

第一次中間取りまとめ

平成 30 年 4 月

実務者検討委員会

(事務局:内閣府知的財産戦略推進事務局)



この報告書は、[クリエイティブ・コモンズ 表示 4.0 国際ライセンス](https://creativecommons.org/licenses/by/4.0/)の下に提供されています。

実務者検討委員会 第一次中間取りまとめ

目次

はじめに.....	2
1. これまでの経緯.....	3
2. 議論の背景.....	4
3. 実務者検討委員会で示されたデジタルアーカイブの構築・連携の現状と課題.....	6
4. デジタルアーカイブ社会の実現に向けた施策の検討のために.....	8
5. 国の分野横断統合ポータル構築.....	16
6. 今後の主要検討課題.....	21
おわりに.....	21

(補足資料)

○共通メタデータフォーマット.....	23
・連携フォーマット(案)について.....	23
・ジャパンサーチ(仮称)の連携フォーマット(案)詳細版.....	27
・利活用フォーマットについて.....	35
・ジャパンサーチ(仮称)利活用のためのメタデータフォーマット仕様.....	43
○デジタルアーカイブアセスメントツール.....	50

(関連資料)

○工程表(平成 29 年9月5日 デジタルアーカイブジャパン推進委員会決定).....	53
○我が国におけるデジタルアーカイブ推進の方向性(報告書) エグゼクティブ・サマリー.....	57
○デジタルアーカイブの構築・共有・活用ガイドライン(ガイドライン)-概要-.....	59

はじめに

現在、我々を取り巻く「サイバー」空間の環境は刻一刻と変化しており、利便性の高いサービスが次々と提供される反面、その担い手の多くが国際プラットフォームであることもあり、「サイバー」空間において日本の長い歴史や伝統に根差した文化の重層性や多様性が埋没してしまうのではないかとの危惧が呈されることも多い。

一方、「リアル」社会においては、ライフデザインや働き方における価値観が多様化し、シェアリングエコノミーに代表されるように必ずしも「所有」にこだわらず必要に応じた様々な価値を「シェア」といった行動が広く行われるようになってきている。

経済活動や社会生活などどの局面においても「サイバー」の存在感は増大する一方であるが、他方で、「リアル」と「サイバー」の結びつきを強化する技術でもある、IoTや人工知能(AI)、ビッグデータといった以前にはなかった新技術の開発・応用も進んできている。こういった新技術の進展は、データを媒介にした異業種同士や供給者と顧客の直接の結びつきを加速させ、コンテンツ¹やデータを距離、時間、費用、言語の制約なく共有することを可能とする。すなわち、モノづくりやコンテンツの作成・発信、マネタイズや評価・価値付けに至るまで誰もが容易に行うことができるようになる。

こういった流れを踏まえ、我が国の企業・大学・行政機関など様々な主体が保有する知的資産²を可能な限りオープンな形でシェアし、その知的資産を利活用することで「サイバー」から「リアル」へ誘導または補完する役割を担い、新たな価値を生む足がかりとなりうるデジタルアーカイブ³の意義について今一度共有した上で、我が国でも、デジタルアーカイブジャパンとして充実させていく必要がある⁴。

従前は、権利による保護も含め、データを独占ないし秘匿することによりその価値の最大化が図られることが多かったが、近年はソフトウェアの開発領域では、GitHubなどに代表されるような開発プロジェクトのための共有ウェブサービスの取組や、学術領域におけるオープンサイエンスの取組などオープンな形でシェアを図ることで価値の最大化や効率化を図る流れが形成されてきている。

こういった流れは、今後さらに拡大されることが予想されるどころ、合わせて、データの提供機関が真正性や正確性を保証した上でオープン化⁵することで、二次創作物の価

¹ 個々の社会・文化・学術情報資源。「デジタルコンテンツ」の他、アナログ媒体の資料・作品等を含む。

² メタデータや写真、動画などのコンテンツに限らず、3Dデータや研究データなどありとあらゆるデータをいう。

³ ここでは、様々なデジタル情報資源を収集・保存・提供する仕組みの総体をいう。デジタルアーカイブで扱うデジタル情報資源は、「デジタルコンテンツ」だけでなく、アナログ媒体の資料・作品も含む「コンテンツ」の内容や所在に関する情報を記述した「メタデータ」や、コンテンツの縮小版や部分表示である「サムネイル/プレビュー」も対象とする。

⁴ 現在国連で提唱されているSDGs(Sustainable Development Goals:持続可能な開発目標)の観点から、いずれは各国で構築されているナショナル・デジタルアーカイブと連携することにより、国際的なデータ提供インフラとしても機能・貢献することも視野にいれておくべきとの指摘もあった。

⁵ インターネットを通じて広く公開されており、商業利用も含めて、目的に応じた活用可能な条件が明示され、手続を要せずに提供されるデータが第三者に自由に利用できるようになっている状態をいう。

値の向上も期待できるばかりでなく、場所や時間を超えて正しい情報を見つけることも可能となる。

こうしたデータやコンテンツの共有基盤となるデジタルアーカイブジャパンの構築には、個々のアーカイブ機関⁶におけるデジタル化やシステム整備、権利処理などへの継続的な取組を基本としつつ、制度的課題の検討や関係者間での調整が必要なこともあり、相応のプロセスと時間が必要となる。平成 29 年度に開催された実務者検討委員会においては、平成 29 年9月のデジタルアーカイブジャパン推進委員会において合意した範囲・事項⁷について、議論や取組を進めてきており、その成果を取りまとめることとした。

1. これまでの経緯

デジタルアーカイブに関する取組としては、平成 29 年4月、「デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会⁸」において、報告書⁹及びガイドライン¹⁰が取りまとめられている。

また、これを踏まえ、「知的財産推進計画 2017」(平成 29 年5月知的財産戦略本部決定)において、知的財産戦略の一環として、2020 年に向け、デジタルアーカイブ構築に係る各種の施策に取り組むことが決定されたほか、「経済財政運営と改革の基本方針 2017」(骨太方針)及び「未来投資戦略 2017」(いずれも平成 29 年6月閣議決定)においても、デジタルアーカイブ施策の必要性や方向性が示されている。

このような施策の必要性及び方向性を受け、平成 29 年9月に第 1 回デジタルアーカイブジャパン推進委員会が開催され、様々な分野におけるデジタルアーカイブ構築の取組について工程表が決定されるとともに、デジタルアーカイブ推進に係る実務的課題に対応するため、実務者検討委員会を設置することが決定された。実務者検討委員会においては、関係省庁等連絡会・実務者協議会の報告書で示された施策等の取組状況について整理するとともに、デジタルアーカイブ構築・利活用に係る課題への対応策について議論を行った。

⁶ アーカイブ/アーカイブズは、公文書館等を指す言葉として一般的に用いられてきたが、ここでは広い意味での記録機関全般を指す。社会・文化・学術情報資源である資料・作品等のコンテンツを収集し、その資源を整理(組織化)し、保存し、提供する機能を持つ機関・団体等をいう。博物館・美術館、図書館、文書館といった文化的施設のほか、大学・研究機関、企業、市民団体、官公庁・地方公共団体等を含む。なお、提供機能が限定的であり、一般への公開を想定していない機関等も含む。

⁷ http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/suisin/dai1/gjjsidai.html
資料1の 8 ページ「デジタルアーカイブジャパンの当面の対象範囲」及び資料4の1ページ「デジタルアーカイブジャパンの実現に向けた工程表(全体)」

⁸ https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/index.html

⁹ 我が国におけるデジタルアーカイブ推進の方向性(平成 29 年 4 月)※「関連資料」中にサマリーを添付 https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/houkokusho.pdf

¹⁰ デジタルアーカイブ構築・共有・活用ガイドライン(平成 29 年 4 月)※「関連資料」中に概要を添付 https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline.pdf

2. 議論の背景

様々なコンテンツのメタデータ¹¹整備やデジタルコンテンツ¹²の拡充をはじめとするデジタルアーカイブに関する取組を進めることは、文化の保存・継承・発展の基盤になるだけでなく、保存されたコンテンツの二次的な利用や国内外に発信する情報の基盤となる重要な取組である。観光、教育、学術、防災など様々な用途での活用を通じて、デジタルアーカイブの構築・共有と活用の循環を持続的なものとし、その便益を国民のものとしていくことで、我が国の社会的、文化的、経済的発展につなげることが重要である。効率的・効果的なデジタルアーカイブの利活用を推進するためには、我が国が保有する様々なコンテンツへのナビゲーションを行い、様々な分野や地域の情報と活用者をつなぐための連携基盤としての役割を果たす「国の分野横断統合ポータル」¹³の構築を行うことも重要である。

現在、我が国の「アーカイブ機関の設置・運営者」は、国、独立行政法人、地方自治体、民間等の多岐にわたっている。デジタルアーカイブの構築は一部で進みつつあるが、世界の動向を踏まえると、メタデータやデジタルコンテンツの提供は十分とは言えず、分野横断的なアーカイブの連携に関する取組についてはほとんど進められていない状況にある。

¹¹ データに関するデータを意味し、「サムネイル/プレビュー」や用語を統制するための語彙等も含むものとして広く定義されることもあるが、ここでは、コンテンツの内容、外形、所在等に関する記述等のデータをいう。図書館における書誌データ、アーカイブ機関の収蔵品等の目録データ、文化財の基礎データ等のテキストデータや URI 参照のデータを指す。

¹² アナログ媒体の資料・作品等をデジタル化した高品質なデジタルコンテンツ、あるいはポーンデジタルの作品（デジタルカメラの写真、電子書籍等）であればそのデジタルコンテンツ自体。

¹³ 我が国が保有する様々なコンテンツのメタデータに関し、分野横断的な検索を可能とするポータルサイト。「知的財産推進計画 2015」において、その構築の必要性が掲げられ、「知的財産推進計画 2017」では、国立国会図書館が、2020 年までに、国の分野横断型の統合ポータル「ジャパンサーチ(仮称)」の構築を目指すものとされている。

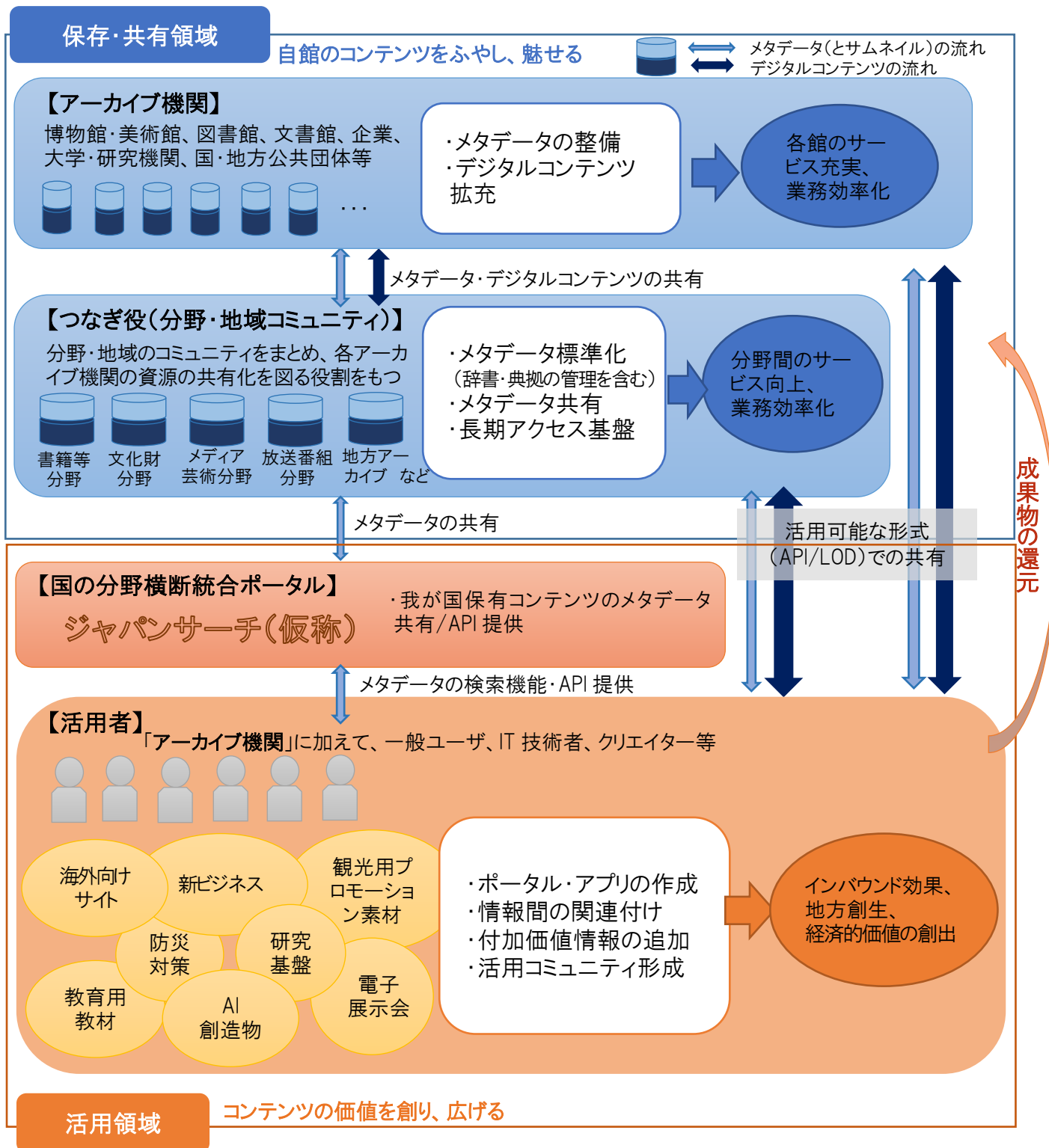


図1 デジタルアーカイブの共有と活用のために

こうした状況の下、関係省庁等連絡会・実務者協議会の報告書では、我が国におけるデジタルアーカイブの取組を推進するために、図1に示すように、分野・地域ごとの「つなぎ役」¹⁴を介して、「アーカイブ機関」と国の分野横断統合ポータル「ジャパンサーチ(仮称)」の間でメタデータの共有を進め、「活用者」¹⁵がジャパンサーチ(仮称)等からメタデータを共有し、さまざまな用途に活用するというサイクルの構築を目指す方向性が示された。また、アーカイブ機関は、関係省庁等連絡会・実務者協議会のガイドラインに沿って、デジタルアーカイブに関する取組(メタデータの整備、サムネイル/プレビュー¹⁶の作成、デジタルコンテンツの拡充、整備したメタデータやサムネイル/プレビューのオープン化、デジタルコンテンツの利用条件表示など)を行うことが望ましいこと、デジタルアーカイブ推進のためには、分野・地域のコミュニティの主体となるつなぎ役の果たす機能・役割が重要となることなどの方向性も打ち出された。

3. 実務者検討委員会で示されたデジタルアーカイブの構築・連携の現状と課題

実務者検討委員会においては、関係省庁及び主要アーカイブ機関から、デジタルアーカイブを構築・利活用していく上での現在の取組状況や今後の課題などについて、次のとおり報告が行われた。

○書籍等分野(国立国会図書館)

書籍等分野のつなぎ役として、「国立国会図書館サーチ」を運用中。資料のデジタル化、メタデータのオープン化¹⁷、デジタルコンテンツの二次利用条件表示の促進等の課題があるが、これらについては国全体としての取組が必要である。

○公文書分野(独立行政法人国立公文書館)

国立公文書館では、「国立公文書館デジタルアーカイブ」と「アジア歴史資料センターデジタルアーカイブ」の2つを運用中。前者はデジタル化の更なる推進、後者はデータベースの拡充が、それぞれ課題となっている。

○文化財分野(文化庁伝統文化課)

文化遺産オンラインを国立情報学研究所とともに運営している。文化遺産オンラインは、国立国会図書館サーチと連携するとともに、独立行政法人国立文化財機構

¹⁴ Europeana の「アグリゲーター」、DPLA の「ハブ」に相当する役割・機能を果たす機関のことをいい、分野・地域コミュニティにおけるメタデータを集約し、API 等による提供を行う機関。メタデータの標準化、用語の統制(辞書・典拠・シソーラスの管理)を行い、コミュニティにおけるメタデータ整備やデジタルコンテンツ作成への支援、長期アクセス保証のための基盤提供、活用コミュニティの形成などの役割を担う。単独の組織が担う場合もあれば、行政、NPO や大学等と分担して行う場合もありうる。

¹⁵ デジタルアーカイブの様々なデータを活用する者。「アーカイブ機関」が自らのために活用することに加え、一般ユーザ、IT 技術者、クリエイターなど、様々な機関・団体・個人がデータの活用者になりうる。

¹⁶ コンテンツの要約又は一部分の表示。縮小した画像(サムネイル)、本文テキストの一部表示や数秒程度の音声・動画(プレビュー)等をいう。

¹⁷ インターネットを通じて広く公開されており、商業利用も含めて、目的に応じた活用可能な条件が明示され、手続を要せずに提供されるデータが第三者に自由に利用できるようになっている状態をいう。なお、ガイドラインでは、PDM、CC0、CC BY、特に CC4.0 の利用条件での公開を推奨している。

の国立博物館所蔵品統合検索システム(ColBase)とも連携しており、文化財分野の他のデータベースとの連携にも努めていく。文化遺産オンラインでは、国指定文化財に係る文字情報は全てデジタル化の上で公開しているが、画像等の資料デジタル化や二次利用条件表示の促進等が課題である。

○文化財分野(独立行政法人国立美術館)

今年度から来年度にかけて、国立美術館の所蔵品の来歴等メタデータの精緻化を進めている。また、公開されている情報資源のゲートウェイを構築する計画で、将来的には、「ジャパンサーチ(仮称)」や文化遺産オンラインとの連携を進める予定である。

○文化財分野(独立行政法人国立文化財機構)

国立文化財機構の4つの国立博物館の所蔵品を検索できる「国立博物館所蔵品統合検索システム(ColBase)」を運営している。メタデータやサムネイルは CC BY¹⁸で公開している。現在、国立国会図書館サーチ、文化遺産オンラインとの連携に向けた調整を進め、ColBase のメタデータ及び画像の充実に努めている。

○メディア芸術／メディアコンテンツ分野(文化庁芸術文化課)

メディア芸術データベースの 2020 年正式公開を目指し、現在、国立国会図書館サーチとの API¹⁹連携を検討中。補助金によるメディア芸術作品のアーカイブ化支援、一部作品のサムネイル画像の表示、利用条件表示にも取り組んでいる。専門的な能力を持つ人材の育成が課題である。

○自然史・理工学分野(独立行政法人国立科学博物館)

生物多様性の分野では、地球規模生物多様性情報機構(GBIF)が世界規模のポータルを運営している。日本では、「S-Net(サイエンス・ミュージアムネット)」というポータルを国立科学博物館が運営している。また、科学史資料や産業技術に関する歴史的資料の所在については、国立科学博物館が運営する「産業技術史資料データベース」で公開している。これら活動の継続・推進・教育、データ等のデジタル化、非生物(岩石・鉱物等の地球科学系、産業技術史系)データへの対応が課題である。

○人文学分野(大学共同利用機関法人人間文化研究機構)

人間文化研究機構に属する6機関等のデータベースを検索できる「統合検索システム nihulNT」を運営し、国立国会図書館サーチとも連携している。nihulNT のライセンスは、現状は各機関に委ねられているが(例えば、国文学研究資料館の歴史的典籍のデータは CC BY-SA²⁰)、ガイドラインに沿ってできるだけオープンにするよう

¹⁸ 原作者のクレジット(氏名、作品タイトルなど)を表示することを主な条件とし、改変はもちろん、営利目的での二次利用も許可される最も自由度の高い CC ライセンス。

¹⁹ Application Programming Interface。オペレーティングシステムやアプリケーションソフトが、他のアプリケーションソフトに対し、機能の一部を利用できるように提供するインターフェースのことをいう。

²⁰ 原作者のクレジット(氏名、作品タイトルなど)を表示し、改変した場合には元の作品と同じライセンス(こ

努めている。データ化の促進とデータの統一性の確保(データクレンジング)が課題である。

○放送番組分野(公益財団法人放送番組センター)

放送ライブラリーで公開する番組のメタデータを検索・閲覧できるシステムを運用しているほか、ドラマのメタデータをジャパンサーチ(仮称)プロタイプと連携(サンプルデータの提供)する作業を進めている。また、公開番組の視聴ニーズに対応するため、図書館などの公共施設や大学の講義の中で利活用する事業を進めている。

○放送番組分野(日本放送協会)

NHK アーカイブスが保存している映像のポータルサイトを公開しており、「みのがしなつかし」「みちしる」など6分野から選択して視聴が可能。放送番組センターによるジャパンサーチ(仮称)プロタイプへのサンプルデータ提供への協力に加え、NHK アーカイブスの一部データとジャパンサーチ(仮称)プロタイプとの連携についても検討を進めている。

上記の各分野における取組のほか、分野横断的な取組として、内閣官房明治 150 年関連施策推進室が「明治 150 年」ポータルサイトを立ち上げ、明治時代に関連するデジタルアーカイブ情報の集約を進めている。

4. デジタルアーカイブ社会の実現に向けた施策の検討のために

実務者検討委員会においては、デジタルアーカイブに関する関係省庁等連絡会・実務者協議会の報告書で示された課題への対応策についても検討が行われた。

(1)「デジタルアーカイブ社会」とは

報告書において提言された「デジタルアーカイブ社会」とは、「デジタルアーカイブが日常的に活用され、様々な創作活動を支える社会・学術・文化の基盤となる社会」と位置付けられる。すなわち、デジタルアーカイブによって、日々生み出される様々なデータが共有され誰でも簡単にアクセスできること、また、それが日常的に利活用されやすい条件で提示されており、誰でも自由に様々なシーンで新しいコンテンツが生み出せる社会である。

平成 29 年 11 月、知的財産戦略推進事務局が産学官のデジタルアーカイブ関係者を集めたフォーラム(産学官フォーラム)を開催したが、そこでは、玉石混交の情報の中から自ら必要とする正確な情報だけが取れる社会や、アーカイブとしてどこかに置かなくとも、自然と蓄積される仕組みが実現される社会を目指すこと、そのためには利活用する人がデータを提供するアーカイブ機関へのリスペクトをもって一定のルールやモラルに

のライセンス)で公開することを主な条件に、営利目的での二次利用も許可される CC ライセンス。

従いながら利活用を行う必要があるといった意見があった。今後も引き続き産学官フォーラム等の機会を通じ、広く社会一般に施策の方向性について周知するとともに、関係者との情報共有、意見交換を行いながら、目指すべき姿と、その実現のために必要とされる課題解決の方向性を探っていくことが必要である。

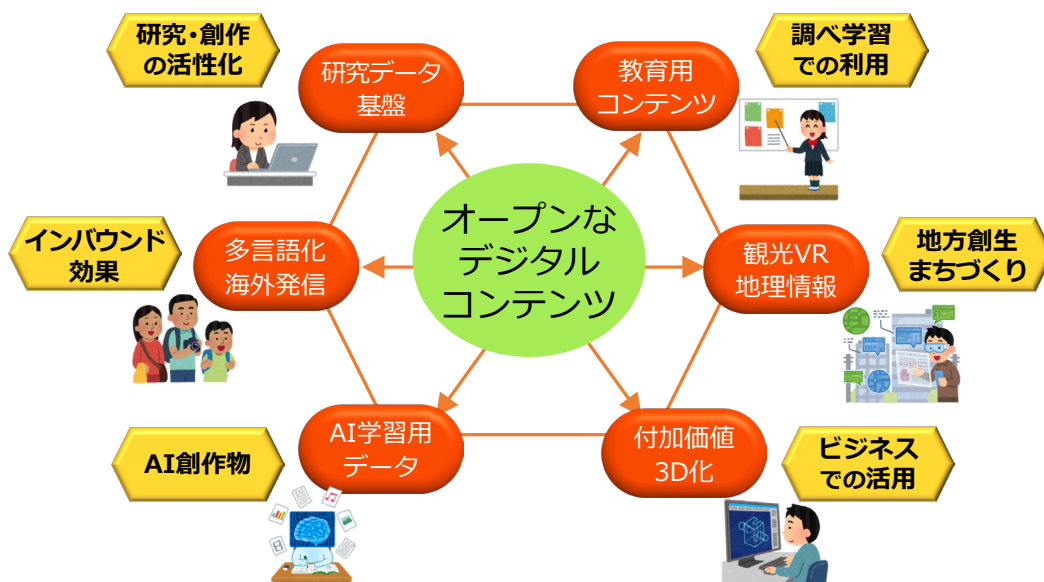


図2 デジタルアーカイブ社会のイメージ(例)

(2) デジタルアーカイブ構築の推進

(メタデータの整備)

デジタルアーカイブを活用していくためには、コンテンツにメタデータを付与し、サムネイル／プレビューの整備を進めていくことが必要となる。しかし、規模の小さいアーカイブ機関においては、人手やノウハウの不足により、これらの整備が進んでいないため、行政やつなぎ役を通じた支援が必要とされている。また、つなぎ役に求められる役割の一つが分野内のメタデータ標準や用語の統制であるが、分野の知識とメタデータの知識双方を有する専門家でなければ分野のメタデータ標準を作ることが難しいといった意見もあった。

これらの課題を解決するためには、当該分野において主導的立場にある大規模なアーカイブ機関が、自らのメタデータ付与の基準や方法を公開していくといった解決策も考えられる。また、AI による顔認識、物体認識、文字列認識などの機能を活用してメタデータを自動付与するといった新たな技術開発が進んでいることから、AI の学習用データの提供を促すことなどにより、これらの技術開発を支援することや、技術の共有化を図るといった取組も考えられる。

（デジタルコンテンツの拡充）

書籍や公文書などの一部分野ではコンテンツのデジタル化が進んでいるものの、全体としてインターネット上で利用できるデジタルコンテンツの量は欧米と比較すると少ないといわざるを得ない。

デジタルコンテンツを拡充していく際には、アーカイブ機関が活用者のニーズを把握すること、限られたリソースの中で何をデジタル化するかという優先順位をつけることが重要である。また、デジタルリマスターなどの技術も活用することでより品質のよいデジタルコンテンツの作成も可能となってきたことから、そのような最新デジタル技術の活用も検討すべきである。

そして、小規模機関においてもデジタルコンテンツの拡充を図っていくためには、小規模機関がコンテンツのデータ整備に責任を持ち、つなぎ役が大きなプラットフォームを用意することで、役割・負担を分担することが望ましいとの意見もあった。

（地域アーカイブの構築）

地域アーカイブの取組は、地域に眠っている多くの資料をデジタル化するというコンテクトで議論されているが、実際には、予算も人的リソースもないため、そもそも貴重な資料の保全が難しい状況にあり、例えば、地方自治体が設置した視聴覚ライブラリーに保管された資料や、市史編さんのために集積された資料などのデジタル化さえも進んでいない。地方自治体に対する調査²¹においても、デジタルアーカイブ構築に関する方針や計画を有するところはほとんどなく、一部アーカイブ機関において所蔵資料のデジタル化やデータ公開に取り組んでいるものの、来館者の増加など顕著な効果がみられるまでには至っておらず、予算や人員、スキルが不足していることを課題として挙げる回答が多かった。

今後、国や地方自治体、大学、企業、NPOなどすべての主体において、地域アーカイブを推進するための基盤となる人材の育成にそれぞれが取り組んでいかなければならない（「(5)人的基盤の整備」参照）。

その上で、地域のデジタルコンテンツを増やしていくためには、地域住民も当事者として参加する環境づくりをすすめ、各アーカイブ機関と地域住民が協働してデジタルアーカイブを構築するプロセスが重要であり、それがその後の利活用の創出にも繋がる。

具体的には、県域・圏域の数か所に、誰でも利用可能なスキャナ等の機器を備えたスタジオやアプリケーションの提供等の支援策を講じることや、地域住民自らが知りたい、利用したいと思えるアプリを提供することによって、自動的に地域のデータがアーカイブされる仕組みを行政側で用意するといった取組を推進していくことが必要である。

²¹「地方自治体における知的財産戦略等の策定・活用に関するアンケート」（平成 29 年9月知的財産戦略推進事務局実施）

(持続可能なデジタルアーカイブの構築)

これまでに構築されたデジタルアーカイブの中には、保存されることなく消えていったものも多く、これら消えゆくデジタルアーカイブの長期的なアクセス保証も大きな課題の一つである。

予算面では、構築当初からサーバシステムなどのリプレイス・運用保守に係る経費を確保する方法を考える必要がある。その際、基金化・クラウドファンディング利用などを検討し、ビジネスとして成り立つような形を模索することも考えられる。

システム面では、固定 URL は必要ではあるものの、無効になれば意味がなくなるため、何十年か後の人たちが使うという意識を持って、情報が長期的に有効であるよう工夫すべきとの意見があった。この点、ガイドラインにおいても言及されているが、長期アクセスを保証する DOI(Digital Object Identifier)²²等の永続的識別子の付与・普及といった取組も考えられる。

また、あるデジタルアーカイブを他機関で引き受けて、従前通り公開していく場合に、著作権者や肖像権者から改めて許諾を得ることが難しく、公開できなくなることがよくあるとの指摘があった。この点については、DOI 等の永続的識別子の付与に加えて、デジタルアーカイブを構築、公開する際に、包括的に権利処理することが望ましいが、いったん公開されたものについて、追加の権利処理を簡便に行えるような仕組みの構築の検討も望まれる。

(3)データの公開・オープン化の促進

データの公開・オープン化、アーカイブ機関側にインセンティブが働かないとの意見がある一方、その存在が、利用者だけでなく、広く認知されることで、コンテンツそのものの価値の向上に寄与するとともに、アーカイブ機関のブランド力の向上に寄与するとの意見もある。実際、アメリカのメトロポリタン美術館やオランダのアムステルダム国立美術館などが、自館や自館所蔵資料の価値向上のために、相当数の所蔵資料の高精細画像を CC0²³で公開している。

アーカイブ機関によっては、著作権や肖像権等が問題となって、所蔵コンテンツを全て無料で公開することが難しい場合もあるため、各アーカイブ機関の判断で適切な利用

²² コンテンツの電子データに付与される国際的な識別子。ISO 26324:2012 として国際標準化されている。DOI の登録には国際 DOI 財団が認める DOI 登録機関への参加が必要であり、我が国の DOI 登録機関としてはジャパンリンクセンター(JaLC)がある。なお、DOI はデータの管理主体が変わっても識別子自体は変わらないため、この観点からも持続可能なデータの提供を行うことに資するものである。
<https://japanlinkcenter.org>

²³ 全ての著作権等の権利を放棄することを意味する。これは、著作権に基づいて訴訟を起こす権利、逸失利益等が出て不法行為に基づく損害賠償を求める訴訟を起こす権利も含めて放棄し、著作者人格権など放棄できない権利については行使しないことを約束するといったことを意味する。

条件を設定できるように全体のシステムを構築していく必要がある。ただし、所在情報だけであっても公開することに価値はあるとの意見もあり、メタデータの整備・公開・オープン化にできる限り対応することが求められる。

地域アーカイブのコンテンツの公開については、プライバシーにも配慮し、公開の範囲を当該コミュニティに限定することについて許諾を得るといった方法も選択肢の一つとして検討すべきである。また、教育研究目的に限定した利用や、公開後一定の年数を経過したものについては、権利処理を要することなく一定の範囲内での利用(特定の施設内での限定公開等)を可能にするなどのゆるやかな合意形成を社会全体で図れないかといった意見もあった。

また、サムネイル/プレビューの作成・公開については、美術や写真の著作物を展示するアーカイブ機関における利用促進に係る制度改革に向けた審議が進められているところだが、さらなる利活用促進のためには、それ以外の著作物を所蔵するアーカイブ機関についても、自館の所蔵資料/収蔵品のサムネイル/プレビューを公開できるような制度が期待されるとの意見があった。

オープン化に関しては、多くのコンテンツホルダーからは、CC0 や CC BY といった条件で公開することにより、どのように利活用されたかが分からなくなるおそれがあるとの指摘があった。しかし、CC0 や CC BY といったクリエイティブ・コモンズ・ライセンスは、著作権の権利情報による利用条件を簡潔に示すもの²⁴であり、典拠表示などを利用条件とすることを否定しているものではないことから、過度なおそれをもって、オープン化を回避することは望ましくないものと考えられる。他方で、こういったおそれに対して、活用者は、適用されているライセンスや利用条件をよく確認し、順守しなければならないのはもちろんのこと、コンテンツ自体の価値をさらに高め、データ提供者にとってもメリットにつながる形で活用することが求められる²⁵。その上で、加工したデータもオープン化の循環に乗せることでより一層のデータ提供環境の向上も図られることとなる。また、データを集約・提供する機関においては、サイトポリシーなどで、典拠等の情報を明らかにするよう活用者に求めることで、データ提供者の懸念を払しょくすることも可能となると考えられるところ、ジャパンサーチ(仮称)においてもそのような対応を進めていくことで、一層のオープン化が図れるのではないかといった意見もあった。

²⁴ なお、クリエイティブ・コモンズ・ライセンス等の著作権ライセンスは、著作権者自ら保有する権利の利用許諾を行うものであるため、著作権保護期間が満了しているコンテンツや、著作物としての要件を満たさない創作性のないデータなどに適用したとしても、その二次利用に著作権法による保護を超えた追加的な制約を課すものではないことに留意が必要。(ガイドライン p.19 より)

²⁵ 欧米では”be culturally sensitive”(文化に配慮する)の観点から、法律などの明示されていなくとも、社会規範に沿った形でデータを活用することが求められるとのこと。

(4)利活用の促進

(市民参加型のアーカイブ)

デジタルアーカイブの利活用を進めるに当たっては、アーカイブ機関だけでなく、地域住民等が協働で構築するプロセスが重要である。市民参加型のアーカイブ構築を通じて、一人一人の市民がデジタルアーカイブの使い手であると同時に作り手でもあるという意識を持てるようにし、アーカイブの作成と利活用の循環を生み出すことが可能となる。市民に見えるよう地域がボトムアップの形で関与することが地域の活性化にも繋がるほか、アーカイブされたコンテンツについても、様々な情報が付加されて新たな価値が生まれてくる。特に様々な活用事例が周知されることにより、地域コミュニティにおいて無価値とされていた資料の価値について再考され、それが他のコンテンツのメタデータの整備やデジタル化に繋がるといった循環が期待される。

また、例えば、自然史・理工学系のデータの多くは、一点ではあまり価値は無いが、大量に集めることで価値が出るという性質がある。それを一人で集めていくことは不可能であるが、皆で同じ粒度のデータを集めることで価値あるデータを生み出すことが可能となる。実際に、欧米やオーストラリアなどでは、市民科学(シチズンサイエンス)の活動によって、生物・自然観測のデータ収集が活発に進められている。

現在、市民参加型デジタルアーカイブの作成環境を提供するソフトウェアもオープンソースで開発・公開されつつあり、これらをうまく活用することでデジタルアーカイブのエコシステム(循環)を生み出していくことが考えられる。過度なコストをかけずに、適時適切なデジタルアーカイブが構築される、いわばエコシステムといえるような仕組みを作ることができれば、既存アーカイブの長期運用のみならず、新たなデジタルコンテンツの作成や発信、利活用も含めた取組を持続的なものとするのが可能と考えられる。国や地方自治体には、市民参加による地域アーカイブの構築や活用の取組、そのための個々の継続的な活動を後押しすることが求められている。

(具体的な利活用のシナリオ)

デジタルアーカイブの利活用を促すためには、単にデータを集約するだけでなく、その使い方(シナリオ)を具体的に示すことや、利活用しやすいプラットフォームでデータを公開することを検討することが必要である。その際、単にデータを公開するだけでなく、よりユーザに身近でかつ意外性のある見せ方を工夫することが重要である。例えば、江戸料理レシピを料理レシピサービス「クックパッド」で写真付きで公開することで、予想もしなかった大きな反響があったという事例がある。また、民間事業者や団体、大学等においても利活用モデルの模索を行っていることが多く、産学官での協力・連携も検討していく必要がある。

その他、個別分野における利活用のポイントは以下のとおりである。

・教育分野での利用を促進するためには、メタデータに学習指導要領に記載されてい

る文言や教科書の見出しを付与するなど、単元との対応関係を明確にし、教員に
使ってもらいやすくする工夫をすると、効果的である。

- ・観光分野においては、地域のコンテンツをデジタル化し、観光に生かす取組は既に
各地で始まっている。提供が容易なデータから漸次オープン化することによって、
観光客の誘客から地方創生に繋げていくことが可能となる。
- ・地域における利用に当たっては、市民を巻き込んでデジタルアーカイブを構築してい
くプロセスを模索することが必要である。
- ・防災利用の観点からは、災害時だけでなく、平時においても、デジタル情報を共有・
利活用し、状況認識の統一を図っておくことが必要である。
- ・その他、デジタル化された歴史資料等を回想法²⁶に利活用することも考えられる。

（今後の活用の可能性）

現在審議が進められている著作権法の柔軟な権利制限規定の下では、所在検索サ
ービスや情報分析サービスなどの利用が可能になる。これらのサービスが権利制限規
定の創設により展開しやすくなれば、デジタルアーカイブの利活用の幅が一層広がるこ
となるため、その動きを注視する必要があるとの意見があった。

例えば、文字情報を含む画像のテキストデータ化によって、その検索利便性は向上
する。また、ジャパンサーチ(仮称)で集約されるメタデータに含まれるコンテンツの所在
情報から大量のデジタルコンテンツにアクセスし、AI がそれらのコンテンツを集約・解析
することで、新たなデータを生み出したり、時代や地域を超えた関連性の高いコンテン
ツの発見をしたりすることも可能となる。こうしたことにより、ストック化(死蔵)してい
たデジタルコンテンツをフローに戻すこと(活用)がより容易になると考えられる。

（5）人的基盤の整備

デジタルアーカイブの利活用を進めるためには、自館の所蔵コンテンツに付加価値を
見出し、発信するキュレーターや、各地域の事情を分かった上でデジタルアーカイブを
構築し、国際標準とすり合わせていくことのできる人材の育成が必要不可欠である。こ
うした人材には、デジタルアーカイブの構築に関する知識はもちろんのこと、コーデ
ィネートや企画、マネジメントに関する能力、著作権法や個人情報保護法、さらには、
人権やプライバシーの権利等に対する理解も求められており、一朝一夕で養成でき
るものではない。

アーカイブ機関の外側にいる第三者の専門人材を活用する仕組みや、経験を積んだ
スタッフを一定程度広域的に活用する仕組みなども含め、多様な方策を検討する必要
がある。また、地域のアーカイブにおいては、アーカイブ機関だけでなく、大学や企業、

²⁶ 認知症やうつ病のリハビリテーションの一つとして活用する方法で、過去の出来事や思い出を想起したり、語ったりすることで病状の進行を遅らせることを目的とするもの。

NPO などが連携してその地域固有の課題と一緒に取り組んでいくことも考えられる。こうした取組と AI などの新技術を活用することによって、限られた人的リソースを適切な業務²⁷に振り向けることが可能となり、現場の負担軽減にも役立つと考えられる。

(6)意識啓発

多くのアーカイブ機関では、来館者数や貸出し数を業績評価における重要な成果指標としており、所蔵品の高精細な写真や動画といったデジタルコンテンツをインターネット上に公開すると、来館者数が減るのではないかという懸念があると指摘されている。しかし、デジタル時代においては、むしろ、優れたコンテンツを発信することにより、多くの情報の中に埋没することなく文化的資産を継承・保存・活用していくことが可能となるのであり、所蔵品やアナログ情報を適切に保存するだけでなく、デジタル情報を保存、発信していく重要性について認識を共有することが必要である。「デジタルアーカイブ」ということばが我が国とは異なる意味で使われることのある海外においては、「デジタルキュレーション」や「デジタルプリザベーション」といったことばで、貴重な文化的・学術的デジタル資源の収集と提供、将来に向けた保存が表されることが多い。貴重な資源の収集、提供、保存はあらゆるアーカイブ機関に共通のミッションであり、デジタルアーカイブの開発と運営はそのミッションをデジタル化された現代の情報環境の中で遂行していくことに他ならないという意識を持つことが必要である。このような認識を周知していくためにも、ガイドラインを効果的な形で普及させ、アーカイブ機関の現場での活用を促す必要がある。

(7)評価の枠組みの整備

来館者数やアクセス数等の従来型の指標だけでは、デジタルアーカイブの取組が適正に評価されないため、関係省庁等連絡会・実務者協議会では、評価指標に関しても議論が行われた。

本委員会は、デジタルアーカイブの構築・共有・活用のための活動に関して、組織的な取組からシステム面も含めてバランスよく自己点検・評価するための指標・ツールについて検討し、「デジタルアーカイブアセスメントツール」(補足資料参照)を作成した²⁸。このツールは、各組織において求められる水準が異なることを踏まえ、項目ごとに三段階のモデルを用意し、アーカイブ機関やつなぎ役が、自らのミッションや役割を認識したうえで、必要な項目を選べるようにしている。このツールを使って、自らの立ち位置を確認することが可能となることが考えられる。

²⁷ 従来からの業務だけでなく、デジタルアーカイブの構築・発信などの業務についてもより積極的に行うことが可能となると考えられる。

²⁸ 補足資料「デジタルアーカイブアセスメントツール」参照。

標準モデル：小規模な機関で行うことが推奨される水準

先進モデル：各アーカイブ機関がそれぞれのミッションや事業に照らして目指す水準

つなぎ役モデル：つなぎ役の役割を持つ機関が目指す水準

今後、このツールの改善の必要性を確認しつつ、本委員会において、引き続き普及に向けた検討を行うとともに、内容の見直しが必要になった場合には適宜検討を行っていく必要がある²⁹。

5. 国の分野横断統合ポータル構築

2020年を目標に、国の分野横断統合ポータルとして「ジャパンサーチ(仮称)」を国立国会図書館が中心となって構築するために、本委員会で検討が行われた方向性と、今後一層の連携促進を図る上での課題を示す。

(1) 連携の方針

国の分野横断統合ポータルであるジャパンサーチ(仮称)とアーカイブ機関との連携は、原則として、つなぎ役を通じて連携することが望ましい。ただし、現時点においてメタデータ集約を行うつなぎ役が明確ではない分野³⁰・地域が多い。そこで、それらの分野・地域については、次の条件に当てはまるアーカイブ機関との直接連携を検討することとする。

- ・国の機関であり、当該分野におけるコンテンツを幅広くカバーしているアーカイブ機関
- ・公益に資する目的のため、当該分野におけるコンテンツを幅広くカバーしているアーカイブ機関
- ・唯一性・独自性の高いコンテンツ群を塊として扱う分野・地域を代表するアーカイブ機関
- ・その他(本実務者検討委員会において適当と認められるアーカイブ機関)

(2) つなぎ役の明確化と支援

誰がつなぎ役であるか明確化することは、ジャパンサーチ(仮称)の構築に向けて大変重要な課題であると同時に、各分野・地域におけるデジタルアーカイブの構築・共有・活用の促進のためにも、必要な取組である。

つなぎ役には、ジャパンサーチ(仮称)と各アーカイブ機関を結ぶこと(メタデータの集約)以外にも、次のとおり、メタデータの標準化やオープン化の推進、分野内における人的基盤の構築、法的な課題への対応など様々な役割が求められている。

²⁹ 本論点に関する議論の中では、デジタルアーカイブに係る基本法を整備することで、デジタルアーカイブ構築・利活用の推進を図るとともに、アーカイブ機関内での取組促進を促すべきとの意見もあった。

³⁰ 「知的財産推進計画 2017」に記載があるのは、書籍等、文化財、メディア芸術、放送番組のみ。

表1：つなぎ役の役割一覧

- (ア) 分野/地域の独自性を反映したポータルを整備・提供
- (イ) (分野/地域における、以下同)メタデータの集約、API 提供
- (ウ) メタデータの整備推進
- (エ) メタデータの標準化、用語の統制(辞書・典拠・シソーラス)
- (オ) デジタルコンテンツ等の二次利用条件³¹の整備・オープン化の推進
- (カ) 所蔵資料/収蔵品等のデジタル化のための技術や法務上の業務支援
- (キ) コンテンツの長期保存・永続的アクセス保証(データホスト)への協力
- (ク) 意識啓発・人材育成
- (ケ) 活用促進のための取組

※『我が国におけるデジタルアーカイブ推進の方向性』(デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会、平成 29 年4月、p.21)より抜粋

つなぎ役は一つの機関が全てを担う必要はない。行政が中心となって産学官が連携して役割を整備し、つなぎ役を支援していくことが求められる。特に地域のつなぎ役については、地方自治体が主体的に推進することが重要であり、社会教育施設や大学の役割と併せて考えていくという観点も必要との意見があった。

具体的には、つなぎ役は、実務者検討委員会での議論等を踏まえながら、毎年の「知的財産推進計画」の中で明確化していくことも考えられる。

(3) 共通メタデータフォーマット³²の策定

ジャパンサーチ(仮称)にデータを提供する機関の作業負担をできる限り軽減しつつ、さまざまな分野の多様なデータの利活用を促進することを可能にするために考えたものが、ジャパンサーチ(仮称)の共通メタデータフォーマットである。

共通メタデータフォーマットには、ジャパンサーチ(仮称)との連携のためのフォーマット(連携フォーマット)と、集約されたメタデータを利活用するためのフォーマット(利活用フォーマット)の二つがある(補足資料「共通メタデータフォーマット」参照。)。全体のデータの流れのイメージは図 3 のとおりである。

³¹ 平成 29 年4月「デジタルアーカイブの構築・共有・活用ガイドライン」デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会 15～16 ページにおいて、既に利用条件の基本的な考え方は示されているため、これを前提として議論を進める必要がある。

³² 「共通メタデータフォーマット」とは、これまで、知的財産推進計画2017において「標準メタデータフォーマット」と呼称していたジャパンサーチ(仮称)上で利用するメタデータのフォーマットと同義であり、各分野・地域共通で使うメタデータフォーマットであることを明確にするために、今後は、この呼称を使用することにする。

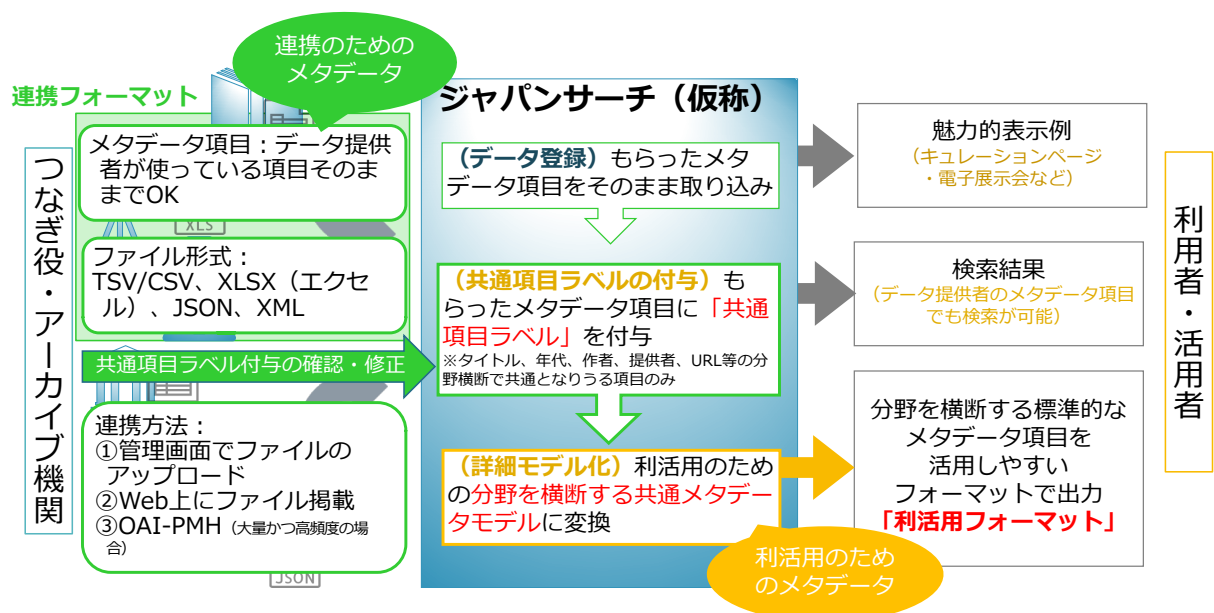


図3 ジャパンサーチ(仮称)におけるメタデータの連携と利活用まで

(連携フォーマット)

多種多様なメタデータ連携フォーマットのメタデータ項目については、各データ提供機関が持っているメタデータ項目をそのままの形でジャパンサーチ(仮称)に提供可能とする。必須項目は管理用のIDと名称/タイトルだけとし、その他は任意とする。ただし、名称/タイトルのヨミや英語名、最終更新日、URL等のデータをもっている場合は、その項目に共通項目ラベル(表2参照)を付与し、それらを同じラベル名で検索することを可能にする。表2に挙げた項目以外の項目は、各アーカイブ機関が示した項目名をそのまま用いることとなる。連携フォーマットについては補足資料を参照のこと。

なお、ファイル形式は、様々な形式に対応する。連携方法は、ジャパンサーチ(仮称)で用意する管理画面からファイルをアップロードする方法や、ウェブ上にファイルを掲載する方法を利用することを主として想定する。

(利活用フォーマット)

多種多様なメタデータをまとめて利活用するに当たって、シンプルで使いやすいものがあると同時に必要十分な情報を提供するためにはどのようなモデルがよいかといった視点から、Europeana等の海外事例も踏まえて、利活用フォーマットを検討した。

ジャパンサーチ(仮称)では、連携フォーマットによる連携後に、データ提供機関から提供されたメタデータ項目の一部について、ジャパンサーチ(仮称)のシステム側で、分野共通に利用可能なメタデータモデルに変換し、「利活用フォーマット」で出力できる仕組みを用意する。利活用フォーマット の概念整理及び実際のフォーマットについては補足資料を参照頂きたい。

なお、利活用フォーマットの主要項目は表 3 のとおりである。未だ分野の標準的なメタデータフォーマットがない場合に、どのような情報をメタデータとして整備する必要があるかを検討するに当たって、参考になるものである。分野の必要に応じて検討することが望ましい。

表 2: 連携フォーマットの共通項目ラベル一覧

	項目名	内容
必須項目	ID	提供元データ内での一意の ID(管理番号)
	名称/タイトル	レコードの名称。検索結果表示に使われる
あれば必須で付与する項目	名称/タイトルヨミ	レコードの名称の読み(カタカナ・平仮名共)
	名称/タイトル英語	レコードの英語名称又はローマ字
	最終更新日	データの最終更新日(日付型)
	URL	レコードのリンク先の URL(つなぎ役と提供元共)
	サムネイル画像 URL	サムネイル画像の URL
	コンテンツ URL	デジタルコンテンツの URL
	提供者	オリジナルのコンテンツの提供者
	コンテンツの権利表示	データベース定義の情報と異なる場合のみ
	寄与者(人物) ※	作者、発行者、出演者など
	時間(日付) ※	制作年、対象時期など
場所 ※	発行地、制作地など	

※あれば原則付与するが、分野の特性に応じて付与しないこともありうる。

表 3: 利活用フォーマットの主要項目一覧

基本項目	内容
タイプ	コンテンツの基本区分(書籍、文化財など大きく情報を区分する枠)
名称	タイトル、別名、読みなど検索対象とする名前
寄与(者)関係	コンテンツに寄与した人/組織(作者、発行者、出演者など)
場所関係	場所に関する情報(発行地、制作地など)
時間関係	時間に関する情報(制作年、対象時期など)
主題・区分	主題・分類/各分野のキーワード的共通認識のある区分(国宝、ドキュメンタリーなど)
識別子	コンテンツを特定するための識別子(ISBN など)
言語	コンテンツの記述言語
サムネイル画像	コンテンツの特徴を確認するための画像(提供元とは別にサムネイルを保持する場合)

6. 今後の主要検討課題

以下の課題については、実務者検討委員会において、引き続き検討を行っていくこととする。

- (1) 利活用モデルの模索やそのための制度的課題の整理
- (2) 新技術を活用したデジタルアーカイブの構築の在り方の検討
- (3) 長期利用保証の在り方の検討
- (4) ジャパンサーチ(仮称)上での共通メタデータフォーマットを踏まえた、各分野におけるメタデータの在り方の検討
- (5) 各分野・地域におけるつなぎ役の役割や分担の明確化、つなぎ役に対する国の支援策の検討
- (6) メタデータ等のオープン化の実施、コンテンツの二次利用条件表示の促進策の検討(望ましい権利表記の共有等)

おわりに

デジタルアーカイブ社会の実現により、文化の保存・継承のみならず、観光、地方創生、教育研究、ビジネスへの利用など、様々な効果が期待できるが、それに向けた本格的取組は始まったばかりである。

本実務者検討委員会においては、今後も、引き続き、6に掲げた課題を中心に、課題整理や、その解決に向けた対応策について議論を進め、各アーカイブ機関やつなぎ役の取組が円滑に進むよう必要な施策の方向性について検討をしていくこととする。

補 足 資 料

連携フォーマット（案）について

基本的な考え方

◆連携フォーマット

- データを提供する連携機関のメタデータモデルをそのままの形（オリジナルのモデル）で受け入れる。
- 共通項目ラベルを付与することで、分野横断的な表示・検索を可能にする。

◆連携仕様

- 原則、ファイルベースでの連携を行う（OAI-PMHを必須としない。）
- 多様なファイル形式に対応する。

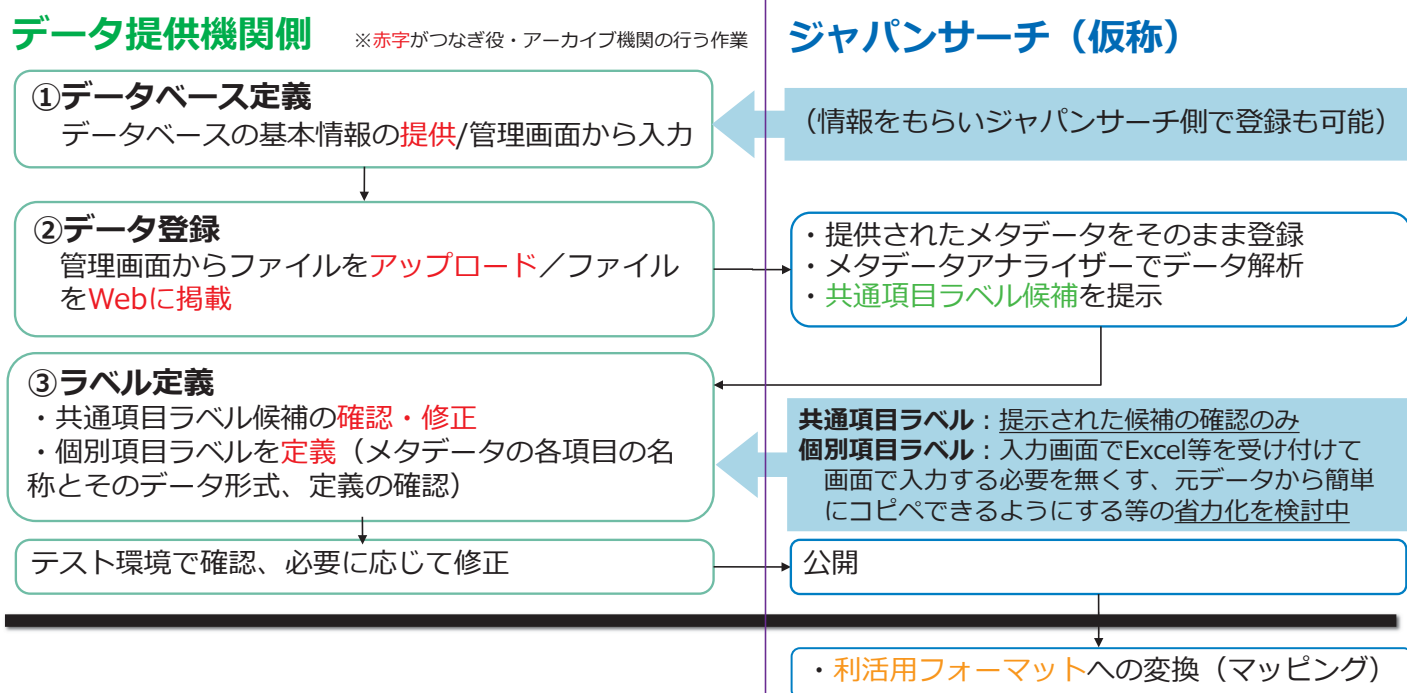
◆利活用フォーマット

- メタデータの利活用促進のため、分野共通で利活用可能なメタデータモデル（利活用フォーマット）を定義する。
- メタデータは、オリジナルモデルと利活用フォーマットの両方で保持する。



多様なデータ形式に対応、連携機関の作業負担の軽減を図りつつ、充分利活用できる仕組みを用意する

データ連携の流れ（イメージ）



連携フォーマット ①データベース定義

◆データベース基本情報の提供

必須項目	ID	データベースのID（ジャパンサーチ側で付与）
	名称（日/英）	データベースの名称
	タイプ	データベースが扱うコンテンツの分野。選択式
	メタデータの権利表示	例外の場合のみ記載。 原則は、CC0（著作権のあるもののみCC BYでも可）
	コンテンツの権利表示	対象のデジタルデータの権利情報。テキスト
	コンテンツの権利区分	選択式のコンテンツの権利情報
	組織名（日/英）	データベースの所有者の名称
任意項目	説明（日/英）	データベースの説明
	URL	データベースのURL
	組織URL	データベースの所有者のURL
	カテゴリ（サブタイプ）	データベースが扱うコンテンツの主題。テキスト

連携フォーマット ②データ登録

◆メタデータ項目

メタデータ項目は自由だが、最低限の必須項目（ID、タイトル等）がある（⇒詳細は「③ラベル定義」参照）

◆連携方式 ※OAI-PMHは、大量かつ高頻度の場合に想定

ファイルのアップロード		管理画面から、手動でファイルをアップロードする方法
ファイルをWebに掲載	ファイル取得	Web上にファイルを掲載し、管理画面でそのURLを指定すると、ジャパンサーチ（仮称）側がファイルを取得しに行く方法
	ファイル定期取得	上記のファイル取得を、指定した日時に自動実行し、データを自動更新する方法

◆対応ファイル形式

TSV・CSV	文字コードはUTF-8。同じデータ項目であれば複数ファイルも可。
XLSX	複数シート・結合セルには対応できない。XLS非対応。
JSON	1行1レコードのJSON Lines形式を推奨。他の構造にも対応予定。
XML	1行1レコード形式を推奨。他の構造にも対応予定。

連携フォーマット ③ラベル定義

◆共通項目ラベルを付与する項目

必須項目	ID	オリジナルデータ内での一意のID（管理番号）
	名称/タイトル	レコードの名称。検索結果表示に使われる
あれば必須で付与する項目	名称/タイトルヨミ	レコードの名称の読み（カタカナ・平仮名共）
	名称/タイトル英語	レコードの英語名称又はローマ字
	最終更新日	データの最終更新日（日付型）
	URL	レコードのリンク先のURL（つなぎ役と提供元共）
	サムネイル画像URL	サムネイル画像のURL
	コンテンツURL	デジタルコンテンツのURL
	提供者	オリジナルのコンテンツの提供者
	コンテンツの権利表示	データベース定義の情報と異なる場合のみ
	寄与者（人物）※	作者、発行者、出演者など
	時間（日付）※	制作年、対象時期など
場所 ※	発行地、制作地など	

※あれば原則として付与するが、分野の特性に応じて付与しない場合もありうる。

◆個別項目ラベルの定義

（オリジナル）ソースデータの項目名とそのデータ形式、定義の確認を行うのみ。

ジャパンサーチ（仮称）の連携フォーマット（案）詳細版

目次

1	基本的な考え方.....	28
1.1	連携フォーマット	28
1.2	連携仕様.....	28
1.3	利活用フォーマット.....	28
2	連携仕様.....	28
2.1	メタデータ項目	28
2.1.1	各メタデータ項目の値	28
2.2	連携方式.....	29
2.3	ファイルフォーマット	30
3	連携に係る作業の流れ.....	31
3.1	データベース定義	31
3.2	データ登録	32
3.3	ラベル定義	33
3.3.1	共通項目ラベル付与	33
3.3.2	個別項目ラベル定義	34
3.4	データ登録、公開	34

1 基本的な考え方

1.1 連携フォーマット

- ・ ジャパンサーチ（仮称）（以下「ジャパンサーチ」とする）は、連携のための単一のメタデータモデルは定義せず、データを提供する連携機関のメタデータモデルをそのままの形（オリジナルのモデル）で受け入れる。
- ・ 受け取ったメタデータは、ジャパンサーチが実装する管理ツールを用いて処理することで、全項目を対象とする単純な検索ができるようにする。
- ・ 異なるデータベースのメタデータの一覧表示や横断検索を可能にするため、「共通項目ラベルの付与」を行う。作業負荷を軽減するため、自動で候補を推定するとともに、必要最低限の項目に絞るものとする。

1.2 連携仕様

- ・ 連携を容易にするため、原則、ファイルベースでの連携を行う（OAI-PMH は必須としない）。
- ・ 多様なファイル形式に対応する。

1.3 利活用フォーマット¹

- ・ メタデータの利活用促進のため、オリジナルのメタデータモデルとは別に、分野共通で利活用可能なメタデータモデル（＝利活用フォーマット）を定義する。
- ・ メタデータは、利活用のためのモデルに変換するが、オリジナルのモデルも保持する。
- ・ 利活用メタデータモデルの生成のためのマッピング作業は、連携後に、国立国会図書館が行う。必要に応じて連携機関の確認を得る。

2 連携仕様

2.1 メタデータ項目

連携機関のメタデータモデルをそのままの形で受け入れる。メタデータモデルは原則として自由であるが、最低限の必須項目（①オリジナル（ソース）データの一意の ID、②名称/タイトルの 2 項目）がある等の制約を設ける。

必須項目の詳細については、3.3.1 参照。

2.1.1 各メタデータ項目の値

各メタデータ項目の値は、下表にあるデータ種別に対応する。基本は全て文字列として扱われるが、ラベル定義（3.3 参照）の際にデータ種別を指定することも可能とする。

¹ 利活用フォーマットについては、補足資料参照。

表1 メタデータ項目として対応可能な値

データ種別	内容
文字列	文字列としての検索が可能。改行については、「 <code>\n</code> 」にエスケープされ、UI に表示される際には BR タグに変換される。項目名の最後に「 <code>_s</code> 」が追加される。
HTML	文字列のヴァリエントで、UI で表示される時に HTML として表示される。XSS を防ぐため以下のタグ・属性しか利用できない（データ登録時に消去される）。 H、P、BR、DIV、SPAN、TABLE、TR、TH、TD、attr:style 項目名の最後に「 <code>_h</code> 」が追加される。
URL	リンクとして表示される URL。項目名の最後に「 <code>_u</code> 」が追加される。
日付	日付として変換される（UnixTime として保持する）。変換に失敗した場合、null になる。変換対応フォーマットは別途用意する資料（技術情報を集約した詳細版を作成予定）に示す。 項目名の最後に、「 <code>_d</code> 」が追加される。
真偽値	真偽値。JSON の true/false に対応。「true」「TRUE」「1」が true として、それ以外の全ての値は false として解釈される。 項目名の最後に、「 <code>_b</code> 」が追加される。
数字	数字（小数点含む）。JSON の number に対応。数字に変換できないエラーデータが入っている場合、0 が入る。 項目名の最後に、「 <code>_n</code> 」が追加される。

2.2 連携方式

ファイルによる連携を基本とし、表2のと通りの連携方式の選択を可能とする。ファイル連携の場合は、連携・更新の都度、全データが格納されたファイルを用いて連携を行う。

表2 ファイルの連携方法

方法	説明
①ファイルアップロード	対応可能なファイルフォーマット（2.3 参照）のファイル（20MB まで）を、管理画面から手動でアップロードする方式。
②ファイルを Web に掲載し、ジャパンサーチがファイル取得する	管理画面でファイルの URL を入力し、ジャパンサーチから取得しに行く方式。ベーシック認証、ダイジェスト認証にも対応可能。応答の形式が対応ファイルフォーマット（2.3 参照）であれば、API の URL でもよい。
③ファイルを Web に掲載し、ジャパンサーチがファイルを定期取得する	ファイル取得と同様だが、ジャパンサーチ側で定期自動実行する。実行時間の指定が可能。
④ハーベスト用 API (OAI-PMH)	データが大容量、かつ高頻度で更新を行う場合の連携方法を想定。※ただし、試験公開版での実装は想定しない。

2.3 ファイルフォーマット

TSV、CSV、XLSX の表形式ファイルフォーマット²、及び JSON、XML の構造化ファイルフォーマットに対応する。

また、複数のファイルを zip 形式で圧縮したファイルにも対応する。(なお、指定されたファイルフォーマット以外のファイルが含まれている場合には、エラーとなる)。

表3 対応ファイルフォーマットの種類と制約

ファイル形式	制約
TSV・CSV	UTF-8 でエンコーディングされたファイルのみ受け付ける。 CSV の場合、RFC4180 に準拠していれば問題はなく、ダブルコーテーション (") で囲まれた範囲であれば改行を含んでもよい。 TSV についても、エスケープ等は CSV に準拠する。
XLSX	1 ファイルにつき 1 シートのみ (複数シートには非対応)。セルの結合等にも対応できない。XLS 非対応のため、古いファイル等に注意が必要。
JSON	1 行 1 レコードとする JSON Lines ³ 形式を推奨する。 (例) JSON Lines {"id":"0001", "title":"タイトル 1", ... } {"id":"0002", "title":"タイトル 2", ... } 但し、ルートをお array とし、1 レコード 1 オブジェクトとする形式、ルートを object とし、key:Value の Value を 1 レコードとする形式にも対応可能とする。 なお、JSON ファイルは原則としてそのままの形で問題ないが、以下の制約がある。 ・ pair の key に使えるのは、アルファベット小文字、数字、アンダーバー (_) のみ。それ以外の文字が含まれる場合、削除される。 ・ 1 つの array の中に、異なる種類の Value が存在してはならない。
XML	1 行 1 レコードとする形式を推奨する。 (例) 1 行 1 レコードの XML <root><id>001</id><title>タイトル 1</title>...</root> <root><id>002</id><title>タイトル 2</title>...</root> 但し、XPath 等で、ルートの下に複数のレコードを続けていく方式にも対応可能とする。

表形式ファイルフォーマット (TSV・CSV、XLSX) では、列を項目、行をメタデータの単位として解釈する。1 行目はヘッダとして指定することを可能とする (アルファベット小文字・数字のみ使用可、重複不可)。これらの項目名は、ジャパンサーチのシステム内部でのみ利用され、検索結果詳細画面等に表示される項目の名称は、ラベル定義 (2.3 参照) で

² 表形式のデータを作成する際のポイントについては、『デジタルアーカイブの構築・共有・活用ガイドライン』(平成 29 年 4 月) p.39 を参照。

https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline.pdf

³ <http://jsonlines.org/>

別途定義する。

構造化ファイルフォーマット（JSON、XML）では、その構造を原則としてそのまま取り込むこととする。入れ子構造にも対応するが、検索結果詳細画面等で、その入れ子の単位で表示することはできず、入れ子を展開したフラットな形式（「key：値」）での表示とする。外部 API で取得した場合は、入れ子構造での表示も可能とする。

3 連携に係る作業の流れ

ジャパンサーチとの連携作業は、大まかに、(1) データベース登録・基本情報の提供→(2) データ登録→(3) ラベル定義（共通項目ラベル付与、個別項目ラベル定義）→(4) 公開の手順に沿って行われる。作業分担は以下の表のとおり。

表 4 登録から公開までの流れ

作業内容	連携機関	ジャパンサーチ
(1) データベース登録	・管理画面から、データベースの基本情報を入力	(情報をもってジャパンサーチ側で登録も可能)
(2) データ登録	・管理画面からメタデータファイルをアップロード又はファイルを Web に掲載	
(3) ラベル定義	・共通項目ラベル候補の確認・修正 ・個別項目ラベル定義（メタデータの各項目の名前、データ形式、定義の確認のみ）	・（メタデータアナライザ ⁴ による）データの自動解析により、共通項目ラベル候補を提示
(4) 公開	・テスト環境での確認・修正	・公開（公開後、利活用フォーマットへの変換作業）

3.1 データベース定義

連携機関には機関 ID・パスワードが発行され、管理画面へのアクセスが可能になる。連携機関は、この管理画面から、ジャパンサーチと連携するデータベースの基本情報について、入力を行う。これをデータベース定義という。ジャパンサーチ側で基本情報をもって代わりに入力することも可能とする。

設定が必要な項目は表 5 の通り。

⁴ ジャパンサーチ側のシステムにおいて、データ種別や項目の充足率等から共通項目ラベル（名称/タイトルや ID 等）の候補を推定し、提示する仕組み。

表5 データベース定義のための項目一覧

項目名	種別	意味	制約
ID	必須	データベースの ID	アルファベット小文字、数字で4文字 (ジャパンサーチ側で付与)
名称	必須	データベースの名称	
名称 (ヨミ)	必須	データベースの名称のヨミ	DB 一覧の表示順のため
名称 (英語)	必須	データベースの名称 (英語)	
説明		データベースの説明	100 字まで (100 字以上は折り畳み表示)
説明 (英語)		データベースの説明 (英語)	
タイプ	必須	データベースが扱うコンテンツの種別。選択式。	データベース当たり一つが推奨だが、複数選択も可
カテゴリ (サブタイプ)		データベースが扱うコンテンツの種別。選択式のタイプを補うための情報。テキスト。	
メタデータの権利表示	必須	メタデータの権利情報について、例外的場合のみ記載。原則は、CC0 (著作権性のあるもののみ CC BY でも可)	クリエイティブ・コモンズライセンス、政府標準利用規約等。権利情報について記述されている外部資源へのリンク (URL) も可
コンテンツの権利表示	必須	対象のデジタルデータがある場合の権利情報。例外がある場合は (メタデータの) 共通項目で定義する。	
コンテンツの権利区分	必須	コンテンツの権利情報だが、検索・絞込み用に選択式になっている。	選択項目としては、クリエイティブ・コモンズライセンスのバリエーションその他が考えられるが、要検討。
URL		データベースの URL	
組織名	必須	データベースの所有者の名称	
組織名 (英語)	必須	データベースの所有者の名称 (英語)	
組織 URL		データベースの所有者の URL	

3.2 データ登録

連携機関は、任意の連携方式 (2.2 参照) によってメタデータファイルを登録する⁵。

⁵ 登録されたデータは、ジャパンサーチのシステム側で JSON Lines 形式に変換される。

3.3 ラベル定義

連携機関は、管理画面を通じて、共通項目ラベル・個別項目ラベルの定義を行う。この定義作業を行うことで、ジャパンサーチ上での一覧表示や横断検索が可能になる。

3.3.1 共通項目ラベル付与

3.2 で登録されたメタデータは、ジャパンサーチのシステム（メタデータアナライザー）により自動的に分析され、「共通項目ラベル」候補が自動的に提示される。連携機関は、提示された候補が適切かどうかを確認し、必要に応じて管理画面で修正を行う。

表6 共通項目ラベル一覧

項目名	種別	意味	制約
ID	必須	オリジナル（ソース）データの一意のID。レコードのURIに使われる。	オリジナル（ソース）データ内で一意であること。 アルファベット大文字小文字、数字のみで構成されていること。
名称/タイトル	必須	レコードの名称。検索結果の表示に使われる。	
名称/タイトルヨミ	あれば必須	レコードの名称の読み。	カタカナ・平仮名問わない
名称/タイトル英語	あれば必須	レコードの英語名称又はローマ字	
最終更新日	あれば必須	データの最終更新日	日付型であること。
URL	あれば必須	レコードのリンク先のURL	つなぎ役のメタデータ表示先、提供元のメタデータ表示先の両方共
サムネイル画像URL	あれば必須	サムネイル画像のURL	
コンテンツURL	あれば必須	デジタルコンテンツがある場合にそのURL	
提供者	あれば必須	オリジナルのコンテンツの提供者	当面は、ID等では無く、文字列とする。
コンテンツの権利表示	あれば必須	対象のデジタルデータがある場合、その権利情報が、データベース定義の情報と異なる場合のみ	クリエイティブ・コモンズライセンス、政府標準利用規約等。権利情報について記述されている外部資源へのリンク（URL）も可
寄与者（人物）※	あれば必須	対象の作成に関わった人（作者、発行者、出演者等）	複数可
時間（日付）※	あれば必須	対象に関連する時間（制作年、対象時期等）	複数可
場所※	あれば必須	対象に関連する場所（発行地、制作地等）	複数可

共通項目ラベルとは、ジャパンサーチと連携している全てのデータベースに共通するメタデータ項目（名称/タイトル、提供者等）に付与するラベルである。ラベルを付与することで、検索結果表示が分かりやすくなる、検索の絞込みができるようになるなどのメリットがある。

共通項目ラベルには、「必須項目」（必須でラベルを付与する項目）、「あれば必須項目」（オリジナルのメタデータに対応項目がある場合に、必須でラベルを付与する項目）の 2 種類がある。具体的な項目は表 6 のとおり。

なお、※のものについては、該当するものがあれば原則として付与するが、分野の特性に応じて、付与しない場合もありうる。

3.3.2 個別項目ラベル定義

連携機関は、メタデータ項目全体について、各項目の内容をシステムが正しく認識できるよう、各メタデータ項目に関する名称や説明、インデックス方法、データ種別等の情報を付与する作業を行う。これを、個別項目ラベル定義という。

※個別項目ラベル定義に当たっては、入力画面で Excel 等のインプットファイルを受け付けて画面で入力する必要をなくす、元データからコピー&ペーストできるようにする等の運用を検討中。連携機関は、各メタデータ項目の名称、データ種別、定義の確認を行うのみでよいようにする。

表 7 個別項目ラベルの定義項目

項目名	種別	意味	制約
名称（日本語）	必須	詳細画面等に表示される、その項目の名称	
名称（英語）		英語画面に切り替えた時の名称。未定義の場合、検索結果は日本語が表示される。	
説明（日本語）		その項目が何を意味するかの説明。定義があれば、詳細画面でユーザが見ることが可能。	
説明（英語）		同、英語。	
格納種別/インデックス方法	必須	以下から選択（デフォルトは通常） 通常：項目が格納され、検索・表示され、公式 API でも提供される API を除く：通常と同様だが、API でデータを取得した際には項目が除外される。 非登録：項目はジャパンサーチに登録されない。	
データ種別	必須	2.1.1 のデータ種別から選択（デフォルトは文字列）	

3.4 データ登録、公開

連携機関は、ラベル定義の作業が終了すれば、テスト環境で実際のメタデータの提供状況を確認することができる。必要に応じて、連携機関による管理画面での修正又はジャパンサーチ側での修正後、一般公開の運びとなる。

利活用フォーマットについて

1. 基本データモデルの考え方
2. ソース情報分離モデル
3. 共通情報の主要項目
4. 扱いやすさと適切さのバランス
5. 寄与(者)関係の共通情報表示例
6. 時間関係情報
7. 時間関係の共通情報表示例
8. 場所関係情報
9. 場所関係の共通情報表示例
10. 提供＝アクセスのための情報
11. 提供/ソース関連の共通情報表示例
12. ライセンス情報の明示
13. 共通情報整備の課題

1 基本データモデルの考え方

■ ソース情報の分離と利用者のタスクに基づく共通情報

- 提供元からのソースデータは来歴情報明記の上でそのまま保持
- 利用者の4つのタスクの観点で共通情報を生成
 - ▶ **Find**：発見＝タイトル、キーワードなどによる検索（標目、アクセスポイント）
 - ▶ **Identify**：識別＝示されている対象が何なのか、求めている（既知の）コンテンツかどうか判断できる情報
 - ▶ **Select**：選択＝（検索結果などの）選択肢のうち、対象が求めている（未知の）コンテンツかどうかを判断できる情報。識別より広い意味での記述
 - ▶ **Obtain**：取得＝示されている（選択した）対象を取得する手段（の情報）の提供

■ 共通情報とタスク

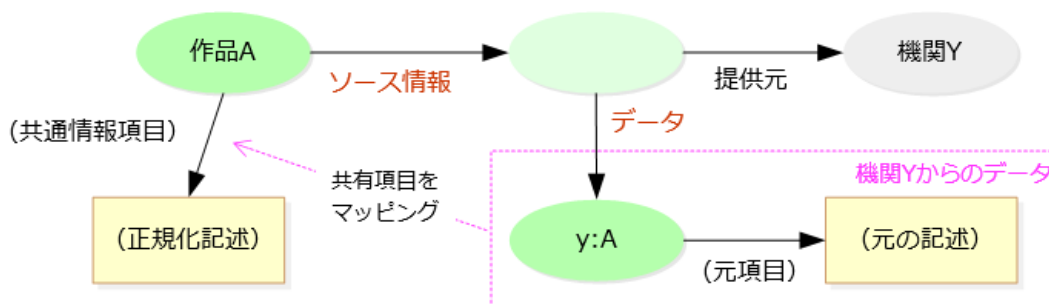
- いつ、どこで、だれが、何を（→特に発見タスク）を基本にシンプルな項目設定
- 単純プロパティと構造的プロパティの併置（識別、選択タスク）
- 提供＝アクセス情報の充実（取得タスク）
- アプリケーションによる活用も踏まえた情報形態（Linked Data）

※ジャパンサーチ利活用フォーマットの検討に関する詳細な情報は、下記参照。
<http://www.ndl.go.jp/jp/dlib/standards/jpsformat.html>

2 ソース情報分離モデル

■ 共通情報とソースデータ/情報

- 利用者タスクに基づき、共通利用や付加価値を持つ再利用のための項目をマッピング
- 共通情報項目の1つとして「ソース情報」を持ち、ソースデータをそのまま保持
 - ▶ ソースデータはJSONに変換して保持し、単純検索、全文検索などでも利用



■ Europeana、DPLAとの関係

- Europeana、DPLAともに収集データとそれを整理した共通データを分離して扱っている
 - ▶ Europeanaは提供元に一定の標準モデル準拠を求める。DPLAは複数形式を許容
- Europeanaは値の正規化、Linke Data化を重視。ジャパンサーチはより多様な活用が可能なモデルを工夫

3 共通情報の主要項目

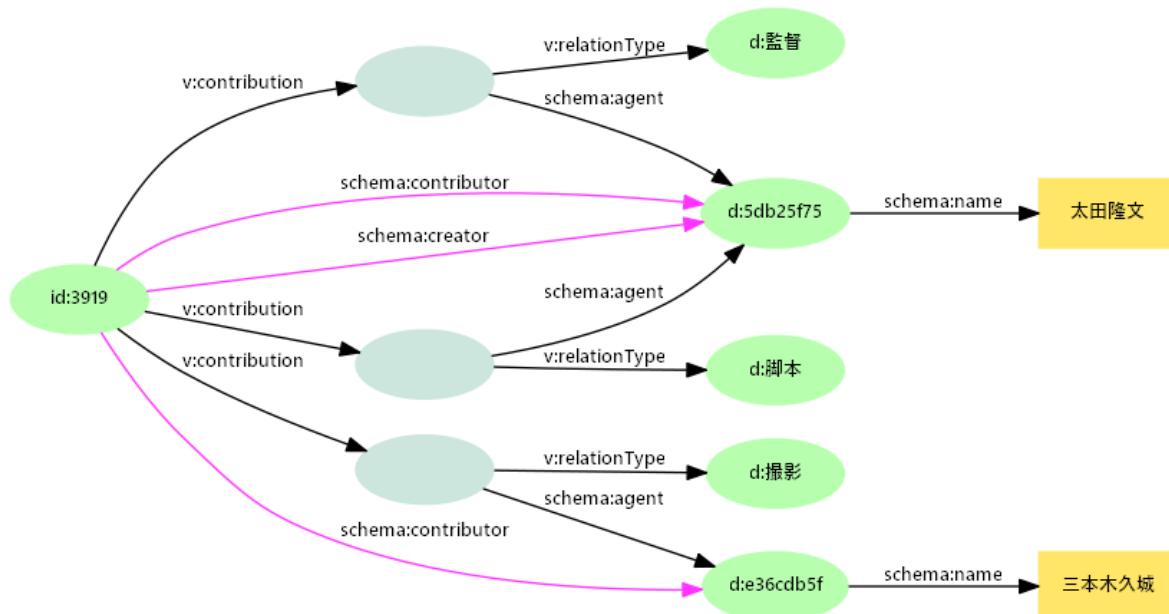
■ 共通情報の基本項目

基本項目	内容
タイプ	コンテンツの基本区分（書籍、文化財など大きく情報を区分する枠）
名称	タイトル、別名、読みなど検索対象とする名前
寄与(者)関係	コンテンツに寄与した人／組織（作者、発行者、出演者など）
場所関係	場所に関する情報（発行地、制作地など）
時間関係	時間に関する情報（制作年、対象時期など）
主題／区分	主題・分類／各分野のキーワード的共通認識のある区分（国宝、ドキュメンタリーなど）
識別子,言語	コンテンツを特定するための識別子（ISBNなど）、コンテンツの記述言語
サムネイル画像	コンテンツの特徴を確認するための画像（提供元とは別にサムネイルを保持）
記述	コンテンツの物理的特徴・素材等の記述、個別項目に収録できない情報
上位コンテンツ	当該コンテンツがその一部である上位コンテンツ（公文書などの資料階層）
提供情報	コンテンツの提供・アクセス、ライセンスに関する情報（構造化して記述）
ソース情報	ソースデータ（ジャパンサーチ(仮称)が連携フォーマットで受け取ったデータ）とその提供者に関する情報（構造化して記述）

4 扱いやすさと適切さのバランス

■ 単純プロパティと構造化プロパティの併用

- 監督、脚本=太田隆文；撮影=三本木久城というメタデータがある時
- **単純プロパティ**：作者は共通のプロパティ creator で調べたい。関与した人は一括して contributor で調べたい（発見）→原則としてSchema.orgを使用
- **構造化プロパティ**：監督、照明といった元データの項目情報も生かしたい（識別、選択）→必要に応じて独自語彙



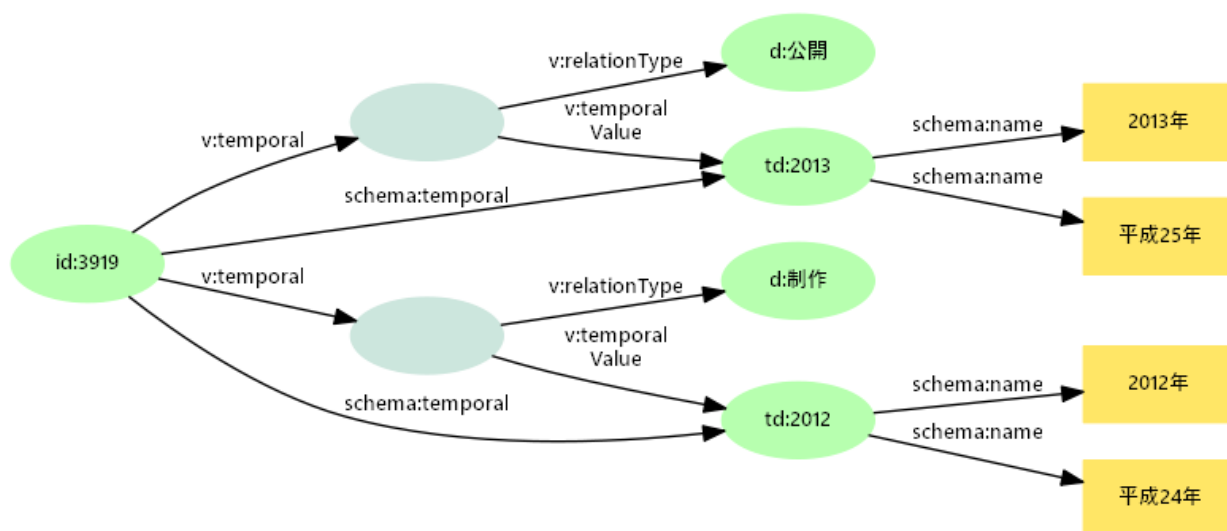
5 寄与(者)関係の共通情報表示例

v:contribution	schema:roleName	撮影	schema:agent	ndlent:00444960 (James Joyce)
	schema:agent	d:今村互	v:relationType	d:制作
	v:relationType	d:staffs		
v:contribution	schema:roleName	編集	schema:agent	ndlent:00040168 (丸谷オー)
	schema:agent	d:太田隆文	v:relationType	d:翻訳
	v:relationType	d:staffs		
v:contribution	schema:roleName	平田俊夫	schema:agent	ndlent:00049772 (永川玲二)
	schema:agent	d:並樹史朗	v:relationType	d:翻訳
	v:relationType	d:casts	schema:agent	ndlent:00078566 (高松雄一)
v:contribution			v:relationType	d:翻訳
			schema:agent	d:日本障害者リハビリテーション協会
			v:relationType	d:出版

6 時間関係情報

■ 「いつ」に関する情報を集約する

- 制作、収集、内容（対象時代）を区別せずに時間関係として扱う（発見）
- 構造化プロパティで関係の違いを示す（識別、選択）
 - ▶ 値は年を単位に正規化し、元記述が月日まで含むなどの場合は構造化情報に保持する



- 時間を文字列ではなくURIで表し、西暦と和暦、時代区分の対応などを「時間オントロジー」としてまとめる
 - ▶ 時間範囲も `td:1868_1913` のようなURIを介して開始年、終了年を構造化することで、単一年の値と同様に処理できる（DPLAでも時間範囲は構造化）

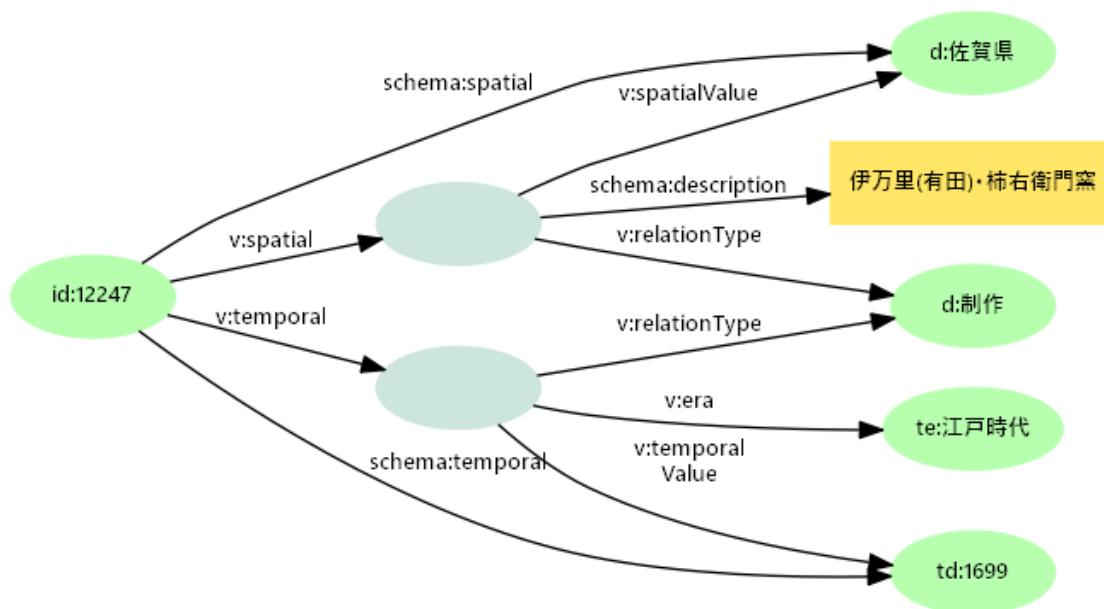
7 時間関係の共通情報表示例

v:temporal (2) ☒	v:relationType	d:公開 ☒
	v:temporalValue	td:2013 ☒ (⇒ 平成25年)
v:temporal	v:relationType	d:制作 ☒
	v:temporalValue	td:2012 ☒ (⇒ 平成24年)
v:temporal	v:era	
	v:temporalValue	td:1903 ☒ (⇒ 明治36年)
te:明治時代 ☒		
schema:name (2) ☒		"1868~1913年"
schema:name		"明治時代"
owl:sameAs		td:1868_1913 ☒
schema:description		放送開始日:2013/12/29
v:relationType		d:放送 ☒
v:temporalValue		td:2013 ☒ (⇒ 平成25年)

8 場所関係情報

■ 「どこで」に関する情報を集約する

- 時間軸と同じく、空間（場所）関係を一括して扱う（発見）
- 関係の違いは構造化プロパティで示す（識別）
- 同じ関係（たとえば「制作」）により、同じ関係を持つ時間、場所、関与者などが結びつく
- 場所を都道府県単位（国単位）で正規化→地域活性化のためのデータ利用



9 場所関係の共通情報表示例

v:spatial	v:relationType	d:制作	
	v:spatialValue	d:伊万里(有田)・柿右衛門窯	
v:spatial	v:relationType	d:出土	
	v:spatialValue	d:佐味田宝塚古墳 (奈良県北葛城郡河合町) 出土	
v:spatial	schema: description (2)	採集場所：都道府県 (日本語) :静岡県	d:関連地名
	schema: description	採集場所：詳細 (日本語) :梨本 南伊豆	d:加賀
	v:relationType	d:採集	d:関連地名
	v:spatialValue	d:日本	d:越中
v:spatial	v:relationType	d:関連地名	
	v:spatialValue	d:能登	

10 提供 = アクセスのための情報

■ コンテンツにアクセスするための情報

- 提供者（保管者）の情報
- 提供者のコンテンツ概要ページなどへのリンク
 - ▶ アクセス情報とソース情報の提供者は、一致する場合もあれば、異なる場合（つなぎ役がソース情報提供者）もある
- 提供者においてコンテンツを確認するための識別子（請求記号など）

■ デジタルコンテンツにアクセスするための情報

- 画像や動画など、コンテンツをデジタル化したもののURL
 - ▶ デジタルコンテンツURLは複数となる（さらに提供者も複数）場合もある
- デジタルコンテンツ利用のための技術情報（フォーマットなど）
- デジタルコンテンツ利用のためのライセンス情報

11 提供/ソース関連の共通情報表示例

w:accessInfo	schema:url	< http://www.narahaku.go.jp/collection/225-0.html >
	schema:description	"機関管理番号:225-0"
	v:contentHolder	d: 奈良国立博物館
	v:digitalObject	< https://colbase.nich.go.jp/uploads/collection_item_images/thumbnail/feb2ffcff8c589f119dd60ee8c5ebcf.jpg >
w:sourceInfo	id:p14813	
	rdf:type	c: アーカイブ情報
	schema:provider	< https://colbase.nich.go.jp/ > (ColBase:国立博...)
	schema:url	< https://colbase.nich.go.jp/collectionItems/view/5b7114ef7684b844c73356fa61c11915/225-0 >
	schema:dateModified	"1970/01/01"
schema:distribution	source:colbase-14813.json	

12 ライセンス情報の明示

■ コンテンツ自身のライセンス

- メディア芸術におけるライセンス記述の標準化と一括提供

■ デジタル化したコンテンツのライセンス（案）

- 提供元毎などに標準ライセンスを決めて一括記述
 - ▶ 統合プラットフォームで情報が得られるものは原則としてCC0にするなど
- 特別なライセンスが必要な例外のみ、オブジェクトの情報として記述
 - ▶ ブラウザなどでの表示の際には、標準ライセンスの場合もアイコンなどで明示する

■ メタデータに関するライセンス（案）

- 原則として共通情報はすべてCC0
- ソースデータも統合プラットフォームで提供できるものはCC0

13 共通情報整備の課題

■ データの品質と正規化

- 提供元データの記述方法のばらつき、表記の揺れについて、どこまで厳密に正規化するか
→ 発見タスク、特に「いつ」「どこで」「だれが」については可能な限り正規化し、URIとして扱う。Linked Dataも念頭に
- 識別、選択情報は原則として正規化せず、ソースデータを生かす
 - ▶ プロパティも無理にマッピングせず、汎用プロパティに導入句付きの値で
- 単一フィールドに複数值が入ると正確な処理が困難
 - ▶ 区切り文字による複数值、（）を用いた読み、URLと表示名の併記など

■ 利活用のための提供データの調整と補完

- 名前（ラベル）は必須。提供されない場合はIDなどから生成する
- コンテンツの基本区分である**タイプ**は、選択のために必須。あらかじめ定義したタイプが提供される方向へ調整したい
- ライセンスは、提供元ごとに基本（デフォルト）ライセンスを設定したい
- 場所情報はサンプルデータでは提供例が少ないが、利活用に有益。提供機関の所在地を適用することも

ジャパンサーチ（仮称）利活用のためのメタデータフォーマット仕様

利用者タスクに対応して整理した分野横断モデルのデータ（共通アーカイブ情報）を記述するためのプロパティを定義する。

利用者の発見タスクを容易にするためには、プロパティ項目は単純であることが望ましい。一方で識別や選択タスクのためには、プロパティは適切な詳細度が必要である。そこで、(1)基本的な記述には単純なプロパティを用いるとともに、(2)「いつ」「どこで」「だれが」（および上位コンテンツ）について提供データの詳細を構造化記述するプロパティを併用する。(3)またコンテンツの提供情報とソースデータ情報の関係を整理し、それぞれを構造化する。

以下の見出しに挙げるプロパティは、共通アーカイブ情報を主語にした記述に用いる。接頭辞`rdf:`、`rdfs:`、`schema:`はそれぞれRDF、RDFスキーマ、Schema.orgの名前空間、`v:`は本仕様案で独自に定義する語彙を表す。

1 基本記述プロパティ

データモデルの予備知識無しに利用できるシンプルな記述のため、Schema.org語彙を中心にした基本記述プロパティを定める。

1.1 基本的な型と識別

1.1.1 `rdf:type`

説明	文書、絵図、音声、工芸などのコンテンツの基本区分に相当する基本情報。共通アーカイブ情報型のサブクラスとして扱う。
値	URI（別途、分野横断統合プラットフォームで定義する基本タイプを用いる）

1.1.2 `rdfs:label`

説明	一覧などに表示しコンテンツを識別する名前。提供されない場合はIDなどから生成する。言語タグは用いない。
値	文字列

1.1.3 `schema:name`

説明	タイトル、別名、読みなど検索対象とする名前。読み、英語名称は言語タグで区別する。同じ言語タグの名前が複数ある場合、リテラル値が <code>rdfs:label</code> と一致しないものは「別名」に相当する。
値	文字列（必要に応じて言語タグ付き）

1.1.4 `schema:identifier`

説明	ISBNなど広く共有されている識別子
値	文字列（導入句、もしくはデータ型付き。あるいはURN）

※RDFグラフにおいては、共通情報に対応する「コンテンツ自身」は共通情報URIに`#work`を加えて識別し、そこから`foaf:isPrimaryTopicOf`（※要検討）で共通情報と関連付ける（項目の複雑化を避けるため当面は共通情報のプロパティとはしない）。

1.2 「いつ」「どこで」「だれが」の汎用記述

次のプロパティは制作、発行などの役割の違いを区別しない汎用プロパティとし、役割を示す構造化記述とセットで用いる。

1.2.1 schema:contributor

説明	コンテンツに寄与した人／組織。出演者等も含む。creator publisherを構造化記述：寄与者[関係=制作 出版]のショートカットとして記述した場合は、このcontributorは省略する。
値	URI (v:contributionのschema:agent値と一致)

1.2.2 schema:temporal

説明	コンテンツの制作、出版、内容などに関する時間。時間の意味の違いは構造化プロパティで判断する。
値	URI (v:temporalのv:temporalValue値と一致)

1.2.3 schema:spatial

説明	コンテンツの制作、出版、内容などに関する場所。場所の意味の違いは構造化プロパティで判断する。
値	URI (v:spatialのv:spatialValue値と一致)

1.3 慣用的なプロパティ

「だれが」「いつ」について、多くのスキーマで用いられる項目（役割）は個別プロパティを用意し、可能な範囲で併記するものとする。

1.3.1 schema:creator

説明	作品制作の中心となった人／組織。構造化記述：寄与者[関係=制作]のショートカットとして位置づける。
値	URI (v:contributionのschema:agent値と一致)

1.3.2 schema:publisher

説明	出版者、製作会社、配給など。構造化記述：寄与者[関係=出版]相当のショートカットとして位置づける。
値	URI (v:contributionのschema:agent値と一致)

1.3.3 schema:dateCreated

説明	制作時の時間リテラル。構造化記述：時間[関係=制作]のリテラル値。
値	文字列

1.3.4 schema:datePublished

説明	出版・発行時の時間リテラル。構造化記述：時間[関係=出版]相当のリテラル値。
値	文字列

1.4 識別と選択

次のプロパティは主として識別、選択のために用い、構造化記述の対象としない（上位コンテンツは、コンテンツ内ページなど部分指定が必要な場合は構造化記述と併用しても良い）。

1.4.1 schema:about

説明	主題、分類、区分（キーワード含む）。主題や内容のジャンル、指定区分など、共通認識があり検索結果の絞込などに利用できる用語
値	URI（典拠URI、もしくは特別名前空間によるキーワードURI）

1.4.2 schema:inLanguage

説明	作品の記述言語
値	URI（id.loc.govによるISO639-2のURI）

1.4.3 schema:image

説明	コンテンツの特徴を確認するための画像。提供元の画像とは別に、統合プラットフォームが一定サイズの画像を複製して保管する。
値	URI

※Europeanaでも詳細情報ページ用のサムネイル（400px幅）は独自に保存している模様。

1.4.4 schema:description

説明	簡潔な説明文ほか、個別項目に収録できないが有益な情報
値	文字列（元項目名を導入句として加えた値）

1.4.5 schema:isPartOf

説明	上位コンテンツ、コレクションなどへのリンク
値	URI（v:partOfのv:source値と一致）

2 構造化記述プロパティ

基本プロパティで記述した内容について、より詳細な情報を提供するため、並行して構造化記述のプロパティを用意する。構造化記述プロパティの値は空白ノードとし、その空白ノードに「構造化値」で示すプロパティを記述する。relationTypeの値は、「制作」「発行」など標準的なものはあらかじめ定義して用い、同じ役割の「いつ」「どこで」「だれが」を結び付けられるようにする。

なお、この構造化記述および次の提供・ソース情報記述については、参考としてTurtle構文によるプロパティの用例を示す。値の記述に用いている接頭辞は、便宜的に付与した仮のものである。

2.1 v:contribution

説明	寄与関係：制作に関与した人／組織の関係	
構造化値	v:relationType	寄与のタイプ。領域を超えて一般化できる制作、編集、翻訳など〔URI〕
	schema:agent	寄与した人物・団体〔典拠URI／IDなどから生成するURI〕
	schema:description	領域固有の役割名、キャストの配役など補足記述〔文字列〕
	schema:url	関連リンク〔URI〕

※人物・団体の名前は典拠URIから辿ることができるようにする。

```

:id3919
  schema:name "朝日のあたる家"@ja ;
  schema:creator dbpedia:太田隆文 ;
  v:contribution [
    schema:agent dbpedia:太田隆文 ;
    v:relationType d:制作 ;
    schema:description "監督"
  ] ...

```

2.2 v:temporal

説明	時間関係：作成、公開、発見、主題などコンテンツに関する時間	
構造化値	v:relationType	時間関係のタイプ。制作、主題など〔URI〕
	v:temporalValue	時間関係における時間（範囲）を示す〔西暦年単位で正規化したURI〕
	v:era	時間関係における時代を示す〔時間オントロジーで定義するURI〕
	schema:description	月日ほか補足情報〔文字列〕

```

:id113315
  schema:name "絹本淡彩鷹見泉石像〈渡辺華山筆〉"@ja ;
  schema:temporal td:1837 ;
  v:temporal [
    v:era te:江戸時代 ;
    v:temporalValue td:1837
  ]
  ...
  td:1837
    schema:name "天保8年" ;
    ...
  te:江戸時代
    schema:name "江戸時代", "1615~1869年" ;
    owl:sameAs td:1615_1869 ;
    ...

```


2.3 v:spatial

説明	場所関係：作成、公開、主題などコンテンツに関係する場所・地域	
構造化値	v:relationType	場所関係のタイプ。制作、主題など〔URI〕
	v:spatialValue	場所関係における場所を示す〔国／都道府県レベルで正規化したURI〕
	schema:description	より詳細な地名・住所など補足記述〔文字列〕
	schema:geo	緯度経度などの位置情報〔schema:GeoCoordinates〕

```

:id4652721
  schema:name "Phtheochroa pistrinana"@en, "セジロホソハマキ"@ja ;
  schema:spatial d:静岡県 ;
  v:spatial [
    v:spatialValue d:静岡県 ;
    v:relationType d:採集 ;
    schema:description "採集場所：都道府県（日本語）：静岡県",
      "採集場所：詳細（日本語）：梨本 南伊豆"
  ] ...

```

2.4 v:partOf

説明	上位コンテンツとの関係（掲載誌のページ）などを記述する	
構造化値	v:relationType	上位コンテンツとの関係のタイプ。掲載など〔URI〕
	v:source	上位コンテンツ関係における上位コンテンツ本体を示す〔URI〕
	v:selector	上位コンテンツ内の特定部分を示す〔文字列〕
	schema:description	補足記述〔文字列〕

※上位コンテンツの一部（ページ）が対象であることを示す必要がある場合に構造化する。

```

:idI5983376
  schema:name "ひらがな日本美術史(82)ルネサンスになる前のもの--喜多川歌麿筆「婦人相學
  十躰 ポツピンを吹く娘」"@ja ;
  schema:isPartOf m:芸術新潮200112 ;
  v:partOf [
    v:source m:芸術新潮200112 ;
    v:relationType d:掲載 ;
    v:selector "掲載ページ:118~124"
  ]
  ...
m:芸術新潮200112
  schema:name "芸術新潮2001年12月号",
  schema:description "巻:52", "号:12", "通号:624" ;
  schema:isPartOf dbpedia:芸術新潮 ;
  ...

```

3 提供・ソース情報記述プロパティ

コンテンツの提供・アクセス情報とソースデータ情報の関係を構造化して記述する。

3.1 v:accessInfo

説明	コンテンツの提供・アクセスに関する情報	
構造化値	schema:provider	コンテンツ（に関する情報）の提供者〔別テーブルで管理する識別URI〕
	v:contentHolder	コンテンツの保管者（提供者と別の場合）〔URI〕
	schema:url	コンテンツの紹介ページやアクセス情報が記載されたページ〔URI〕
	v:digitalObject	コンテンツのデジタル画像、音声動画など。全てライセンス情報を持つ〔URI〕
	v:contentRights	コンテンツ自身の権利記述〔文字列〕
	v:contentId	提供元が付与する識別子〔文字列〕
	schema:description	補足記述〔文字列〕

※提供者のURIは別テーブルで管理し、提供者の画像のデフォルトライセンスなどを記述する。
url、digitalObjectの値には、さらにそれを主語として画像、音声などの型（rdf:type）を付与する。
可能であればフォーマット（schema:fileFormat）も示したいが、提供情報の範囲では難しいかもしれない。
デジタル画像、音声動画などを主語として、schema:licenseでそのライセンスを記述する。

```
:id12247
  schema:name "色絵唐花福寿文反皿"@ja ;
  v:accessInfo [
    schema:provider dbpedia:九州国立博物館 ;
    schema:description "機関管理番号:G51" ;
    v:digitalObject
      <https://colbase.nich.go.jp/.../103dbc63d6af71b21cf57bae0204b75c.jpg> ;
    schema:url <http://collection.kyuhaku.jp/gallery/16428.html>
  ] ;
  ...
<https://colbase.nich.go.jp/.../103dbc63d6af71b21cf57bae0204b75c.jpg>
  a dcmitypes:Image ;
  schema:license <http://creativecommons.org/license/...> ;
  ...
```

3.2 v:sourceInfo

説明	収集したデータ及びその提供元に関する情報	
構造化値	schema:provider	ソース情報の提供者（アグリゲータ）〔別テーブルで管理する識別URI〕
	schema:url	集約元情報の公開ページ〔URI〕
	v:sourceData	プラットフォームが保持・提供するソースデータ〔URI〕
	rdfs:seeAlso	ソース情報がRDFである場合〔URI〕
	schema:description	補足記述〔文字列〕
	schema:dateModified	収集元データの更新日、もしくは収集日〔日付型〕

※提供者のURIは提供・アクセス情報と同じテーブルで管理する。

```
v:sourceInfo [  
  schema:provider <https://colbase.nich.go.jp/> ;  
  schema:url <https://colbase.nich.go.jp/collectionItems/view/.../16428> ;  
  schema:dateModified "2018-03-01" ;  
  v:sourceData <https://.../colbase-12247.json>  
]
```


デジタルアーカイブアクセスメントツール

5. 持続可能性の担保	※ガイドライン pp.11-13 2.4.4長期アクセスの保証のために	長期にわたり安定的にデータを保存・管理するため、個別のコンテンツを判別し認識できる識別子(重複しない一意の管理番号)を付与している。	当該コミュニティに属する組織・団体に対し、機関識別子の付与・管理を行っている。また、当該コミュニティに対し、コンテンツへの永続的識別子の付与・普及を努めている。
①メタデータ管理(識別子付与)	公開メタデータについて詳細表示ページを用意し、永続的な固定URLでアクセスできる。(一部でも可。つなぎ役等のポータルを通じて提供されている場合を含む。)	公開メタデータ全てについて詳細表示ページを用意し、DOI(デジタルオブジェクト識別子)等の永続的識別子を付与するとして長期アクセス保証を意図して公開している。	当該コミュニティに属する組織・団体に対し、機関識別子の付与・管理を行っている。また、当該コミュニティに対し、コンテンツへの永続的識別子の付与・普及を努めている。
②アクセス保証(メタデータ)	公開しているサムネイル/プレビュー又はデジタルコンテンツは、永続的な固定URLでアクセスできる。(つなぎ役等のポータルを通じて提供されている場合を含む。)	公開しているサムネイル/プレビュー又はデジタルコンテンツは、永続的な固定URLでアクセスできる。(つなぎ役等のポータルを通じて提供されている場合を含む。)	公開しているサムネイル/プレビュー又はデジタルコンテンツは、永続的な固定URLでアクセスできる。(つなぎ役等のポータルを通じて提供されている場合を含む。)
③アクセス保証(コンテンツ)	メタデータも含め、サムネイル/プレビューやデジタルコンテンツの管理において、特定の機	メタデータも含め、サムネイル/プレビューやデジタルコンテンツは、永続的な固定URLでアクセスできる。(つなぎ役等のポータルを通じて提供されている場合を含む。)	メタデータも含め、サムネイル/プレビュー又はデジタルコンテンツは、永続的な固定URLでアクセスできる。(つなぎ役等のポータルを通じて提供されている場合を含む。)
④データ移行性	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。
⑤データ管理	データの更新(修正・削除等)作業や抽出作業の取得等を簡便に行うことができるようにしている。	データの更新(修正・削除等)作業や抽出作業の取得等を簡便に行うことができるようにしている。	データの更新(修正・削除等)作業や抽出作業の取得等を簡便に行うことができるようにしている。
⑥保存フォーマット	デジタルコンテンツの保存用で作成するマスターデータについて、特定製品等に依存せず、仕様書等が公開され、かつ広く普及している(国際標準等で定められた)データ形式を採用している。	デジタルコンテンツの保存用で作成するマスターデータについて、特定製品等に依存せず、仕様書等が公開され、かつ広く普及している(国際標準等で定められた)データ形式を採用している。	デジタルコンテンツの保存用で作成するマスターデータについて、特定製品等に依存せず、仕様書等が公開され、かつ広く普及している(国際標準等で定められた)データ形式を採用している。
⑦データ保存	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。	バックアップによる提供データの複製、データ保存場所の分散等により万が一に備えた保存の体制を整えている。
⑧システム安定性	/	/	/
6. 相互運用性の確保	※ガイドライン pp.20-23 3.⑤データ共有の方法	メタデータのダウンロードを可能にしている、又はAPIを提供し、その使用についての説明ページ等を掲載している。	検索用API(Open Search, SRW等)、ハバースト用API(OAI-PMH等)のいずれかを提供している。
①ダウンロード・API(メタデータ)	デジタルコンテンツのダウンロードを可能にしている(一部でも可)。	デジタルコンテンツのダウンロードを可能にしている(一部でも可)。	検索用API(Open Search, SRW等)、ハバースト用API(OAI-PMH等)のいずれかを提供している。
②ダウンロード・API(コンテンツ)	使用する用語について、組織内で統一した辞書・典拠・ソースなどといった管理を行っている。	国際標準を意図し、又はつなぎ役が提供する分野の標準的な辞書・典拠・ソースなどを採用している。(一部でも可)	国際標準を意図し、当該コミュニティにおける共通用語(人名・地名等)を収集し、辞書・典拠・ソースといった管理を行い、それを公開している。または、当該分野の国際標準等の作成・更新や普及に関わっている。
③共通用語	メタデータの一部分の項目にURI(Uniform Resource Identifier)を付与して公開している。	メタデータの一部分の項目にURI(Uniform Resource Identifier)を付与して公開している。	同じ分野内の共通用語(人名・地名等)を収集し、URIを付与している、又はWikidata、DBpedia等に識別リソースを提供している。
④URI	Linked Data	Linked Data対応のためのメタデータセット(RDF、JSON形式など)を公開している、又はURI付与済みのメタデータセットを公開している。	Linked Data対応のために、同分野・地域内のデータに限り、メタデータセット(RDF、JSON形式など)を公開している、又はURI付与済みのメタデータセットを公開している。
⑤公開フォーマット	サムネイル/プレビュー又はデジタルコンテンツは、広く標準的に使われているフォーマットで公開している。	サムネイル/プレビュー又はデジタルコンテンツは、広く標準的に使われているフォーマットで公開している。	サムネイル/プレビュー又はデジタルコンテンツは、広く標準的に使われているフォーマットで公開している。
⑦閲覧環境(コンテンツ)	デジタルコンテンツの提供において共通利用が可能な標準形式(画像の場合はJIF等)に準拠し、特定のソフトウェアのみに依存しない閲覧環境を提供している。	デジタルコンテンツの提供において共通利用が可能な標準形式(画像の場合はJIF等)に準拠し、特定のソフトウェアのみに依存しない閲覧環境を提供している。	デジタルコンテンツの提供において共通利用が可能な標準形式(画像の場合はJIF等)に準拠し、特定のソフトウェアのみに依存しない閲覧環境を提供している。
7. 利活用促進に向けた取組	※ガイドライン pp.24-28 4.データを利活用するに当たって	SNS等を通じて、デジタルコンテンツに関する情報を発信している。また、Web上に問い合わせ窓口を設けている。	SNS等を通じて、デジタルコンテンツに関する情報を定期的に発信し、SNS等を活用してデジタルコンテンツに関する問い合わせに對して活用者とのコミュニケーションもしている。
①情報発信	活用者に向けて、メタデータやデジタルコンテンツの利活用をWebで提供している。	メタデータやデジタルコンテンツの利活用をWeb上で提供するとともに、利活用事例等も収集し、紹介している。	当該コミュニティが提供するデータの利活用事例を収集し、Web上で紹介している。
②利活用事例	デジタルアーカイブ利活用促進を目的としたイベント(ライブトーク等)を開催している。	デジタルアーカイブ利活用促進を目的としたイベント(ライブトーク等)を開催している。	活用者と当該コミュニティをつなぎ、活性化するための取組(ライブトーク等のイベント他)を行っている。
③イベント等	利用統計を取得している。	利用統計を取得し、利用分析を行い、サービスの向上に努めている。	当該コミュニティのためのポータルサイトでの利用統計をコミュニティにフィードバックしている。また、その利用分析を行い、サービス向上に努めている。
④利用分析	デジタルアーカイブで提供しているデータについて、更なる情報の追加や多言語化など、付加価値の付与に努めている(一般参加者によるタグ付け等の取組も含む)。	デジタルアーカイブで提供しているデータについて、更なる情報の追加や多言語化など、付加価値の付与に努めている(一般参加者によるタグ付け等の取組も含む)。	当該コミュニティのデータについて、更なる情報の追加や多言語化など、付加価値の付与に努めている(一般参加者によるタグ付け等の取組も含む)。
⑤付加価値	Linked Dataによって、自らのデータを他機関が提供するデータとの関連付けを行い、その情報を提供している。	Linked Dataによって、自らのデータを他機関が提供するデータとの関連付けを行い、その情報を提供している。	Linked Dataによって、異なる機関間や他のコミュニティが提供するデータとの関連付けを行い、その情報を提供している。
⑥Linked Data	コンテンツ利用に関する情報やヘルプ情報を日本語以外の言語(英語、中国語、韓国語等)でも提供している。	コンテンツ利用に関する情報やヘルプ情報を日本語以外の言語(英語、中国語、韓国語等)でも提供している。	コンテンツ利用に関する情報やヘルプ情報を日本語以外の言語(英語、中国語、韓国語等)でも提供している。
⑦多言語対応	/	/	/

關 連 資 料

デジタルアーカイブジャパンの実現に向けた工程表(全体)

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
検討・実施体制	デジタルアーカイブの連携に関する関係省庁等連絡会議・実務者協議会		デジタルアーカイブジャパン推進委員会・実務者検討委員会 ↑連携 産学官フォーラム				
分野横断メタデータフォーマット			推進委員会において検討・決定	フォーラムなどにおいて周知		不断の見直しと継続的な周知	
つなぎ役とアーカイブ機関に対する支援策			推進委員会においてニーズの洗い出し・検討 実態・ニーズ調査	予算措置の上、デジタルアーカイブ化の支援策を集中的に実施			各アーカイブ機関内での恒常的なデジタルアーカイブ化(メタデータ整備、デジタル化)
ジャパンサーチ(仮称)開発			プロトタイプ構築、要件 など	基本設計・詳細設計、開発 など			
NDLとつなぎ役との連携促進		つなぎ役となる主要アーカイブとNDLとの連携の調整検討・接続		NDLサーチ(又はジャパンサーチ(仮称)プロトタイプ)と接続		ジャパンサーチ(仮称)と接続	

デジタルアーカイブジャパンの実現に向けた書籍等分野における工程表

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
(1)ポータルを整備・提供			【NDL】公共図書館のデジタルアーカイブ等とNDLサーチとの連携の調整検討・接続				【NDL】ジャパンサーチ(仮称)との連携に向けたNDLサーチ連携元機関との合意形成
(2)メタデータの整備推進、標準化・用語の統制			【NDL】既存の分野内メタデータフォーマット(DC-NDL)に沿ったメタデータ整備				【NDL】典拠データを一元的に検索・提供するサービス(Web NDL Authorities)の提供
(3)コンテンツのデジタル化支援			【NDL】資料デジタル化研修の実施、デジタル化に関するガイドラインの公開				
(4)二次利用条件の整備、オープン化の推進				【NDL】NDL作成データの二次利用条件の整備、オープン化の促進			
(5)意識啓発・人材育成			【文部科学省】人材育成のための施策の実施				
(6)コンテンツの長期保存への協力				【NDL】納本制度による紙等の資料の収集、インターネット資料・オンライン資料の収集、他機関デジタルコンテンツの収集			
(7)活用促進のための取組				【NDL】他機関デジタルコンテンツの図書館送信の実施			【NDL】活用促進に関するイベント等の開催、事例共有の推進

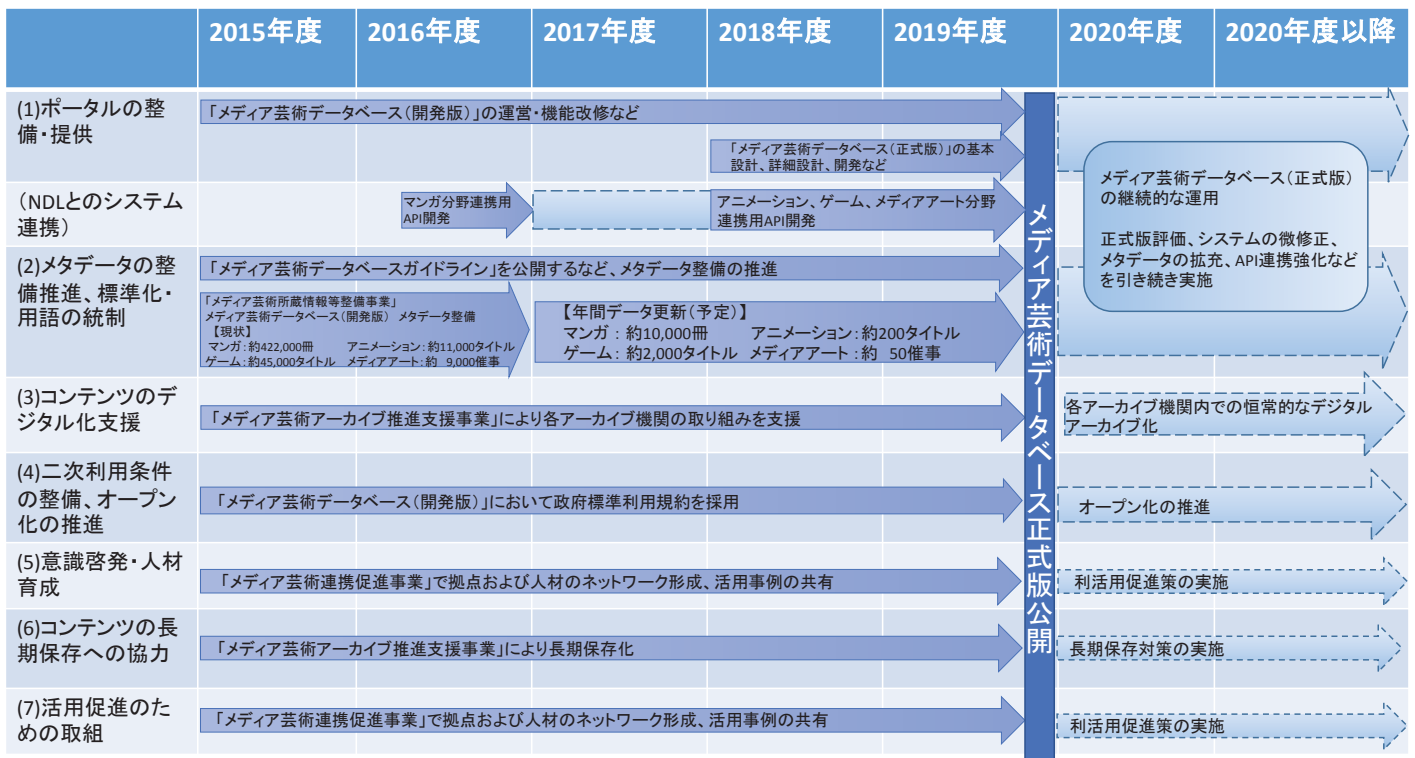
デジタルアーカイブジャパンの実現に向けた(独)国立公文書館における工程表

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
(1)ポータルを整備・提供	国立公文書館デジタルアーカイブ、アジア歴史資料センターデジタルアーカイブと地方公文書館等との横断検索等による連携を実施						
(NDLとのシステム連携)	国立公文書館デジタルアーカイブ、アジア歴史資料センターデジタルアーカイブとNDLサーチとの連携				ジャパンサーチ(仮称)との接続検討	ジャパンサーチ(仮称)と接続	
(2)メタデータの整備推進、標準化・用語の統制	「国立公文書館デジタルアーカイブ」EAD定義を踏まえたメタデータ運用						
(3)コンテンツのデジタル化支援	公文書館等との連携に向けた国立公文書館による技術支援(デジタルアーカイブ・システムの標準仕様書に係る取組み)						
(4)二次利用条件の整備、オープン化の推進					デジタルアーカイブの構築・共有・活用ガイドラインを踏まえた利活用		
(5)意識啓発・人材育成	国立公文書館における研修(デジタルアーカイブ科目を含む)を実施						
(6)コンテンツの長期保存への協力							
(7)活用促進のための取組	施策の検討	利活用促進策の実施(SNS対応等)					

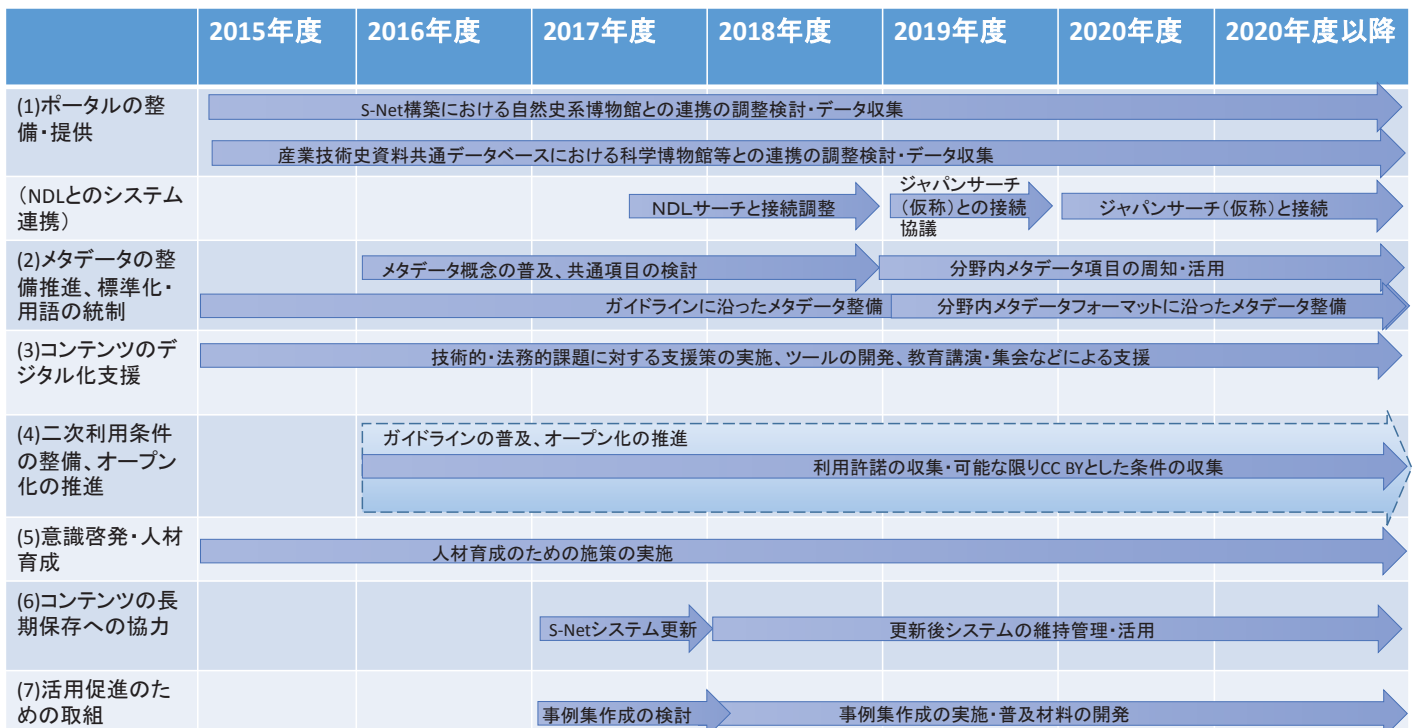
デジタルアーカイブジャパン(仮称)の実現に向けた文化財分野における工程表

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
(1)ポータルを整備・提供	各アーカイブ機関との連携の調整検討・接続						
(NDLとのシステム連携)		NDLサーチとファイル連携	NDLサーチとのAPI連携調整		ジャパンサーチとの接続協議	ジャパンサーチと接続	
(2)メタデータの整備推進、標準化・用語の統制	ガイドラインに沿ったメタデータ整備					分野内メタデータフォーマットに沿ったメタデータ整備	
(3)コンテンツのデジタル化支援	ガイドラインに沿ったデジタル化の推進						
(4)二次利用条件の整備、オープン化の推進				ガイドラインに沿った二次利用条件の整備、オープン化の推進			
(5)意識啓発・人材育成	人材育成のための研修等の実施						
(6)コンテンツの長期保存への協力				つなぎ役のプラットフォームの増強	長期保存対策の実施		
(7)活用促進のための取組	所蔵情報の増加、利用者増加に向けた普及啓発						
	スマートフォンサイトの構築、多言語化						

デジタルアーカイブジャパンの実現に向けたメディア芸術分野における工程表



デジタルアーカイブジャパンの実現に向けた自然科学博物館分野における工程表



デジタルアーカイブジャパンの実現に向けた放送コンテンツ分野における工程表

【放送番組センター】

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
(1)ポータルを整備・提供							
(NDLとのシステム連携)			NDLサーチとの接続調整		ジャパンサーチ(仮称)との接続調整	ジャパンサーチ(仮称)との接続	
(2)メタデータの整備推進、標準化・用語の統制	メタデータの充実・補完						
(3)コンテンツのデジタル化支援							
(4)二次利用条件の整備、オープン化の推進							
(5)意識啓発・人材育成			デジタルアーカイブジャパンの具体的な内容が示された段階で検討				
(6)コンテンツの長期保存への協力	放送法に基づき、放送番組の収集、保管、放送番組に関する情報の収集・分類・整理・保管等を実施						
(7)活用促進のための取組			デジタルアーカイブジャパンの具体的な内容が示された段階で検討				

デジタルアーカイブジャパンの実現に向けた放送コンテンツ分野における工程表

【NHKアーカイブス】

	2015年度	2016年度	2017年度	2018年度	2019年度	2020年度	2020年度以降
(1)ポータルを整備・提供							
(NDLとのシステム連携)		「東日本大震災アーカイブス」API連携	ジャパンサーチ(仮称)との連携を検討				
(2)メタデータの整備推進、標準化・用語の統制	メタデータの充実・補完						
(3)コンテンツのデジタル化支援							
(4)二次利用条件の整備、オープン化の推進							
(5)意識啓発・人材育成			デジタルアーカイブジャパンの具体的な内容が示された段階で検討				
(6)コンテンツの長期保存への協力							
(7)活用促進のための取組			デジタルアーカイブジャパンの具体的な内容が示された段階で検討				

我が国におけるデジタルアーカイブ推進の方向性

エグゼクティブ・サマリー

平成 29 年 4 月

デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会

(事務局:内閣府知的財産戦略推進事務局)

デジタルアーカイブの活用の対象として、観光、教育、学術、防災などの様々な目的が考えられる。こうした活用を通じて、デジタルアーカイブの構築・共有と活用の循環を持続的なものとし、その便益を博物館・美術館、図書館、文書館、大学、企業、市民コミュニティなどの「アーカイブ機関」を通じて、国民のものとしていくことで、我が国の社会的、文化的、経済的発展につなげていくことが重要である。

本報告書は、平成 27 年 9 月に内閣府に設置されたデジタルアーカイブの連携に関する関係省庁等連絡会及び実務者協議会での検討を踏まえ、我が国におけるデジタルアーカイブの構築とその活用促進に関する実務的課題に対する推進の方向性を示すものである。

序章 デジタルアーカイブ社会

デジタルアーカイブは、好きなときに好きな場所から、多種多様な情報・コンテンツへのアクセスを可能とする。また、教育や研究における利用、観光利用、ビジネスでの活用といった様々な活用を通じて、新たな経済的価値を創出し、イノベーションを推進する基盤となる。また、多様なコンテンツへのアクセスがどこからでも可能になり、地域間格差の社会的課題の解決にも資する。こうした基盤を構築することは、国の戦略として重要な取組であり、公的機関がデジタルアーカイブに取り組むことは社会的責務として求められている。

デジタルアーカイブは、活用する者だけでなく、データを提供するアーカイブ機関にとっても、デジタルコンテンツを使ったサービスの充実、来館者数の増加、業務効率化等のメリットがもたらされる。



デジタルアーカイブ社会のイメージ(例)

第1章 現状と課題

1. 諸外国の現状

欧米を中心に、様々な分野・領域のアーカイブ機関が連携して、各機関が保有する多様なコンテンツのメタデータをまとめてインターネットで検索・閲覧できる統合ポータル構築が進んでいる。EU の Europeana や米国の DPLA(米国デジタル公共図書館)では、メタデータの集約等を行う「アグリゲーター」(又は「ハブ」)を中核として連携を進めている。また、これらの統合ポータルは、デジタルアーカイブの活用促進に向けて、メタデータのオープン化(クリエイティブ・コモンズの CC0 等の表示)やデジタルコンテンツへの利用条件表示を進めている。

2. 日本の現状

分野によっては進んでいる部分もあるが、日本全体として見た場合、海外と比べて、デジタルコンテンツの提供は、量的に十分な状況とはいえない。メタデータの整備・公開も十分とはいえない。書籍等分野のように、メタデータの連携が進められている分野もあるが、分野を超えたデジタルアーカイブ間の連携は、全体としては進んでいない。

活用面においても、Europeana や DPLA のように、メタデータを CC0 で提供しているところは見当たらない。デジタルコンテンツへの利用条件表示もほとんど行われていない。また、コピーやダウンロード、メール送信ができない、専用ソフトが必要で汎用性がない、外国語(英語等)に対応していないなど、活用する者のニーズに対応できていない場合が多い。

3. 諸外国の現状を踏まえた日本の課題

諸外国の取組に追いつき、より優れたデジタルアーカイブを提供していくうえで、デジタルアーカイブ構築と連携を推進するための仕組み(インセンティブを生み出す仕組み等)の構築、中小機関及び地方における人的・財政的リソースの不足や技術的・法務的課題への対応、メタデータやサムネイル/プレビュー、デジタルコンテンツのオープン化の推進等が必要である。

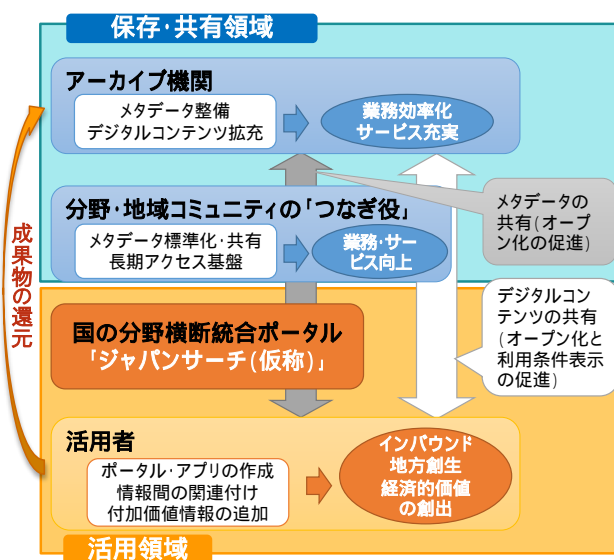
第2章 我が国におけるデジタルアーカイブ推進の在り方

1. 「共有」が支えるデジタルアーカイブサイクル

分野・地域のコミュニティの内外でのデジタルコンテンツの共有は、デジタルアーカイブにおいて、収集・保存と両輪となり、活用を支え推進するための重要な要素である。この保存・共有・活用のサイクルを、自館、地域・分野コミュニティ、さらには、日本国内、世界へと広げることで、活用の幅が大きく広がっていく。

2. デジタルアーカイブ社会の構築

各アーカイブ機関は、メタデータの整備やデジタルコンテンツを拡充する。分野や地域コミュニティごとの「つなぎ役」(Europeana の「アグリゲーター」、DPLA の「ハブ」に相当)は、メタデータをとりまとめて、国の分野横断統合ポータル(国立国会図書館が検討を進める「ジャパンサーチ(仮称)」)と共有できるようにする。活用者は、ジャパンサーチ(仮称)等を通じて、共有されるメタデータやデジタルコンテンツをデータ提供者のメリットにつながる形で、様々な用途に活用することができる。



デジタルアーカイブの共有と利活用に向けて

3. アーカイブ機関に求められる役割

- アーカイブ機関には、以下の役割が求められる。
- ・本報告書とは別途に取りまとめられた「デジタルアーカイブの構築・共有・活用ガイドライン」の採用
- ・人材の確保及び育成
- ・デジタルアーカイブの取組が業績として適切に評価される仕組みの設計(評価指標の見直し)
- ・海外発信の強化(メタデータの英語又はローマ字表記等)

4. つなぎ役に求められる役割

- つなぎ役には、以下の役割が求められる。
- ・分野・地域の独自性を反映したポータルの整備・提供

- ・メタデータの整備推進・標準化・用語の統制
- ・アーカイブ機関におけるメタデータやデジタルコンテンツ等の利用条件表示の推進、オープン化の推進、活用取組の推進
- ・アーカイブ機関におけるデジタルコンテンツ拡充及び保存に対する技術や法務上の業務支援
- ・デジタルアーカイブの評価指標の見直しとアーカイブ機関へのインセンティブ付与
- ・アーカイブ機関の意識啓発・人材育成支援

5. 国や地方自治体等に求められる役割

- 国及び地方自治体は、以下の役割が求められる。
- ・デジタルアーカイブの積極的な活用
- ・デジタルアーカイブに関わる多様な役割を担う人々のコミュニティの醸成
- ・アーカイブ機関の課題解決に必要な人的・財政的支援措置及び技術・法務上の業務支援のためのネットワーク整備等

第3章 今後の国の取組の方向性

今後の国の取組の方向性は以下のとおりである。

- ・「デジタルアーカイブの構築・共有・活用ガイドライン」の策定・普及
- ・保有するデジタル情報資源のオープン化推進
- ・国の統合ポータル「ジャパンサーチ(仮称)」の継続検討
- ・デジタルアーカイブ活用促進のための官民合同フォーラムの設置の検討
- ・つなぎ役の取組支援
- ・アーカイブ機関の人材教育支援(技術的講習会、研修を行う団体への支援等)
- ・アーカイブ機関の取組を促進するためのインセンティブ(各種の助成事業の活用や評価に応じた顕彰等)の検討

第4章 残された論点

国家戦略として、アーカイブ機関の取組をさらに強力にけん引するようなビジョンの構築とその実現のための枠組の継続的な検討が必要である。各アーカイブ機関が無理なくデータを整備・共有・連携できる共通基盤(プラットフォーム)の構築についての検討や、長期利用・永続的アクセスを意識した取組についての検討も必要である。

また、つなぎ役の機能を果たす機関を設定することが困難な分野では、関係省庁や自治体が自らポータルを立ち上げることも考えられる。引き続き、分野・地域ごとに、どのような支援策が必要かを確認しながら、本報告書での課題が解決されているかをフォローアップしていく必要がある。

デジタルアーカイブの 構築・共有・活用ガイドライン - 概要 -

平成29年4月

デジタルアーカイブの連携に関する
関係省庁等連絡会・実務者協議会

ガイドラインの対象・目的

対象

「アーカイブ機関」 (=コンテンツを保有する機関) + 「つなぎ役」 + 「活用者」

広い概念での記録機関全般を指し、**コンテンツを保有している機関すべて**を対象とする。文化的施設（博物館・美術館、図書館、文書館）のほか、大学・研究機関、企業、官公庁、地方公共団体等を含む。

分野・地域コミュニティにおいて、**メタデータの集約と提供**を行い、コミュニティにおけるメタデータの標準化、用語の統制等を行う役割を担う。

デジタルアーカイブ上の様々な**データを活用する者**。自らのデータを活用するアーカイブ機関に加え、一般ユーザ、IT技術者、クリエイターなど。

目的

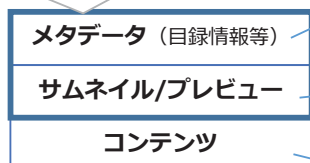
各機関がガイドラインに沿った取組を行うことによって、**我が国のデジタル情報資源を豊かにし、活用者はもちろん、アーカイブ機関自らもその恩恵を最大限に享受できるようにすることを目指す**

(ガイドラインの内容)

- 「アーカイブ機関」が取り組むべきデジタル情報資源の整備・運用方法
- 「つなぎ役」がデジタル情報資源の共有化を促すに当たって取り組むべき事項
- 「活用者」がデジタルアーカイブの利活用に当たって取り組むべき事項

「デジタルアーカイブ」とは、様々なデジタル情報資源を収集・保存・提供する仕組みの総体をいう。「デジタルコンテンツ」のほか、アナログ媒体の資料・作品等も「コンテンツ」に含まれるものとした上で、コンテンツの内容や所在等の情報を記述した「メタデータ」や、コンテンツの縮小版又は部分表示である「サムネイル/プレビュー」も対象とする。

上二つをオープンに（自由な二次利用が可能な条件で）流通させることで、コンテンツの活用が促進される



コンテンツの内容や所在等の情報を記述するデータ。目録・書誌データ、文化財基礎データ等のテキストやID

コンテンツの縮小画像（サムネイル）、本文テキストの一部表示や数秒程度の音声・動画（プレビュー）等

デジタルコンテンツのほか、アナログ媒体の資料・作品等も含む。

図 デジタルアーカイブ連携における流通単位

我が国として目指すべきデジタルアーカイブ推進の方向性（1章）

アーカイブ機関

各種コンテンツ（アナログ含む）のメタデータの整備、資料・作品のデジタル化等によりデジタルコンテンツを拡充する
できる限りオープンな（自由な二次利用が可能な）条件でデジタル情報資源を提供する

つなぎ役（分野・地域コミュニティ）

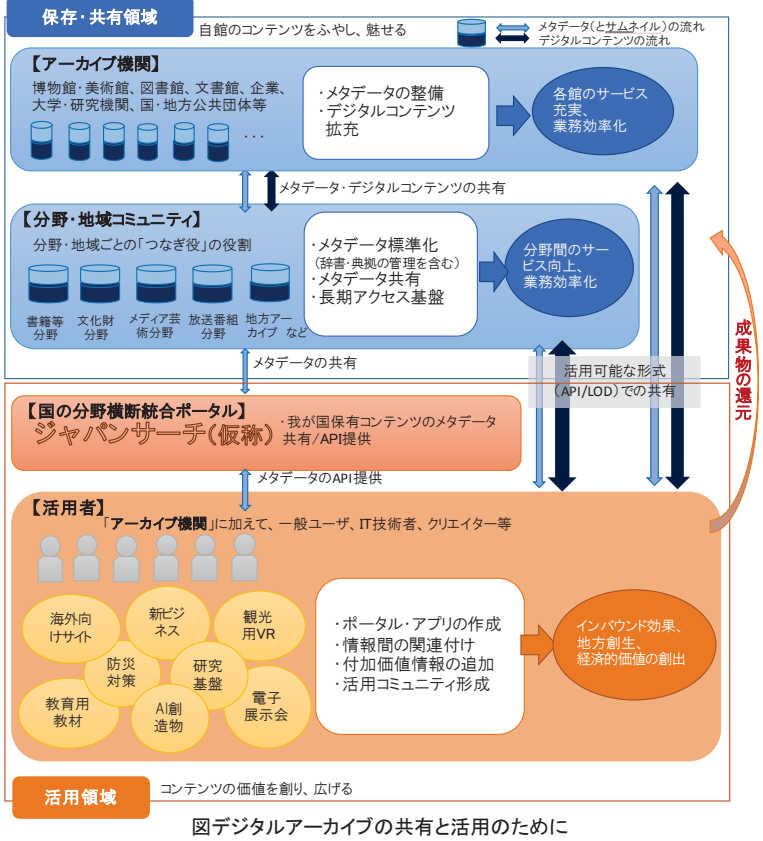
分野・地域のコミュニティをまとめ、各アーカイブ機関が提供する資源の共有化を図る。**分野内・外のメタデータの共有化**に加え、分野内のメタデータの標準化、用語（辞書・典拠）の管理等を行う。コミュニティ内の長期アクセスを保障する基盤提供などの役割も担う

国の分野横断統合ポータル

我が国が保有する様々なコンテンツへの効果的なナビゲーションを提供し、保存・共有領域にある多様な資源とその「活用者」とのつなぎ役を果たす

活用者

保存・共有領域でオープンになったデジタル情報資源に関して、**その価値を一層高める方法で利用や活用を進める**。活用者は、自らの成果をデータ提供者（アーカイブ機関やつなぎ役）に還元する



我が国のデジタル情報資源が効率的に生み出され、国全体として有効に活用されていくことを目指す

デジタルアーカイブのメリット

アーカイブ機関にとって

メタデータの整備やデジタルコンテンツの拡充といった取組は、日々の業務運営はもちろん、災害時の被害状況の把握にも役立つ。情報技術を利用した効果的なサービス展開も可能となる。ホームページでの発信や展示会等での利用など、自らが整備したデジタルアーカイブの最大の活用者は、結局のところ、その機関自身といえる。

活用者や社会にとって

デジタルデータは、時間や場所を問わず利用できるメリットがある。加えて、オープンな（自由な二次利用が可能な）デジタルコンテンツが増えることによって、観光用VRのアプリ提供、教育目的での利用、人工知能（AI）の学習用、新規ビジネスの創出など、様々な人々が様々な目的で活用することが可能となり、社会が活性化する。



デジタルアーカイブの自館でのメリット（例）



デジタルアーカイブ社会における活用（例）

デジタルアーカイブの整備に当たって（2章）～アーカイブ機関が行うこと

(1)メタデータの整備

- 「タイトル（ラベル）」「作者（人物）」「日付（時代）」「場所」「管理番号（識別子）」の5項目について、判明している場合は必須の情報として記述する。このほかは、必要に応じて、分野の事情を考慮した主要な標準（参考資料「確認すべき標準・ガイドライン」等）を参考に整備することが望ましい。
- コンテンツの権利情報や二次利用条件といった情報も併せて整備されることがよい。
- 国際的な共有を考えた場合、多言語化（英語・ローマ字表記）に取り組むことが望ましい。

(2)サムネイル/プレビューの作成

- メタデータの情報を補うため、本文テキストの一部を入力する方法のほか、コンテンツの縮小画像（サムネイル）や、音声・動画の部分抽出（プレビュー）を作成する。

(3)デジタルコンテンツの作成・収集

- 保存用としては、可能な限り高品質なものを作成する。加えて、利用・提供のしやすさを優先して情報量を抑えたものや、発見を助けるためのもの（サムネイル/プレビュー）も同時に作成することがよい。
- 外部に作業委託する場合、デジタル化成果物が自らの所有物となること、また、自ら自由に使えることに加え第三者の活用も可能となるよう著作権の状態について、契約内容の確認を行う。
- 個人所有の写真・動画等を収集する際は、肖像権、プライバシー権等の諸権利に留意しつつ、自らのサービスでの活用に加え第三者の活用も可能となるよう、権利処理を行う。
- 撮影時に自動的に記録された撮影日時・機器・解像度などの情報は削除しないよう注意する。
- デジタルデータ作成時の情報が分かるよう、デジタル化の際のドキュメント等を残しておく。

(4)長期アクセスの保証のために

- 個別の資料・作品の情報を判別・認識できる識別子（重複しない管理番号）を付与する。
- メタデータにURIを付与することが望ましい。URIの付与が自らできない場合は、メタデータの管理ファイルを安定したウェブ上に公開する方法がある。
- システム持続可能性のため、特定の機器（システム、メディア等）に依存しないデータ形式とし、データ移行性を確保する。また、ストレージ機器・システム等のリプレース経費や運用コストを見込んでおく。
- 災害や大規模なシステム障害等への対応可能性を高めるため、データ共有による分散化・複数化を進める。

データを共有するに当たって（3章）～アーカイブ機関とつなぎ役が行うこと

(1)公開ポリシーの考え方

- 自らが作成・保有するデータに関し、著作権等に配慮した上で、公開範囲と二次利用条件を決定する。

(2)二次利用条件の表示方法

- 利用条件の検討においては、権利の状態を確認し、第三者の権利が含まれる場合は許諾を得る必要がある。
- 世界的主流となっている、クリエイティブ・コモンズCC0、クリエイティブ・コモンズ・ライセンス（CC BY、CC BY-SA等）、パブリック・ドメイン・マーク（PDM）などを利用して利用条件を明示する。
- 自由利用以外の場合は、データを利用するための手続をメタデータや提供ページ等で明示的に示す。

(3)望ましい利用条件（オープン化の推進）

- 活用が最大限行われるよう、可能な限りオープン化（自由な二次利用を可能に）することが望まれる。特にメタデータは、国際的な流通・活用の観点から、CC0の採用が望ましい。
- 著作権保護期間が満了しているなど著作権による制限がないものは、PDMなどを用いて自由な利用が可能であることを明示することがよい。
- 公的機関のもの又は公的助成により生成されたデータの利用条件は、以下のとおりとすることが求められる。

データ種別	自らが著作権を保有するものの二次利用条件
メタデータ	CC0
サムネイル/プレビュー	CC0、CC BY、（PDM）
デジタルコンテンツ	CC0、CC BY、（PDM）

CC0とは…著作権法上認められる、その者が持つ全ての権利を放棄して、パブリック・ドメインに提供すること

CC BYとは…原作者のクレジット（氏名、作品タイトルなど）を表示することを主な条件とし、改変はもちろん、営利目的での二次利用も許可される最も自由度の高いライセンス

(4)利用条件表示の検討に当たっての留意点

- 著作権のほか、肖像権、プライバシー権等の諸権利にも留意が必要である。

(5)データ共有の方法

- メタデータの共有のため、①OAI-PMH（ハーベスト用API）、②Linked Dataに加えて、③その他API（SPARQL、検索用API）による連携の仕組みが備わっていることが望ましい。これらの用意が無理な場合は、表形式のデータをウェブ上の安定したところに置く方法でも連携できる。
- サムネイル/プレビューは、そのURLがメタデータ項目の一部としてメタデータとセットで提供されるとよい。
- デジタルコンテンツは、相互運用性を確保し、異なるシステム間においても一緒に利用できる仕組みが用意できるとよい（画像の場合はIIIFに対応する等）。

(1)データの活用における留意点

- アーカイブ機関が提供しているデータに関し、活用者は、コンテンツ自体の価値をさらに高め、データ提供者にとってもメリットにつながる形で活用することが求められる。
- 活用者は、適用されているライセンスや利用条件をよく確認し、順守しなければならない。
- 著作権保護期間が満了しているデータや（権利が放棄されたことを示す）CC0が適用されたデータであっても、データ提供者等の貢献の社会的認知、データの信頼性の担保から、活用者は、出典、データ提供者等のクレジットや元データのURLを示すことが望ましい。また、著作者人格権等の配慮が必要な場合がある。

(2)付加価値情報の付与

- 活用者は、デジタルアーカイブで提供されているデータに関し、付加価値となる情報を追加して利用することが求められる（例：Linked Dataを活用した情報の追加、英語・ローマ字表記の追加）。
- また、元のデータに何の情報を追加したかが分かるような形で活用したデータを提供する。

(3)情報間の関連付け

- 活用者は、分野間で共通する情報（地理情報、時間情報、人物情報等）を用いて、異なるアーカイブ機関間で提供されているメタデータを関連付けていくことによって、メタデータをより豊かにする（例：地図上にデジタルコンテンツをマッピングすることで観光客に役立つアプリの作成、美術作品を作成時間順に並べてそれぞれの所蔵館を示す等）
- 情報の有効な共有のため、つなぎ役は、分野コミュニティにおける用語（辞書・典拠・シソーラス）を統制し、用語にURIを付与することが求められる。また、分野内のメタデータフォーマットの標準化も必要。

(4)活用の結果できた成果物の還元

- 活用者は、（2～3章の）データ提供者としてのアーカイブ機関が行うべきことにも取り組む（オープンな利用条件での提供、Linked Dataによる活用の広がり確保、識別子の付与や長期アクセスの保証等）
- 活用者は、データを使った成果について、Twitter等のSNSやWikipediaなどに積極的に発信する。
- データ提供者であるアーカイブ機関や分野・地域コミュニティに対し（つなぎ役を経由するなどして）、活用者は、付加価値情報や関連付けした情報をフィードバックすることが望ましい。

(5)活用のためのコミュニティ形成

- つなぎ役は、活用を進めるためのコミュニティの形成に寄与し、活用事例の共有の場を設定する。
- アーカイブ機関やつなぎ役は、活用者が使いやすいよう、メタデータに関する解説や、応用の際のヒントになる情報を発信する。

