

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—

中山 正 樹

1. 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2004⁽³⁾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。

2. 「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、ポ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。

3. 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。

4. 「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っていて、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がられ、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ャ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とは」⁽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。

5. 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。

6. 「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。

6.1. 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

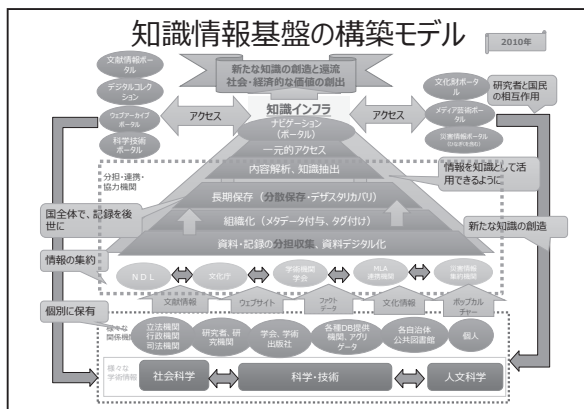


図1 知識情報基盤の構築モデル

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⁾。

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—

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。

6.2. 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

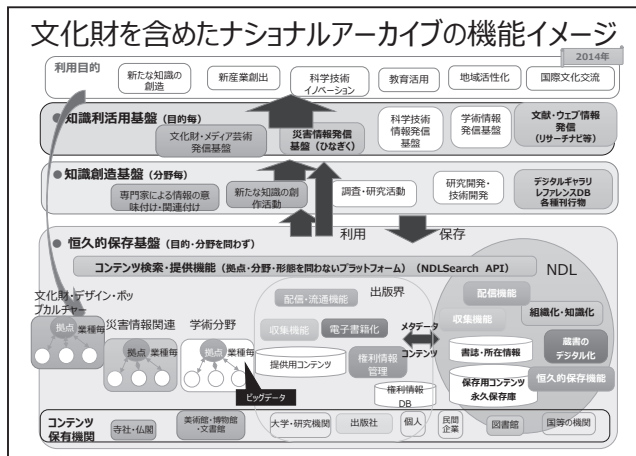


図2 文化財を含めたナショナルアーカイブの機能イメージ

「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⁾。

6.3. 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。

6.4. 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。

6.4.1. 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。

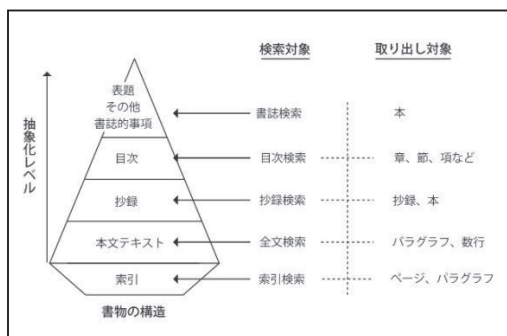


図3 種々の検索システム

(「未来の図書館を作るとは」より)

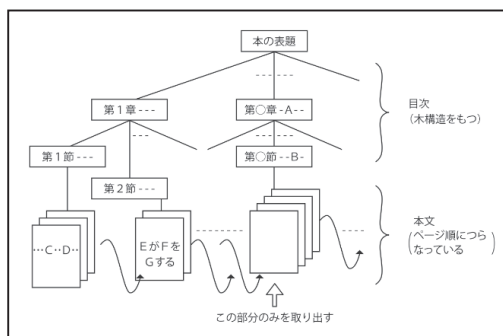


図4 書籍の構造化

(「未来の図書館を作るとは」より)

- 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

6.4.2. 知識創造活動

- 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

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ー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

6.4.3. 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、更に利用者同士が繋がっていく「場」としての機能を期待する。

- 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。

- 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

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、資料場所を探し出し、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。

6.5. 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、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。

6.5.1. 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、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。

7. 「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。

7.1. 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。

7.2. 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

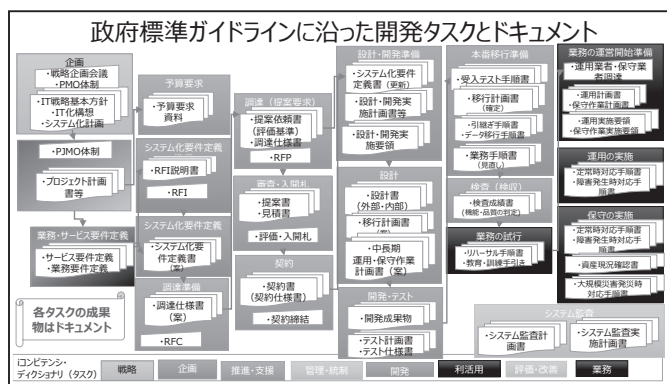


図5 政府標準ガイドラインに沿った開発タスクとドキュメント

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。

7.2.1. 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、よりの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。

7.2.2. 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。

7.3. 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。

7.4. 業務遂行のタスクとスキル・知識の蓄積のスキーム

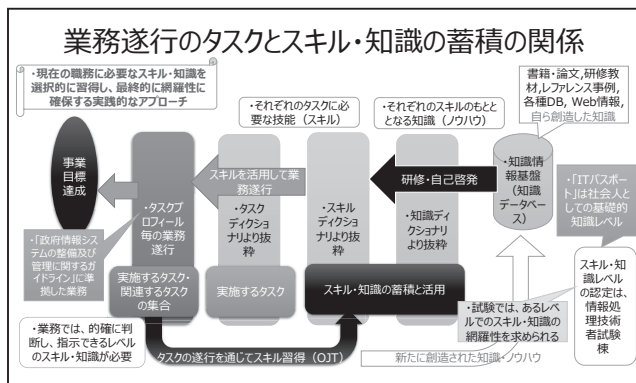


図6 業務遂行のタスクとスキル・知識の蓄積の関係

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。

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。

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、事前を選択的に習得して、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。

7.5. 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。

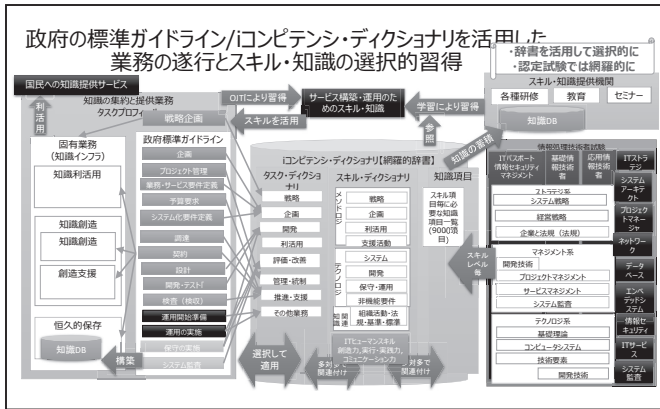


図7 政府標準ガイドライン／iコンピテンシ・ディクショナリを活用した業務の遂行とスキル・知識の選択的習得

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。

「④ IT 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。

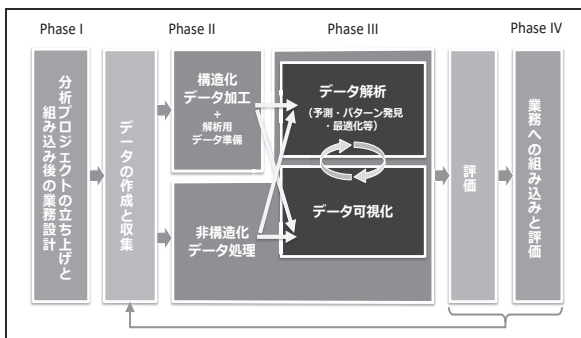


図8 データサイエンス領域のタスク構造

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に示されている。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。

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。

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」、② IT 関連技法などのスキルとして「テクノロジー」、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。

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に示されている。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。

8. 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も、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は人の仕事を奪うよりも、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・吉田暁・中山正樹. Web2.0世代のデジタルアーカイブポータル提供を目指して. 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) バイカレント・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)

(なかやま まさき。2017年9月9日受理)