# 知識インフラとしてのナショナルアーカイブの構築を目指して【詳細】【詳細】

同志社大学大学院総合政策科学研究科 嘱託講師 中山正樹

# ナショナルアーカイブとは

# 知識インフラ、デジタルアーカイブ、ナショナルアーカイブとは

#### • デジタルアーカイブとは

- 一般では、情報をデジタル化して保存し利活 用できる仕組み(保存だけではない)
- NDLが進めてきた電子図書館事業そのもの
- 出版界では、長期保存するという概念はない?
- 出版界の「電子図書館サービス」には、図書館界でのデジタルアーカイブという概念は含まれない?

#### • 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の概念
- 目指すところは、デジタル文化資源全体のナショナルアーカイブと同じ
- 出版物は、知識インフラの中で、最重要視される情報。

#### • ナショナルアーカイブとは

- 国全体でデジタルアーカイブする仕組み
- 各機関が提供するデジタルアーカイブをあたか も1つのアーカイブとして利活用できる仕組み
- 知識インフラの実現形の1つ
  - 電子書籍に絞っては、「電子書籍のナショナル アーカイブ」
  - 文化資源全体で、「デジタル文化資源のナショ ナルアーカイブ」
- インターナショナルアーカイブ
  - 各国のナショナルアーカイブをあたかも 1 つの アーカイブとして利活用できる仕組み

#### デジタル化コンテンツの活用メリット

- 電子書籍のメモ、データベース化
  - SNSでの連携、ソーシャルタギング
    - 専門家に限らず、一般の人の知見を集合知識化
- 著者とのコミュニケーション
  - 人と情報、情報と情報、情報を通じて人と人が関連付け
    - FOAFは、単に友達の友達の輪ではない
- 情報源は、紙の本だけでない
  - TVで大衆娯楽を鑑賞していても、ドキュメンタリーを見ていても、放送大学を見ていても
  - 趣味の中からも
  - まったりと、 珈琲を飲みながら、 TV を見ていても
- 人工知能
  - 情報が思考により、様々な知見、知識となっていく
  - 大量のファクトデータによるシミュレーションだけでなく、個人の思考ルールも機械化されて、自分の頭脳に近くなる
  - 同じ情報を使っても、人によって知見は異なる

- アーカイブは、知識の外部記憶
- 効率化は、目的を達成する前の機械的に 可能な時間
  - 趣味で手作りする時間は大切
    - 趣味で、意識的に時間をかけていることは、重要
    - しかし、もっと効率的にできることを知らないで、時間をかけていることは見直せる
  - 情報を探す時間を、創造する時間に
- 新しいアイデアとイノベーションを見つけ出す ための挑戦
  - 「あらゆる図書館の文献でのリンクからTV番組まで、人知と呼べるすべてのデータが詰め込まれている」

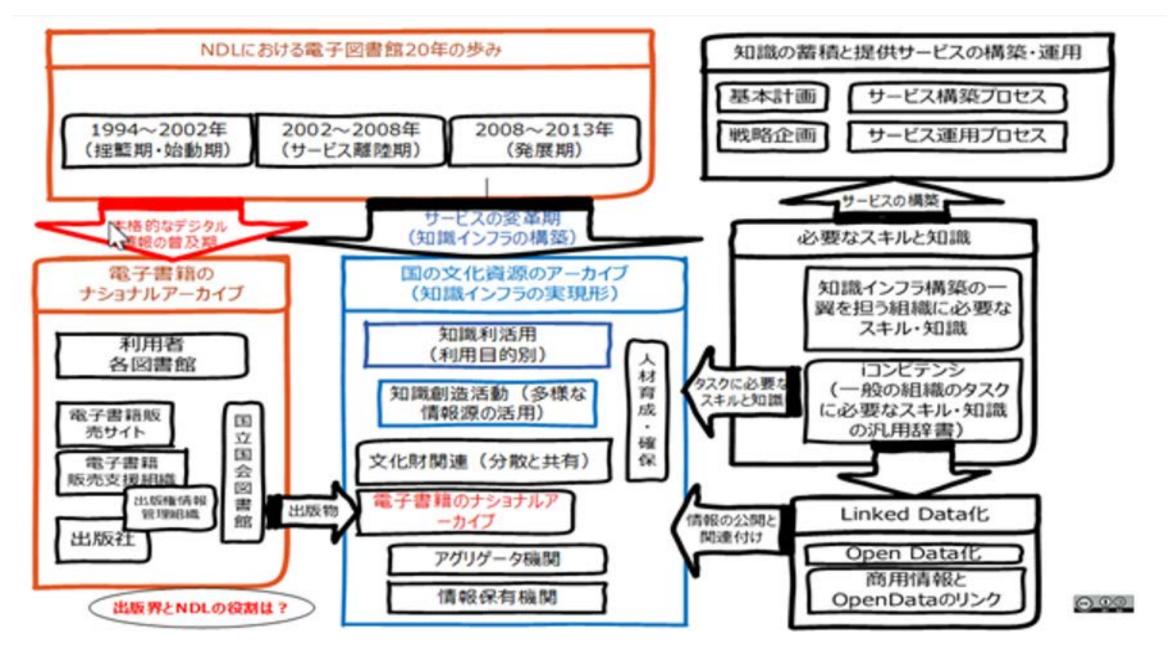
## ナショナルアーカイブで何をできるようにするか

- •情報を探し出す作業の効率化・質の向上
  - 関連付けられた網羅的な情報から、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

# ナショナルアーカイブと構想全体のイメージ



# 書籍分野のナショナルアーカイブ

#### 書籍分野のナショナルアーカイブの概念モデル - 出版界との役割分担 -

ブ機能にも期

待が…

利用者 図書館 一般利用者 (公共・大学・学校他) 電子書籍 絶版資料 出版界 電子書籍販売会社 出版活動 ⑤配信・流通機能 (電子書籍配信) 電子出版支援組織 提供用コンテンツ の支援 (1)コンテンツ生 (1)コンテンツ ②収集•一時保存機能 成機能 生成機能 (提供用コンテンツ管理) (電子書籍) (所蔵資料 デジタル化) 補 ④権利情報・管理情 出版者 報収集•管理機能 (出版情報等管理) 文化資産全 体のアーカイ 権利情報DB

視覚障害者等

保護期間 満了等

国立国会図書館

⑤送信•提供機能 (所在情報、保護期間満了資料等の ネット提供・絶版資料の図書館送信)

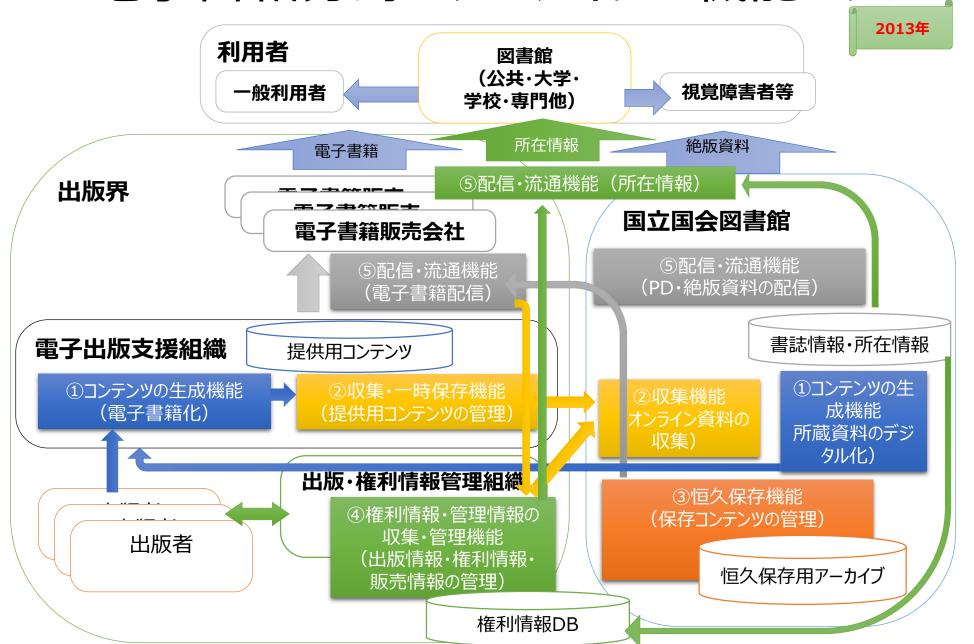
書誌情報·所在情報

②収集機能 (オンライン 資料の収 集)

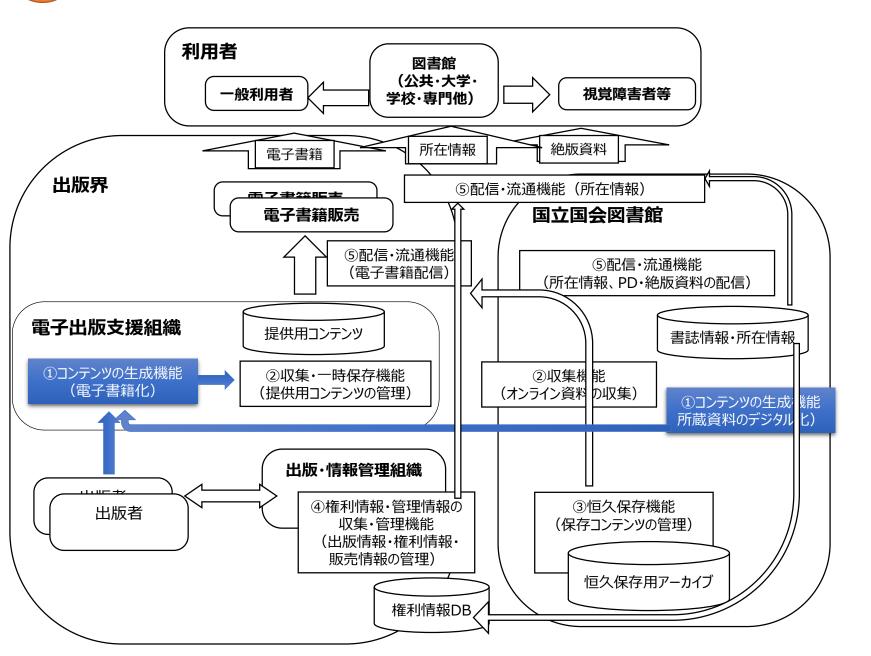
9

③恒久保存機能 (保存コンテンツの管理)

# ☆電子書籍分野のアーカイブの機能モデル

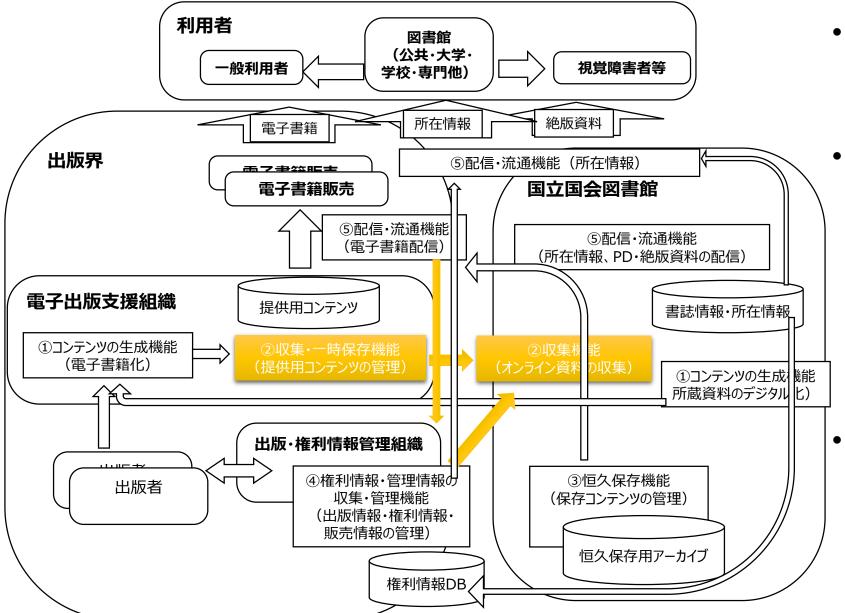


# デジタルコンテンツの生成機能



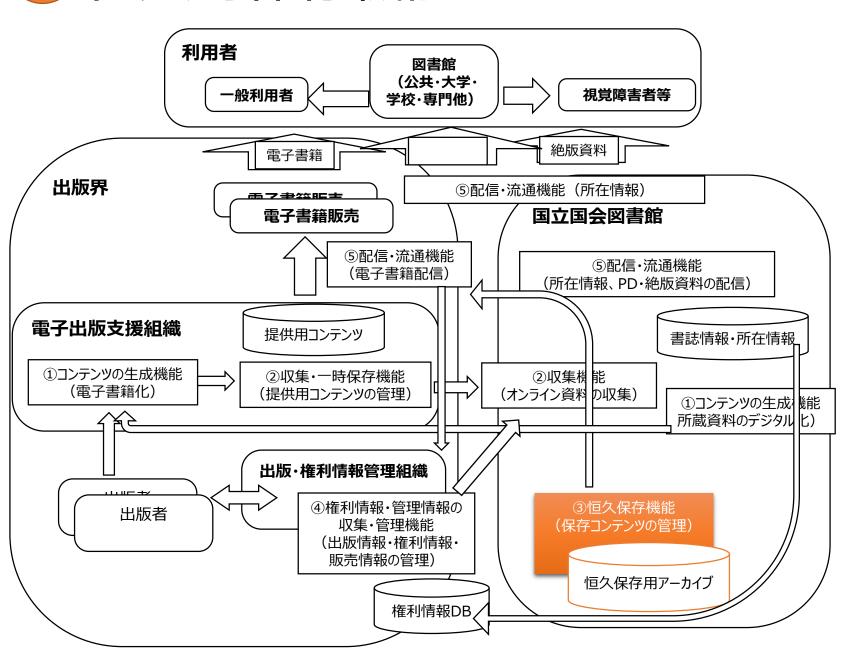
- 出版界
  - 電子書籍化
    - テキスト化
    - EPUB化
- NDL
  - 保存のためのデジタル化
  - 現在はイメージ化
  - 今後は検索のためのテキスト化
- 図書館
  - 郷土資料のデジタル化
- 連携
  - NDLイメージ化データの二 次利用提供
    - 出版界での復刊のために
  - デジタル化仕様の共通化
    - EPUB仕様、画像・音声・動画仕様

## 電子書籍・書誌情報の収集機能



- 電子出版支援組織
  - 販売用コンテンツの保存
    - ビューアに依存しないEPUB
- 出版•権利情報管理組織
  - 出版情報DBの構築
    - 出版・出版サイト情報の提供プロトコル (API) の開放
      - 《書誌、書影、出版情報、出版サイト情
- NDL
  - 全国書誌(国内出版物の総合目録)、 提供元情報の作成
    - 出版情報・提供サイト情報の収集
  - 書誌作成において、出版情報の活用(私
    - 近刊情報、新刊情報
  - 公的機関のウェブ情報の収集の拡大
  - 民間無償オンライン資料収集の拡大
  - 民間有償オンライン資料の収集(未実施)
  - 公的機関の情報のLinked Open Data 化の推進
- 連携
  - メタデータ仕様の共通化、相互利用
    - メタデータ記述要素・記述規則
    - ONYX⇔DublinCore、MARC21
  - 有償オンライン資料の制度化(現在協議中)

# 恒久的保存機能



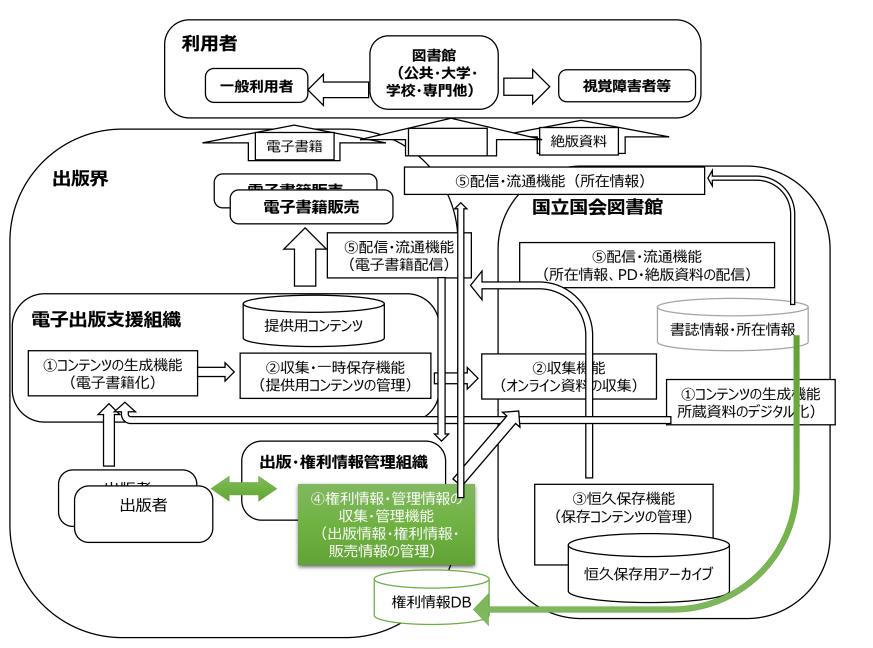
#### • NDL

- 将来に亘って利用を保証
  - 有償・無償に関わらず著作物を、文化資産としてアーカイブし、後世に残す
- ダークアーカイブの役割を持つ
  - 著作権、出版権、肖像権等の権利がある著作物
  - 提供元機関が、維持・提供が困難な事態が発生した場合、提供元機関に提供

#### • 出版界

- 電子書籍のバックアップサイト として活用
- 関係機関との連携
  - あらゆる記憶・記録を、利活 用できる形で、後世に残す
    - 出版界も含め、他の文化資源保存機関と分担して、ナショナルアーカイブを構築
    - 研究機関と連携して、長期 保存技術の研究開発、実 用化実証実験
  - アーカイブ内の情報へのアクセスのための仕様の共通化

# 権利情報・管理情報の収集・管理機能



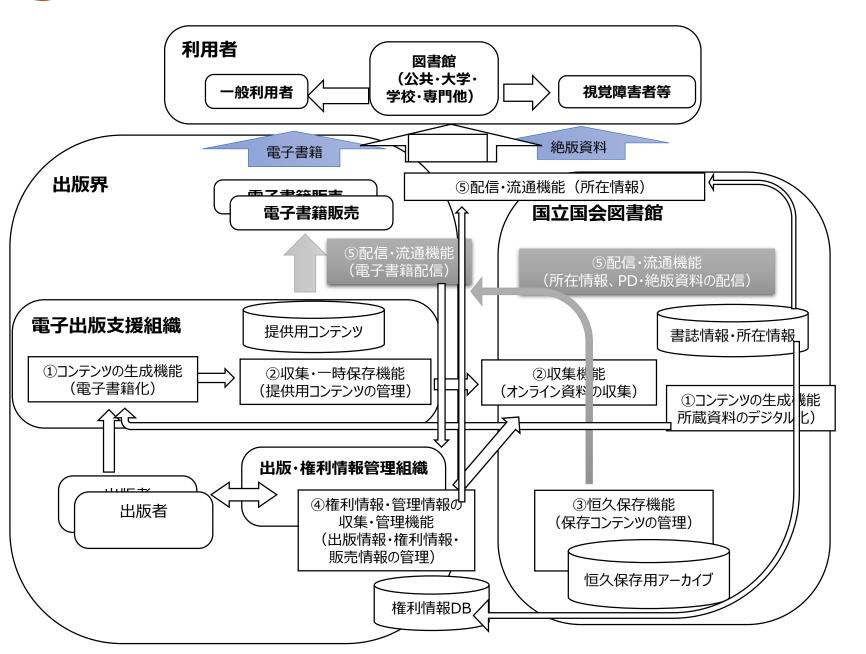
- 出版界
  - 著作物の権利情報の収集・ 管理・提供
    - 著作者情報の管理
    - 著作権、出版権等の権利情報
  - 出版情報の管理
    - ・ 基本書誌、内容紹介、著者 紹介、書影、試し読み、書評 リンク、重版情報、ジャンルコー ド、
    - 出版権登録情報
    - 著作者情報、出版社情報、
    - 著作権情報

#### NDL

- 書誌情報、件名、NDC分類 コード、著作者典拠情報の提供(私見)
- 連携
  - ISBN、著者典拠ID等の永 続的識別子による著作物同 定
  - 著作権情報の共有(私見)



## 配信·流通機能



#### • 出版界

- 各電子書籍販売サイトから、インターネット利用者へ 提供
- 商用の電子図書館サービスサイトから、公共図書館利用者へ提供

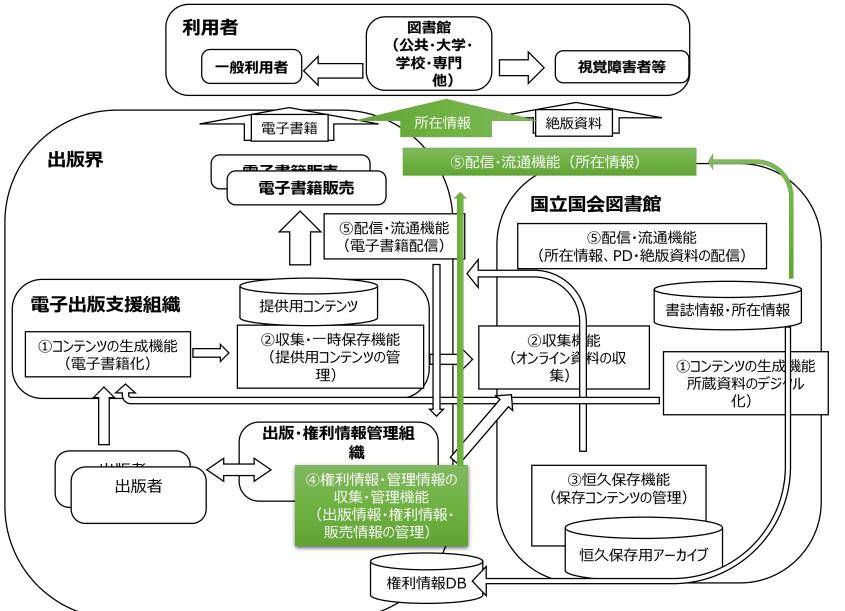
#### NDL

- 著作権切れデジタル化資料をインターネット利用者へ提供
- 絶版デジタル化資料を公 共図書館利用者へ提供

#### 連携

- 利用者の閲覧環境の共 通化(私見)
  - NDLデジタル化資料の商用電子書籍ビューアでの閲覧

# 目録および所在情報の提供



#### 目的

- 全ての出版物の存在を可視化
  - 利用者に対して、所在場所に関わらず、何らかの形で入手可能な著作物を網 羅的に見つけ出せるようにする
- 利用者の選択肢の確保
  - 利用者が必要とする著作物と、その入手先を自由に選択できるようにする

#### • 出版界

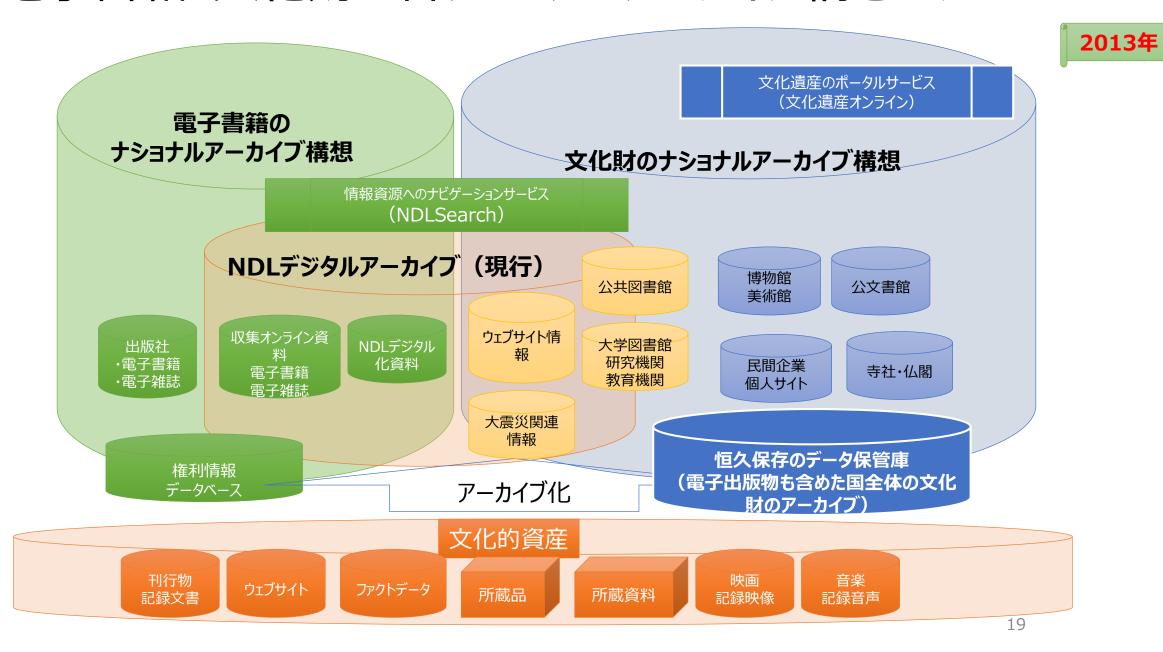
- 絶版になっている出版物も含めて網 羅的に検索できるサービスの提供
- 販売促進
  - 商用出版物を優先表示
- NDL
  - 所蔵していない出版物も含めて網羅的に検索できるサービスの提供
  - 利活用の推進
    - 利用者が入手しやすい提供元を優先表示

#### • 連携

- 利用者視点で、利用者が資料を探し出すために必要な情報を共有化
- それぞれの利用者の目的に応じた検索サービスの構築を容易にする通信 プロトコル、メタデータ仕様の共通化

# 文化情報資産のナショナルアーカイブの構築に向けて

#### 電子書籍・文化財の各ナショナルアーカイブ構想のカバレージ





# ☆各種アーカイブ構築施策の一元化

書籍,電子書 籍,古典籍、メ ディア芸術、 JJAPACON,

個別の情報基盤(個別所管体制)

出版関連

文献・ウェブ 情報関連 ポータル (国立国会 図書館サーチ)

組織化·知識化

収集

デジタル化

権利情報

書籍関連 アーカイブ 文化財 関連

文化財 ポータル (文化 遺産オン ライン)

組織化·知識化

収集

デジタル化

権利情報

文化財

アーカイブ

権利情報

大規模

災害関連

大規模災

害情報

ポータル

(ひなぎ

く・地震関

係ポータ

ル)

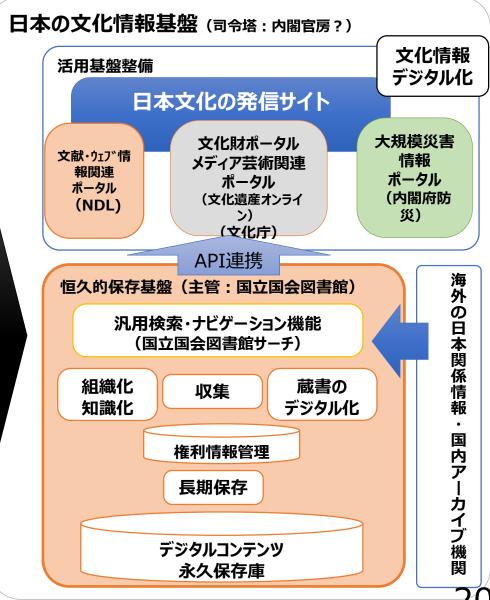
組織化·知識化

収集

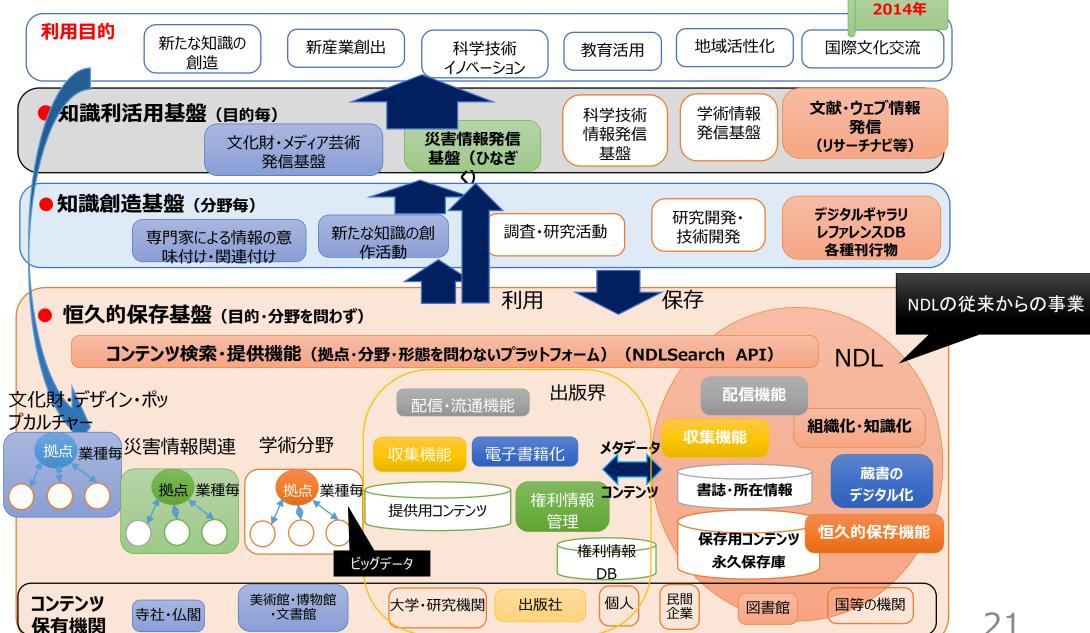
デジタル化

災害関係 アーカイブ (大震災 を含む)

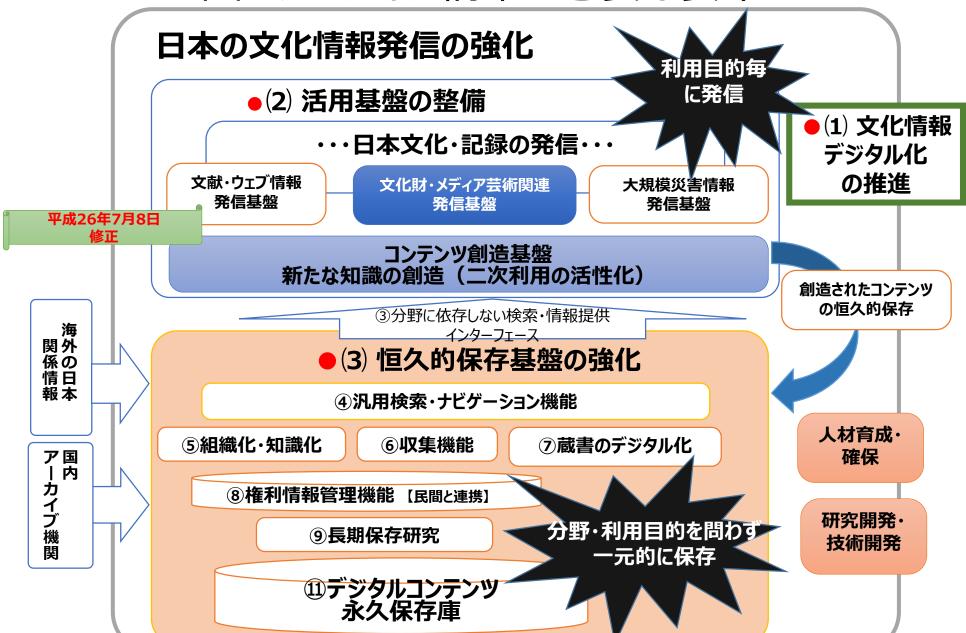
各施策の一本化



# 文化財を含めたナショナルアーカイブの機能イメージ



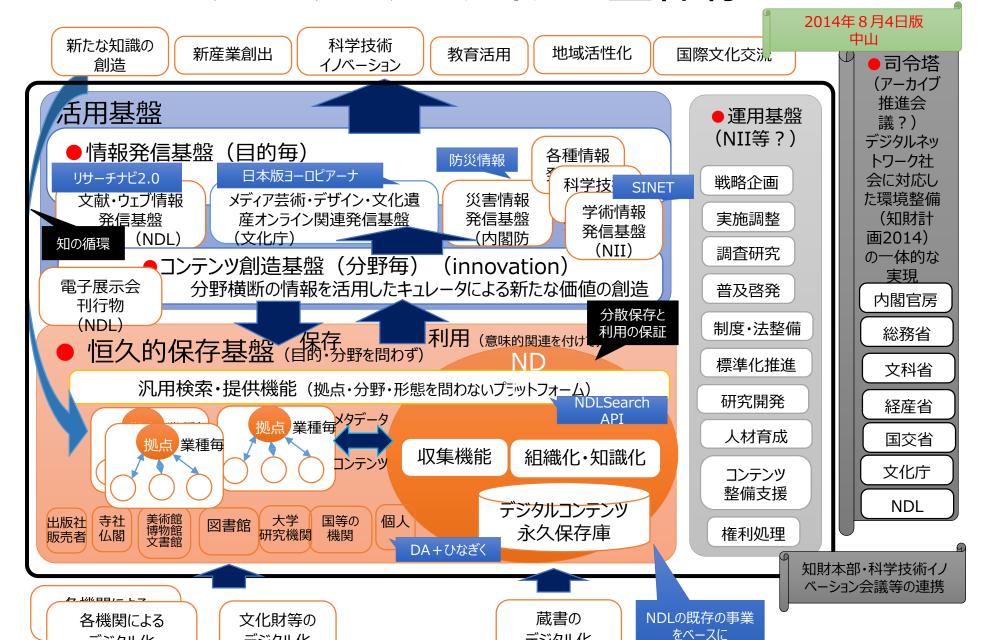
#### 国のアーカイブ構築に必要な要素



# ナショナルアーカイブの概念



#### ナショナルアーカイブの全体像



デジタル化

デジタル化

デジタル化

24

。 ららゆる情報を文化資産として将来にか たって保存し、利活用を保証する 知識の再利用と 新たな知識の創造 Digital Transformationの実現 教育 ●知識インフラの実現形 新しい産業の創出・ 防災 · 減災 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 イノベーション 教養・娯楽 新たな知識の創造 報として集約し、相互に関連付 観光 けて知識化し、将来にわたって 地域活性化 国際文化交流 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 情報発信(利用目的別) 「新たな知識の創造と還流」の 仕組みを構築する デジタルが価 文化財情報サイト 物理的空間 学術情報サイ 科学技術情報 日本の文化の 発信サイト 文献·Web情 人材育成 値創造の仕 報サイト サイト (ファシリティ) 組みを根本 から変えること Cheif Digital を意識した戦 情報探索支援・レファレンスサービス 次世代統計利用システム (statdb.nstac.go.jp) Officer 略企画 (CDO) Cheif Information 情報システ Officer ム化により、 知識創造活動(多様な情報源の活用) (CIO) 組織戦略の 実施責任者 将来構想:戦 場の提供 略企画人材 コミュニティをを繋ぐ ミュニティに参画する サブジェクトライ 複数の情報原・分野にまた ブラリアン・ア-キビスト ● アイデアソン がる情報を関連付ける 情報の組織化・構造化 辞書・典拠・シソーラス き本情報付与 類の作成 ハッカソン データモデル、シンタックス、セマン ティック、メタデータ 同音異義語、同義語 エンベデッドライ ブラリアン 語彙の共通化 IT活用人材 ビッグデータ 文化財関連(分散と共有) (組織を越えた共有) デジタル人材 文献情報 レファレンス情報 トラフィック情報 地名情報 地図情報 災害情報関連 人名典拠 件名典拠 標準化推進人 材 書評 復興の記録写 災害の記録写 出版物 真·映像 真•映像 IPA共通語彙( 観光ガイド Wikipedia 電子書籍 3.11電力消費量 記事情報 書籍 放射線データ 電子雜誌 研究開発人材 分析結果 収蔵品情報 デザイン 気象データ 論文 雑誌 法制化人材 地域資料 電子ジャーナル 名所·旧跡情報 ゲーム 文化財DB ●情報源の多様化→相対的重要性の低下 古典籍 国宝 ・知識創造、アイデアの創出に役立つ情報は、 政府情報 政府刊行物 政府統計情報 過去から現在にわたって発信され続けている情 アニメ 無形文化財 報で、出版物に類するものだけではない。 放送 ●一般社会のITおよびデジタル活用に比べ ●ロングテール て、文献を扱ってきた出版界、図書館界の進 分散データベース マンガ 展は? ● Linked Data化 アグリゲータ機関 観光情報サイ e-G(3.11電力消費量 東京国立博物品 宿泊情報 NDL JST/NII等 データカタロン サイト 人間文化研究 商業出版団体 報道機関団体 イベント情報 サイト 文化遺産オン 機構等 カーリル 国立美術館 (data.go.jp) ライン 政府各種報告書、締計データ 情報保有機関 ● Open Data化 鯖江市、横浜市 オープンデータ憲章(2013 Open 大学図書館 学術出版社 公共図書館 美術館 立法機関 電子行政オープンデータ推進のため 出版社 個人運営 寺社•仏閣 各報道機関 民間企業 博物館 のロードマップ 行政機関 情報サイト (2013年、2015年改訂)

#### ナショナル・アーカイブの検討に当たっての考察

「社会・経済的な価値を創出」を目指して、様々な分野のあらゆる記録を情報として集約、相 互に関連付けて知識化し、将来にわたって利用を保証。 「新たな知識の創造と還流」の仕組みを構築する

- 利用者、権利保持者の双方の利益になるように
- 世界規模でのアーカイブ構築の一翼を担う
- 国としてのアーカイブ構築
  - 縦割り行政の分野単位でなく、国全体のアーカイフとして大きな器の中で、各種アーカイフは分野の1つとして、全体で整合性を持って、効率的、効果的に進めるべき
  - 各分野の分散アーカイブを相互補完しあう形でネットワークを形成。分野を越えて、情報同志を関連付けて知識インフラとして利用できるように。
- 利用目的毎のポータルの提供
  - 分野毎に多様な利用者ニーズに あったポータルを利用できるように

- サービスの高度化
  - デジタル化・収集・組織化・知識化・ 保存・提供の各フェーズの高度化に 資する研究開発の促進と成果の活 用
  - 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の人材の確保・育成
- 法的課題の解決
  - デジタル化・収集・提供に関連する制度 的制約

☆ナショナルアーカイブの各基盤の概念

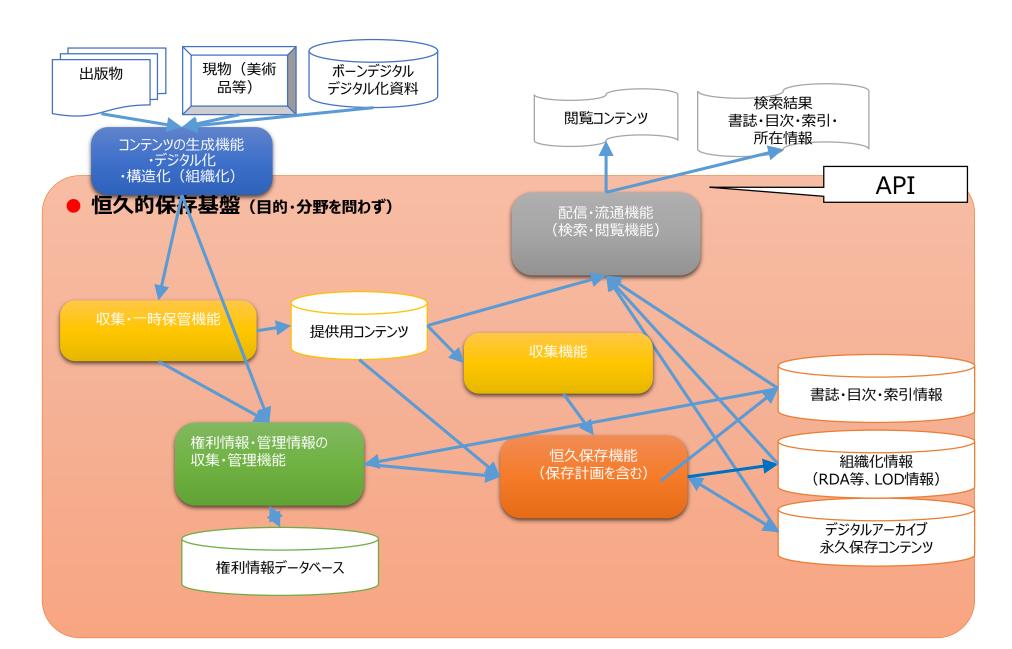
# 恒久的保存基盤とは?

#### NDLが責任を持って構築する部分

- 従来からの「知識インフラ構築」の概念に相当するも
  - 体現形の種別、所蔵場所を問わず分野横断的なコンテンツを組み合わせてりようできるように
  - 知の創造のための素材としての情報及び新たに想像された情報の永久保存
- デジタル化/収集/組織化・知識化/保存/汎用 検索・ナビゲーション
  - 仕組みは、OAISフレームワークに準拠
- デジタル化
  - 紙資料のイメージ画像化は「媒体変換」に相当
  - 様々な体現形の情報を生成するのは「創造」
- 収集
  - あらゆる体現形の情報を収集する
- 組織化(知識化)
  - 各情報が持つメタデータは、そのまま保持する(劣化させ ない)
  - 自動メタデータ付与機能を持つ
  - 情報に永続的識別子を付与する
  - 全文テキスト等を活用して、情報と情報を意味的内容で 関連付けをする(LOD化、セマンティックWeb化)

- 保存(レポジトリ)
  - あらゆる体現形の情報を永久保存する
  - 長期保存のためのマイグレーション機能も含む
- 提供
  - 汎用検索・ナビゲーション (NDLSearchの検索API機能+DAの一次情報提供API機能)
  - データプロバイダ—的機能
  - 各コンテンツホルダーが保有している情報は、あらかじめメタデータを収集、もしくは横断検索して所在場所へナビゲートする
  - 永久保存庫に格納された情報の全てを検索対象として、 一次情報を活用基盤に提供する。
- 恒久的保存基盤の実装
  - 業務・業態の関連機関の種別ごとに拠点があり、それらの拠点とNDLで、分散アーカイブを形成
  - 各拠点間で情報を自動的に持ち合える仕組みを持つ (例えばP2Pネットワーク)

# ☆恒久的保存基盤



#### 2014年8月6日追加

# コンテンツ創造基盤とは?

恒久的保存基盤に蓄積されている複数の情報を素材として活用(参照もしくは組み合わせ)して、二次的情報として、新たなコンテンツ(知識)を創出する

#### 概念

- 分野のアーキビスト、ライブラリアン、レファレンサー、研究者等を含めた 専門家が、各分野の対象領域を越えて、情報を関連付け(知識 化)、情報を組み合わせて新たなコンテンツを創造する。
- 創造されたコンテンツ、関連付けられた情報は、恒久的保存基盤に フィードバクされて蓄積される

#### 文化芸術

- 従来の文化芸術における各分野の対象領域を超えて、伝統文化と 現代的な文化芸術を組み合わせた新たな日本文化の創造
- 異分野の専門家、ユーザ同士が共同でコンテンツ創造できる場
- 専門家、利用者視点でのコンテンツ生成

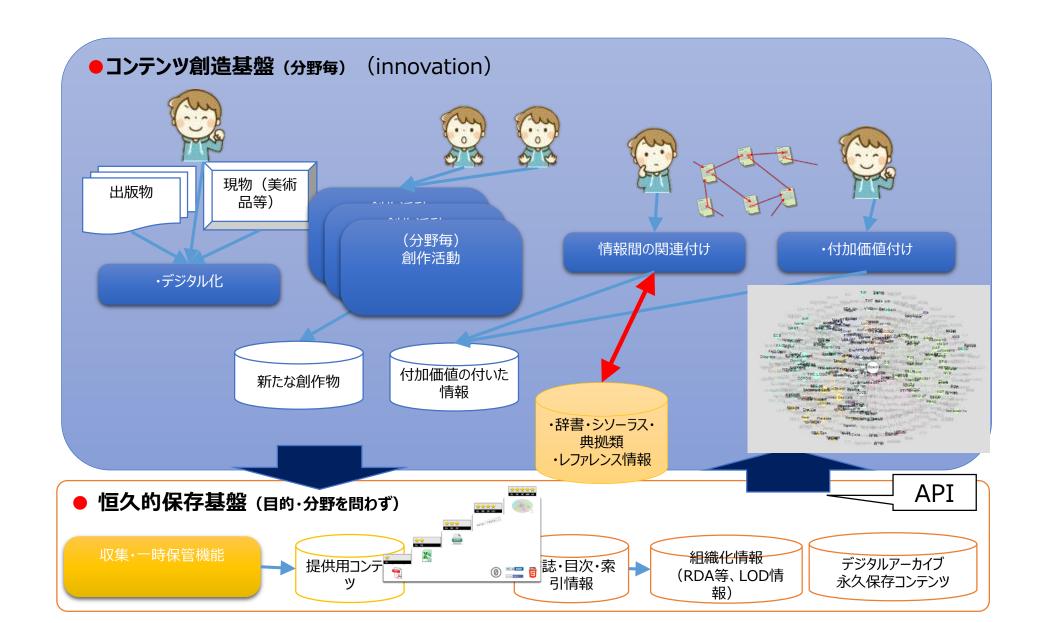
#### • 利用者別

- 研究者向けコンテンツ
- 一般向けコンテンツ
- 高齢者・障害者向けコンテンツ
  - リフロー型コンテンツ
  - 読上げ可能コンテンツ
- 子ども向けコンテンツ
- 創出コンテンツ種別
  - 調査研究(新たな知識の創造)
    - 論文、プレゼン資料、高精細画像
  - 教育用コンテンツ
    - デジタル教科書
  - 教養・娯楽コンテンツ
    - ポップカルチャーコンテンツ

#### 分野別

- 科学技術分野
  - 各種次世代技術開発
- 人文科学
- 国文学·歴史学
- 社会科学
- • •

## ☆コンテンツ創造基盤



# ☆情報を媒介して専門家と専門家を繋ぐ

#### ● コンテンツ創造基盤 (分野毎) (innovation)

ンデータ、LOD関連

Knowledge

Japan

o://okfn.jp/ )

ン、アイデアソン、アンカンファレンス

l Open Data Initiative

データ流通推進コンソーシアム(総務省)

METI プロジェクト (経産省)

ガバメントラボ(経産省)

タログサイト(内閣官房)

◇ ハッカソン

Foundation

- ◆ アイデアソン
- ◇ アンカンファレンス



文化遺産オンライン

- ▶ じんもんこん
- OpenGLAMじんもんこん







人文科学とコンピュータ





OSS Community Dictionary



- ▶ オープンソース関連
- ♦ Code4Lib
- ♦ OSS Community (<u>http://ospn.jp/</u>)













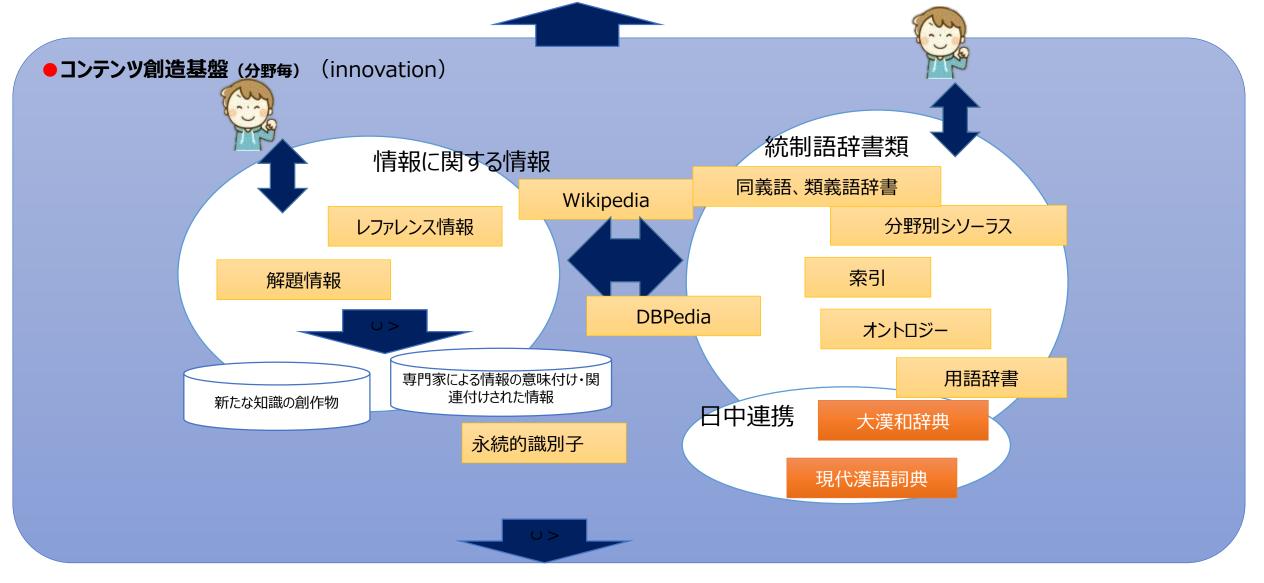




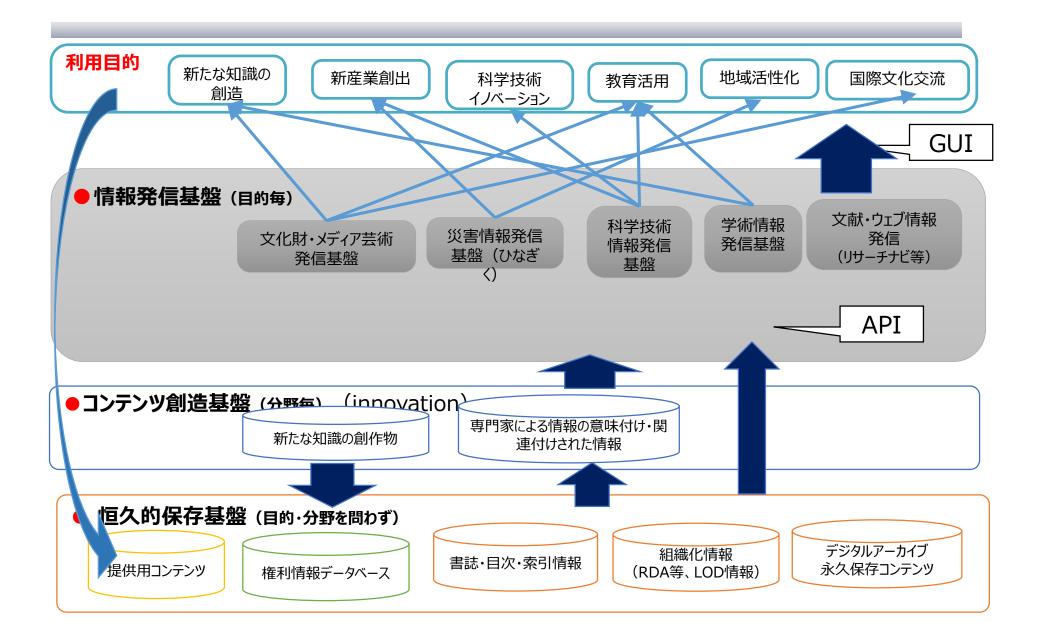
- MLA
- ♦ OpenGLAM

▶ ナショナルアーカイブ構築の司令塔、中核的組織

# ☆人を媒介して辞書と辞書を繋ぐ-



## ☆情報発信基盤



# 「ナショナルアーカイブ」の構築を目指して

新たな 発見と 価値の 創出へ

- ・海外からの注
- ・地方への注目 ·教育/商業
- /防災等への 利用

経済 効果

- ・コンテンツ商用利 用の収益還元
- •集客効果

【コンテンツの活用/創出の促進】

- ・目的別検索 ・付加価値サービス
- ・新規コンテンツ創出

ND

【メタデータの利活用の促進】

- ・目的別、テーマ/ジャンル別の検索プラッ トフォームの提供
- 付加価値情報の付与 (多言語化、画像化等)

#### 国立国会図書館が果たす役割

【国全体の分野横断型統合ポータルの構築】

・国内保有のコンテンツ所在情報を含むメタデータ集約/API 提供・全体標準化、利活用の共通ルール策定協力

【恒久的保存基盤の整備】・アーカイブ構築のための 法整備(孤児著作物への対策等)への協力 ・維持困難 アーカイブのコンテンツ保護 ・各図書館がデジタル化した入 手困難書籍の収集・保存

【領域ごとのアグリゲ

デジタルアーカイブ 連携の関係府省 等間の調整

恒久的

- ・(領域ごと) メタデータ集約
- ・メタデータ標準化
- ・デジタル化/公開支援

·所蔵目録DB化 ・デジタル化=コンテンツ作 成・ウェブ公開の推進

保存基盤

【各機関】

官公庁、企業、美術館、博物館、文書館ほか

電子図 書館構 想の実現 形である が、NDL は全面に 立って推 進役にな らないの か?

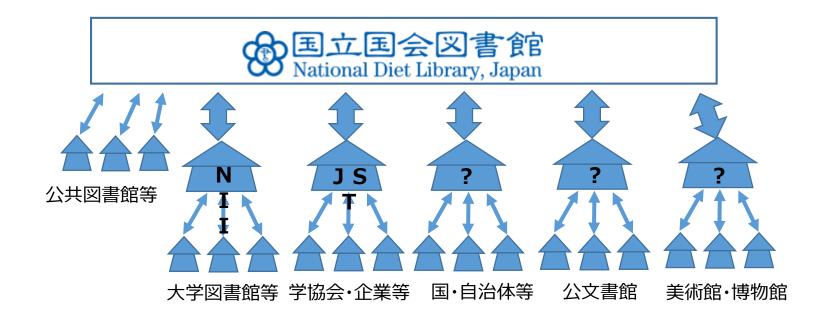
【転用:もっと近くに国立国会図書館 第17回図書館総合展2015年11月10日】

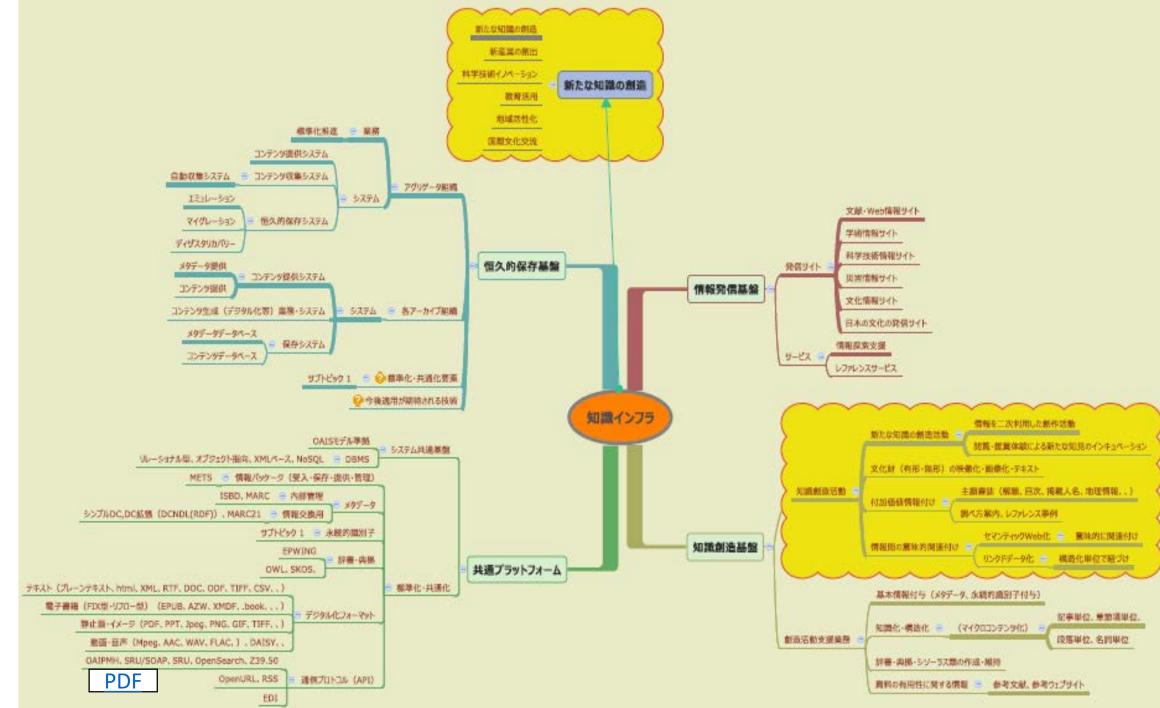
#### 国立国会図書館が果たし得る役割

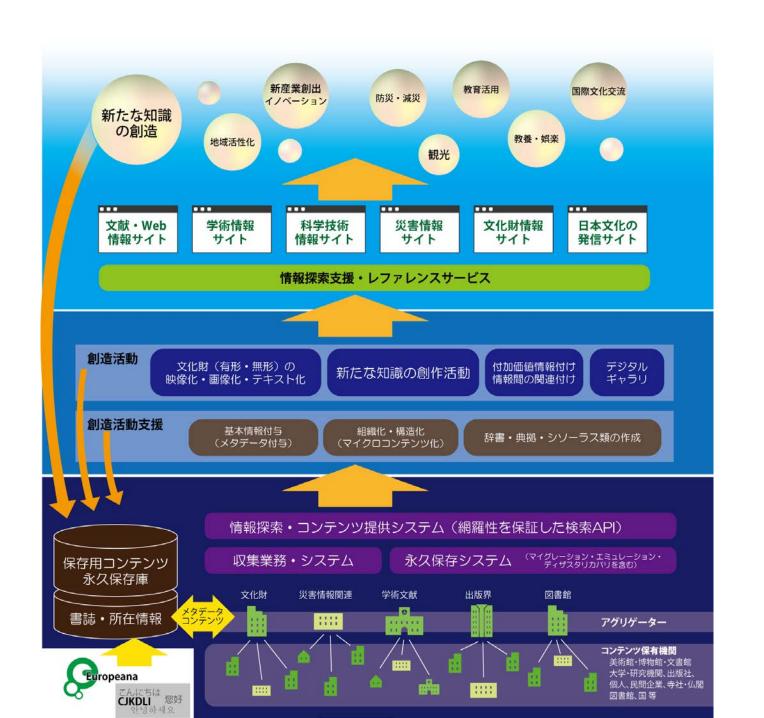
- ➤ 国全体のメタデータ集約/提供における、システム面(統合的ポータル・API提供)での対応
  - 国立国会図書館サーチの機能拡張による実現



- ▶図書館界のアグリゲータ
  - 各図書館の資料デジタル化の支援(手引き公開、研修の実施等)
  - デジタル化資料等の(識別子含む)メタデータ標準化の推進
- ➤ 出版物等のデジタル化データ等への付加価値サービスの創出
  - テキスト化と本文検索、電子展示会
  - N D L ラボでの研究活動







#### ☆文化資産アーカイブ構築の一環で国が支援 【知財計画2015】



。美術館・博物館、大学・研究機関、民間施設の関係者 向け

⊕ 司書や学芸員等現職人材向け

#### 若手の標準化人財

●ビジネスツールとして戦略的に活用する管理職、営業職

・地方ゆかりの文化情報などのコンテンツの収集と利活用 を可能とするデジタルアーカイブ構築を支援する。

(地方におけるデジタルアーカイブ構築支援)

- ・メタデータのオープン化に向けた課題の検討
- ・統合ポータルからデータセットを抽出する機能の普及等

(目的に応じたポータル構築環境の整備)

あらかじめ著作権者が行う意思表示の在り方など必要な 取組の在り方について検討を行う。

利用に係る著作権者の意思表示

知財人財の戦略的な育成・活用

●美術館等での複製、裁定制度の補償金供託の見直し、 裁定を受けた著作物の再利用手続きの簡素化

アーカイブの構築と利活用の促進のための 著作権制度の整備

統合ポータルの構築

分野横断的な検索が可能なポータルサイトの整備 ⊕

アーカイブ構築の手順等についての研修等。⊕

アグリゲーターによる取組

分野毎のメタデータ形式の標準化などの策定、デジタル。 化への協力、メタデータの集約化。

アーカイブ構築の手順等についての研修等。⊕

メタデータ付与やAPIを付した形での公開のための助言。

NDL所蔵資料のデジタル化。、デジタル化データの利活。 用の促進に向けた取組を強化。

文化資源や国宝・重要文化財以外の地域の文化資源 に関するデータの集約。画像掲載率の向上を図る。 多言語化を含め利活用に資する取組を推進する。

地方の博物館・美術館等に対して必要な情報の周知を

メディア芸術等分野

文化財分野

放送コンテンツ分野

書籍等分野

(短期:2015~6,

中期:2017~8)

知財計画2015

#### 知識情報基盤の構築

