

2016年5月4日

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」

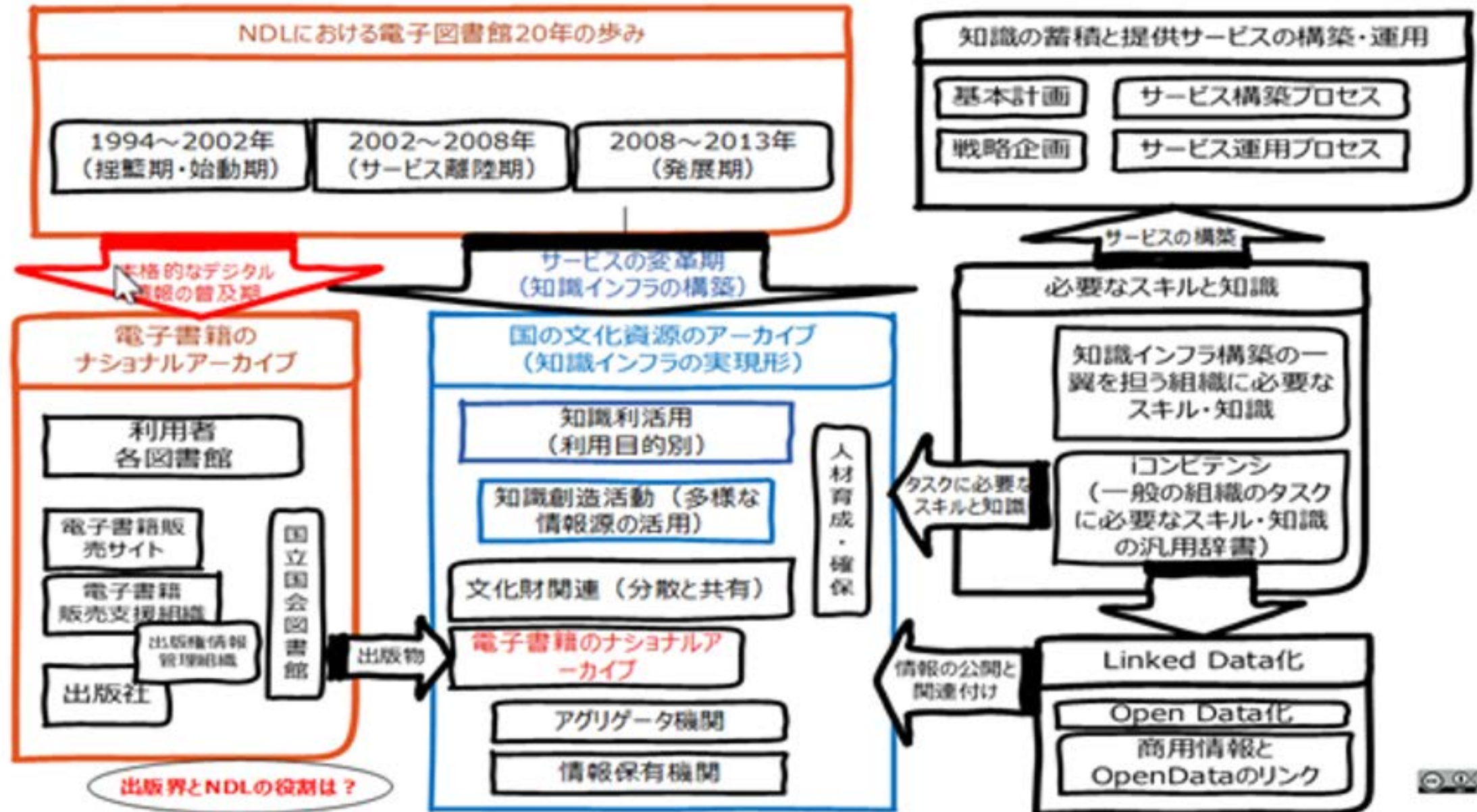
ナショナルアーカイブで何ができるようになるか

- 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な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、今までに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

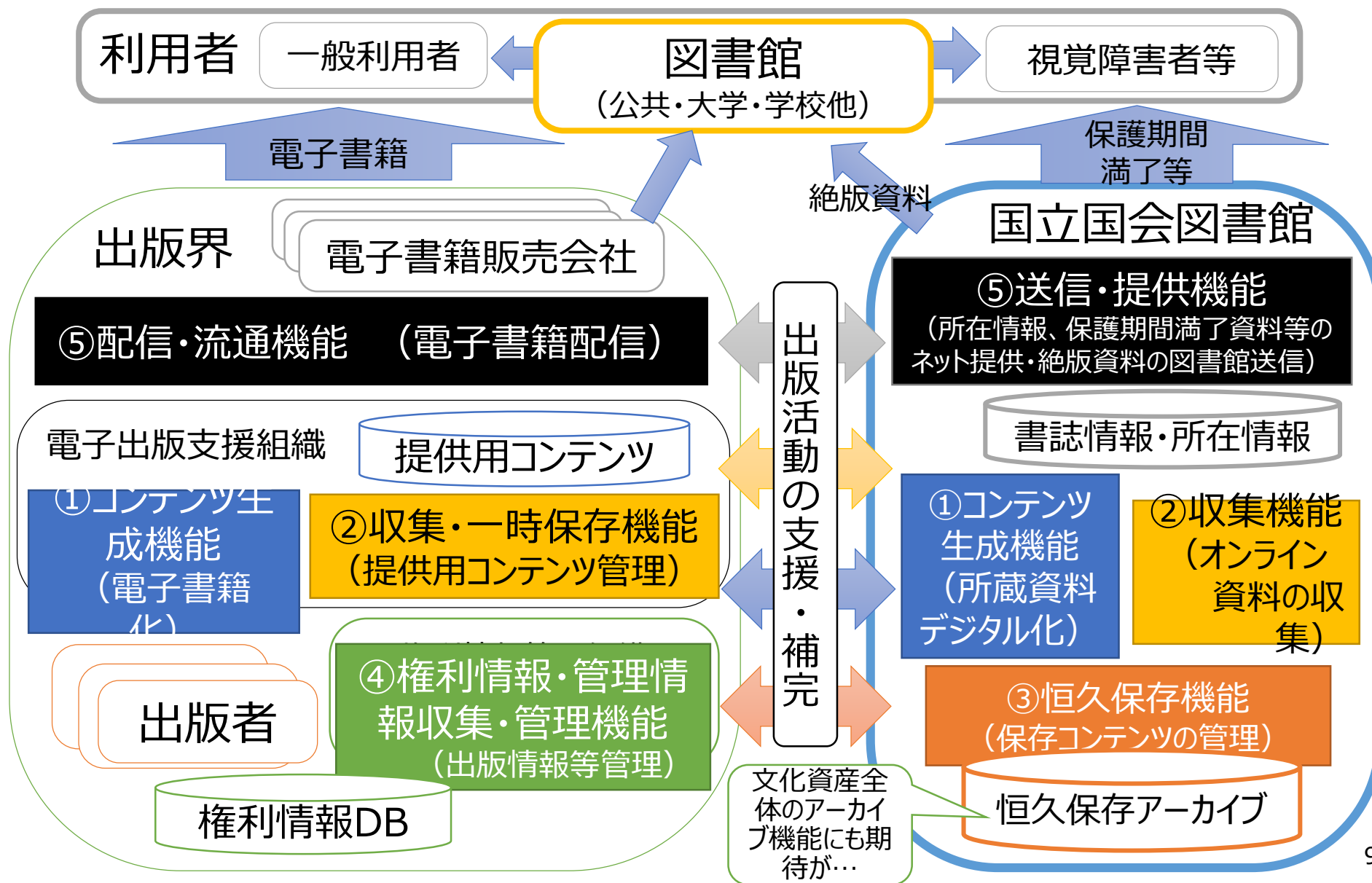
ナショナルアーカイブと構想全体のイメージ



書籍分野のナショナルアーカイブ

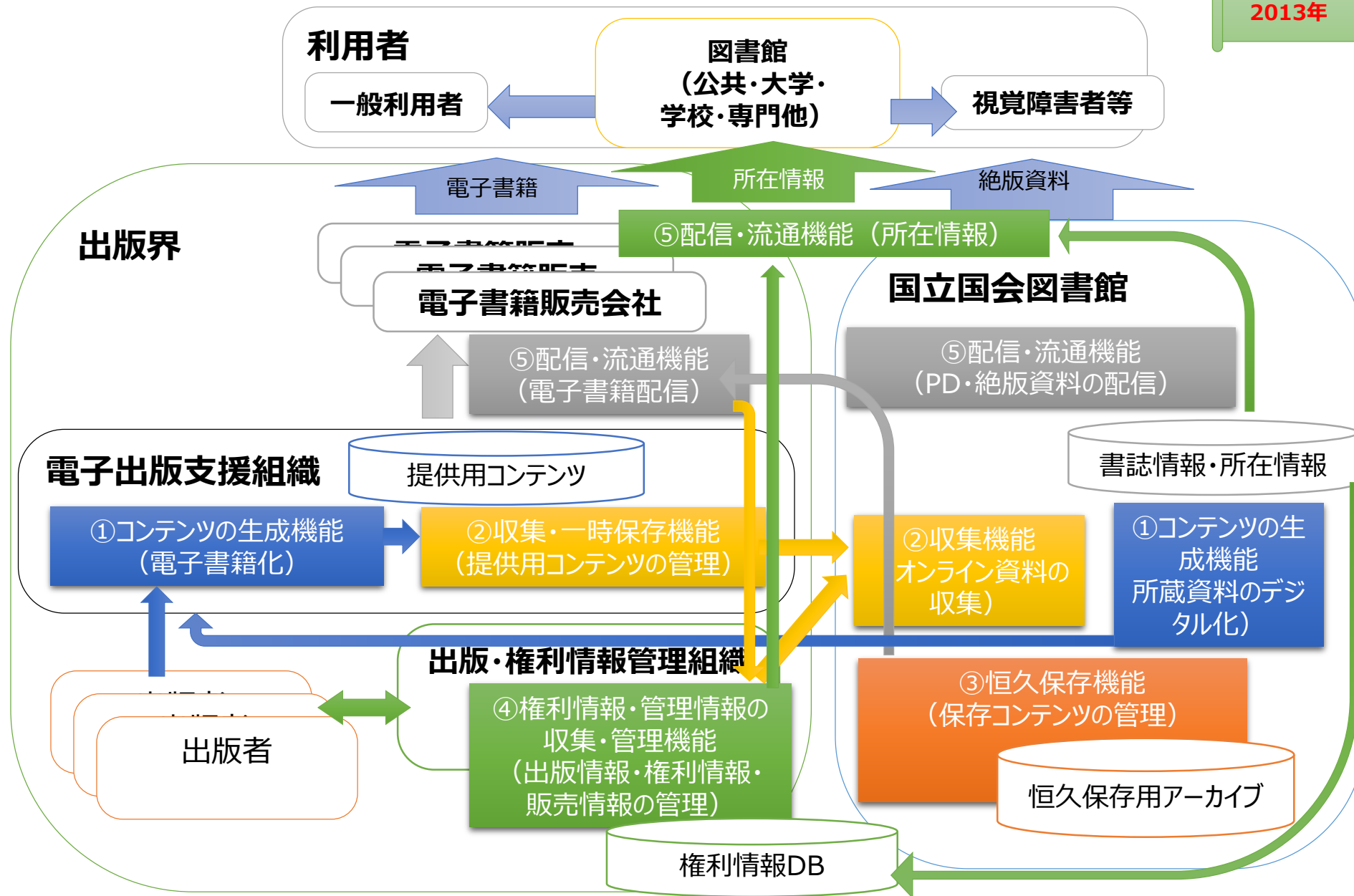
書籍分野のナショナルアーカイブの概念モデル

－ 出版界との役割分担 －

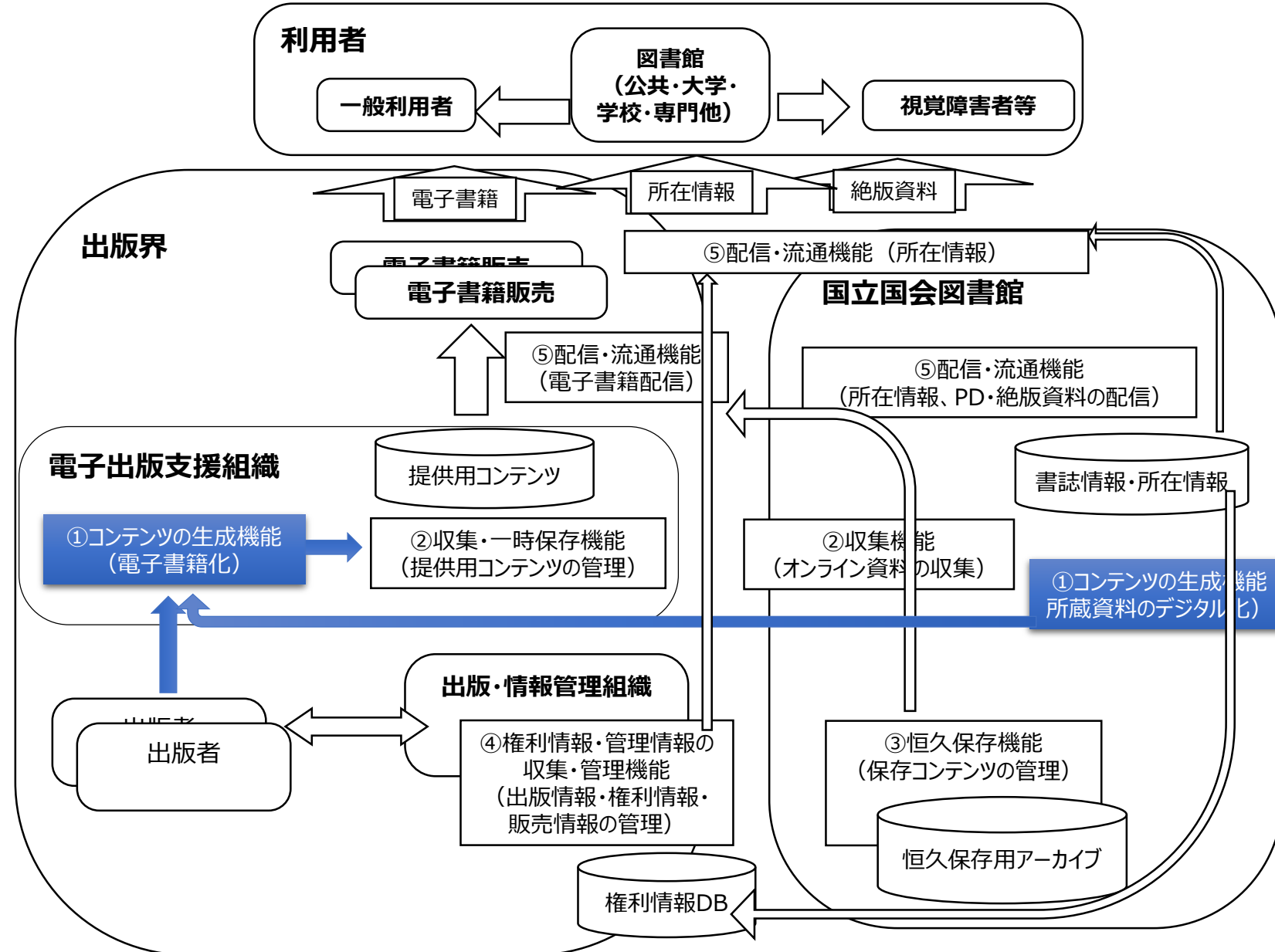


☆ 電子書籍分野のアーカイブの機能モデル

2013年

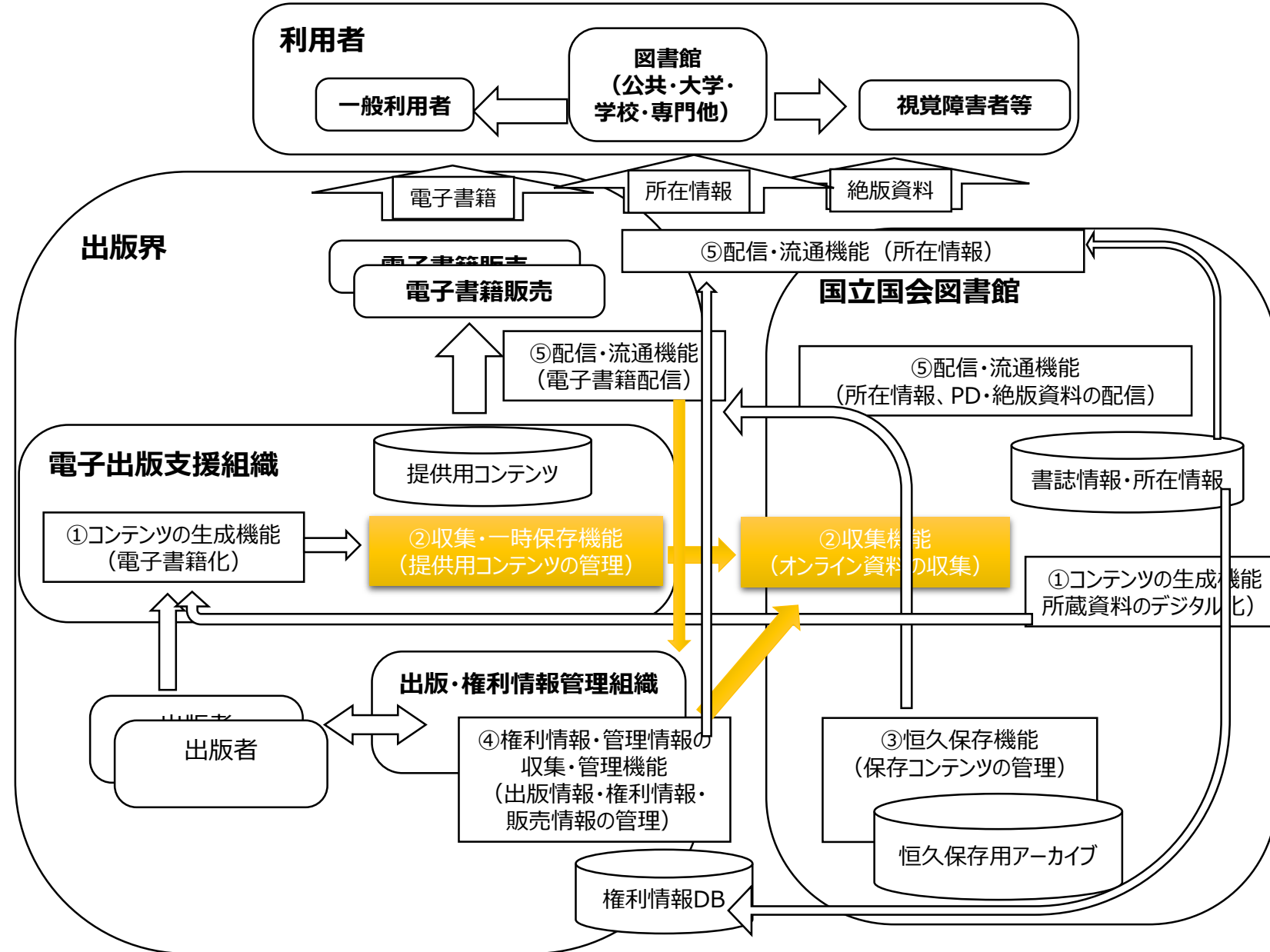


デジタルコンテンツの生成機能



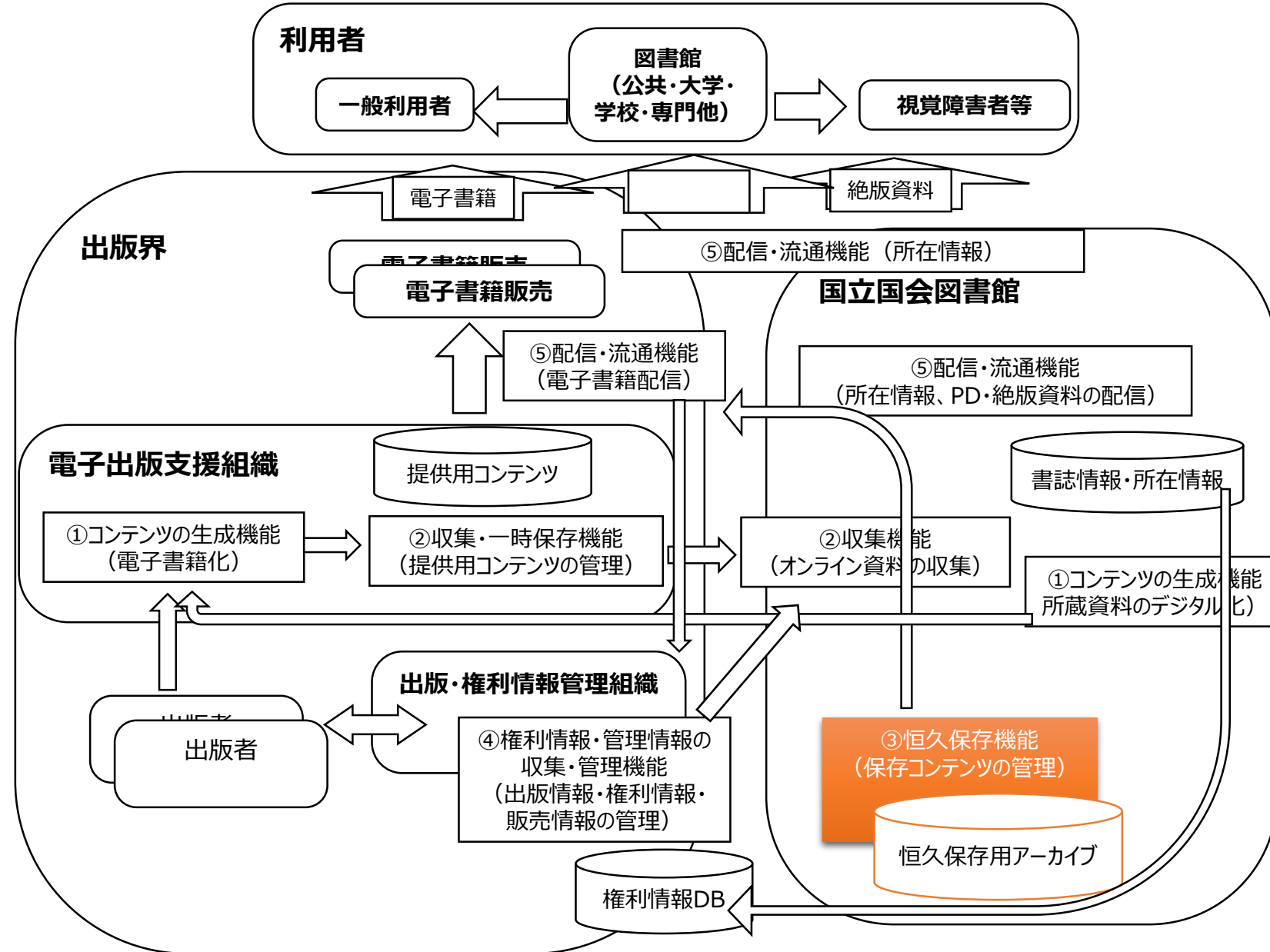
- 出版界
 - 電子書籍化
 - テキスト化
 - EPUB化
- NDL
 - 保存のためのデジタル化
 - 現在はイメージ化
 - 今後は検索のためのテキスト化
- 図書館
 - 郷土資料のデジタル化
- 連携
 - NDLイメージ化データの二次利用提供
 - 出版界での復刊のために
 - デジタル化仕様の共通化
 - EPUB仕様、画像・音声・動画仕様

電子書籍・書誌情報の収集機能



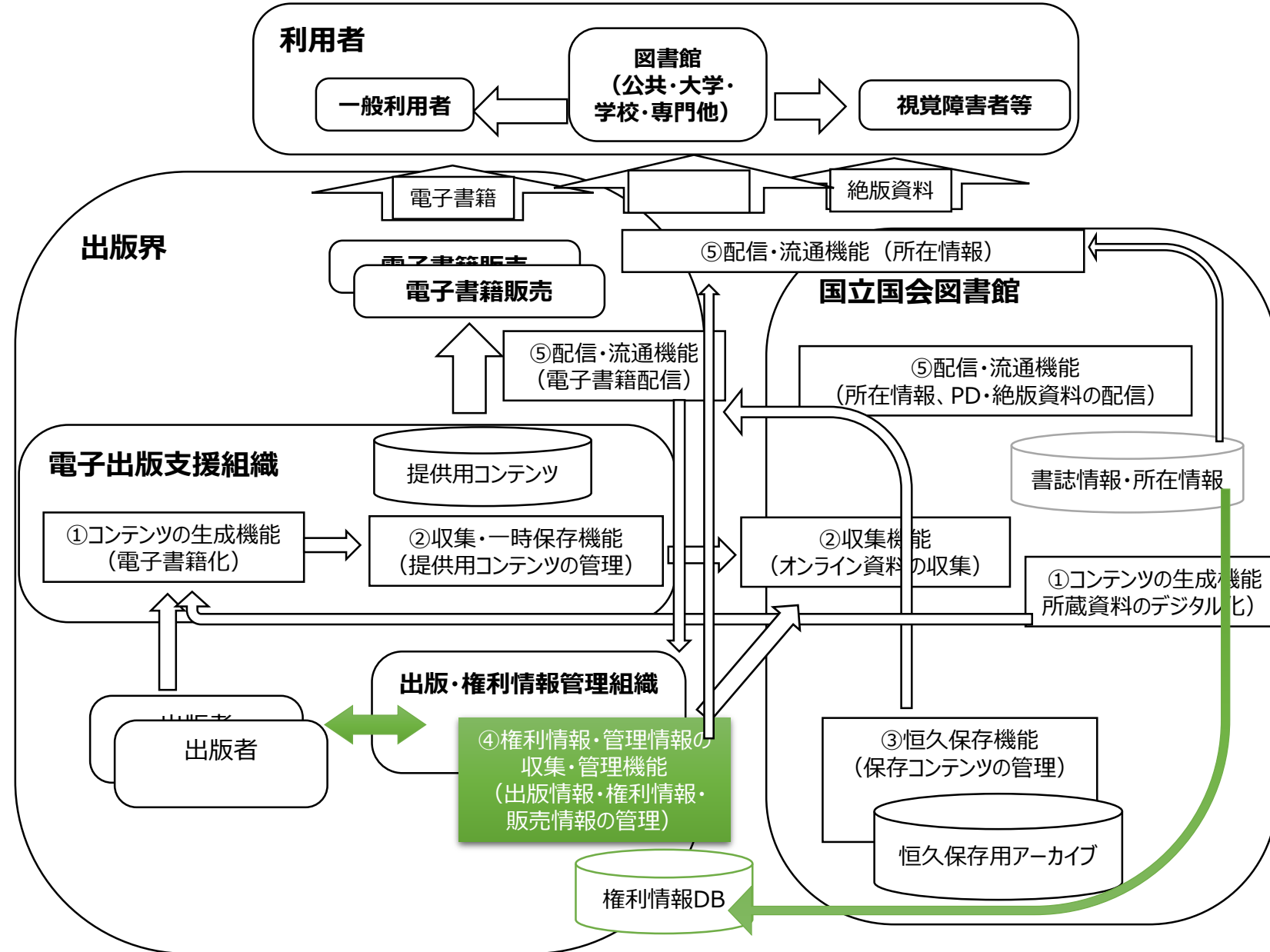
- 電子出版支援組織
 - 販売用コンテンツの保存
 - ビューアに依存しないEPUB
- 出版・権利情報管理組織
 - 出版情報DBの構築
 - 出版・出版サイト情報の提供プロトコル (API) の開放
 - (書誌、書影、出版情報、出版サイト情報)
- NDL
 - 全国書誌 (国内出版物の総合目録)、提供元情報の作成
 - 出版情報・提供サイト情報の収集
 - 書誌作成において、出版情報の活用 (私見)
 - 近刊情報、新刊情報
 - 公的機関のウェブ情報の収集の拡大
 - 民間無償オンライン資料収集の拡大
 - 民間有償オンライン資料の収集 (未実施)
 - 公的機関の情報のLinked 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
 - ONIX⇔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

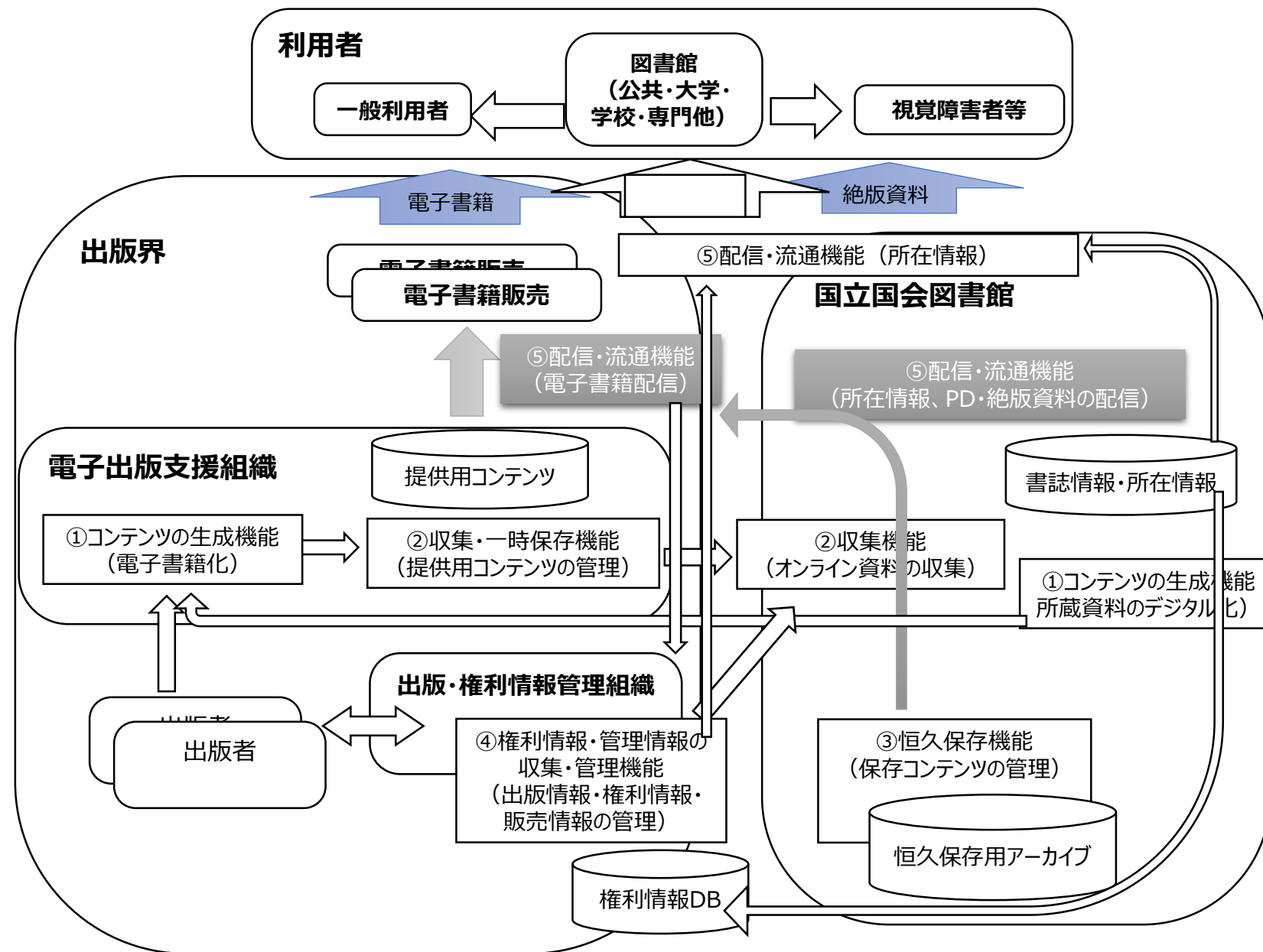
● NDL

- 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

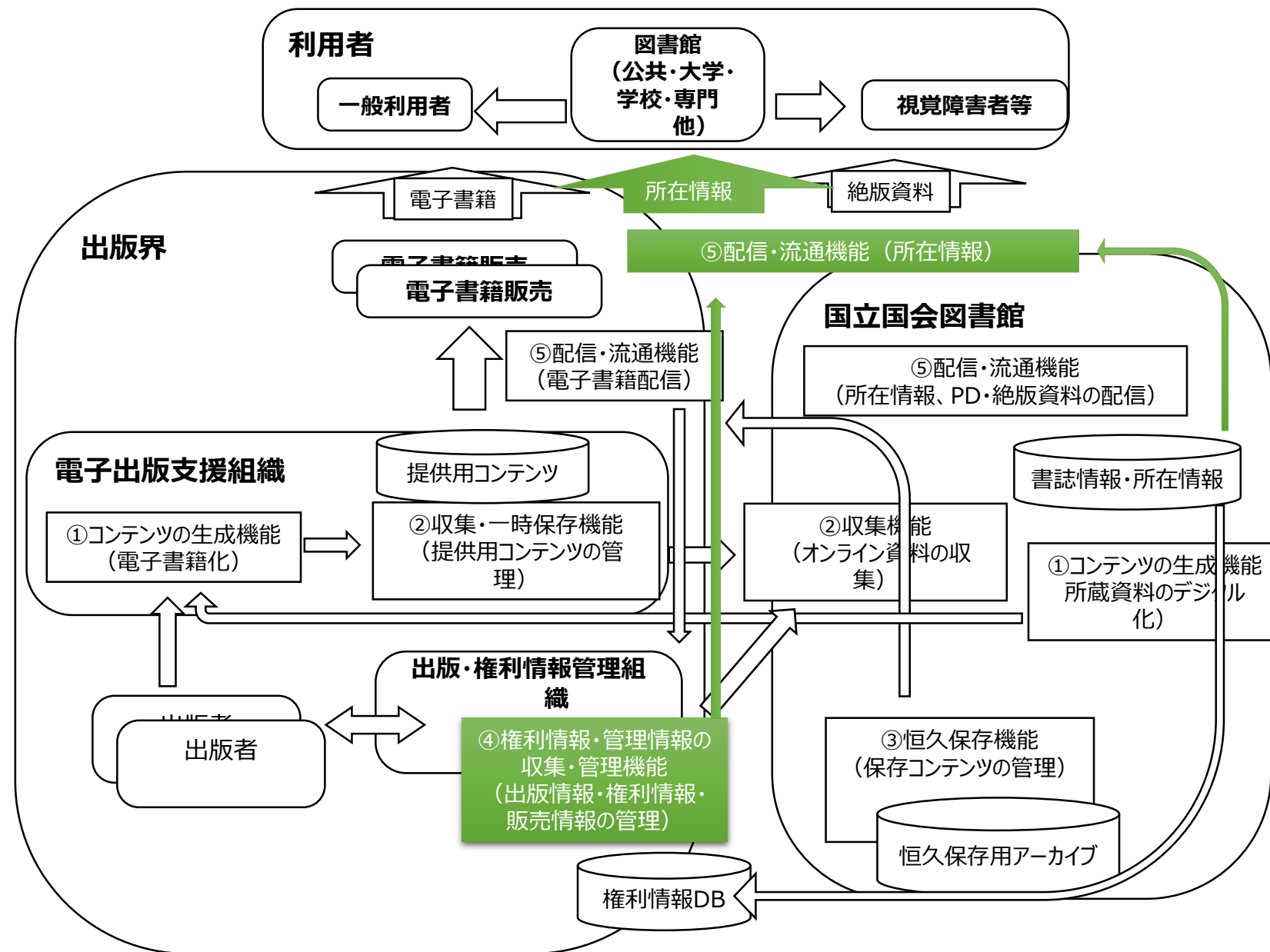
• NDL

- 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



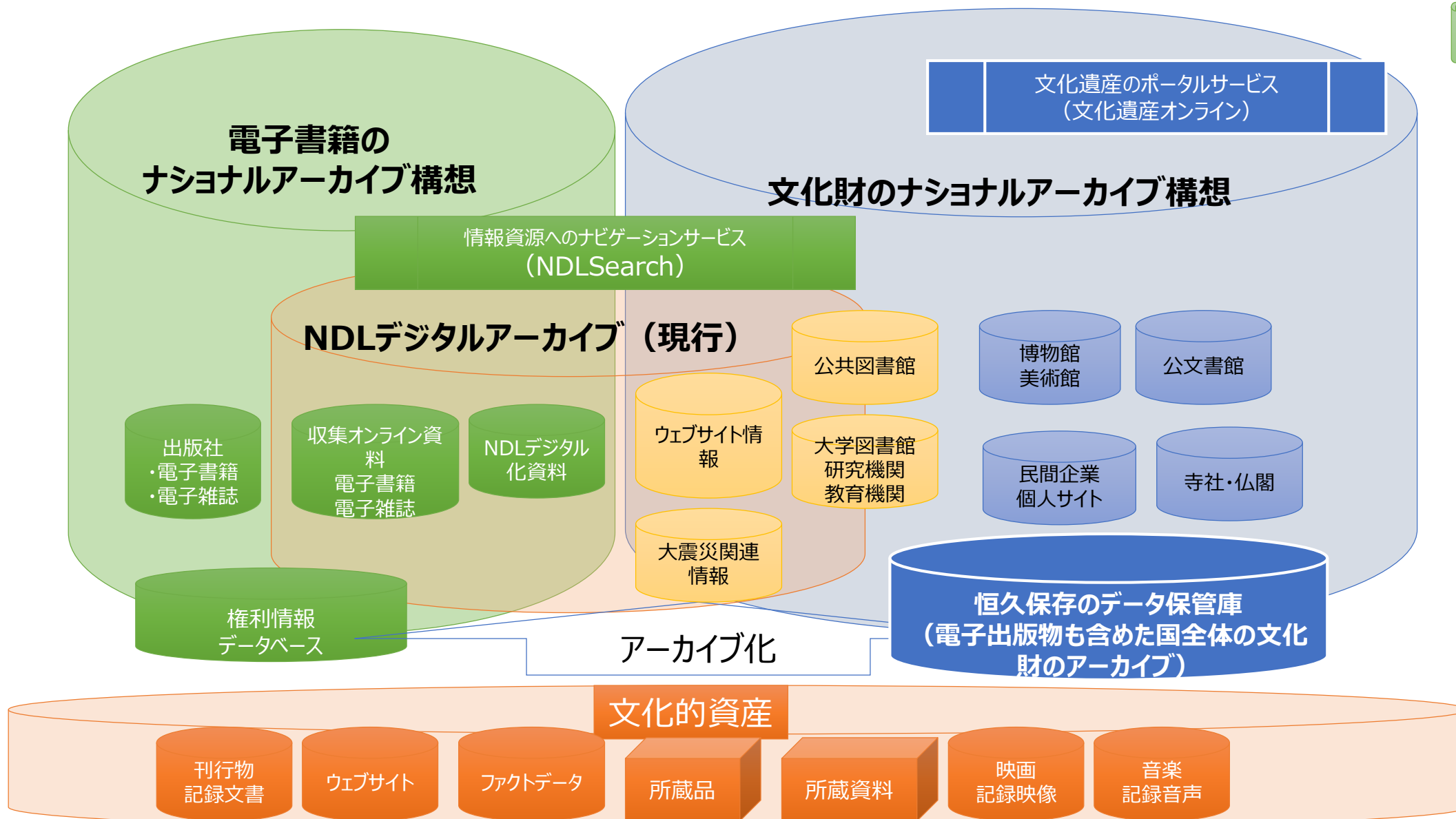
- 目的
 - 全ての出版物の存在を可視化
 - 利用者に対して、所在場所に関わらず、何らかの形で入手可能な著作物を網羅的に見つけ出せるようにする
 - 利用者の選択肢の確保
 - 利用者が必要とする著作物と、その入手先を自由に選択できるようにする
- 出版界
 - 絶版になっている出版物も含めて網羅的に検索できるサービスの提供
 - 販売促進
 - 商用出版物を優先表示
- NDL
 - 所蔵していない出版物も含めて網羅的に検索できるサービスの提供
 - 利活用の推進
 - 利用者が入手しやすい提供元を優先表示
- 連携
 - 利用者視点で、利用者が資料を探し出すために必要な情報を共有化
 - 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ッジ

2013年

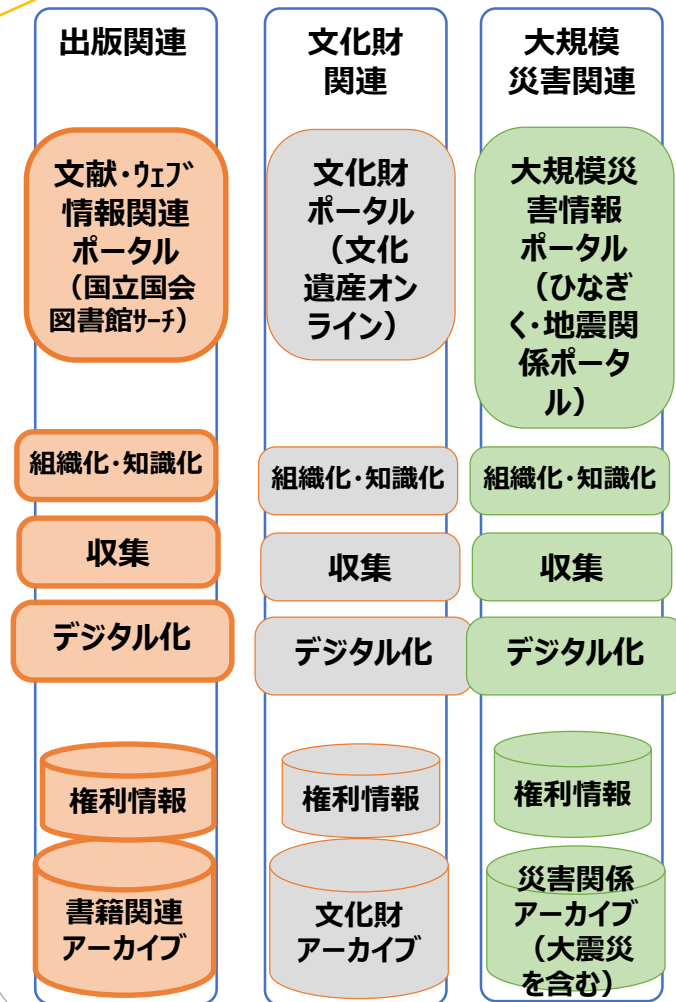


☆ 各種アーカイブ構築施策の一元化

2014年

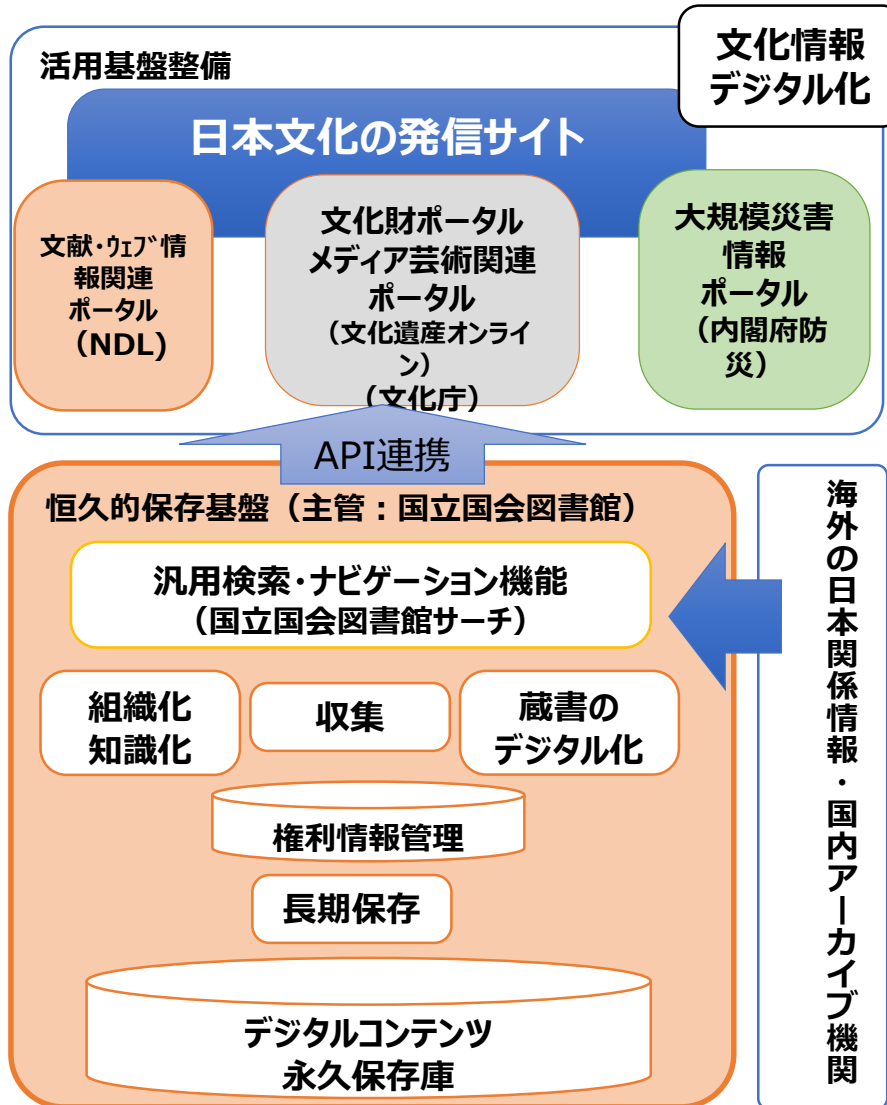
書籍, 電子書籍, 古典籍、メディア芸術、JJAPACON,

個別の情報基盤（個別所管体制）

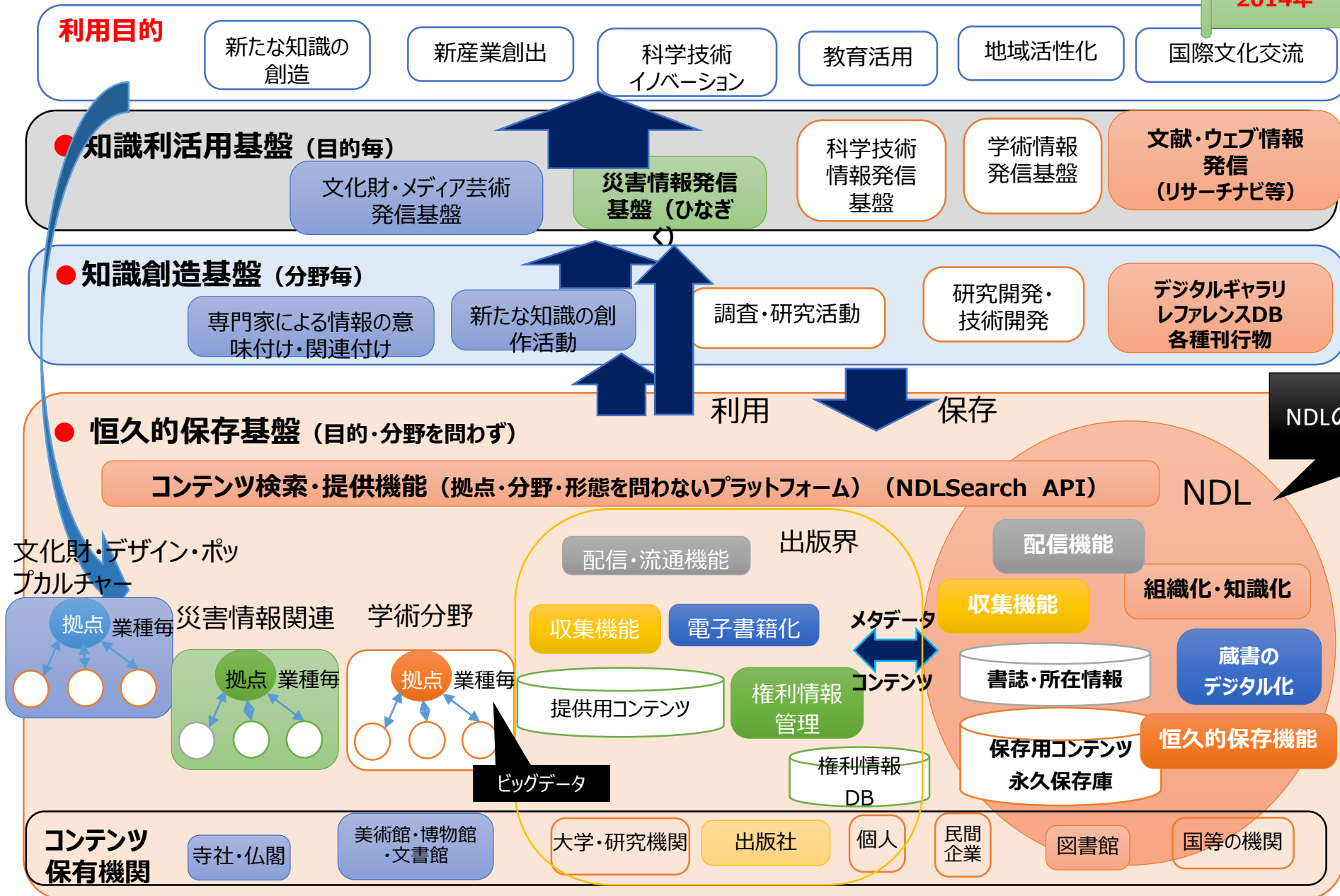


各施策の一元化

日本の文化情報基盤（司令塔：内閣官房？）



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

日本の文化情報発信の強化

●(2) 活用基盤の整備

…日本文化・記録の発信…

文献・ウェブ情報
発信基盤

文化財・メディア芸術関連
発信基盤

大規模災害情報
発信基盤

利用目的毎
に発信

●(1) 文化情報 デジタル化 の推進

平成26年7月8日
修正

コンテンツ創造基盤
新たな知識の創造（二次利用の活性化）

創造されたコンテンツ
の恒久的保存

③分野に依存しない検索・情報提供
インターフェース

●(3) 恒久的保存基盤の強化

④汎用検索・ナビゲーション機能

⑤組織化・知識化

⑥収集機能

⑦蔵書のデジタル化

⑧権利情報管理機能【民間と連携】

⑨長期保存研究

分野・利用目的を問わず
一元的に保存

⑪デジタルコンテンツ
永久保存庫

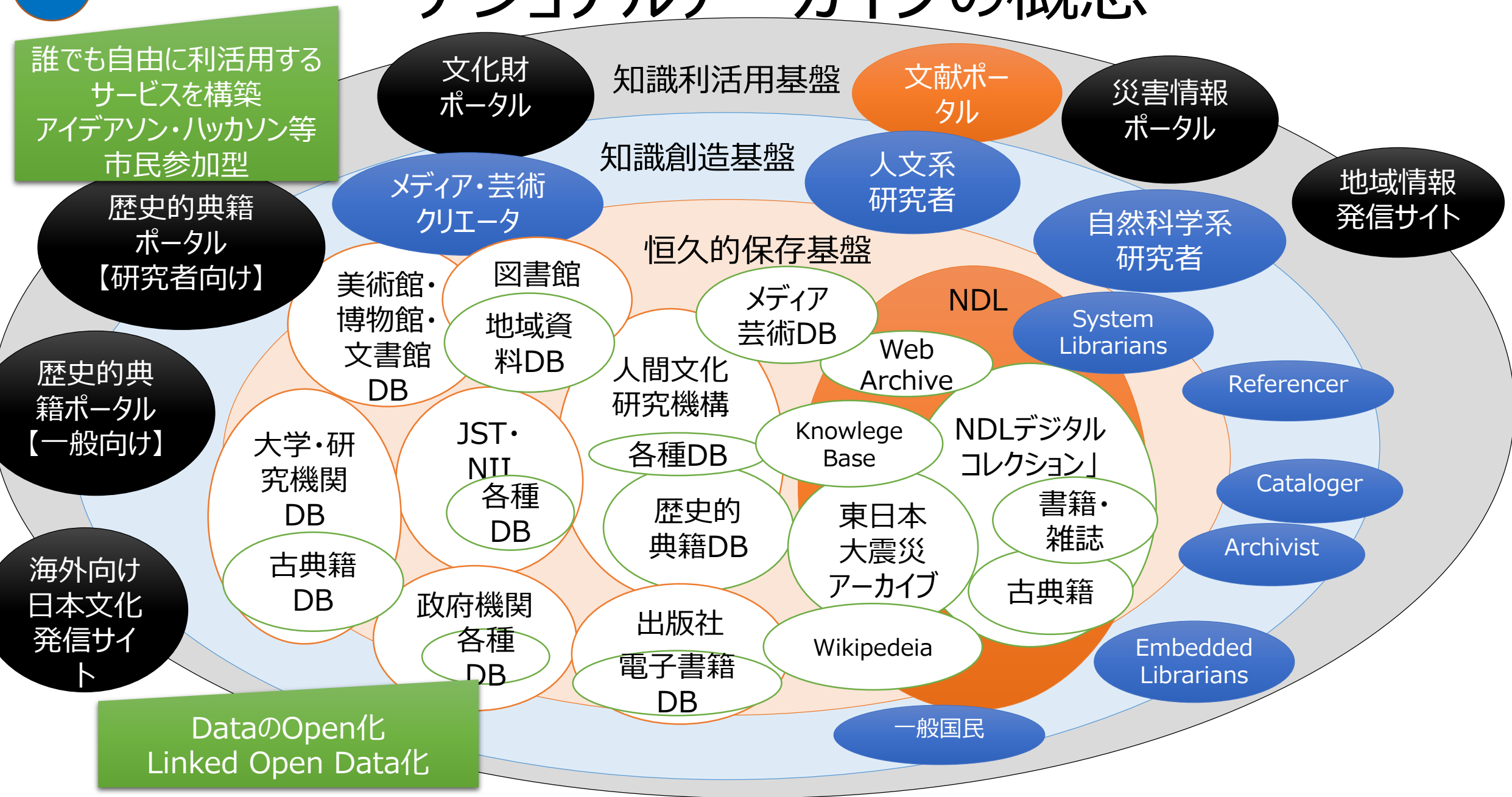
人材育成・
確保

研究開発・
技術開発

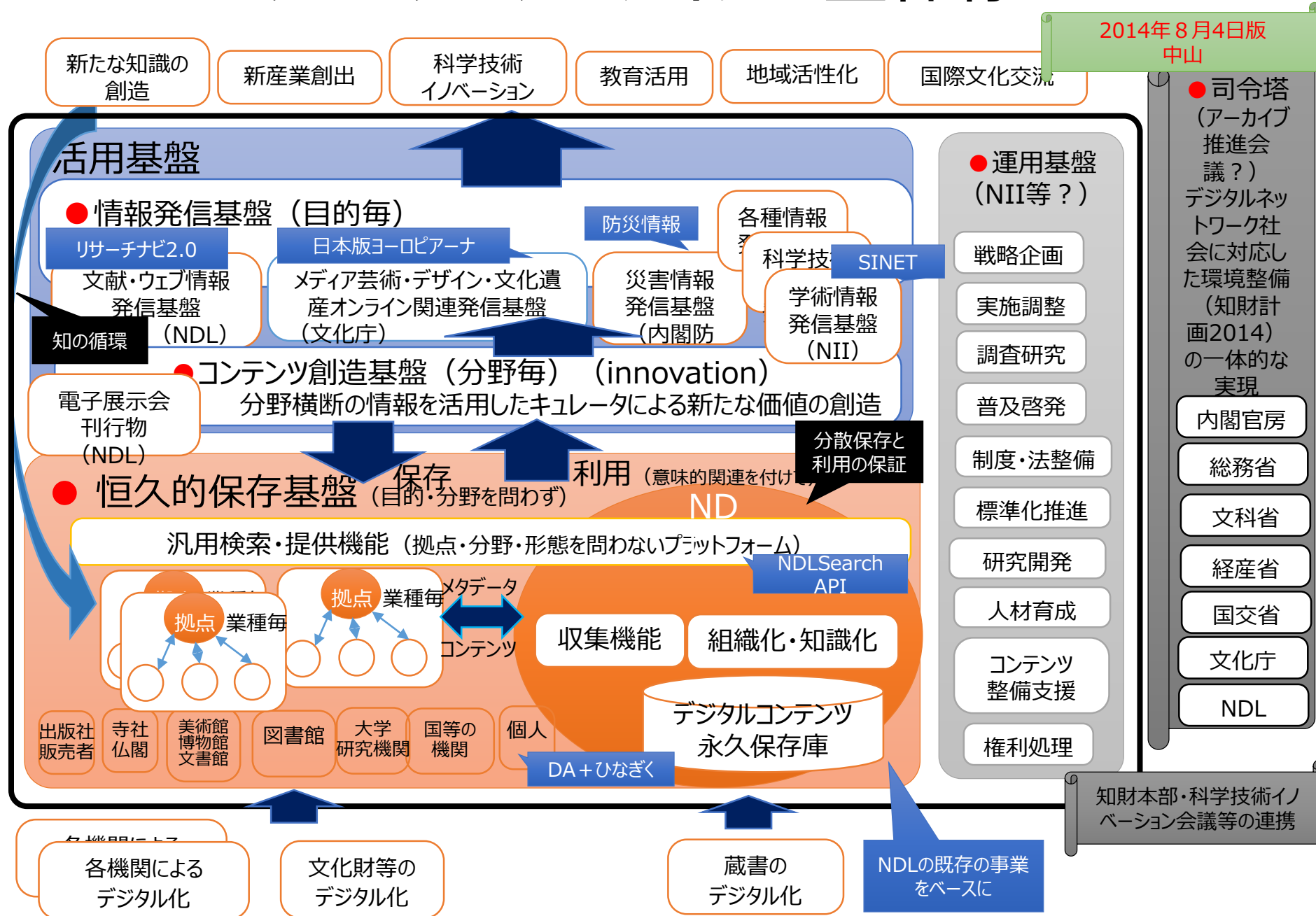
海外の日本
関係情報

国内
アーカイブ機関

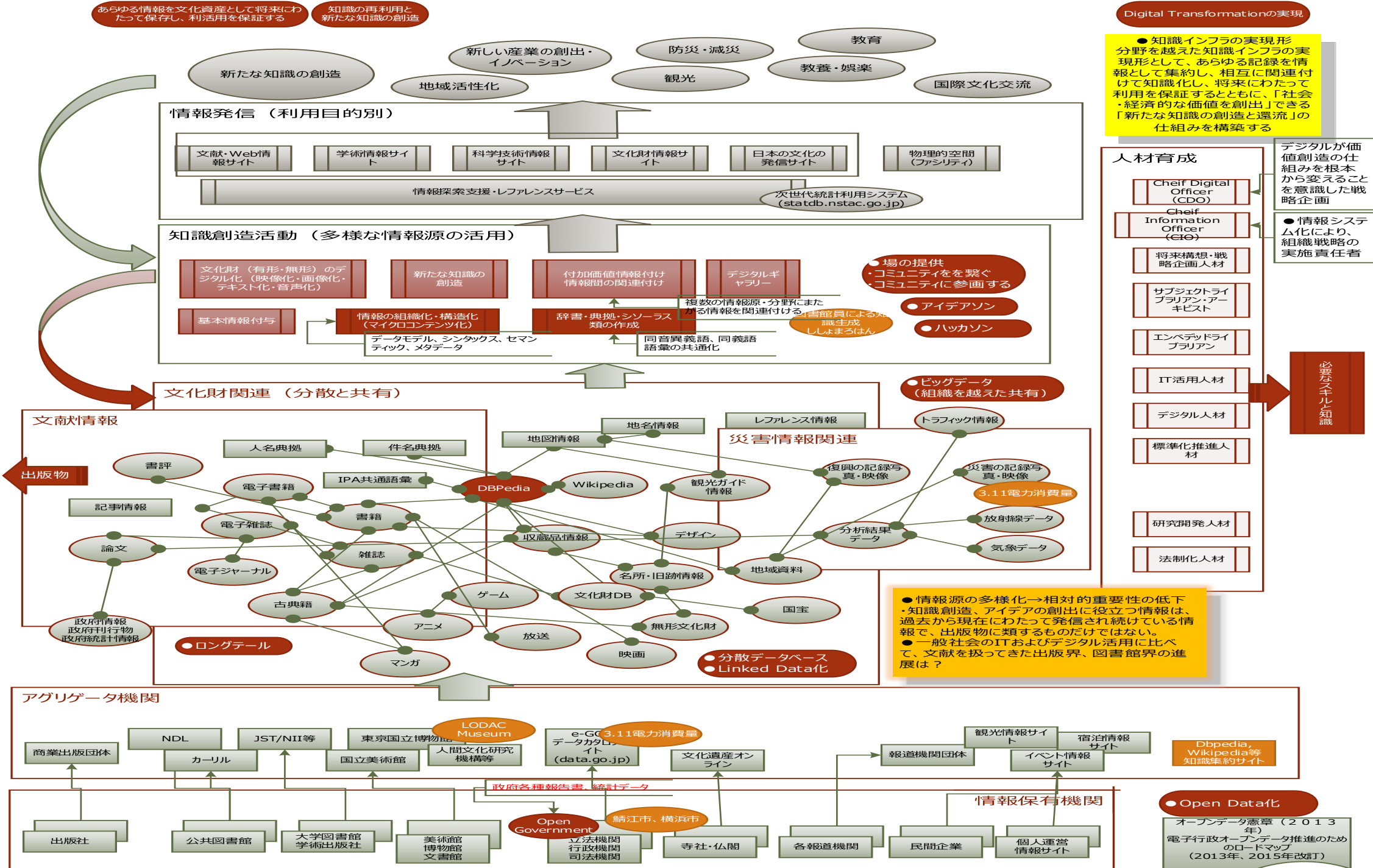
ナショナルアーカイブの概念



ナショナルアーカイブの全体像



国の文化資源のアーカイブ



ナショナル・アーカイブの検討に当たっての考察

2014年6月3日
文化庁有識者会議

「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、国全体のアーカイブとして大きな器の中で、各種アーカイブは分野の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。

• サービスの高度化

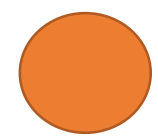
- デジタル化・収集・組織化・知識化・保存・提供の各フェーズの高度化に資する研究開発の促進と成果の活用
- システムエンジニア、デジタルアーキビスト、プリザベーションキュレーター、アーカイブとユーザを繋ぐコーディネータ等の人材の確保・育成

• 利用目的毎のポータル提供

- 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

• 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



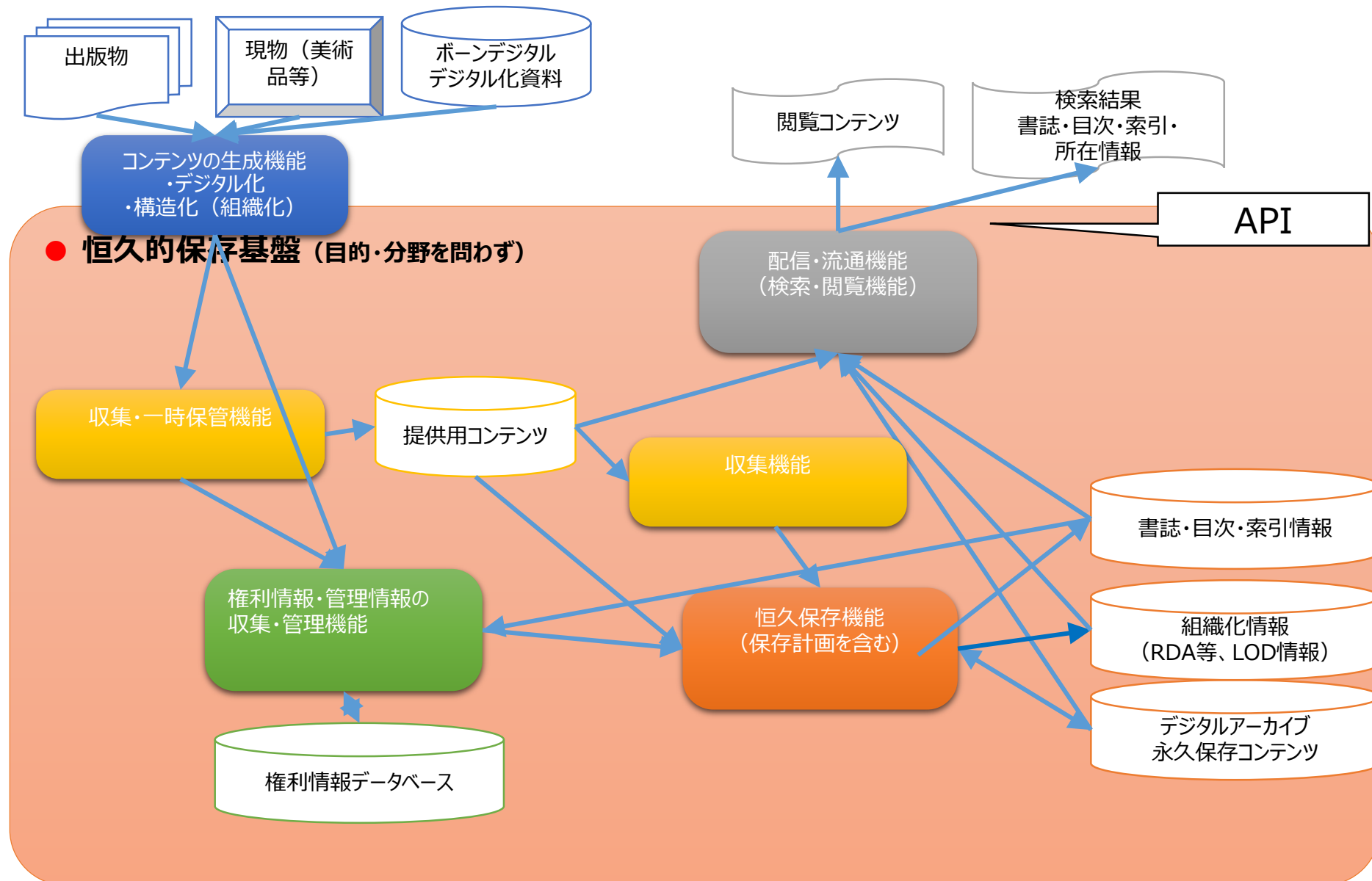
☆ ナショナルアーカイブの各基盤の概念

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のの
 - 表現形の種別、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たりできるように
 - 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
 - 様々な表現形の情報を生成するのは「創造」
- 収集
 - あらゆる表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）
 - あらゆる表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（NDL Searchの検索API機能＋DAの一次情報提供API機能）
 - データプロバイダー的機能
 - 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



コンテンツ創造基盤とは？

2014年8月6日追加

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ッ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

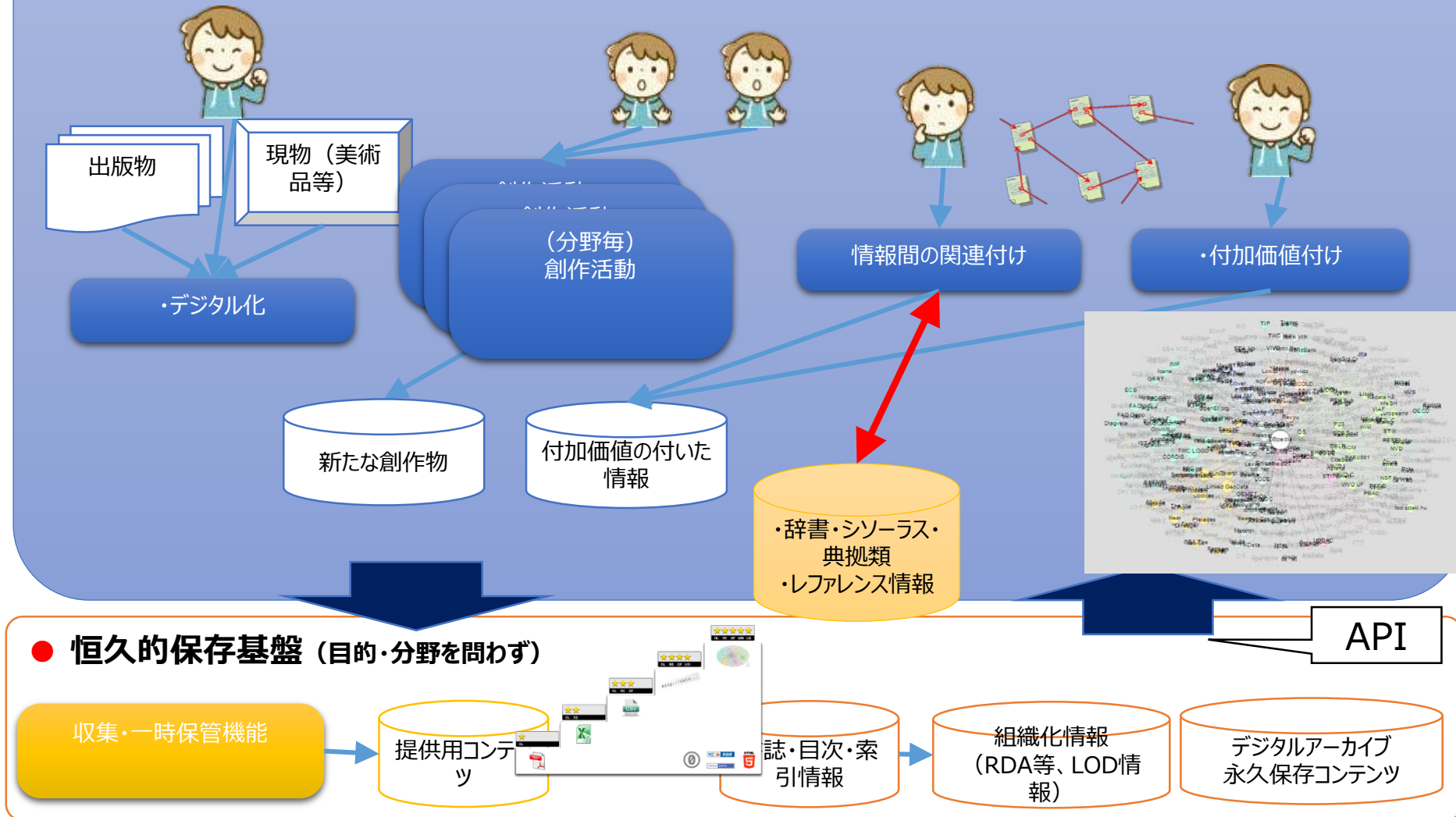
- 調査研究（新たな知識の創造）
 - 論文、プレゼン資料、高精細画像
- 教育用コンテンツ
 - デジタル教科書
- 教養・娯楽コンテンツ
 - ポップカルチャーコンテンツ

- 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

●コンテンツ創造基盤（分野毎）（innovation）



☆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 Foundation Japan
p://okfn.jp/)
ン、アイデアソン、アンカンファレンス
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
◇ アイデアソン
◇ アンカンファレンス

●●● 一般財団法人
●●● デジタル文化財創出機構
●●● Society for Digital Heritage

文化遺産オンライン
Cultural Heritage Online

➢ じんもんこん
➢ OpenGLAMじんもんこん

◇ SINET (Science Information NETwork) (文科省)

➢ ナショナルアーカイブ構築の司令塔、中核的組織
人 中間官民知的財産本部

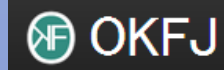


コンテンツ
創造基盤

人文科学とコンピュータ

Code4Lib JAPAN

OSS Community Dictionary



➢ オープンソース関連
◇ Code4Lib
◇ OSS Community (<http://ospn.jp/>)

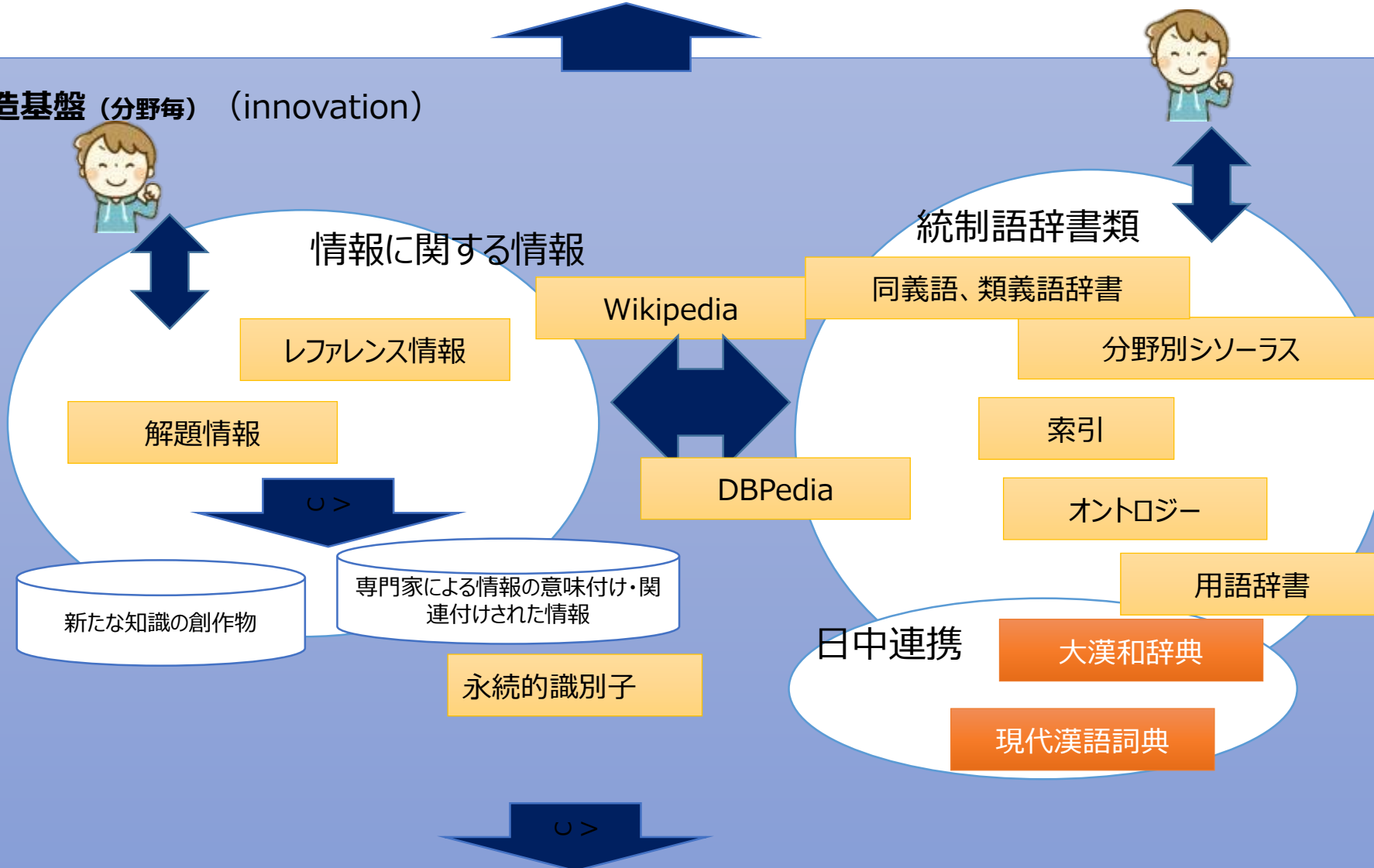


➢ MLA
◇ OpenGLAM
➢ 電子書籍 電子雑誌

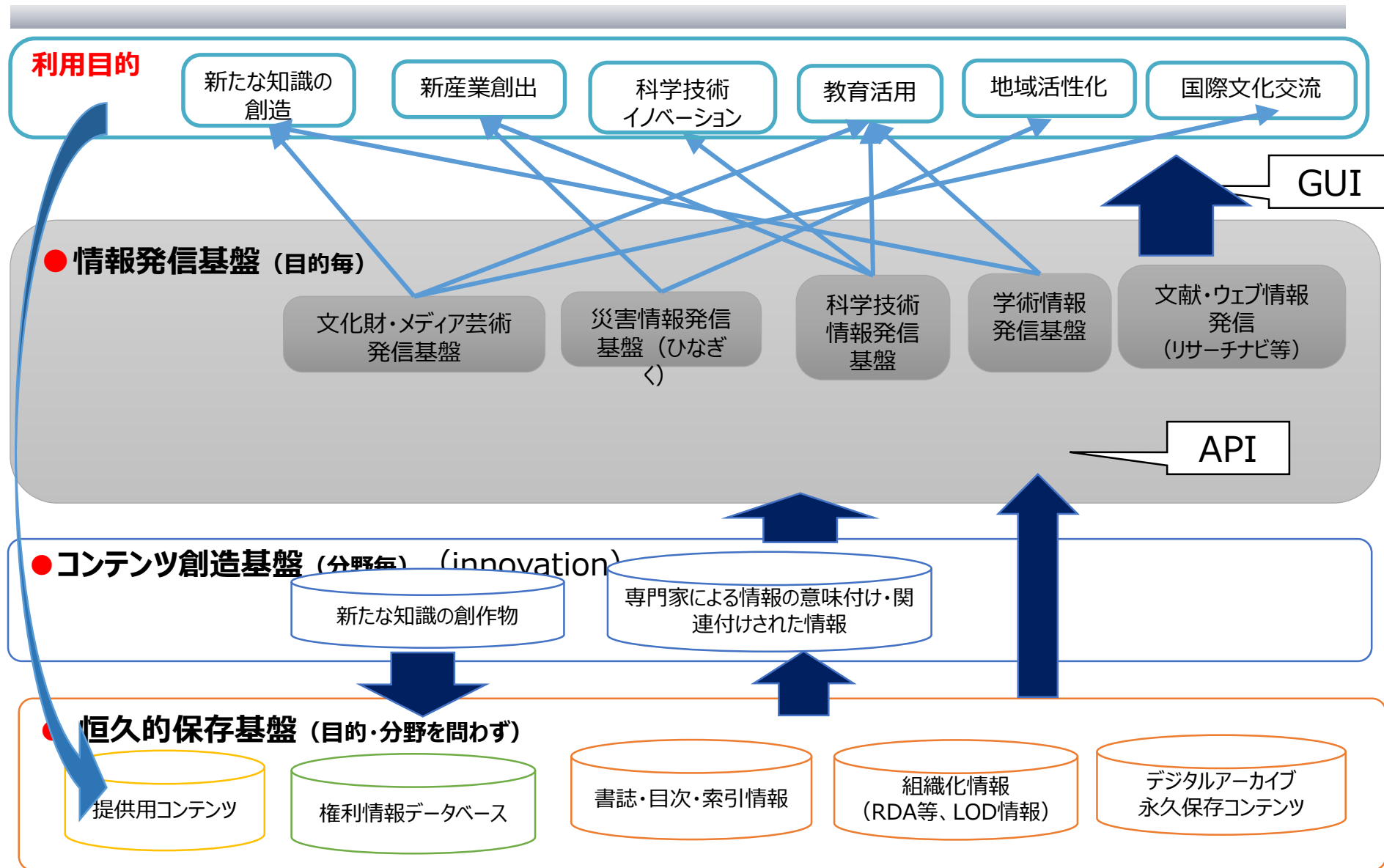


☆ 人を媒介して辞書と辞書を繋ぐー

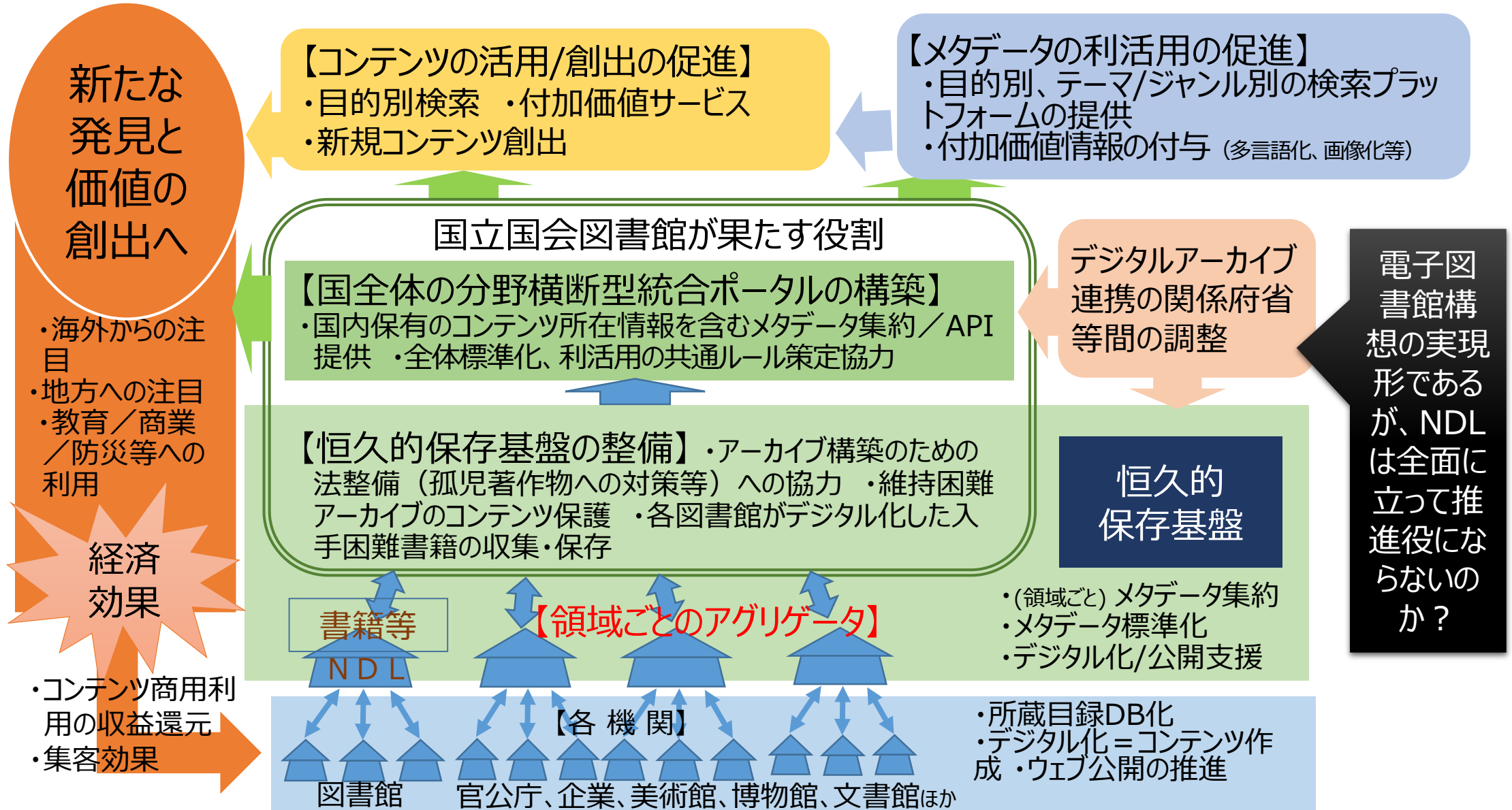
●コンテンツ創造基盤（分野毎）（innovation）




☆情報発信基盤

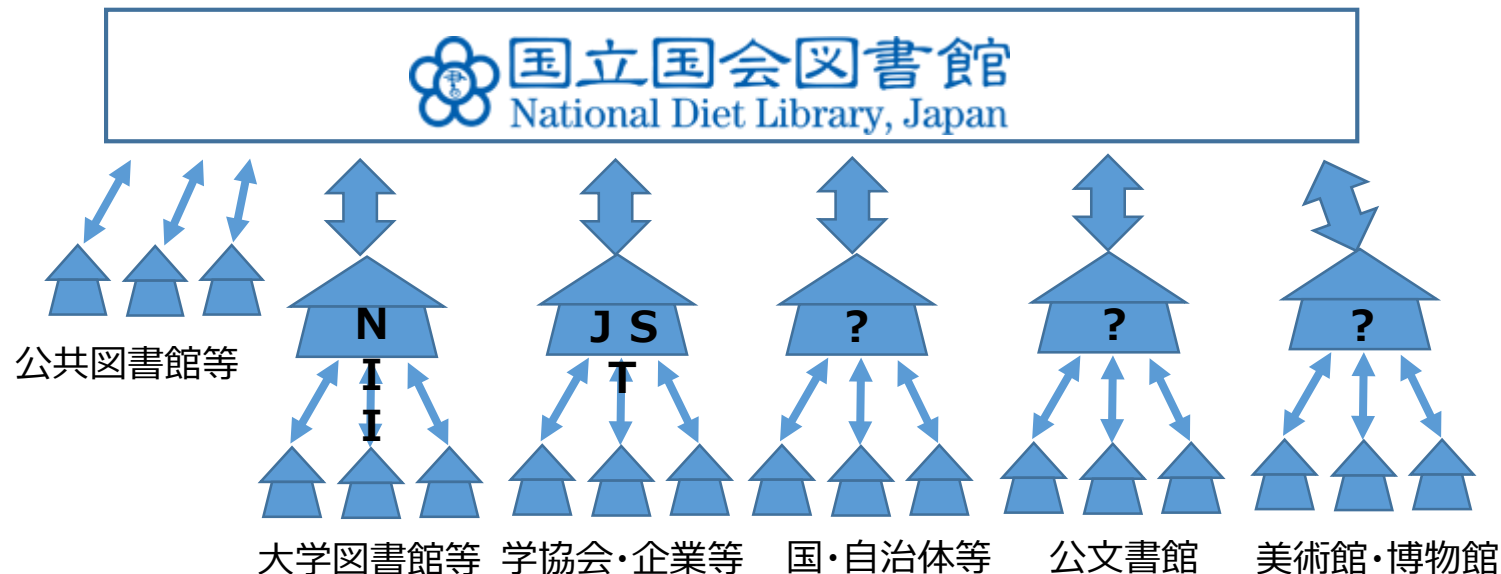


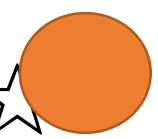
「ナショナルアーカイブ」の構築を目指して



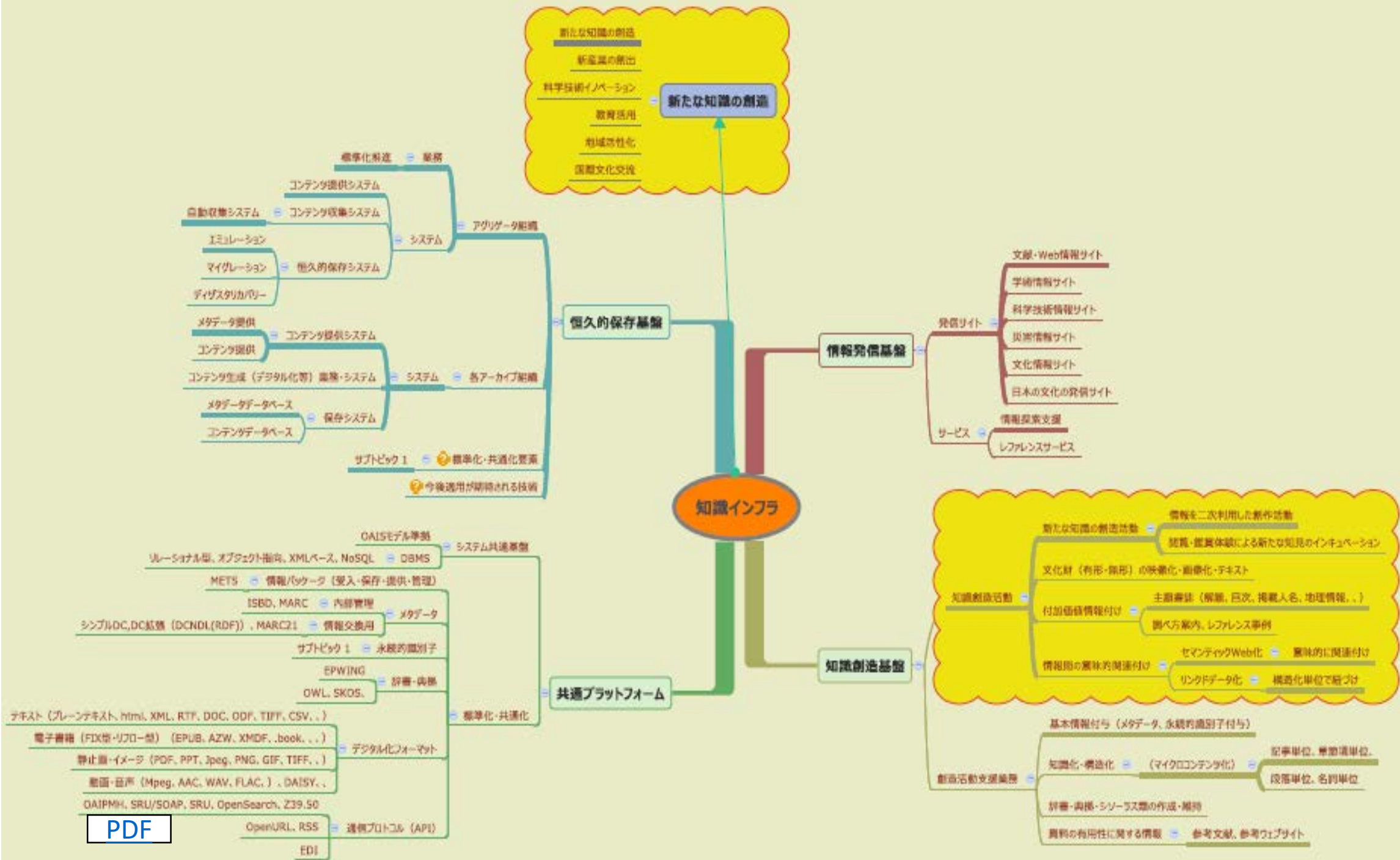
国立国会図書館が果たし得る役割

- 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
 - NDLラボでの研究活動

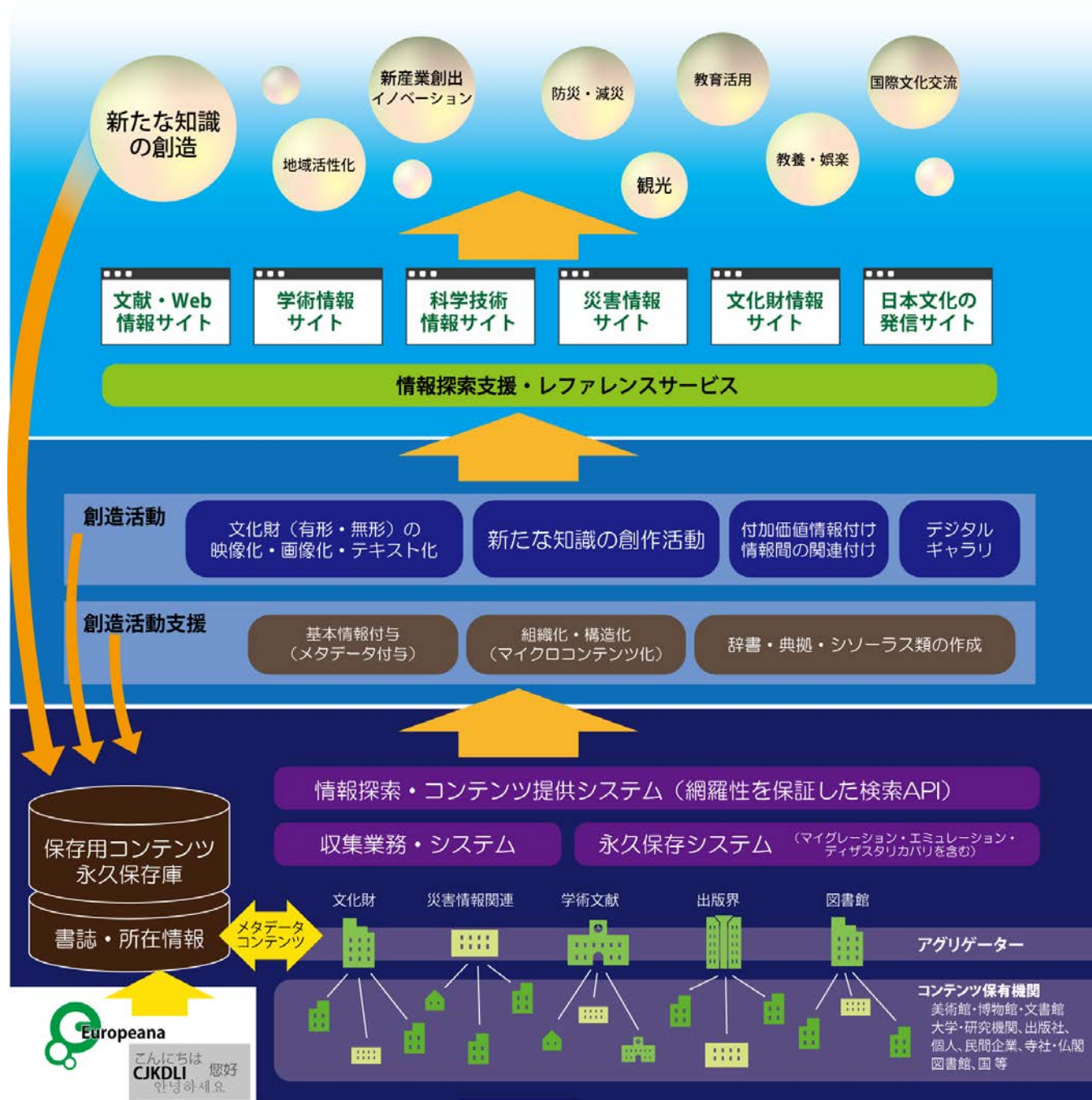




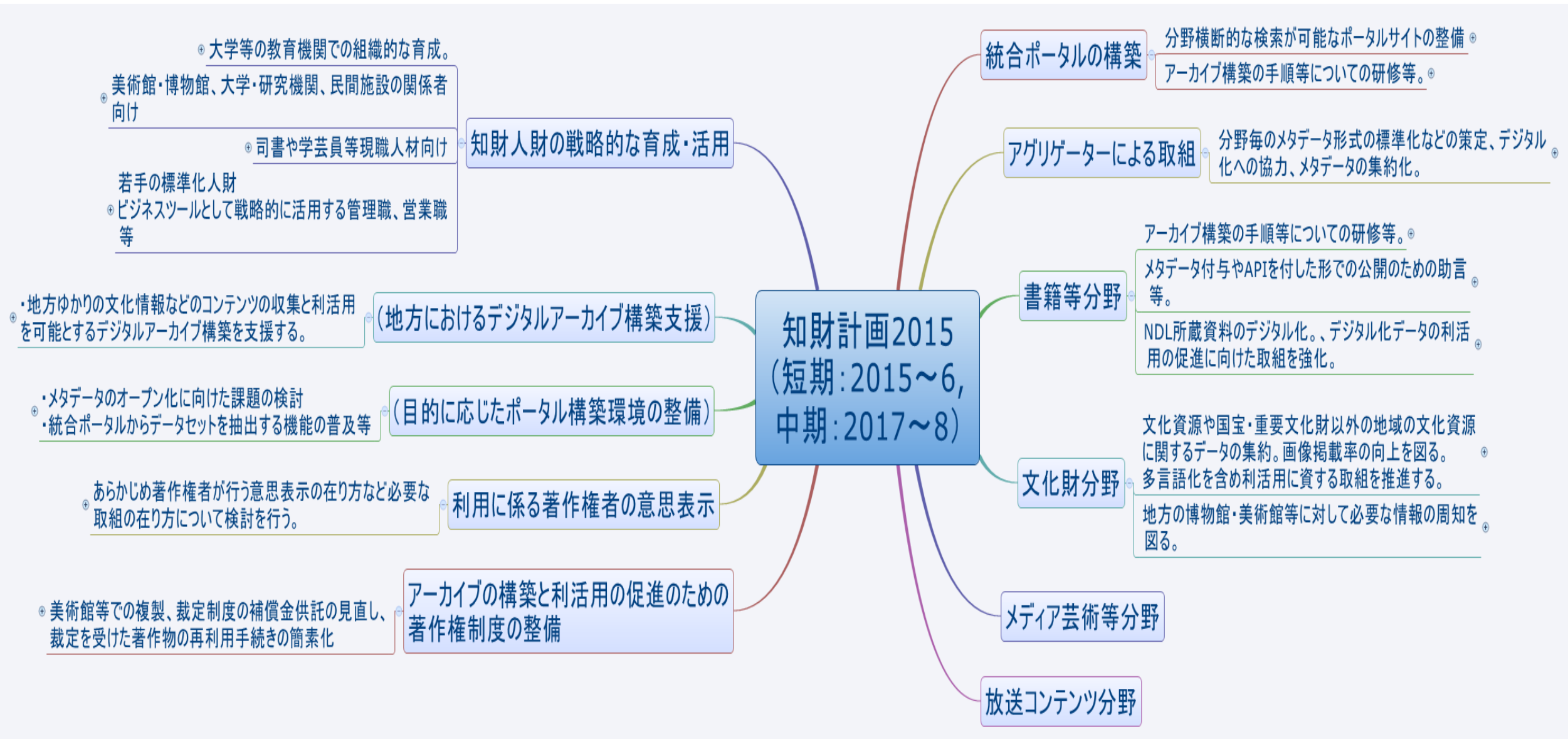
知識インフラ構築の構成要素



☆文化財を含めたナショナル アーカイブの構築イメージ



☆文化資産アーカイブ構築の一環で国が支援 【知財計画2015】



知識情報基盤の構築

