

デジタルアーカイブの現状と課題

ARC Days 2021

2021年7月31日

福島幸宏

慶應義塾大学文学部 図書館・情報学専攻 准教授

fukusima-y@keio.jp

<https://researchmap.jp/fukusima-y/>

自己紹介

- 学生・院生時代は日本近現代史を専攻
 - 高知県出身・島根大学・京都府立大学大学院・大阪市立大学大学院
 - 神社史、地域社会に関心／自治体史調査・フィールドの調査を経験
- MLAの職員・研究者として
 - 京都府立総合資料館歴史資料課／庶務課 2005年4月～2015年3月
 - 京都府立図書館企画調整課 2015年4月～2019年3月
 - 東京大学大学院情報学環特任准教授 2019年4月～2021年3月
 - 慶應義塾大学文学部准教授 2021年4月～
- 京都府職員として関わったこと
 - 京都府行政文書（重要文化財）の管理運用 → 20世紀以降の紙資料で初の重文指定資料／科研による検討／修理事業の開始
 - 戦時期・戦後の行政文書の公開 → 京都の戦時期／占領期研究の進展
 - 文化庁 指定文化財（美術工芸品）企画・展示セミナー修了 → 学芸員としての中核的な研修
 - 「[京都市明細図](#)」の公開 → 京都の街歩き事業の基盤に（福島ほか2012）
 - 「[東寺百合文書（国宝）](#)」のweb公開 → CC BYで公開／Library of the Year 2014大賞受賞／世界記憶遺産に（福島2014）
 - 京都府立図書館サービス計画策定 → 図書館協議会の設置／評価基準を検討
 - 都道府県図書館の横断検索システムの高速化 → カーリルのシステムを導入
 - [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](#)
[メタデータのオープン化等検討ワーキンググループ構成員](#) → ジャパンサーチへ
 - 京都府立図書館貴重書コレクションの構築 → IIIF+DOI+CC0 の組み合わせは国内初
- 現在
 - 慶應義塾大学文学部准教授／東京大学大学院情報学環客員准教授
 - これからの学術情報システム構築検討委員会委員
 - 日本歴史学協会常任委員／デジタルアーカイブ学会理事／日本アーカイブズ学会委員 など
 - MLA（博物館・図書館・公文書館）連携・知識情報基盤・デジタルアーカイブについて検討

本日の構成

- 1 はじめに
- 2 デジタルアーカイブの現状
- 3 デジタルアーカイブの課題
- 4 おわりに

Ⅰ はじめに

デジタルアーカイブ：現在のいくつかの定義

- 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**
 - 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2017/2020
- 有形・無形の文化財をデジタル情報として記録し、劣化なく永久保存するとともに、ネットワークなどを用いて提供すること。最初からデジタル情報として生産された文化財も対象となる。
 - 日本図書館情報学会用語事典編集委員会編2020
- デジタルアーカイブというのは、さまざまな記録知とか記憶知、文書も入りますけれども、文書だけではない、記録や記憶が集積している場所です。その中には、証言とかネット情報とか写真とか映像とか地図とか設計図とか脚本とか録音とかノートとか、いろいろなものが入ってくる
 - 吉見俊哉2017
- 「デジタルアーカイブ」の定義や要件を現時点では厳格に定めることは困難であり、むしろ今後の発展の可能性を縛ることになり望ましくない
 - 柳与志夫2020

デジタルアーカイブ：現段階の福島の定義

- 要件としては以下（スリムモデル（後述）+ α ）
 - 利用規約の明示：基本中の基本であるとともに、二次利用促進のため
 - 機械可読性の担保：データ流通・二次利用促進のため
 - 環境に依存しないデータ移行性の担保：特定の環境・システムに依存しないため
 - アクセシビリティの確保：多様な環境での利用を可能とするため
 - 真正性の確保：データの由来と改変の経緯の記録
 - 永続性の保障：システムではなく社会的な仕組みに依拠して
- 社会が遺すことを選択した／すべき知識情報基盤としてのデジタルデータとそれにまつわる仕組みの総体。真正性や永続性の確保、万人へのアクセス保障がその要件となる。
 - 文章で定義せよ、となれば上記の通り
- デジタルアーカイブは、未だ独自の領域を確立できてない
 - 概念としても
 - 学問分野としても

2 デジタルアーカイブの現状

2010年代以降の状況

- アーカイブ／デジタルアーカイブの試みの〈氾濫〉
 - 1995と2011のインパクト:被災写真資料(失われた風景)への注目／アーカイブに「現在を記録する」意味が追加
 - 利活用への大きなシフト(古賀2017)
 - 現在も各地での構築が日々行われている
- オープンデータと参加型編集
 - オープンデータ:クリエイティブコモンズライセンスを活用した展開(福島2014)(澤谷2018)
 - 参加型編集:WikipediaTOWN(是住2015)(福島2018)
- デジタルヒューマニティーズの展開
 - 歴史学分野での進展(後藤・橋本2019)により理論的・実践的に一定のまとまりが見えている段階
 - 古文書翻刻のクラウドソーシングである「みんなで翻刻」(2017)
- 統合／横断の動き
 - 組織内:京の記憶アーカイブ(2015)／東京大学学術資産等アーカイブズポータル(2019)／Keio Object Hub(2021)
 - 組織外:越後佐渡デジタルライブラリー(2011)／群馬県立図書館デジタルライブラリー(2021)

オープンデータ

- デジタルアーカイブと非常に親和性が高い
 - ライセンス優先／機械可読性重視の点
- オープンデータの定義
 - 「オープンデータ基本指針」（高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議 決定：2017年5月30日）
 - 二次利用が可能な利用ルールで公開
 - 機械判読に適したデータ形式
 - 無償で利用できるもの
- キーワード
 - 二次利用
 - 機械判読
 - オープンデータの5つ星
 - クリエイティブ・コモンズ・ライセンス

「オープンデータの5つの段階(出典:★)」と、データ形式

段階	公開の状態	データ形式 例	参考) Linked Open Data 5star	
1段階	オープンライセンスの元、データを公開	PDF、JPG	OL – Open License (計算機により参照できる(可読))	人が理解 するための 公開文書 (編集不可)
2段階	1段階に加え、コンピュータで処理可能なデータで公開	xls、doc	RE – Readable (Human & Machine) (コンピュータでデータが編集可能)	公開文書 (編集可)
3段階	2段階に加え、オープンに利用できるフォーマットでデータ公開	XML、CSV	OF – Open Format (アプリケーションに依存しない形式)	
4段階	Web標準(RDF等)のフォーマットでデータ公開	RDF、XML	URI – Universal Resource Identifier (リソースのユニーク化、Webリンク)	機械判読 可能な 公開データ
5段階	4段階が外部連携可能な状態でデータを公開	LoD、RDF スキーマ	LD – Linked Data (データ間の融合情報が規定。検索可能)	

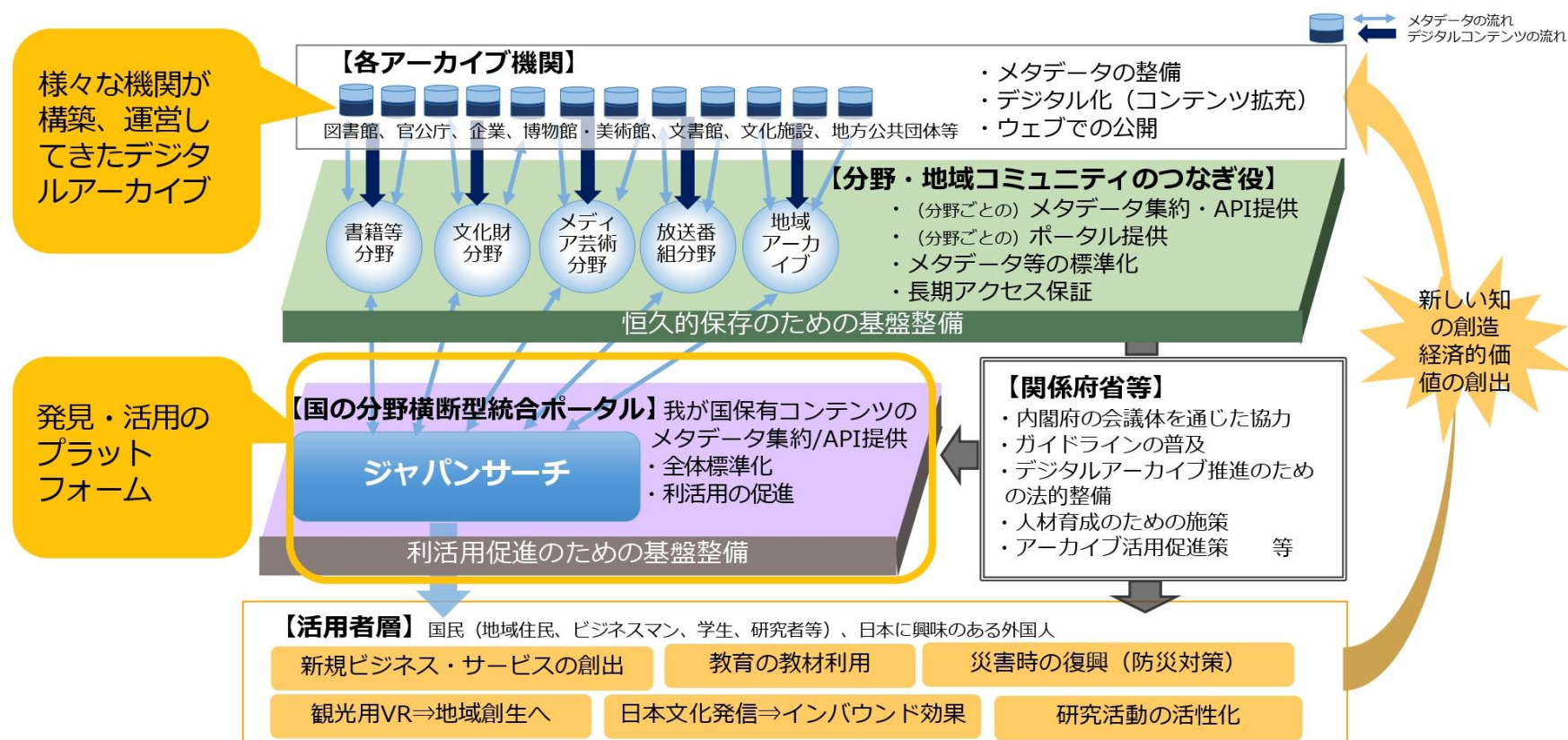
オープンデータの5つの段階

出典：★ Open Dataのサイト (<http://5stardata.info/>) およびTim Berners-Lee氏のLinked Dataに関する提言ページ (<http://www.w3.org/DesignIssues/LinkedData.html>) を参考に作成。

所与の条件としてのジャパンサーチ

- 2020年8月に正式公開された、「我が国が保有する多様なコンテンツのメタデータをまとめて検索できる「国の分野横断型統合ポータル」」
- ここへの対応を行えば、他の様々なポータルへのデータ連携が容易に
- 今後は、地域や分野の「つなぎ役」の充実が重要に

「蓄積」から「発見」「活用」へ



デジタルアーカイブのガイドライン

- 「我が国が目指すデジタルアーカイブ社会の実現に向けて」（2020年8月）
 - 知的資産のシェアと利活用により新たな価値を創生する社会基盤としてのデジタルアーカイブジャパン（デジタルアーカイブ社会の実現）を推進
 - 幅広い知識や理解を要する人材育成には、育成環境や財政基盤等の課題がある
- 「デジタルアーカイブの構築・共有・活用ガイドライン」（2017年4月）
 - メタデータの整備、長期アクセスの保証、望ましい利用条件、データ共有、データ活用、成果物の還元、コミュニティ形成を重視
 - 「活用できる表形式のデータとは？」を付録として添付
 - この点は繰り返し課題に：総務省「[統計表における機械判読可能なデータの表記方法の統一ルールの策定](#)」（2020）
- 「デジタルアーカイブにおける望ましい二次利用条件表示の在り方について」（2019年3月）
 - 国際的に普及しているパブリック・ドメイン・ツール及び CC ライセンス。特に、CC0、CC BY を強く推奨する。
 - Rights Statements からは、著作権あり、著作権ありー教育目的の利用可、著作権なしー他の法的制限あり、著作権なしー契約による制限あり、著作権未評価のマーク。
 - 日本独自の表示としては、裁定制度により利用された著作物であることがわかるマーク（著作権未決定-裁定制度利用著作物）。
- 「デジタルアーカイブのための長期保存ガイドライン」（2020年8月）
 - 媒体の議論に重点を置かない
 - ドキュメンテーション、管理の仕組みの構築、組織的対応による真正性の確保
 - 分散保存の強い推奨
- 「デジタルアーカイブアセスメントツール（改定版）」（2020年8月）
 - デジタルアーカイブの自己点検のためのツール
 - 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（各機関のミッション等の必要に応じて目指す水準）／つなぎ役モデル（分野・地域コミュニティのつなぎ役の役割をもつ機関が目指す水準）

スリムモデルの提案

- 「内閣府知的財産戦略本部 デジタルアーカイブの連携に関する実務者協議会メタデータのオープン化等検討ワーキンググループ」において提示（福島2016）
 - 2017年4月公表の「デジタルアーカイブの構築・共有・活用ガイドライン」に一部反映
- 発想の背景
 - デジタルアーカイブの消滅／構築主体の持続性自体が課題となっているという意識
 - 簡易さ／持続可能性 をキーワードに考える
 - 「これだけは」という要件をどう考えるか
- スリムモデルの要件
 - 利用規約の明示
 - 機械可読性の担保
 - 環境に依存しないデータ移行性の確保
 - アクセシビリティを保障
- どこに労力を割くかを検討する
 - モジュール的なシステム利用（冨澤ほか2018）（[江草2018](#)）（[中村・高嶋2021](#)）
 - アーカイブズにおいても収集と活用にリソースを集中する（福島・天野2019）

技術周りの動向

- フォトグラメトリ
 - ノートルダム大聖堂のデジタル的な復元
 - フランス国立図書館リシュリュー館の天井画
 - 石造物3Dアーカイブ
- データ処理技術との組み合わせ
 - GIS (Geographic Information System) 分野との積極的な連携
 - IIIF (International Image Interoperability Framework) (永崎2016)
- くずし字認識
 - 古典籍収集事業等によるデータ集積: まずは学習アプリ等
 - 画像認識技術+学習データ → くずし字認識アプリ「みを(Miwo)」の開発
 - 学習データの拡充が課題: 版本等に留まっている側面も
- テキストの構造化
 - TEI (Text Encoding Initiative): テキストの構造とメタ情報を記述
 - 近世ヨーロッパ史における帳簿分析／延喜式への応用 (小風・後藤2019)

現在の焦点：写真のカラー化／カラー写真の発見

- 沖縄県公文書館の「写真が語る沖縄」
 - NARAからの収集写真、地方写真、県所蔵写真などを横断で検索
- 北摂アーカイブスの活動
 - 大阪府豊中市+大阪府箕面市の共同運営
 - 実質は「地域フォトエディター」による活動
- 白黒写真のカラー化とアプリ化（庭田・渡邊2020）
 - 渡邊英徳（東京大学）+広島女学院高等学校による展開
 - 自動着色+聞き取りによる補正+新たな写真の発掘+展示会の開催など
- 占領期写真の収集
 - 衣川太一（写真収集家／神戸映画資料館）によるオークション収集
 - 展示「戦後京都の「色」はアメリカにあった！カラー写真が描く<オキュパイド・ジャパン>とその後」
- 肖像権ガイドラインの公開
 - デジタルアーカイブ学会（2021年4月19日）



→資料収集の段階／色を巡る動向：デジタルアーカイブ未然の状況／肖像権の議論も展開

現在の焦点:対象コンテンツの拡大

- 動画資料
 - 映画／テレビ／[動画](#)
- ボーンデジタルの論文・書籍・情報
 - 電子ジャーナル／電子書籍／[統計資料](#)
- 空間自体の情報化
 - 建築／地域／地球／宇宙空間
 - デジタルツインの試み:[東京都デジタルツイン実現プロジェクト](#) (2021)
- 人間の情報行動の集積
 - 位置情報
 - 情報収集行動
- →文化遺産のデジタル化という出発点からくる、ボーンデジタル資料やデータ類へのアプローチの弱さをどう位置づけるか

デジタルアーカイブの達成

- 「失敗」の経験
 - 維持：継承の仕組みを伴わない構築は避ける
 - 単独・複雑なシステムは極力導入しない／実験と永続的運用を分ける必要
- 情報技術自体についてはそれぞれで追求されていく段階
 - もともと最先端技術の導入は少ない：〈枯れた技術〉での展開が可能な領域
- 各種ガイドラインの充実
 - 読み解けば大きな失敗はない状況に
 - バックボーンとしてのジャパンサーチの存在
- 新たな事例の集積
 - 蓄積の場としてのデジタルアーカイブ学会(2017)
 - ジャパンサーチと教育利用(大井・渡邊2020)
 - 福井県文書館の取組み：授業で使う文書館資料－新学習指導要領を見据えた活用例－
 - おうちミュージアムの展開
- →課題はありつつも技術的な問題はクリアが可能な状況／仕組みの構築が課題

3 デジタルアーカイブの課題

デジタルアーカイブをめぐる当面の課題

- デジタルアーカイブ学 or デジタルアーカイブ領域の設定
 - 固有の課題の周縁部分の集合体？／独自の領域を持てる？
 - 評価指標や情報を取扱う上での倫理の確立：デジタルアーカイブからMLAへの投げ返しが可能なように
- 基盤整備
 - 情報活用のための法制度整備：利用者・現場の運用者の立場から
 - 地域の情報環境の一層の改善
- 人材養成
 - 〈仕組み全体〉を構築できる存在（福島2011・2018）
 - マインドセットとスキルセットの転換
- 学校教育への貫入
 - GIGAスクール構想：児童生徒1人1台のPC＋ネットワーク環境
 - 新学習指導要領（2022年度新入生適用）や共通テストでも重視
- 文化情報資源という視野の確立による課題の解消
 - 博物館・図書館・公文書館（MLA）等が対象とする資料（国立国会図書館2010）
 - ボーンデジタルも含め「地域」の〈資料総体〉、というところまで踏み込みたい

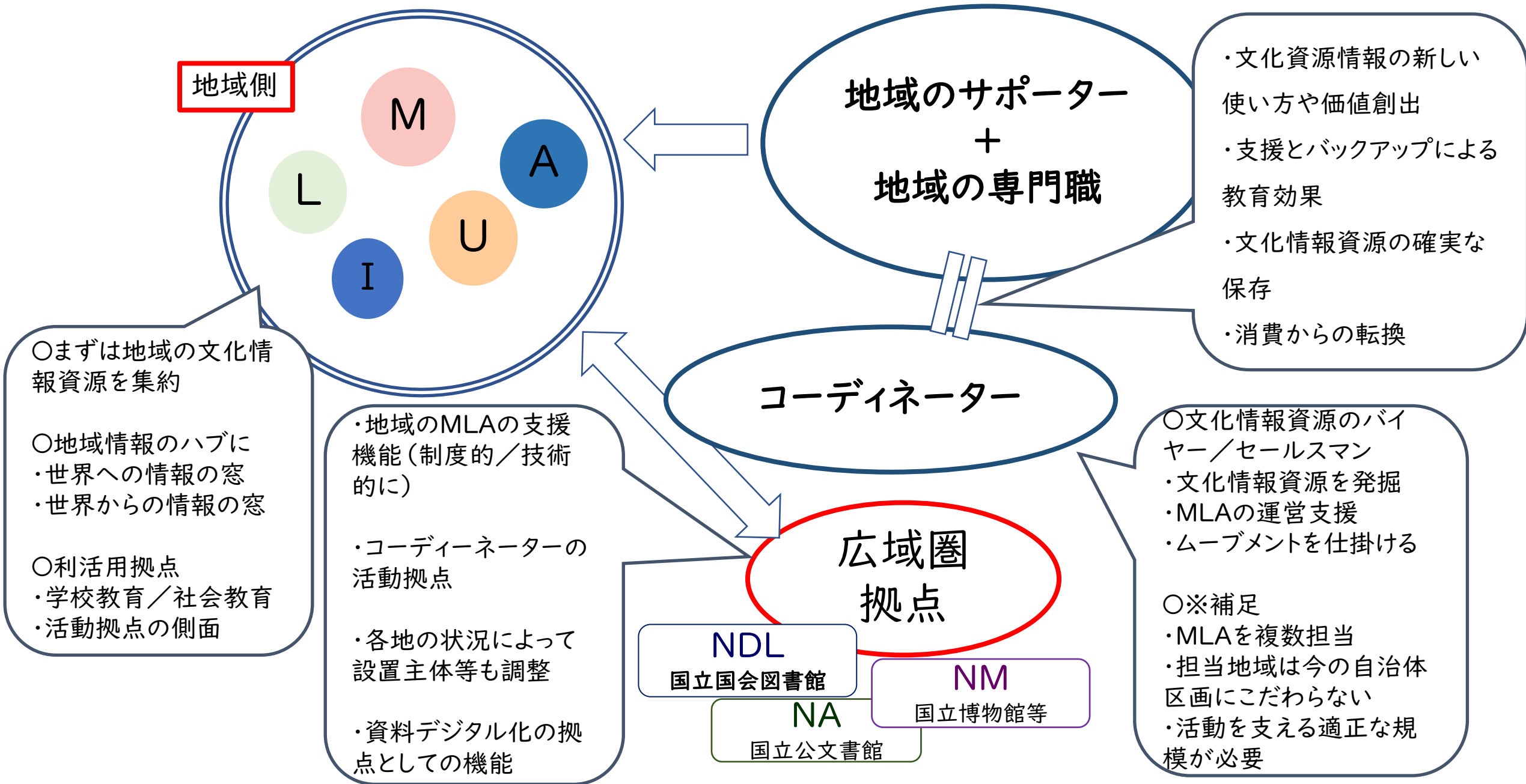
文化情報資源

- 文化資源の定義
 - 「ある時代の社会と文化を知るための手がかりとなる貴重な資料の総体」(文化資源学会設立趣意書(2002年6月12日採択))
 - 「その時代をともに生きる人々が心豊かに生きるための精神的な共有財産としての「文化」をつくるための資源」(「文化資源の広域的活用に関する研究」2003年3月 地方自治研究機構)
- 博物館・図書館・公文書館(MLA)等が対象とする資料
 - 当初は街作りの素材という側面が強調
 - デジタル化の効果に注目し、MLAが対象とする資料を指す用法が定着しつつある
- 「地域」の〈資料総体〉、というところまで踏み込みたい
 - ボーンデジタルについての議論が指摘がありながらも組み込まれていない
 - 文化情報資源へ

立命館大学「文化情報学専修」

- 様々な形態の文化を保存・理解・継承・活用、さらには新しい文化創生を実現。
- デジタル時代の今、歴史・文化・芸術を対象とした研究にもデジタル技術を導入した研究手法が急激な勢いで導入されてきています。
- これを文化情報学と称し、専修では様々な形態の文化を保存・理解・継承・活用、さらには新しい文化創生を実現するための研究をします。**歴史都市や文化財・文化遺産、出版メディアなどを対象にしたデジタルアーカイブを文化資源（情報）と位置づける**ことで生れる様々な可能性について、国境や時代という境界を越えた視点で研究を進めます。
- 専修では、国内外の著名な博物館や文化遺産保護組織へのインターンシップや、学内の歴史都市防災研究センターやアート・リサーチセンターで実施されている研究プロジェクトに参加し、実践的な研究環境の中で問題意識を持ち、論文をまとめていくという新しいスタイルのカリキュラムを提供します。
- <http://www.ritsumeai.ac.jp/gslt/introduce/major/major14.html/>

情報・資料面の融合と機能面の融合



4 おわりに

参考文献

- 江草由佳2018「移行しやすく使いやすいデジタルアーカイブの構築:教育図書館貴重資料デジタルコレクションの経験から」情報知識学会誌, 28-5, p. 367-370
- 大井将生, 渡邊英徳2020「ジャパンサーチを活用した小中高でのキュレーション授業デザイン:デジタルアーカイブの教育活用意義と可能性」デジタルアーカイブ学会誌, 4-4, p.352-359
- 古賀崇2017「日本におけるデジタルアーカイブのゆくえを探る:国際動向を踏まえた「より深い利用」に向けての展望」『情報の科学と技術』67(2)
- 小風尚樹, 後藤真2019「『延喜式』へのTEI適用と日本史資料のテキストデータ共有・流通」『国立歴史民俗博物館研究報告』218
- 国立大学図書館協会2019『大学図書館におけるデジタルアーカイブの利活用に向けて』(国立大学図書館協議会)
- 国立国会図書館2010『文化・学術機関におけるデジタルアーカイブ等の運営に関する調査研究』(国立国会図書館)
- 後藤真・橋本雄太編2019『歴史情報学の教科書』(文学通信)
- 是住久美子2015「ライブラリアンによるWikipedia Townへの支援」 <http://current.ndl.go.jp/cal847> (20210729確認)
- 澤谷晃子2018「大阪市立図書館デジタルアーカイブのオープンデータの利活用促進に向けた取り組み」 <http://current.ndl.go.jp/cal925> (20210729確認)
- 田山健二2018「ADEACの取り組み」『デジタルアーカイブ学会誌』2-4
- 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2020「3か年総括報告書 我が国が目指すデジタルアーカイブ社会の実現に向けて」
https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/pdf/r0208_3kanen_houkoku_honbun.pdf (20210729確認)
- 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2017「デジタルアーカイブの構築・共有・活用ガイドライン」
http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/index.html (20210729確認)
- 富澤かな, 木村拓, 成田健太郎, 永井正勝, 中村覚, 福島幸宏2018「デジタルアーカイブの「裾野のモデル」を求めて」『情報の科学と技術』68(3)
- 永崎研宜2016「今、まさに広まりつつある国際的なデジタルアーカイブの規格、IIIFのご紹介」 <http://digitalnagasaki.hatenablog.com/entry/2016/04/28/192349> (20210729確認)
- 中村覚, 高嶋朋子2021「持続性と利活用性を考慮したデジタルアーカイブ構築手法の提案」『デジタルアーカイブ学会誌』5-1
- 日本図書館情報学会用語事典編集委員会編2020『図書館情報学会用語事典 第5版』(丸善出版)
- 庭田杏珠・渡邊英徳2020『AIとカラー化した写真でよみがえる戦前・戦争』(光文社)
- 福島幸宏2011「地域拠点の形成と意義」『デジタル文化資源の活用—地域の記憶とアーカイブ』(勉誠出版)
- 福島幸宏, 赤石直美, 瀬戸寿一, 矢野桂司2012「『京都市明細図』を読む—いくつかの素材の提示として—」『メディアに描かれた京都の様態に関する学際的研究 平成23年度京都府立大学地域貢献型特別研究 (ACTR) 研究成果報告書』(京都府立大学)
- 福島幸宏2014「京都府立総合資料館による東寺百合文書のWEB公開とその反響」 <http://current.ndl.go.jp/cal561> (20210729確認)
- 福島幸宏2016「ガイドラインに要れるべき要件(福島構成員資料)」
http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/meta_data/dai2/siryou3_3.pdf (20210729確認)
- 福島幸宏2017「ウィキペディアタウンをMLAの立場から考える」 <http://magazine-k.jp/2017/07/11/wikipediatown-for-mla/> (20210729確認)
- 福島幸宏2018「これからの図書館員像—情報の専門家／地域の専門家として」『現代思想』46-18
- 福島幸宏, 天野絵里子2019「アーカイブズ構築のスリムモデル」(Code4Libジャパン2019報告資料)
- 柳与志夫2020『デジタルアーカイブの理論と政策』(勁草書房)
- 吉見俊哉2017「なぜ、デジタルアーカイブなのか? —知識循環型社会の歴史意識」『デジタルアーカイブ学会誌』1-1