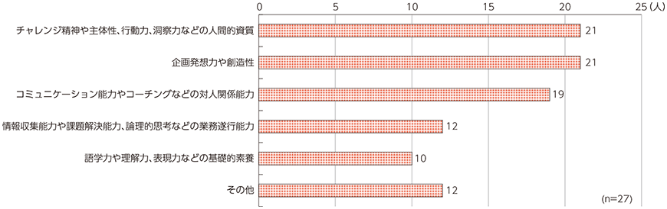
情報収集能力や課題解決能力、論理的思考などの業務遂行能力

チャレンジ精神や主体性、行動力、洞察力などの人間的資質

企画発想力や創造性

語学力や理解力、表現力などの基礎的素養

コミュニケーション能力やコーチングなどの対人関係能力



[（総務省「ICTの進化が雇用と働き方に及ぼす影響に関する調査研究」）](http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/html/nd270000.html)

ルーティンワーク、マニュアル化された仕事は、アルバイト、非常勤、外部委託、そして人工知能に置き換わる

## ~~ビジネスプロセスにおける業務とシステムの役割分担の見直し~~

~~AIの実用段階に達した今、今まで人が担ってきた部分の作業も、精密化すると一部が人工知能を活用したほうが効率的なことが多々ある~~

## 業務担当の役割と資質

Librarian、 cataloguers、 reference librarian、 Embedded librarian

業務として、

人工知能により自動出力された内容の評価、補正

選書

関連付けに必要な典拠類の構築

情報に関する基本情報付け（メタデータ付与）

情報に関する付加価値情報付け

情報間の関連付け

分類・主題情報の付与

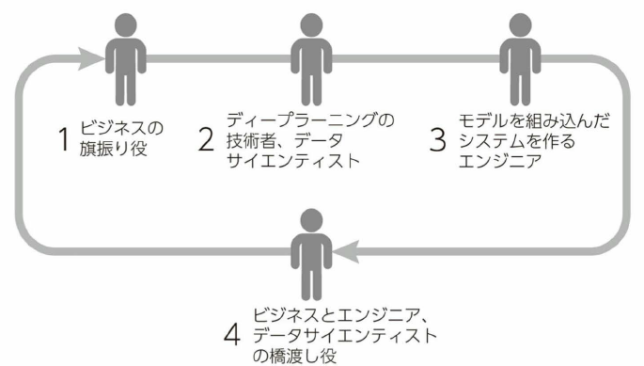
人工知能を活用したレファレンス

### 集いの場（intellectual commons）の運営

研究グループの中に図書館的業務のできる人を置くことが必要。embedded librarian（研究協力図書館員）と呼ぶようになって来た。【長尾先生】

## IT担当部門

【グーグルに学ぶディープラーニング（日経ビッグデータ）】



l .ビジネスの旗振り役2 ディ－プラ一二ングの技術者、データサイ工ンテイスト3. モデルを組み込んだシステムを作る工ンジ二ア4. ビジネスと工ンジ二ア、データサイ工ンテイストの橋渡し役

### .ビジネスの旗振り役

CIO

CDO

**データサイエンティスト**

技術

ビジネス

分析

コミュニケーション能力

主流のプログラミング言語でプロトタイプを開発でき、数学、統計学、確率論、コンピュータサイエンスの確かな素養のある人材

かつ、ビジネス課題への感覚が鋭く顧客に共感できること

### ビジネスと工ンジ二ア、データサイ工ンテイストの橋渡し役

4の橋渡し役が既存の組織にはない、一方で重要な役割を果たす人。工ンジ二アと人工知能の技術を理解しながらビジネスとつないでいく人

## ~~図書館サービスの構築・運用に従事する人に必要な知識とスキルは？~~

~~一気に人工知能に置き換わらなくても、少なくとも職員から、アルバイト、非常勤、外部委託へとシフトしていくものが多い。その中で図書館職員としての仕事を見極める必要がある~~

## 図書館情報システムに係わる人材に必要な知識とスキル

### 図書館システムの構築に必要な知識とスキル

AIシステムの回答精度を上げるための知識とスキル

### 図書館情報システムを活用したサービスの運用に必要な知識とスキル

場としての図書館

**対人関係能力**

知識情報保有機関としての図書館

**AIシステムに的確な情報をインプットするための知識とスキル**

## 一般的な情報化推進部門

### 網羅的なスキル構成

#### メソドロジ

ITビジネス活動の様々な局面で発揮される手法、方法などで、発揮される対象領域が広く、汎用性、応用性が高いスキルカテゴリ

#### テクノロジ

ITビジネス活動の様々な局面で発揮されるIT関連技法などで、対象領域が特定されるものが多いスキルカテゴリ

#### 関連業務知識

ITビジネス活動の様々な局面で活用される、メソドロジ、テクノロジ以外の関連業務知識。業務固有のスキル。

#### ITヒューマンスキル

markerpriority-1

企業・組織が IT に関するタスクを実行する際に必要となるスキル。人工知能が実用化される状況において、人間として特に重要なスキル。

markerpriority-1

**●実行力・実践力**

俯瞰力・深耕力・革新力・継続力

効果を上げるために、実行・実践環境や状況を適切に捉える力

効果的継続の実行と新しい取り組みや新領域へ挑戦する力

**●創造力**

問題発見力・問題分析力・仮説設定力・論理思考力・概念化力

状況を認知して問題を発見し、見極め、解決案を策定する「価値の創造・問題解決」を着実に遂行する力

複雑な状況や問題に対して、論理的思考により概念の形成、判断の構築、命題設定を行う力

**●コミュニケーション力**

自分の考えを伝える力・相手の考え方を理解する力・共感を呼ぶ力

情報の獲得や更なる情報の要求や内容の確認、他者への情報提供、他者に影響を与えたり、協働への引導する力

### 一般的なシステム構築のタスク

#### 戦略

事業戦略把握・策定支援

要求（構想）の確認

新ビジネスモデルへの提言

事業戦略の実現シナリオへの提言

IT製品・サービス戦略策定

市場動向の調査・分析・予測

IT製品・サービス戦略の策定

#### 企画

IT戦略策定・実行推進

基本方針の策定

IT化計画の策定

IT戦略実行マネジメント

システム企画立案

markerpriority-1

システム化構想の立案

システム化計画の策定

業務・システム要件定義

ITサービス要件定義

情報セキュリティ要件定義

#### 開発

システム要件定義・方式設計

運用設計

移行設計

基盤システム構築

アプリケーションシステム開発

ソフトウェア製品開発

Webサイト開発

システムテスト

移行・導入

ソフトウェア保守

ハードウェア・ソフトウェア製品導入

ファシリティ設計・構築

プロジェクトマネジメント

#### 利活用

markerpriority-1

サービスデスク

IT運用コントロール

システム運用管理

Webサイト運用管理

ファシリティ運用管理

サービスマネジメント

#### 評価・改善

システム評価・改善

IT戦略評価・改善

IT製品・サービス戦略評価・改善

事業戦略評価・改善支援

システム監査

資産管理・評価

#### 管理・統制

システム監査

資産管理・評価

事業継続マネジメント

情報セキュリティマネジメント

品質マネジメント

契約管理

コンプライアンス

人的資源管理

#### 推進・支援

マーケティング・セールス

再利用

再利用資産管理

再利用資産管理の準備

ソフトウェア開発プロセスの標準化手法

ソフトウェアエンジニアリングの標準化手法

論理思考力

概念化力

自分の考えを伝える力

再利用資産の保管と管理

ソフトウェア開発プロセスの標準化手法

ソフトウェアエンジニアリングの標準化手法

論理思考力

継続力

共感を呼ぶ力

再利用施策管理

再利用施策の計画

ソフトウェア開発プロセスの標準化手法

ソフトウェアエンジニアリングの標準化手法

問題分析力

仮説設定力

概念化力

共感を呼ぶ力

再利用施策の実施

ソフトウェア開発プロセスの標準化手法

ソフトウェアエンジニアリングの標準化手法

問題分析力

仮説設定力

革新力

調達・委託

標準の策定・維持・管理

新ビジネス・新技術の調査・分析と技術支援

#### 業務

ビジネス目標の決定

状況の評価

目標の決定とプロジェクト計画の策定

データの理解

データマイニングのためのデータの準備

モデリング

モデリング手法の選択

テスト設計

モデルの構築

モデルの評価

評価

結果とモデルの展開

ビジネスでの活用と評価

ビジネスでの活用

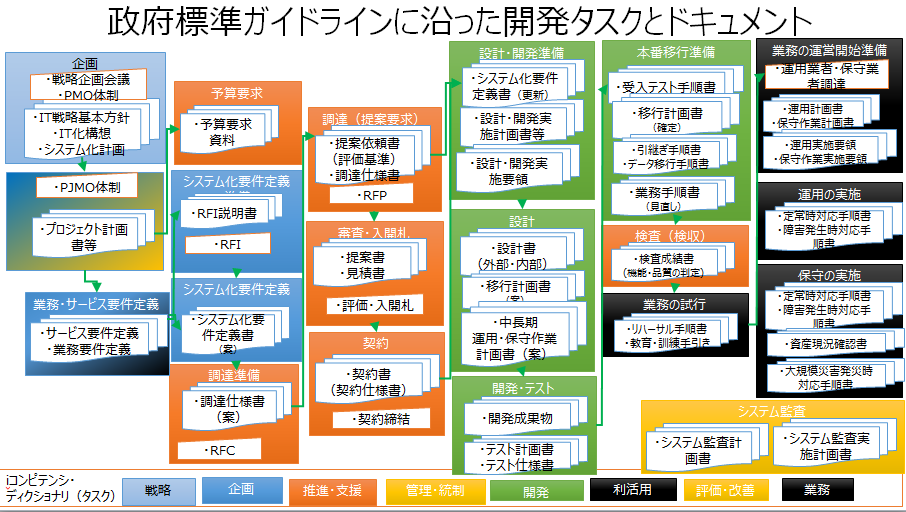
ビジネス成果の確認

### 政府標準ガイドラインに沿ったシステム構築と運用

「政府情報システムの整備及び管理に関する標準ガイドライン」（政府標準ガイドライン）は、業務の効率化及び高度化、情報セキュリティを含む情報システムの運用リスクへの適切な対応等、具体的な取組を政府横断的に進めるため、情報システムの標準的な整備及び管理について、その手続・手順に関する基本的な方針及び事項並びに各組織の役割等を定める体系的な共通のルールとして策定し、政府機関全体でこのルールに沿って運用されている。

これは政府機関における調達の共通ルールであるが、知の共有化を進める公的機関でも適正な調達を行う際の手引きとして参考にし、調達手続き、調達仕様の内容を必要に応じて取捨選択して、システム構築。運用を進めることにより、重要事項の考慮漏れをなくすことができ、発注者と受注者の認識の齟齬を減らすことができる。

### 政府標準ガイドラインに沿った開発タスクとドキュメント



*政府標準ガイドラインは概要編、実務者手引書等で構成される*

この図は、政府標準ガイドラインに沿って、組織としての事業計画に基づいた、業務・サービスの企画段階から、運用・保守、その後のシステム監査まで様々な業務（タスク）があるが、その各工程でのドキュメントを抜き出したもの。膨大な種類の仕様書類があるように見えるが、テンプレート的にデフォルトとなる記述内容も示されており、計画書・企画書をベースに、必要な手続き、仕様内容を加筆訂正していく形でドキュメント化していく。

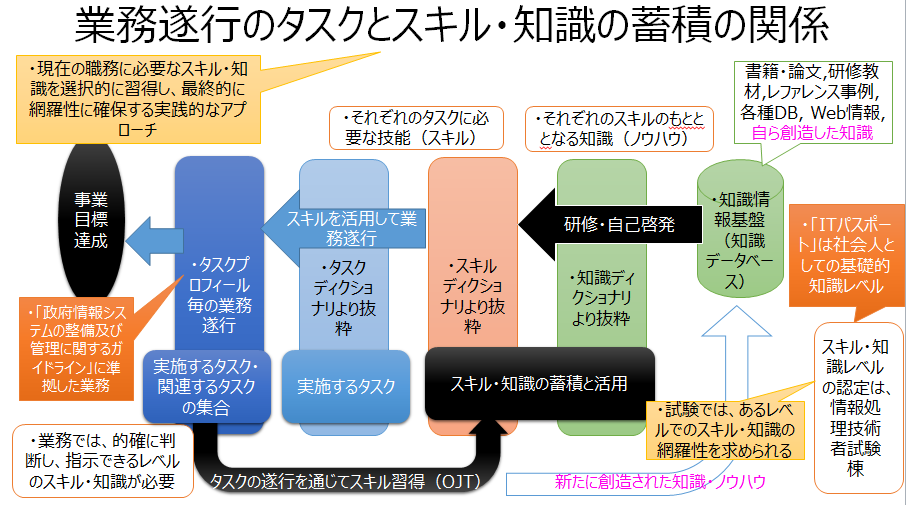
個々の組織が、デジタル知識基盤に対応したデジタルアーカイブシステムの構築を、企画・設計・構築・運用保守を、適正な内容と費用で行うためには、標準的な調達手順（プロセス）で行うことが重要で、「政府の調達標準ガイドライン」を参考にしたプロセスと成果物の規定により、適切な技術、パッケージを提供して漏れのない仕様で開発を進めやすくなる。

### iコンピテンシ・ディクショナリ（iCD）

i コンピテンシディクショナリ（iCD）は、情報処理振興機構（IPA）が作成した、システム開発におけるタスクとスキル・知識を体系的に洗い出して、それぞれ、タスクの種類を列挙したものが、「タスクディクショナリ」、スキルの種類を列挙したものが、「スキルディクショナリ」、知識の種類を列挙したものが、「知識ディクショナリ」として辞書化したものである。個々のタスクに必要なスキル、スキルに必要な知識が紐付けられている。また、事業を実施するために、様々な業務があり、その業務も複数のタスクにより遂行される。その、タスクの固まりが「タスクプロフィール」として例示的に示されている

各組織が人材育成について検討する際、事業の内容に合わせて、「タスクディクショナリ」からタスクを選択すると、そのタスクを遂行するために必要なスキル、知識が提示される。タスクを担当する人材には、そのスキル・知識を絞り込んで習得するようにすることにより、短期間に効率的に人材育成ができる。

### 業務遂行のタスクとスキル・知識の蓄積のスキーム



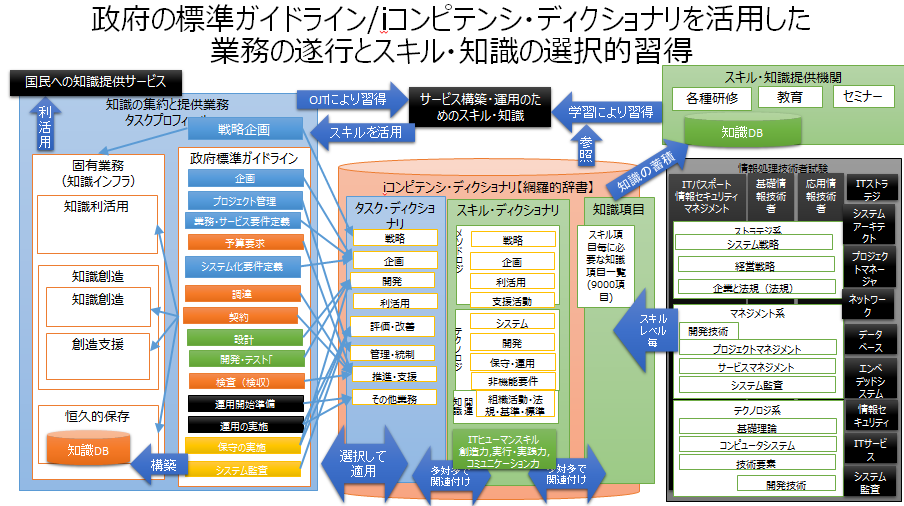
全てのスキルと知識を身に付いていれば、どんなタスクもこなせるが、そんな人はいない

１つのタスクを実施するためには、様々なスキルが必要であり、１つのスキルは、経験と様々な知識により身に付く。現在の職務に必要なスキル・知識を選択的に習得し、最終的に網羅性に確保する実践的なアプローチが、iコンピテンシ・ディクショナリの考え方である。

業務を遂行するためには、的確に判断し、指示できるレベルのスキル・知識が必要であり、業務に必要なスキル・知識を、事前に選択的に習得して、業務を遂行することが望ましいが、業務を遂行する前に持っていないスキル・知識は、タスクの遂行を通じて習得（OJT）することで、業務を円滑に進められるようになる。

*人材の知識のレベルを評価するものが、情報技術者試験制度である。ある一定レベルの知識を網羅的に習得していることは評価認定できるが、知識の網羅性を確保することは困難であり、また、その知識を活用した実践的なスキルは評価できないのが現状である。*

### iコンピテンシ・ディクショナリを活用した業務の遂行とスキル・知識の選択的習得方法



政府の標準ガイドラインで示された開発プロセスでの業務をタスクとして、iコンピテンシ・ディクショナリで具備すべきスキル・知識の選択の概念を示した図。

知的情報基盤の構成する各機関が、「政府の調達標準ガイドライン」に沿った開発プロセス（タスク）を遂行できるスキルと知識を持った情報システム関連の担当者の育成を効率的に行うために、個々のタスクに必要なスキルと知識の項目が示されている「iコンピテンシ・ディクショナリ」（2016年情報処理振興機構（IPA））を参考にして、選択的にスキルと知識を身に付けることにより、実践的な人材を効率的に育成する手法を取ることが効率的である

*●*

*知的情報基盤の構成する各機関でのデジタルアーカイブ事業を念頭にタスクプロフィールを仮定し、今後説明するデジタルアーカイブ構築プロジェクトを、政府標準ガイドラインに沿った業務jの工程（タスクプロフィール）で想定し、その中の個々のタスクに必要なスキルを選択*

*・そのスキルを身につけるためのベースとなる知識を選択*

*●*

*・サービス構築・運用のためのスキル・知識は、右上の「スキル・知識提供機関」でのセミナー等により習得*

*・「スキルを活用」して、左の政府標準ガイドラインに沿ったタスクを遂行*

*・タスクを遂行することにより、「OJTによりスキルを習得」*

*●*

*・右の黒い部分は、情報処理技術者試験の分類、左から右に向かってレベルが高くなる*

### データサイエンス領域のスキル標準

政府の調達ガイドライン、iコンピテンシ・ディクショナリ、データサイエンス領域のスキル標準「ITSS+を参照のこと。

iコンピテンシ・ディクショナリの補足として活用

従来のiコンピテンシ・ディクショナリでは、まだに辞書化が十分でない領域であったが、知識インフラの構築は、ビッグデータや人工知能の活用が必須であり、典型的なデータサイエンス領域として位置づけられる

### 今後の図書館サービスの実現のためのタスクと必要なスキル

#### 文化資源の収集・保存・修復・公開のスキル

図書館が扱う情報資産は、文献だけではない。

#### 専門分野に関する知見（文化・芸術・学術）

#### 文化資源を取り扱うための知識・技能

保存・修復技術

文化資源に価値を見出し、情報として記述するカタロガーとしての知識・技能

文化資源の価値を顕在化させて共有するための企画・発信するキュレーターとしての知識・技能

文化資源と人々をつなぎ、新たな価値を創出するコーディネータ、エンベデッドライブラリアンとしての知識・技能

文化資源を扱う活動の使命を明らかにし、その達成に向け経営資源を配分し、事業を統括するマネージャーとしての知識・技能

#### デジタル技術を活用したアーカイブ化のための知見

文化資源を取り扱う様々な局面でITを活用し、文化資源をデジタル化し情報メディアに乗せていく技術を有する

著作権をはじめとする知的財産権、肖像権、契約など各種法律分野に関する知識を有する

#### 文化資源を情報として収集・組織化・保存し、公開することを実現するシステムの開発・運用管理の知識・技能

効率的・効果的なシステム開発を行うシステムエンジニア

基礎となる学問分野の知識

**A：自然言語分野**

**B：音声・音楽分野**

**C：画像・映像分野**

**D：コンピュータ・ソフトウェア、情報通信**

**E：知識工学、人工知能**

**F：図書館学、図書館情報学**

文化資源の保有機関にも、先進技術の研究開発および実用化を目指す研究者

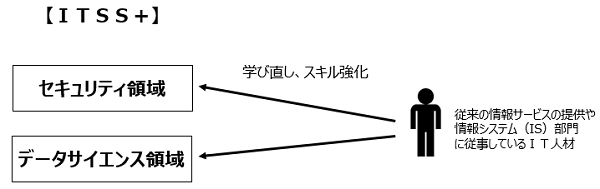
適用が期待される次世代技術【更新中】

markerflag-red

ブロックチェーン

# データサイエンス領域のスキル標準「ITSS+」の適用

## 位置付け



iコンピテンシ・ディクショナリの補足として活用

従来のiコンピテンシ・ディクショナリでは、まだに辞書化が十分でない領域

知識インフラの構築は、ビッグデータや人工知能の活用が必須であり、典型的なデータサイエンス領域として位置づけられる

markerstar-red

## スキルカテゴリ

#### ビジネス

課題背景を理解した上で、ビジネス課題を整理し、解決する。

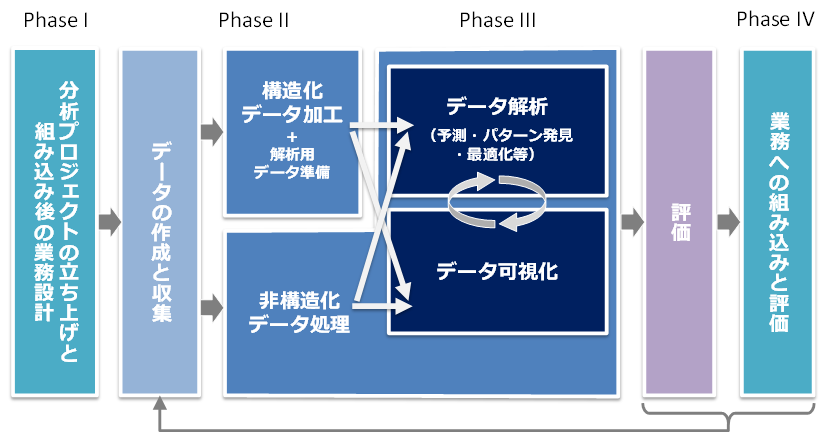
#### データサイエンス

情報処理、人工知能、統計学などの情報科学系の知恵を理解し、活用する。

#### データエンジニアリング

データサイエンスを意味のある形に使えるようにし、実装、運用する。

## タスク構造



## タスク

### 分析プロジェクトの立ち上げと組み込み後の業務設計

前提条件の明確化

目標の明確化

推進体制設計

計画の承認

環境整備

組み込み後の業務設計

### データの作成と収集

データ分析設計

データ収集

### 構造化データ加工

データ加工

データ利用環境の構築

### 解析用データ準備

入力変数データと目的変数データの作成

### データの準備

### データ解析

データ集計、抽出

性質・関係性の把握

検定

予測・判別

グルーピング

最適化

シミュレーション

### データ可視化

可視化の企画とデータ収集

2軸チャート化

多次元の可視化

関係性の可視化

地図上の可視化

挙動・軌跡の可視化

ダイナミックな可視化

リアルタイム可視化

### 非構造化データ処理

方針検討

言語処理

（データ収集）

（辞書構築）

（構造解析）

（特徴量変換）

（教師あり分析）

（教師なし分析）

（情報検索）

（文書生成）

画像処理

（類似画像推定）

（画像認識）

（画像復元）

音声/音楽処理

（本人認証や話者識別）

（感情分析）

（テキスト化）

（音楽分析）

### 評価

モデル評価

分析評価

### 業務への組み込みと評価

業務への組み込み

業務で活用するためのソリューション開発を行う（小規模な表計算ソフトでのツール開発や大規模ビッグデータシステムへの組み込みなど）

組み込み結果の評価

データ分析結果活用による業務改善

データ分析結果を活用して、業務上の問題点や課題を明らかにする

明らかになった業務上の問題点や課題に対する対応策を検討する

検討された対応策について、業務への反映と改善を図る

ビジネス上の効果の把握とさらなる改善

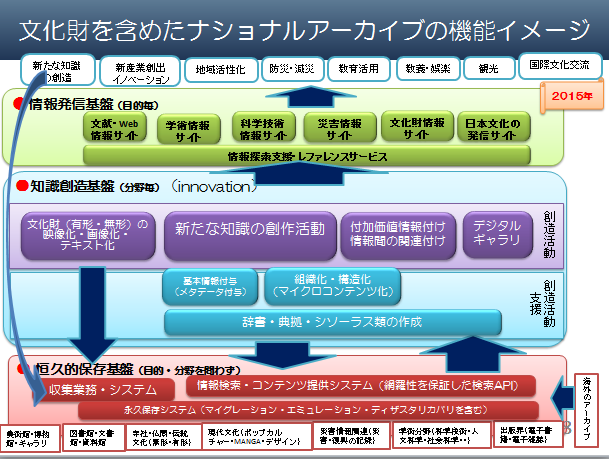
ステークホルダーからの意見や要望を収集し、さらなる改善に活かす

### [詳細は、データサイエンス領域のスキル標準「ITSS+」の適用【別シート】参照](https://www.ipa.go.jp/jinzai/itss/itssplus.html)

markerarrow-right

# 「見たことのない図書館を考える（2015年中山）」の実現性は？

2014年度同志社大学図書館司書課程講演会



## ナショナルアーカイブで何をできるようにするか

### 情報を探し出す作業の効率化・質の向上

網羅的な情報から、利用者の属性、スキル、利用場所に応じた的確な情報を絞り込んで提示

対話及びあいまいな条件による本文情報への的確なナビゲーション

### 情報を探し出せるようにするための作業の効率化・質の向上

主題分類単位の検索で網羅性を確保

専門家、図書館員等のノウハウの形式知化・DB化

可能な限り自動化

メタデータ付与、組織化、構造化、本文情報間の関連付け

### 新たな知識創造のコミュニティを構築

人と情報の関係、情報と情報の関係をリンクさせ、人と人を関連付け

## ナショナルアーカイブで何が変わるか

### 新しい発想により、様々なイノベーションが期待できる

有用な情報が網羅的に関連付けられて利用可能になることにより、今までは困難であった新しいサービスやビジネスが生み出される可能性がある

### 国民による創造的な活動の促進

情報を探すための工数を、創造的な活動に時間に振り向けることができる

利用可能な限られた情報に基づいた研究が、網羅性の高い情報が利用可能になることにより、より高度な研究へシフト

情報に紐づいた人同士のコミュニティにより創造活動が活性化する



## 図書館員等に求められるもの

### 利用者の情報探索支援の変化

情報の探し方は、IT技術の進展とともに変化する

図書館員の経験や勘に基づく判断のレベルは、ITの進展とともにアップする

人間ならではの仕事の価値

### より専門性の高い知識・ノウハウ

データベース検索技術者（サーチャー）的な業務は減少

機械的に可能なレファレンス依頼は減少

所蔵館を越え、本文内容で、より専門性の高い依頼へシフト

専門家との人的ネットワークが必要

### 高度な情報組織化のスキル

機械的（自動的）な組織化情報、リンク情報を修正するために、より専門性の高い知識を求められる

### 高度なIT技術の利用スキル

図書館システムを活用したデジタル化コンテンツの扱いのために、利用者以上のITリテラシーを持つことは必須。

更に、より高度な情報処理技術が必要となる

### デジタルコンテンツ、システム関連

システムエンジニア、デジタルアーキビスト、プリザベーションキュレーター、アーカイブとユーザを繋ぐコーディネータ等

# 次世代サービス構築及び提供に当たって留意してほしいこと

## 効率化、人はより創造的な業務へ

答えが１つでアルゴリズムが明確な業務は、システム化が容易⇒人工知能化が容易

情報システムは、図書館員の仕事を効率化・高度化させるもの。そのシステムの開発には図書館員の力が必要。図書館員もITの知識が必要⇒システムライブラリアン

## 保守的な組織にありがちな前例主義・横並び主義からの脱却

自分が利用者だったらどうなっていてほしいか

「民間はできるが公的機関だからできない」ということはない

組織規則、内規は、変えられる

## 与えられた権限には、実施の責任と義務を負っている

使命を果たすために与えられた権限には「実施の責任と義務がある」

できる範囲をコツコツとではなく、できる部分は責任を持って実施

できない部分は、他にどのようにしてもらいたいかも、責任を持って提示

## サービス提供者は利用者以上のITスキルを持つ

技術に進展は早い。若い利用者のIT利用技術は高い。

レガシーサービスの利用者のITリテラシーと高めることも、サービス提供者の役割

来館者のスキルが高くないのは、対応するサービス提供者のITスキルが高くないことも要因

## One of themの組織としての役割を考える。Give＆Takeの協力関係を想定する【単館主義からの脱却】

国全体がなすべき施策の中で、個々の組織の事業の進め方を考える

自分でできないことは、どのようにしてもらいたいかを提示（メリットも示す）

個別案件の利害調整でなく、文化の発展に向けて、ビジネスモデル全体での役割調整

創造力を持ってサービスの設計・構築・提供を考える

## 外部の動き（特に商用サービス）を知る組織外の情報を組織内へ

他の図書館、出版界、他の業種、業態の人との交流の場への参画、主催も積極的に

## 今後10年を見据える

世の中の動向を想定する。⇒自分の将来を見る

今の延長で可能な範囲での実施でなく、今後10年の人工知能技術の進展、その時点の社会のニーズを見据え、国全体で何をする必要があるかを考える

その時、情報提供機関の役割は？サービス提供者の役割は？

# おわりに

文化情報資源を中核とした「知識インフラの構築と利活用」のサービスの方向性を見極めて、提供する機関の1つとしての図書館情報サービスの構築と提供を考えていくことが重要と考える

2040年代には、人工知能が意志や感情を持って人間を超える「シンギュラリティ―時代」が来ると言われて、また10年後にはどんな新しい技術が確立しているかわからないが、現時点での人工知能の実用化レベルでも、社会は大きく変革する。  
その社会の中での図書館サービスは、従来の延長線上での業務を行っていては、有益なサービスを提供する機関としての存立が危ぶまれる

人工知能により、人の仕事が奪われるのではなく、より人間らしい仕事にシフトしていく。 仕事が奪われるとしたら、むしろ、人工知能を活用した省力化と新しいサービスに取り組まなかったために事業を継続できなくなる機関・組織であろう。

**～～～～～～～～～～～～～～～～～～～～**

1. **標題（和文および英文）**

電子図書館サービスからナショナルアーカイブの構築へ

－LOD化によるデジタル文化財の利活用を目指して－

From digital library services to the construction of National Archives

－Aiming at utilization of digital cultural heritage by making LOD－

1. **著者名（日本語およびローマ字）**

中山　正樹

NAKAYAMA Masaki

1. **著者の所属機関名および部署名（和文および英文）**

元国立国会図書館　専門調査員・電子情報部長

（同志社大学大学院総合政策科学研究科　嘱託講師）

Former Director General of Digital Information Department of the National Diet Library

1. **著者の所属機関の住所（和文および英文）、E-mail アドレス**

nakayama@slis.doshisha.ac.jp

1. **著者抄録（和文および英文）**

筆者は、国立国会図書館に在職中、電子図書館サービスシステムの設計・構築と運用に携わり、今後の方向性としてのナショナルアーカイブ構想の立案に関わってきた。

国立国会図書館は、電子図書館構想を策定し、「電子図書館サービス」を進展させてきたが、今後は、「電子図書館サービス」から、図書館の枠を越えてあらゆる分野の情報を関連付けて利活用する基盤として「知識インフラ」の構築へ発展させ、その一翼を担うこととした。その実現形が、「ナショナルアーカイブ」の構築である。

「電子図書館サービス」の設計・構築において、筆者が意識的に適用してきたセマンティックウェブに関連する技術を振り返る。また今後、ナショナルアーカイブの構築において、関係機関が連携して何を目指し、LOD化技術を含めたセマンティックウェブの技術を、どのように体系的に適用していくべきかを展望する。

The author、 during his tenure at the National Diet Library、 had been involved in the design、 construction and management of digital library service system、 and in making plans of National Archives initiative.

The National Diet Library、 formulating “Electronic Library Concept”in 1998、 has developed digital library services、 and is now seeking to build "knowledge infrastructure" as the basis for associating the information in all fields beyond the framework of library and utilizing them.

This paper reflects on the technology related to the Semantic Web、 which the author had consciously applied in the design and construction of digital library service system、 and also puts in perspective what goals the NDL should aim at、in cooperation with related institutions for the construction of the National Archives、 and how to systematically apply the technology related to the Semantic Web including LOD.

1. **著者付与キーワード（和文および英文）**

ナショナルアーカイブ、 デジタルアーカイブ、 知識インフラ、 セマンティックウェブ、 Linked Open Data、 クラウドコンピューティング、 WebAPI電子図書館、 国立国会図書館、 国立国会図書館サーチ、

National archive、 Digital archive、 knowledgeinfrastructure、 Semantic web、 Linked Open Data、 Cloud computing、 Web API、 Digital library、 Natonal Diet Library(NDL)、 NDL Search

# はじめに

筆者は、民間企業、政府関係の独立行政法人において、情報システム化により、業務サービス及びソフトウェア開発の効率化に取り組んできた。2002年に国立国会図書館（NDL）に入館して以降は、電子図書館事業の企画立案、図書館サービスシステムの構築・運用に携わり、2011年より電子情報部長および情報化統括責任者（CIO）として、電子図書館事業、情報システム全般を統括し、次世代図書館サービスの構築に向けて検討を進めてきた。筆者の信条は、一貫して、労働集約的な業務を情報システムにより効率化して、人がより創造的な業務へ移行できるようにすることである。

図書館に限らず、様々な業種・業態の組織において、大量の情報の中から効率的に適切な情報を取り出す情報提供システムが構築・運用されている。情報システムの観点では、図書館システムは、そのような情報システムの中での１アプリケーション分野であり、情報提供サービスとしても、利用者にとってOneOfThemのサービスであると考えている。

筆者は、そのような基本的な考え方で、NDL以外の情報資源、技術ノウハウを活用して、利用者の情報取得、知識創造を効率化するサービスを、如何にして効率的に構築・運用できるようにするかを念頭に進めてきた。次世代図書館サービスシステムは、従来型の図書館の延長線で考えるのではなく、世の中の多くの情報提供サービスの中で信頼性の高い情報を提供する機関としての責任と義務を果たせるサービスの提供を目指すべきとしてきた構想を論述する。

昨今、クラウドコンピューティングの普及に伴って、ウェブ上の情報を意味的に理解できる形で公開・共有を目指すセマンティックウェブ技術の中で、構造化された情報同士をリンクさせるための技術や方法論の総称としてのLinked Open Data(LOD)が注目されてきた。この概念は、NDLが1990年代に構想した電子図書館構想の「地球規模の知的財産を誰でも容易に利用できるようにする」という理念の実現のために、電子図書館サービスの構築の仕組みの考え方や、NDLが個別に適用してきた要素技術の集合としてみることができる。

筆者が、電子図書館サービスの構築において、セマンティックウェブ、LODの考え方を、どのように適用してきたか、今までの歩みを顧みるとともに、今後、電子図書館サービスの発展形としての国の文化財全体の利活用を目指したNDLを中核とした情報保有機関によるデジタル文化財のアーカイブ構築において、LODの考え方がどのように適用されていくべきかの方向性について、技術的な観点にフォーカスして考察する。

# 電子図書館20年の歩み

1980年代から、海外の多くの先進的な図書館同様、日本でも電子図書館事業に取り組み、1994年に我が国で最初の大規模な電子図書館の実証実験プロジェクトがNDLにより実施された。その後、NDLは、2002年から本格的なサービスとして離陸し発展させて、現在に至っている。

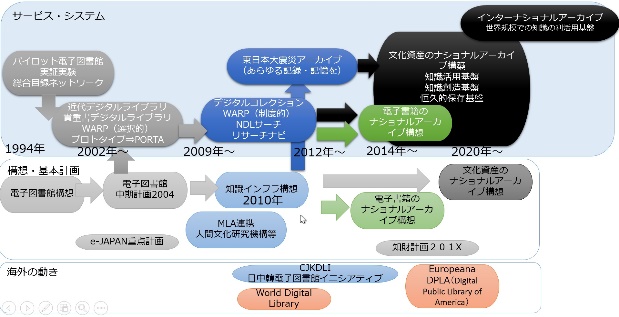


図1電子図書館に関連する構想とサービスの変遷

## 電子図書館構想と実証実験

1994年1月、通商産業省（現：経済産業省）の高度情報化プロジェクト事業の一環で、情報処理振興事業協会（現情報処理推進機構（IPA）は、慶應義塾大学の湘南藤沢キャンパス(SFC)内に、情報基盤センター（CII）を設置した。1）当時としては高速のネットワーク（6Mbps）、大容量サーバ（4TB）を配備して、各省庁の白書のSGML化や、各種統計情報を串刺しにして連関を分析する「新産業創造データベース事業」、学校教育の高度化を図るための教育ツールとしてのオーサリングシステム、エグゼキュータシステムを提供する「100校プロジェクト」とともに、NDLと連携協力して、パイロット電子図書館プロジェクトを開始した。筆者は、本事業には、IPA側の担当として参画した。

パイロット電子図書館プロジェクトの目的は、21世紀の高度情報社会において、地球規模の知的財産を誰でも容易に利用できるように、地球上に広く分散して個々に収集・蓄積されている知的資源を、空間的・時間的制約を越えてアクセス可能とする環境を提供するための実証実験であり、NDLは、全国公共図書館総合目録ネットワークプロジェクトとパイロット電子図書館実証実験プロジェクトを実施した。

総合目録ネットワーク実験は、出版物の総合目録と図書館間の相互貸借のためのシステムとして、NDLを含む58館の政令指定都市および都道府県立図書館が参加して実施された。共通のメタデータ記述規則としてMARCベースの総合目録共通フォーマット、通信プロトコルとしてファイル転送規約を設定し、各図書館からの大量の初期データ、週次の差分データを受け取ることにより、1、400万件の書誌データがRDBMSの１つであるOracleデータベースに蓄積された。ISBN等の識別子がないメタデータも多く、タイトル、著者等で同一の書籍のメタデータであることを認識する機械的な同定作業を行った。所定の時間内に処理するためには、レスポンスを低下させないデータベースのテーブル設計、SQL言語によるプログラミングでは様々な工夫が実証実験された。

パイロット電子図書館実証実験では、NDL所蔵の貴重書、明治期刊行図書・雑誌、出版社から提供を受けた資料等を、マイクロフィルムからのスキャニング、デジタルカメラでの撮影により、当時としては高精細画像（4、000×5、000ピクセル）で、tiffおよびJpegフォーマットで約1、000万ページ（CD-ROM約2000枚分）のコンテンツをデジタル化し、画像データベースに登録。個々の画像データにメタデータを付与してメタデータデータベースを構築して、効率的な検索・利用方法、ユーザインタフェース（UI）、電子化データの高度利用、効果的な電子図書館の構築支援の手法等について実証実験を行った。その成果であるシステム及びデジタル化コンテンツは、その以降の電子図書館サービスのベースとして引き継がれている。

## 国立国会図書館中期計画2004

NDLは、パイロット電子図書館プロジェクトの成果を踏まえて、2002年10月に開館された関西館を拠点として、近代デジタルライブラリー、インターネット資源の選択的収集事業（WARP）、各種の電子展示会を公開・提供した。

2003年には、e-Japan重点計画2003、e-JAPAN戦略Ⅱ加速化パッケージ、（内閣官房IT戦略本部）において、「国のデジタルアーカイブ構想」、「ジャパン・ウェブ・アーカイブ構想」の実現を、また2004年には、e-Japan重点計画2004において「国立デジタル・アーカイブ・ポータル構想」を一層推進することが明記された。

　このような動きを踏まえて、NDLは、2004年2月に「電子図書館中期計画2004」（中期計画）を策定した。この中期計画において、デジタルコンテンツを広汎な利用者に提供するために、NDLが国のデジタルアーカイブの重要な拠点となるということ、また国内外の多様な利用者層の需要に応じ、日本のデジタル情報全体へのナビゲーションを行う総合サイトを構築し、利用者がワンストップで利用できるようにすることを目指すこととした。2）

### アクションプランとしての3つの柱

１つ目は、 「デジタル･アーカイブの構築」 である。これには、書籍のデジタル化の推進と、インターネット情報の収集と恒久的保存の２つがある。２つ目は、「情報資源に関する情報の充実」で、レファレンス情報などナレッジの蓄積を蓄積し提供提供すること、3つ目は、「デジタル･アーカイブのポータル機能」である。

### 電子図書館中期計画2004の実現イメージ

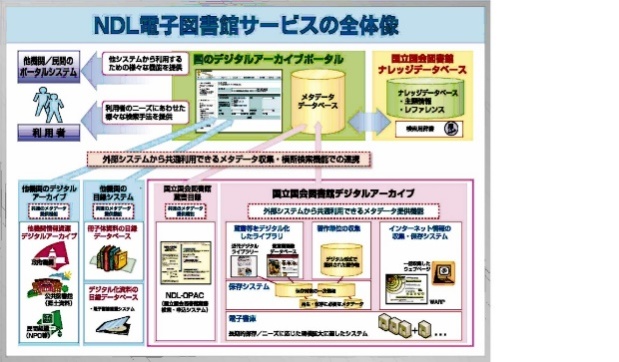


図2 NDL電子図書館サービスの全体像

図2は、2004年にNDLが作成した電子図書館サービスの実現のイメージである。

右下の「デジタルアーカイブ」は、デジタル化したリソースやインターネット情報などを蓄積し、長期保存する。

その左は、既存システムのNDL-OPACなどの「蔵書目録」は、これらが書誌情報やデータベース内のメタデータを提供する機能を備えることにより、デジタルアーカイブポータルから統合的に利用できるようにする。

上の「ポータル」は、これらNDLの資源だけでなく、左下にある、他機関のデジタルアーカイブや目録などの有用情報も合わせて案内できるようにするもの。

左上の「ポータルのインタフェース」は、GUIにより利用者へのニーズに合わせた検索手段の提供と、他システムがポータルの機能を利用するためのAPI機能を持つ。

また、右の、「主題情報、目次情報、解題情報、レファレンス情報など検索を支援するための情報、知識としての「ナレッジデータベース」である。

2004年に描いたイメージの実現形として、ポータルを発展させたものが、PORTAであり、また、現在の「国立国会図書館サーチ（NDLサーチ）」である。

## デジタルアーカイブの構築

資料のデジタル化に関しては、NDLは、2009年5月から大規模なデジタル化を開始し、2011年までに、冊子体としては230万冊、約2億枚の画像をデジタル化した。また、インターネット資料の収集に関しては、NDLは、2010年4月に国等のウェブサイトの制度的な収集を開始した。さらに、いわゆる電子書籍・電子雑誌に類するオンライン資料は、2013年1月から、無償でDRMによるアクセス制限が掛かっていない資料についてのみの収集を開始した。デジタル化および収集した資料の提供に関しては、著作権保護期間の満了もしくは著作権者から許諾を得た資料は、インターネット上に公開している。デジタル化した資料のうち、市場で入手困難な資料に関しては、公共図書館等で閲覧が可能になっている。ここでは詳細は割愛するので、他の文献を参照していただきたい。本稿では、保存システムの機能として、メタデータに関連した施策として、「長期保存システム」のみ紹介する。

### デジタル情報の長期保存（OAIS、 情報パッケージ(SIP、 AIP、 DIP)）

電子情報を長期的に保存するためのシステムの構築に当たっては、NDLは、この長期保存システムとして必要な機能やプロセスを抽象的に定めたモデルを参照しつつ構築している。

その一つとして、OAIS（Open Archival Information System）レファレンスモデル(ISO 14721:2003)モデルがある。

このレファレンスモデルに沿ったパッケージとして、米Hewlett-Packard 社のDspace、IBM社のDIASがあった。受入機能として適切なメタデータの付与、アーカイブ（保存）機能として定期的なエラーチェックや保存媒体の更新、データ管理機能としてメタデータの管理、保存管理・運営機能として長期保存計画の立案、提供者との交渉、マイグレーションの実施、さらに保存計画機能として旧式化の検知、対応策の提示、アクセス機能として再生環境の保存、エミュレーションが想定されている。

このうち、保存計画を実現する要素技術であるフォーマットレジストリやフォーマット識別ツールの開発が海外で行われている。

情報パッケージとしては、MODSをベースにしたメタデータをMetsで括ったSIP（受入れメタデータ）、AIP（保存メタデータ）、DIP（提供メタデータ）を想定した。

## デジタルアーカイブポータルのプロトタイプ

NDLは、中期計画での3つ目の柱であるポータル機能について、2004年10月から開発を開始した。様々なデジタルアーカイブ内の情報を統合検索する仕組みの実用性を検証するために、デジタルアーカイブポータルのプロトタイプシステムとして、2005年7月に試験公開し、機能の検証を開始した。これが、現在の分野を越えたナショナルアーカイブ構想の原点であったと考える。

### 開発当初のサービス要件とシステム化

#### サービス要件定義において

* 利用イメージとして、利用者の情報探索行動の全般をサポートすることを目指すこととした。  
  情報探索行動として、問題の定義、情報ニーズの識別、情報探索戦略の策定、情報の獲得（情報源の咀嚼、情報の抽出）、情報の統合（情報の組織化、提示）、評価（成果の評価、プロセスの評価）等のステップを想定した。このようにして目的の情報の存在を知るだけでなく、情報そのものの閲覧および入手先までナビゲートを目指すこととした。
* それを実現するために、各デジタルアーカイブの構築の在り方にも言及した。交換用の標準的なメタデータ記述要素、記述規則の適用、交換用APIの実装を求めることとした。また、巨大な知識ベースとして「意味ある情報資源」として利用できるように、本文内容を組織化、検索できるように研究開発、技術開発が進むことも求めた。

#### システム化において

* 先進的かつ将来標準的な仕様となることが見込まれる技術を適用すること。
* オープンソースソフトウェア（OSS）・パッケージソフトを活用することとして、特に、適用事例が多いOSSを活用し、可能な限り新規開発はせず適用したソフトウェアのカスタマイズは必要最低限とすることとした。
* Webサービスとして他サービスから利用しやすいものとすることとして、各々の機能は独立したWebサービス機能として提供し、図書館界のみならず、デジタルアーカイブの世界で標準となり得る仕様の採用を目指した。

## PORTAの構築

NDLは、プロトタイプでの検証を踏まえて、実用システムとして「PORTA」を構築し、2007年10月に正式公開した。3)

### システムアーキテクチャ

NDLは、システム基盤として、Web2.0時代のサービス指向アーキテクチャ（SOA）を適用し、仮想OS上でOSSを組み合わせ、Dublin Core（DC）に準拠した国際標準のメタデータ記述要素、記述規則、メタデータ交換の共通API（OAI-PMH、SRW（SRU/SOAP）、RSS、Z39.50等）を利用して、商用を含めた外部サービスとマッシュアップによるサービス連携と、コンテンツマネジメントシステム（CMS）による機能追加の容易性の確保を目指した。先進的な検索機能として連想検索検索エンジン（GETA）等を利用したあいまい検索も実装した。

### 統合検索対象の拡大

統合検索対象は、画像イメージ形式で提供している近代デジタルライブラリーとテキスト形式で提供している青空文庫の統合検索から始め、同時代の資料が分散保存されている国立公文書館アジア歴史資料センターとNDL-OPAC、地域情報をまとめているアグリゲータサイトとしてデジタル岡山大百科等を、ワンストップで検索しコンテンツに辿り着けるようにした。

### 統合検索のために苦労した点

* データ提供機関に対して、統合検索への賛同を得ること。大きな意義は、各機関のコンテンツの利活用が促進されることであり、また、利用者の利便性が向上することであったが、サービスが横取りされると誤解されることが多々あった、また、サイトへのアクセスが増えることを不安視され、なかなか賛同が得られなかった。
* また、連携には、提供機関のサイトで標準プロトコルの実装が必要であったが、ほとんどのサイトで、外部提供インタフェースを持っておらず、各サイト側に実装が必要であったが、開発ベンダーが法外な開発費を提示し、その費用が捻出できない状況であった。
* さらに、メタデータのマッピングの調整においては、それぞれの機関において、 Dublin Core（DC）をベースにした記述要素を適用できても、記述規則が異なり、統合しても内容として認識できない状況であった。特に、横断検索においては、メタデータの表記のゆれをカバーした検索はできず、調整には、膨大な時間を費やした。汎用的なメタデータのマッピング、記述規則等のルールの普及が必要であった。

### インキュベータとしてのPORTAの役割

筆者は、PORTAは唯一の統合検索サービスではなく、PORTAがインキュベータとなって、各利用者ニーズにあった様々な統合検索サービスが出現することを期待した。

* まず、継続して、デジタルコンテンツの可視化を目指すこととし、インキュベータの役割として、各データプロバイダに検索やサービス連携のためのAPIが実装されることの普及啓発活動を継続するとともに、外部機関においてPORTAの外部APIを利用した様々なサービスが出現することを期待した。
* また、利用者の情報探索行動において、統括可能情報の網羅性を高めるために、様々な機関から提供されるデータベースを分散デジタルアーカイブと位置付け、コンテンツの体系的な構築と、網羅性の高い統合検索サービスの実現を目指すこととした。
* さらに、単なるキーワードでの検索サービスではなく、意味的に関連付けるセマンティックウェブ・サービスでの連携を目指すこととした。

### Web3.0の時代を見据えて適用すべき技術と進め方

筆者は、今後の機能強化のため、セマンティックウェブ技術を中核として、適用すべき技術として3つに分類した。

#### 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の適用を想定した。

#### 管理を容易にする技術

* メタデータの付与を自動化・省力化する機能

書誌的事項は、情報の作成元が作成することを基本としつつ、管理、検索に十分なメタデータが付与されていない場合は、多少精度が低くてもメタデータは可能な限り自動付与する機能、自動付与されたメタデータの確認作業を省力化する技術の適用を想定した。

* セマンティックウェブ技術の活用

当時、集合知を利用するWeb2.0から、意味的情報を相互利用するWeb3.0が叫ばれた時期であり、Web2.0の要素として、Blog、Forksonomy、Ajax、Mashup等の技術、更に

Web3.0の要素として、RDF、SKOS、オントロジー、GRDDL、SPARQLの技術の適用を想定した。

* 蔵書管理指向からサービス指向へ

将来的には、目録を利用者側の視点から見直すモデル（FRBR）に基づくメタデータの再構築も想定した。

* 検索を容易にする技術

従来型のキーワードマッチングによる検索だけでなく、キーワードサジェストやクラスタリング検索等により、検索候補を推定する技術を活用した検索機能も想定した。

#### PORTA構築の進め方

筆者は、進め方として、図書館の枠を超えた国際標準、業界標準を積極的に適用するとともに、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を進め、関係機関との共同構築、オープンソースの利用、単独でのトータルシステムを目指すのではなく、各機関の研究成果を組み合わせてマッシュアップによるサービス構築を想定した。

まずは、海外ではIIPC、国内では様々な技術の研究開発性を保有しているJST、NII、NICT、AIST、IPA等の政府機関、大学の研究室との連携、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）との連携によるウェブ協調型のサービスの提供を目指した。

## トータルな次期図書館システムを目指して

2009年当時の冊子体資料を扱う従来型の図書館システムの最大の問題は、運用・保守作業、機能拡張コストが肥大化して、機能拡張により新たな利用者ニーズに応えるための機能拡張ができなくなったことである。そのために、NDLは、下記のような方針により、全面リニューアルすることとした。

* 紙・デジタルを問わずトータルな次期図書館システムの構築と提供の実現を目指す。
* NDLサーチを、図書館ポータル、次世代OPAC、PORTA、ゆにかねっとを統合した入口機能とする。
* 新規開発を極力なくす。
* 実現すべき機能とレベルを明確にしたシステム化要件定義書により、見積もり内容を的確に判断する。
* 特定ベンダーしか提案できないRFPにならないようにして、競争原理が働く調達を行う。
* 従来型業務・サービスは、パッケージで実現できない業務は、業務の内容や流れを見直し、可能な限りパッケージの仕様に合わせる。

　　この方針に基づきシステム開発を進めた。

冊子体刊行物の収集・組織化業務及び、蔵書検索・申込システム「NDL-OPAC」を、パッケージに切り替えることによる開発・運用コストを大幅に削減し、デジタル化資料の来館者サービスシステム、国立国会図書館サーチ等、デジタル時代の次期図書館利用者サービスの基本機能を充実させた。

## 国立国会図書館サーチ（NDLサーチ）

NDLは、PORTAの後継の情報探索サービスシステムとして、2012年から正式に提供を開始した。のちに提起される「知識インフラ」のフレームワーク作りのベースとしたサービスであり、筆者が振り返るとLinked Data、Linked Open Dataと同一の方向性を指向したものとなっていた。

NDLは、システムの構築に当たっては、様々な視点を持つ専門家の参画のもとで、サービス要件を確定させ、サービス要件を満たす機能を実装するためのシステム化要件を定義した。段階的な実装を想定し、OSSを含めて、その時点での最適な適用技術、パッケージを示す技術標準適用ガイドラインを踏まえて、開発を進めた。

### 機能の概要

NDLは、FRBRのモデルに沿って同一著作、原作から様々な媒体や形式に派生して発行されている著作物を書誌事項で体系的に把握できるように著作単位でグルーピング表示する機能、検索キーワードが明確でなくても想定されるキーワードサジェスト機能、障害者向け機能、検索ワード、外部サービスを活用して内容表示を日中韓英に翻訳して表示する機能、個人ごとの検索要件、検索範囲をあらかじめ特定できるパーソナライズ機能、スマートフォン対応機能等を実装した。

NDLは、交換用メタデータとして、冊子体出版物以外の情報への適用を想定して従来の仕様をよりシンプルにしたDC-NDL(Simple)と、セマンティックウェブに適合した記述モデルであるRDFをデータの記述に使用し、FRBRのWork、Manifestation、Itemを体系的に定義したDC-NDL(RDF)を交換用メタデータ記述要素・規則として定義した。

NDLは、メタデータハーベスト、横断検索のためのWebAPIを設定し、そのAPIの実装を働きかけることにより、連携先の一層の拡大を図るとともに、外部システムでメタデータを利活用のためのWebAPIとして、検索結果や新着書誌情報をFeedly等のニュースリーダで確認できるように、RSS配信する機能、DC-NDL(RDF版)、JSON形式でダウンロードする機能等の実装を強化し、「情報ハブ」の役割を提供した。4)

また、著者名典拠、件名典拠情報に関しては、国立国会図書館典拠データ検索・提供サービス（WebNDLA）として、キーワード検索、NDC分類記号検索により検索できるほか、NDLサーチと連携し、関連キーワードの提示、典拠コントロールされた各標目による書誌データの再検索など多様な検索機能を提供した。また、全典拠データの個別ダウンロード、国立国会図書館件名標目表（NDLSH）の収録範囲となる典拠データを、RDF/XML形式、 RDF/Turtle形式、 JSON形式で、ダウンロードすることを可能とした。

### サービスの状況

NDLサーチでは、人間文化研究機構の6構成機関の研究資源共有化システム（nihuINT）、国立美術館の4つの美術館が所蔵している作品の総合目録、商用の電子書籍ポータル（hon.jp）との連携等のPORTAが担ってきたデジタル情報のポータルとしての役割を継続するとともに、ゆにかねっとが担ってきた各地の公共図書館蔵書の総合目録としての役割を引き継いだうえ、NDL-OPAC、インターネット資料収集保存事業（WARP）、国立国会図書館デジタルコレクション、国会会議録検索システム、リサーチ・ナビ（調べ方案内）、レファレンス協同データベース等のNDLの主要データベースを統合検索し、図書館、美術館、博物館、文書館、政府機関、新刊書籍の出版予定情報としてJPOの近刊図書情報、Japan Knowledge等の商用データベースなど、200DB、約1億件の書誌・メタデータを横断的に検索できるようにした。検索対象が必ずしもデジタル化されてインターネットで閲覧可能なコンテンツでなく、所在情報により入手手段へのアプローチに留まるものも多いが、今後各機関によるデジタル化が進めば、インターネットで閲覧できるコンテンツへのナビゲーションも充実すると思われる。

また、統合検索のほかに、検索キーワードを使った外部サービス検索として、WorldCat、WebCat Plus、Wikipediaでの再検索、更に入手手段として、各種オンライン書籍・電子書籍販売サイト、古本屋ポータルサイトへのナビゲートも行った。

最近の進展としては、公共図書館からのデータ提供方式のOAI-PMH / DC-NDL(RDF) への切替、ジャパンリンクセンター（JaLC）との連携によりDOIでの固定URLでの検索、WebNDLAからVIAFへのナビゲートも可能になっている。

### 連携方針・連携モデル

統合検索先の拡大においては、連携可能なところから進め、網羅性の確保のため、分野ごとの中核的なデータベースに標準的な連携機能の実装を働きかけてきたが、標準仕様が浸透してきた状況において、NDLは、改めて連携方針・連携モデルを提示し明確にした。

#### 連携実施計画策定の課題・背景

* 日本におけるメタデータ提供のプラットフォームとしての認知度が高まりつつある今、連携可能な範囲での拡張ではなく、目標と計画を設定して内外に示す必要がある。
* ナショナルアーカイブ（構想）におけるコンテンツ検索・提供機能を担うことを見据える必要がある。
* 図書館等の情報機関が保持するメタデータの集約及びAPIを通じた一般への提供により、政府が推進する「公共データの民間開放（オープンデータ）」の一翼を担う必要がある。

#### 連携方針・連携モデル

次の5項目を連携方針の柱とする。

* 日本の刊行物及び刊行物と同等の内容を有するコンテンツの網羅を志向。
* 公的機関、学術研究機関、図書館・文書館・博物館・美術館等の文化機関が作成し、インターネット上で提供している一次情報（コンテンツ）、二次情報（メタデータ）及び参考情報等を対象とする。
* 一般利用者にとって有用性が高いコンテンツを持つシステムを優先。その際、一次情報の入手までの障壁が低いシステムを重視。
* APIを実装したシステムを優先。メタデータ授受に用いるフォーマットとして、DC-NDL(RDF)を推奨。
* 効率的に検索対象を拡大するために、個別のシステム（=data provider）との連携よりも、それらを集約した統合検索サービス（=aggregator）との連携を優先。

つまり、APIは普及期にありインキュベータとしての役割を終了し、既にAPIが実装されているシステムとの連携を拡大し、人間文化研究機構、JST、NII等のように、各所管において、統合データベースを提供しているサービスとの連携により、効率的に網羅性を確保するということである。

## 公共的書誌情報基盤

情報内容として整理・要約されている出版物に関する情報の共有化と利活用は最も重要視される。

公共的書誌情報基盤は、出版界が作成する出版情報（販売促進情報）を活用した図書館での書誌作成の省力化を目指すもので、出版界と図書館界が持つ著者、著作の書誌・書評等の情報の共通識別子によるLinked Data化により、出版情報の網羅的なデータベースの構築と共有が期待できる。

### 公共的書誌情報基盤の整備（2010年）

NDLが、出版関係機関と協力し、我が国を代表する標準的な書誌情報を作成・提供する公共的基盤を整備するものである。出版文化の基礎となる質の高い出版・書誌情報が、無償もしくは廉価にて、迅速またタイムリーに読者、利用者に届けられることを目的とするものであった。5)

また、この事業によりNDLの納本事務がより網羅的かつ円滑に行われることが期待された。

当時の想定は、次のようなものであった。

* 近刊情報は、出版物のISBNとNDL書誌IDが関連付けられるように、出版社からISBNが付与された書誌情報が、JPOを経由してNDLに送られ、JPNO、 書誌IDを付与して、近刊情報として公開する。その近刊情報をJPOを経由して取次に送ることにより、共通の識別子で近刊情報が管理される。
* 新刊情報は、取次において、JPO近刊情報と現物を突合し、出版情報追記して、新刊情報とする。その情報をNDLにおいて、近刊情報を新刊情報として置換え、最低限の書誌事項を追記して新刊情報として公開すること。

しかしながら、NDLでの実装においては、納本された新着情報の提供は、出版情報をインプロセス情報として提供することにより早まったものの、出版界と図書館界での情報の共有には繋がっていない。

### 全国書誌情報の利活用に向けた超党派勉強会での検討（活字文化議員連盟　2015年9月）

図書館などで広く活用されているMARCのもととなる書誌データを国民が無料で利用できるように政策や予算なども含め総合的な施策を検討するもの。

背景としては、NDLが作成する書誌データは、民間の書誌データを活用することで迅速な提供に向けて取り組んでいるが、民間の提供スピードには及ばず、公共図書館などでの利活用が進まない現状があり、2016年春をめどに課題のたたき台をとりまとめるとされている。

筆者が期待する結論としては、近刊情報、新刊情報が図書館で活用され、また、図書館での書誌情報、典拠情報が出版界での出版物に関するデータベースで利活用されることにより、出版情報を活用した書誌作成が効率化されるとともに、出版物の検索の網羅性が確保されることにより利用者が出版物を見つけやすくサービスの提供が容易になることである。

## 知識インフラの構築を目指して

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

「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

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。6)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。7)

### 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、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。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

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が期待される。8)

### 知識情報基盤の構築モデル

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。

「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。9)

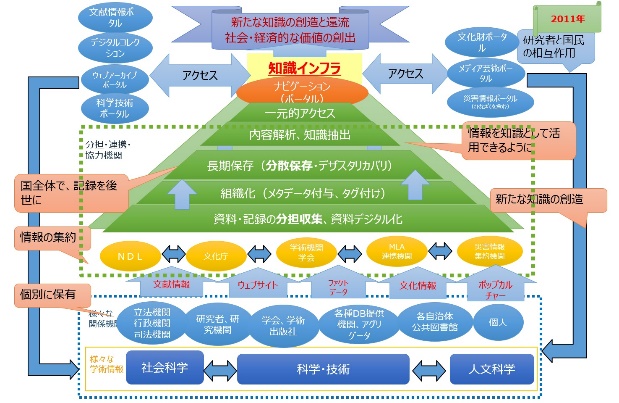


図3 知識情報基盤の構築モデル

## 東日本大震災アーカイブ

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。10)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。11)

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

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。

#### 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。

#### 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。

### 東日本大震災アーカイブのシステムイメージ

NDLは、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。

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。

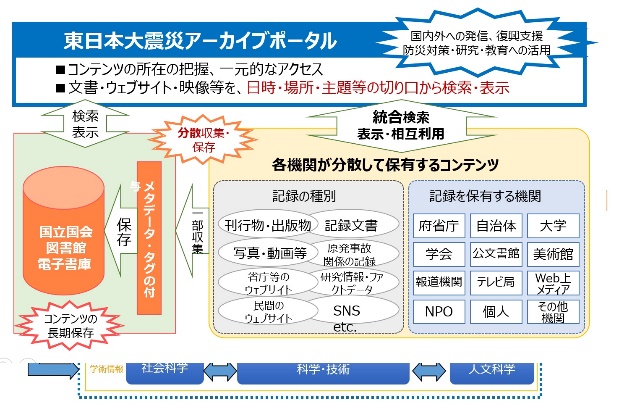


図4 東日本大震災アーカイブのシステムイメージ

* 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

構築・運用においては様々な課題が顕在化した。

#### 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。

#### 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）。

#### 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。

#### 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。

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。

# 今後10年で目指すところ（2015年～2024年）

NDLは、このような歩みで電子図書館事業を進展させてきたが、筆者は、今後10年を踏まえて、社会の要請、情報技術の実用化動向を想定した図書館サービスの姿と、その実現に向けて実施すべき事項を明確にし、速やかに対応していく必要があると考える。

東日本大震災アーカイブを踏まえて、分野を越えた知識インフラの実現形として、あらゆる記録を情報として集約し、相互に関連付けて知識化し、将来にわたって利用を保証するとともに、「社会・経済的な価値を創出」できる「新たな知識の創造と還流」の仕組みを現実的に構築していく時期と考える。

## 「知的財産政策ビジョン」での今後10年の目標

我が国では、「知的財産政策ビジョン」（2013年6月7日知的財産戦略本部）等において、文化資産のデジタルアーカイブ化を推進する政策も含めて、今後10年を見据えた知的財産関連の政府の取組としての目標を掲げた。12)政策としては、従来の事業モデルの延長線上での「改善」だけでなく、事業モデルそのものを創造・転換する「イノベーション」を創出すること、あらゆる分野の知的財産を対象としたビッグデータも含めたクラウド化、オープン化、更にユーザが共同でコンテンツを制作するクラウドソーシングの普及が見込まれている。そのための知財人財の育成・確保、コンテンツの権利処理の円滑化、電子書籍の普及促進、データの収集・蓄積・分析による多様な付加価値の創造の研究開発、日本の伝統や文化に根ざした魅力あるコンテンツ・製品などの発掘・創造等の取組も示された。

## 電子書籍ナショナルアーカイブ

有形の出版物を中心とした物としてのアーカイブ機能の延長に、無形の情報としての電子書籍のナショナルアーカイブがある。

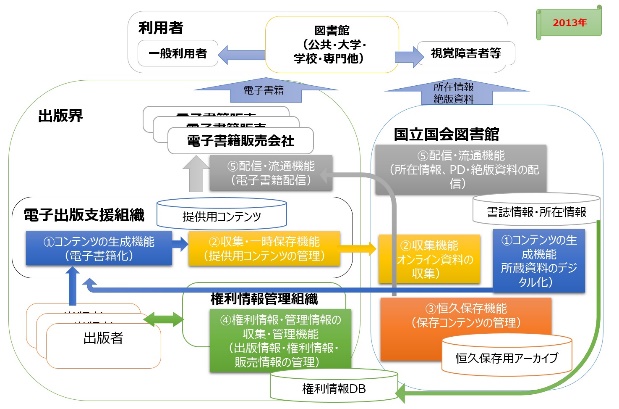


図5 電子書籍分野のアーカイブの機能モデル

### 電子書籍分野のアーカイブの役割

筆者は、電子書籍分野のナショナルアーカイブでの必要な機能を5つに分解し、それぞれの出版界と図書館界の役割分担を例示した。

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」13）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。

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。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。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、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

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

　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。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。

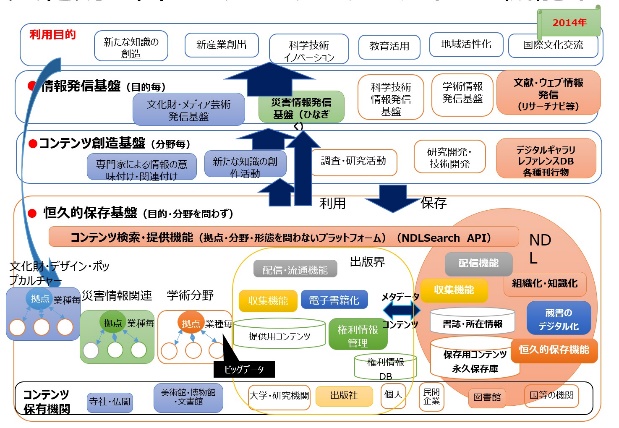
## 国の文化資産のアーカイブ

以下は、筆者がNDL在職中に、知識インフラの実現形として国のデジタル文化財のアーカイブシステムの基盤として想定したものであり、現時点において、NDLを含めて関係機関のコンセンサスが得られたものではない。

### アーカイブに関連した国の活動の方向性

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）14)、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）15)、学術情報の公開と共有の拡充に関する計画（文部科学省）16)、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。18)

図6 文化財を含めたナショナルアーカイブの機能イメージ

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。

### 恒久的保存基盤

恒久的保存基盤は、恒久保存と利活用のための共通プラットフォームとして、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し、各機関の情報を相互に持ち合って、障害、災害に備えるとともに、情報のマイグレーションを行うことにより、将来にわたって利用を保証する仕組みである。その分散アーカイブを集合して、あたかも１つのアーカイブとして見えるようにして網羅性、完全性を確保し、個々の情報同士を意味的に関連付けて、情報間のネットワークを構築することを想定する。

このようにネットワーク化された情報に対して、分野を越えて網羅性を保証した検索GUIとして、本文の全文検索、あいまい検索、シソーラス検索などを組み合わせた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ことである。

### 知識創造基盤

知識創造基盤は、それぞれの分野の専門家のみならず、広く国民も含めて、様々な分野の網羅的な知識を活用して、新たな著作物を創造する場である。創造活動を支援する基盤として、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を想定する。

新たな知識を創造する活動は、まず、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がある。関連付けて利用できる情報の幅が広がるため、より高度な創造性が期待できる。

また、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が含まれる。

ここで生成された情報は、新たな知識として恒久的保存基盤に蓄積されていくことを想定する。

### 知識利活用基盤

「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、知識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基盤である。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GUI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もので、レファレンスサービスによる情報探索支援、オンラインレファレンスなども含まれる。

### 運用基盤

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。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。

# 文化財でのLOD化の目指すところ

従来は、NDLは、同分野の情報の網羅性を意識して、各機関の情報に付けられたメタデータによる網羅的な情報探索、情報検索を効率的に実践できることに注力してきた。今後の国の文化財全体のナショナルアーカイブでは、異分野の様々な情報も含めて情報を関連付けることにより、思わぬところで自分の興味を喚起する情報に出会え、新たな発想の知識の創出が可能になる環境の実現が重要であると筆者は考える。

Linked Data化は、様々な情報を関連付けて知識として再利用するための具体的な実現方法を示したものであり、LOD化が進むことにより、あらゆる情報が関連付けられて利活用が可能になる。筆者は、LOD化の動きが、分野を越えて進展することにより、文化財全体のアーカイブがナショナルアーカイブとして実現されていくものと考える。

## クラウドコンピューティングの世界でのサービスの連携（2007年想定）

図7は、筆者が2007年に想定した、クラウドコンピューティングの普及を見越した将来像である。

　様々な機関から提供されているWebサービスが、連携することにより、より効果的なサービスとなっていく。サービスが連携できる情報提供環境として、当時の大きなパラダイムシフトの一つとしてのクラウドコンピューティングは、今や個人の情報管理のインフラとして普及している。

クラウドコンピューティングは、ネットワークコンピューティング、グリッドコンピューティング、SaaS（Software as a Service）、Webサービス等の技術を融合させた基盤である。

個別の組織が、単独でコンピュータ資源（ハードウェア、ソフトウェア）や情報を抱えてサービスを提供し、利用者が、個別のサービスを利用するような形ではない。クラウドコンピューティングの普及により、様々な機関の情報の共有だけでなく、個別のサービスを組み合わせて利用することが容易に実現できる環境が整った。

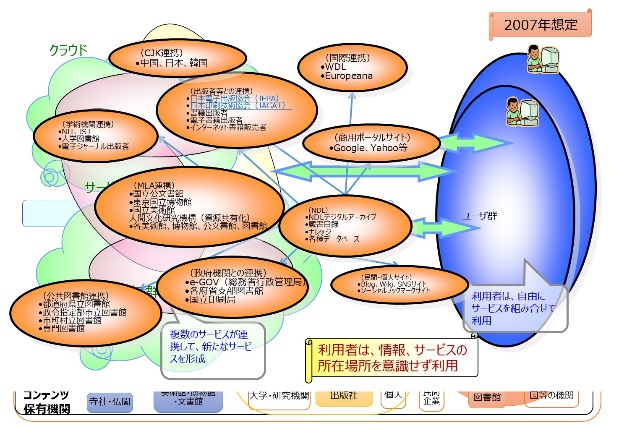
各情報保有機関のサービスも、このパラダイムシフトに対応していかなければ、孤立したサービスになっていく。クラウドコンピューティングが普及期に入り、また、デジタルトランスフォーメーションの考え方が一般化してきた社会において、各機関が保有するデジタル情報の管理の仕方、顧客サービスのあり方を再考する必要がある。今後ますます、情報と情報が情報群、サービスとサービスがサービス群、人と人がユーザ群として有機的に関連付けられていく。

いわゆるWeb2.0の時代では、パーソナライズ、人と人の関係を活用して情報を選択してきた。また、サービスも先進的な機関が提供するWebサービスを個別に利用してきた。Web3.0の時代以降は、それに加えて、情報と情報、サービスとサービスの関係を組み合わせて、ユーザ群と情報群が多対多の関係で、より的確な情報の活用が可能になってくる。

従来のように個別のデータベースを検索し、コンテンツを閲覧するような形から、クラウドコンピューティングの世界での複数のリソースから得た情報に、分析情報などの付加価値を付けて、利用者へ助言するような内容を知識として提供するサービスが実現されつつある。例えば、近年のカーナビの情報は、単に目的に行くルートを情報として提供しているのではなく、様々な手段により動的に変化する状況を集約し、その結果を踏まえてのリルート情報や、ルート上の様々な位置でのスポット情報などを合わせて提供している。ナショナルアーカイブの枠組みの中で、Linked Open Data化が進むと、更に利便性の高いサービスとして発展する。

このようなインターネットの世界での情報提供の状況の変革が、NDLで進めているウェブアーカイブ、情報探索サービスのあり方にも大きく影響してくる。従来のウェブアーカイブは各機関のウェブ情報を収集保存してきているが、このようなサービスの形をそのまま保存することは困難であり、また、関連付けられたメタデータ等、どのようにしてサービスの中にある情報を収集していくのかを考える必要がある。

クラウドコンピューティングの普及により、想定したサービスが現実になってきたが、今後、NDLは、情報を収集し将来にわたって保存する使命をどのように果たしていくか、考えていく時代が来ているように思える。

図7 クラウドの世界でのサービスの連携

## 文化財のアーカイブ構築のために何に留意するべきか

最後に、ナショナルアーカイブの構築を目指す中で、司令塔的な役割を果たす組織および文化財を保有する各機関は何に留意すべきか、筆者が想定する事項を列挙する。

* 今後10年を見据えて、個別の組織の論理ではなく、利用者視点でサービスを構築する。
* アーカイブの網羅性の確保のために、業種業態毎のアグリゲータ機関を設置してメタデータを集約する。更に、Linked Data化するために、情報を一意にする識別子を付与する。
* メタデータはもとより、コンテンツをOpen Data化する。
* 分野を越えたアーカイブ構築のために、メタデータの記述要素、記述規則および情報交換の通信規約（API）は、共通仕様を実装する。
* メタデータおよび情報内容の用語表現の違いを吸収する辞書を構築する。
* アーカイブの構築・運用に必要な人材を育成・確保する。

これを実現するするために、国の役割は次の点にある。

* 知的財産政策ビジョン、知財計画に掲げた施策を確実に実施する。
* 文化資源のアーカイブの利活用に必要な制度を整備する。
* アーカイブの構築・運用に必要なスキル・知識を習得するための体系的なカリキュラムと教材を作成するとともに、教育体制を確立する。

# まとめ

NDLは、今までは「電子図書館構想」（1998年）の実現を目指し、具体的な実施の柱立てとして「電子図書館中期計画」（2004年）を策定し、デジタルアーカイブ構築、情報に関する情報の充実、デジタルアーカイブポータルの構築を行ってきた。

筆者は、電子図書館としての知識・情報・文化の新しい基盤を含めて、図書館の枠を越えた基盤作りが、「知識インフラ」であり、知識インフラの実現形の１つが「ひなぎく」、知識インフラ構築の全体の実現形が、「文化資源のナショナルアーカイブ」と位置付けている。1990年代から目指す方向性は明確であり、構築に当たっては、セマンティックウェブ技術の適用を常に意識して進めてきており、具体的な実現手段として浸透しつつあるLinked Data化は、ナショナルアーカイブの構築においても、実現手段として中核的な技術の一つとなると考える。

# おわりに

個々の文化財の保有機関でのデジタル化が進む状況において、文化的資産をあらゆる人々が将来にわたり享受、活用できるようにし、人々の創造的な活用に貢献するためには、個々に情報資源の「見える化」をするだけでなく、全体のそれぞれの施策が同一の方向性を持って、相互に資源を補完し合って、社会全体でより効率的な利活用の保証に取り組む必要があり、ナショナルアーカイブは重要な基盤の役割を果たすことになる。ナショナルアーカイブの構築の先には、世界レベルでの「インターナショナルアーカイブ」があり、その構築へと発展することを目指していくべきと考えている。

今後10年のデジタル情報化の進展を見据えると、現行の図書館業務は根幹から変革が求められると思われる。知識創造のための情報の多くがインターネットにより、入手できる状況において、知識創造を支援する図書館の役割と必要な機能を再検討し、速やかに対応していかなければ、情報提供機関としての図書館は存立意義を失う可能性がある。

～～～

■参考文献

1. 田屋裕之.「パイロット電子図書館システム」の現状と課題.出版ニュース.1996、 96 1/上・中、p.14-17
2. 中山正樹.国立国会図書館電子図書館中期計画2004の実施に向けて.情報の科学と技術. 2004、Vol.54、No.9、p.453-460.
3. 久古聡美・吉田曉・中山正樹. Web2.0世代のデジタルアーカイブポータルの提供を目指して.情報管理. 2006、Vol.49、No.6、p.313-323.
4. 国立国会図書館. “機能概要≪国立国会図書館サーチについて”. 国立国会図書館サーチ、 　<http://iss.ndl.go.jp/information/function/>、　(accessed 2016-1-15).
5. 国立国会図書館. “日本全国書誌の在り方に関する検討会議について”. 国立国会図書館インターネット資料収集事業(WARP). <http://warp.da.ndl.go.jp/info:ndljp/pid/8703998/www.ndl.go.jp/jp/library/data/jnbconf_20100303.html>、 (accessed 2016-1-15).
6. 内閣府科学技術政策・イノベーション担当. “第4期科学技術基本計画（平成23年8月19日　閣議決定）”. 内閣府. <http://www8.cao.go.jp/cstp/kihonkeikaku/4honbun.pdf>、(accessed 2016-1-15).
7. 国立国会図書館. “第三期科学技術情報整備基本計画”. 国立国会図書館インターネット資料収集事業(WARP). <http://dl.ndl.go.jp/view/download/digidepo_8262067_po_basic_plan03.pdf?contentNo=1>、 (accessed 2016-1-11).
8. 長尾真.未来の図書館を作るとは.長尾真、 2012、 27p
9. 中山正樹. 国立国会図書館におけるデジタルアーカイブ構築－知の共有化を目指して－. 情報管理. 2012、Vol.54、No.11、p.715-724.
10. 内閣官房東日本大震災復興構想会議. 復興への提言～悲惨のなかの希望～. 2011-06-25、 第12回東日本大震災復興構想会議. 57p. <http://www.cas.go.jp/jp/fukkou/pdf/kousou12/teigen.pdf>、 (accessed 2016-1-15).
11. 国立国会図書館. “東日本大震災アーカイブ構築プロジェクト”. 国立国会図書館.<http://kn.ndl.go.jp/static/about/project>、(accessed 2016-1-15).
12. 内閣官房知的財産戦略本部. “知的財産政策ビジョン”.首相官邸政策会議.<https://www.kantei.go.jp/jp/singi/titeki2/kettei/vision2013.pdf>、(accessed 2016-1-15)
13. 長尾真. “「電子図書館と電子出版の今後」JEPAセミナー”. 電子出版協会.<http://www.slideshare.net/JEPAslide/20150303-drnagao-45401397>、 (accessed 2016-1-15)
14. デジタル文化財創出機構.“「国立デジタル文化資産振興センター(仮称)」設立構想提言”.文化庁文化関係資料のアーカイブに関する有識者会議.<http://www.bunka.go.jp/seisaku/bunkashingikai/kondankaito/bunka_archive/01/pdf/shiryo_5-2.pdf>、(accessed 2016-1-15)
15. 文化庁.「文化芸術の振興に関する基本的な方針（第4次基本方針）」.文化庁文化審議会.<http://www.bunka.go.jp/koho_hodo_oshirase/hodohappyo/pdf/2015052201.pdf>、(accessed 2016-1-15)
16. 日本学術会議情報学委員会. 我が国の学術情報基盤の在り方について －SINET の持続的整備に向けて－.日本学術会議.<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t192-2.pdf>、(accessed 2016-1-15).
17. 「アーカイブ立国宣言」編集委員会 編. アーカイブ立国宣言. ポット出版、 2014、272p.
18. 中山正樹.国立国会図書館のサービスシステムの歩みと新たな方向性の模索－電子図書館事業20年の歩みー. 国立国会図書館月報. 2015、Vol.648、p.18-24.

**■図表、写真のキャプション**

図1 電子図書館に関連する構想とサービスの変遷

図2 NDL電子図書館サービスの全体像

図3 知識情報基盤の構築モデル

図4 東日本大震災アーカイブのシステムイメージ

図5 電子書籍分野のアーカイブの機能モデル

図6 文化財を含めたナショナルアーカイブの機能イメージ

図7 クラウドの世界でのサービスの連携

≪討議報告≫

LOD化によるデジタル文化財の利活用を目指して

**（発表：中山正樹）**

**和中　幹雄（大阪学院大学）**

フォーラム当日の中山氏による発表題目は、「LOD化によるデジタル文化財の利活用を目指して」であった。本誌掲載の論文ではそれは副題目となり、発表内容をより端的に示す題目を用いている。この発表の趣旨は、「国立国会図書館（NDL）の電子図書館事業の20年間を情報技術の観点から振り返り、今後のナショナルアーカイブに向けた展望と課題を概観したもの」ということができる。

フォーラム当日のプレゼン資料には、「NDL電子図書館サービスの全体像」「長期保存システムの概念モデル」「国立国会図書館サーチのコンセプト」「文化財を含めたナショナルアーカイブの機能イメージ」「ナショナルアーカイブにおけるLOD化」などなど、数多くのモデル図やイメージ図で溢れ、個々のスライド一枚を取ってみても、優に一つの発表内容となるものが多く見られた。そのため、予定された時間で終了できず、発表の趣旨の理解も不十分だったように感じられた。しかしこの点については、本誌掲載の論文を読めば氷解するであろう。

彼の発表が自分の職歴の紹介から始まるのも故なしとしない。NDLの電子図書館事業は、1994年に情報処理振興事業協会（IPA）との協力で実施したパイロット電子図書館プロジェクトによって始まる。彼はこの事業にIPA側の担当者として参画する。その後、関西館が開館し電子図書館サービスが開始される2002年にNDLに入館し、2004年に、デジタル・アーカイブ構築とデジタル・アーカイブポータルの考え方を明示した「電子図書館中期計画2004」策定の中心メンバーとなり、その後、PORTAからNDLサーチの開発に携わり、2015年3月末に、電子図書館事業や情報システム全般を統括する立場にある者として定年退職を迎える。このように、NDL電子図書館事業の20年と中山氏の図書館システムに関わるキャリアは全く重なるからである。

1994年から始まる現在までの20年間のシステム構築に当たっては、セマンティックウェブ技術の適用を常に意識して進めてきたと彼は述べている。つまり、LODも含めたセマンティックウェブの要素技術をどのような考え方で個別に適用してきたかを振り返ることにより、今後の方向性を示したのが本発表であったと言える。

情報技術の観点からの発表ではあるが、特に今後のナショナルアーカイブ構築に向けた課題整理において、内閣に設置された知的財産戦略本部による「知的財産政策ビジョン」などの国の政策を背景とした提言が多く見られる。彼は、発表を終わるにあたって、次のように述べている。「LOD化されたナショナルアーカイブ構築に当たっては、それらの意義・理念の確認、認識の共有、草の根的な試作など、やれるところからやるという段階から、大所高所の方針・計画に基づいた、具体的な実用サービス構築を行うという次の段階に移行する必要がある。そのためには、司令塔が必要である」と。

我が国にどのような司令塔が可能であろうか。

彼が示したモデル図やイメージ図は、これからの図書館を考える上で示唆に富む情報が数多く含まれている宝の山と言える。

司会者の仕切りが悪かったこともあり、討議時間はほとんど取れず、質疑もNDLサーチの個別問題に終始したことは残念であった。当日のプレゼン資料にあるさまざまなモデル図やイメージ図を基にまとめられた本誌掲載の論文をお読みいただき、今後、活発な議論がなされることを期待したい。

NDLサーチをめぐる質疑は、次のようなものであった。

Q.ディスカバリーサービスのNDLサーチで、NDL-OPACの代替ができるのか？

A．NDLサーチのGUIは、ナショナルアーカイブの情報発信基盤（知識利活用基盤）で全ての情報を検索対象とする。来館利用者向けのNDL-OPACは、現在はALEPHで提供しているが、今後は、NDLサーチで、あらかじめ来館者向けに当館所蔵の図書・雑誌、デジタル化コンテンツに絞り込んだGUIを用意すれば、現在のNDL-OPACと同等のサービスが提供できる。現に、児童書OPACは、NDLサーチで、GUIを変えて児童書に絞り込んで提供している。

Q．新聞総合目録がNDLサーチに取り込まれて使いづらくなった。

A．総合目録と言われていた、ゆにかねっと、児童書総合目録、新聞総合目録は、全て統合しNDLサーチでGUIを提供するとしてサービスを構築したが、新聞総合目録は、提供する情報のレベル、検索手法が他と大きく違い、相当無理があったのは確かである。

Q．NDLサーチの検索結果のデフォルトの「適合度順」はよくわからない。

A．書誌情報における検索キーワードの個数、出現頻度、データベースごとの優先度を数値化して、数値の高いものの順に表示している。ロジックの完成度はまだ高くない。

【2017年】AIを活用した「知の共有化」システムの方向性  
―知の共有化システムの創生の歩みー

2017年10月3日

元国立国会図書館　専門調査員・電子情報部長

中山正樹

# はじめに

# 概要

## 「知の共有化」とは

## デジタルトランスフォーメーション(デジタル変革)

## 「未来の図書館を作るとは」【長尾先生】を読み返して

## 知的情報資源のビッグデータとしての活用の可能性

# 電子図書館20年の歩み

## 電子図書館構想と実証実験

## 国立国会図書館中期計画2004

## デジタルアーカイブの構築

## デジタルアーカイブポータルのプロトタイプ

## PORTAの構築

## 国立国会図書館サーチ（NDLサーチ）

# 知識インフラの構築を目指して

## 「知の共有化」システムの方向性

## 知識インフラの概念

## 東日本大震災アーカイブ

## 電子書籍ナショナルアーカイブ

## 文化情報資源のナショナルアーカイブ

# 今後10年を見据えた国等の施策*（2017年～2024年）*

## 「知的財産政策ビジョン」での今後10年の目標

# 「知の共有化」システムの構築業務と人材の資質とスキル

## 政府標準ガイドラインに沿ったシステム構築と運用

## 政府標準ガイドラインに沿った開発タスクとドキュメント

## iコンピテンシ・ディクショナリ（iCD）

## 業務遂行のタスクとスキル・知識の蓄積のスキーム

## iCDを活用した業務の遂行とスキル・知識の選択的習得

# *今後10年で目指すところ（2015年～2024年）*

## 「知的財産政策ビジョン」での今後10年の目標

## 国の文化資産のアーカイブ

# 文化財でのLOD化の目指すところ