

歴史系博物館の データの統合化が持つ課題

国立歴史民俗博物館 後藤真

流れ

1. nihuINTの説明と横断検索のメリット
2. 国立歴史民俗博物館（歴博）が進める事業である総合資料学
3. 両者の事業実務および、歴史系資料のデジタル化を行った経験から

nihuINT

- nihuINT: (nihu INTegrated Retrieval System)
 - 人間文化研究機構統合検索システム
 - 公開：2008年4月 更新：2012年5月
 - <http://int.nihu.jp/>
- 多種多様なデータベースを一元的・網羅的に検索
- レコード数（2016年2月時点）：合計600万以上（他機関含まず）
- 対象DB（2016年2月時点）：合計 175
 - 人間文化研究機構：146（各機関: 138, 地域研究拠点: 8）
 - 国立国会図書館：13
 - 2010年7月14日よりシステム連携開始
 - PORTA → NDL Search (2012年1月)
 - 京都大学地域研究統合情報センター 16
 - 2015年5月よりシステム連携開始

DBの分類

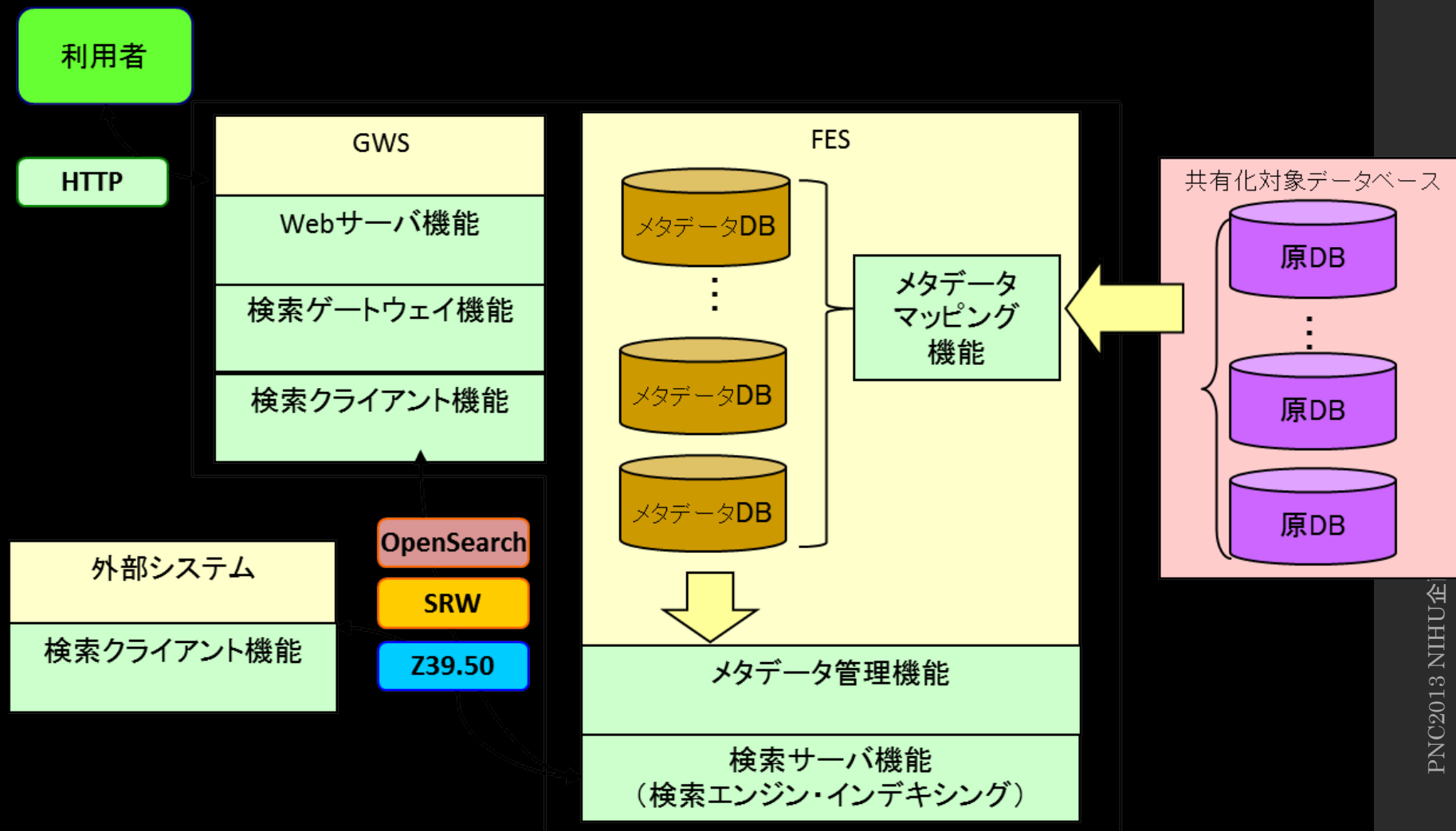
(確認時期は前スライドと少しずれます)

・機関別・種類別・研究分野別（6種類，28分野）

機関	DB数	種類	数
歴博	42	所蔵資料目録	41
民博	20	所蔵図書・雑誌目録	9
日文研	31	本文・フルテキスト	11
地球研	7	事項・ファクト	48
国文研	25	画像・映像・音響	8
国語研	6	研究文献目録	21
地域	7	その他	13
NDL	13		

研究分野	DB数	研究分野	DB数
情報学	3	日本史	58
生活科学	2	東洋史	3
科学社会学・科学技術史	3	考古学	17
文化財科学	1	人文地理学	7
地理学	7	文化人類学・民俗学	44
環境学	4	法学	4
地域研究	6	政治学	4
哲学	11	経済学	6
芸術学	23	経営学	4
日本文学	38	社会学	4
言語学	10	教育学	6
日本語学	12	地球惑星科学	2
日本語教育	3	基礎生物学	5
史学一般	13	総合	23

nihuINTシステム概要

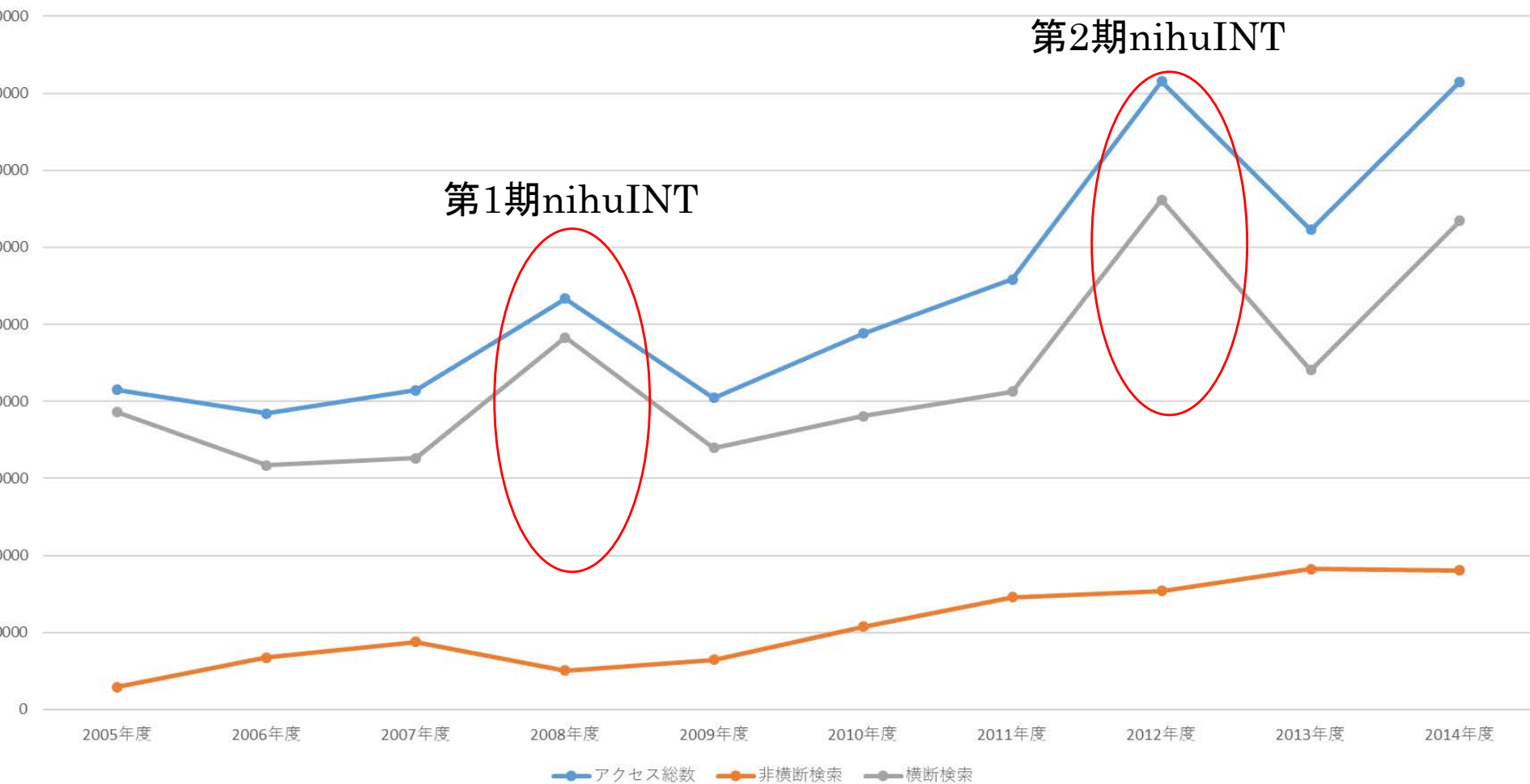


メタデータ方式

- ダブリンコアのもっとも基本的な部分のみを採用
 - Identifier (識別子)
 - Keywords (種別・主題)
 - Title (名称・題名)
 - Who (人名や機関の名称)
 - When (時間に関する情報)
 - Where (場所に関する情報)
- 一部のデータベースには博物館コアを採用

横断検索のメリット（スケール）

横断検索前後の歴博のDBアクセス数



総合資料学の概要

総合資料学

研究機関の資料研究の課題を解決する
情報環境・研究モデル・社会還元の総体的研究

連携研究の必要性

歴史学をはじめ
人文系諸学

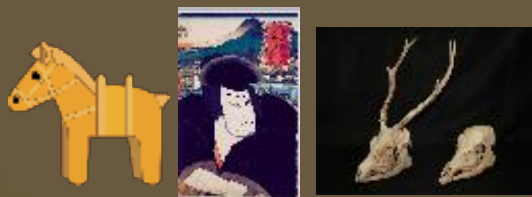
「危機」

社会的要請への対応
分析データを用いた歴史像
の複眼化

自然科学など
理系諸学

学問データへの
より多様な意義付け
記述的情報によるデータ
の複眼化

多種・多様・大量の
大学・博物館等資料



連携研究の実現

文理連携研究
チームによる
地域資料の総合的
調査・研究

大学・博物館等と
連携した、
地域の歴史・文化
の再構築



研究のための自由なアクセス

- 研究と資料情報の循環を実現
- 博物館型オープンサイエンスの構築
- 粒度の細かい地域史・地域文化研究を実現

基盤を支える博物館等情報の災害時バックアップ



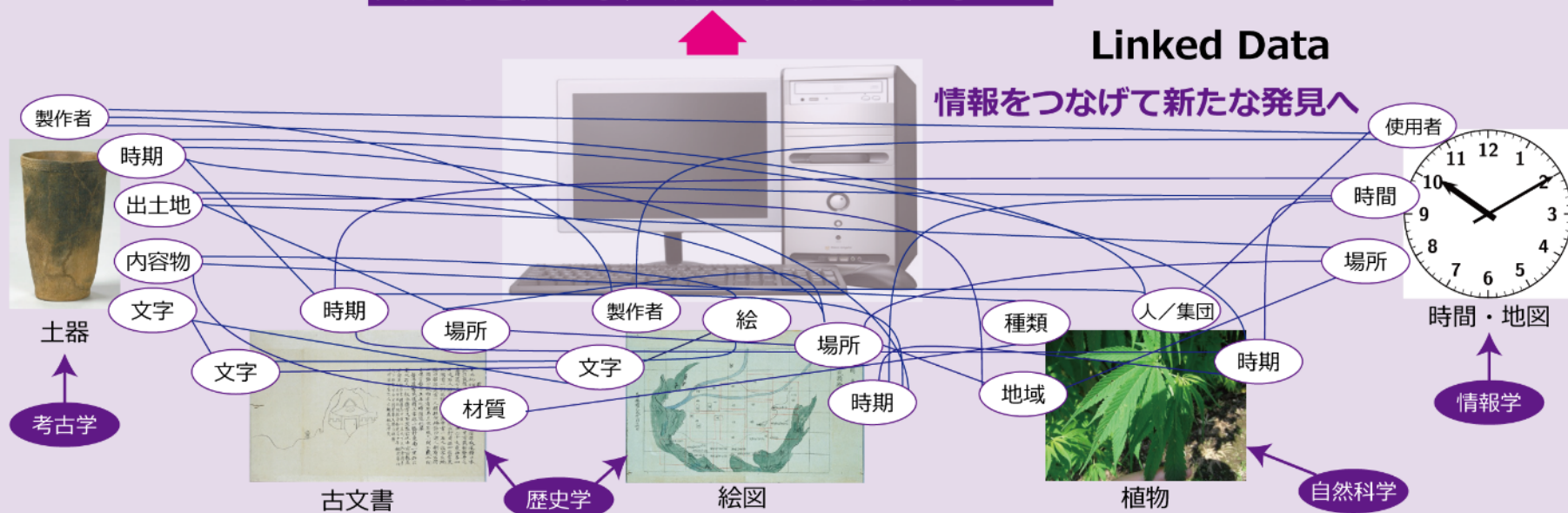
大学共同利用機関法人 人間文化研究機構

国立歴史民俗博物館

National Museum of Japanese History

① アクセスデータ構築のワークショップ

異分野を横断した研究の実現を支える基盤



発掘による
出土状況
▶埋葬されていた
所有者の身分

近世考古学

文献史学

技術史学

金属材料学

小判



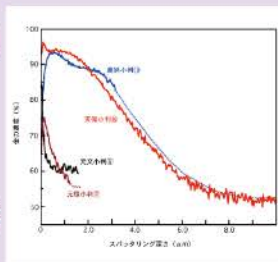
小判所絵図
▶「色付」の実態

金位并金吹方手続書

▶金色に仕上げる技術「色付」

オージェ分光分析

▶「色付」による金濃度の深さ
方向分布



分析化学

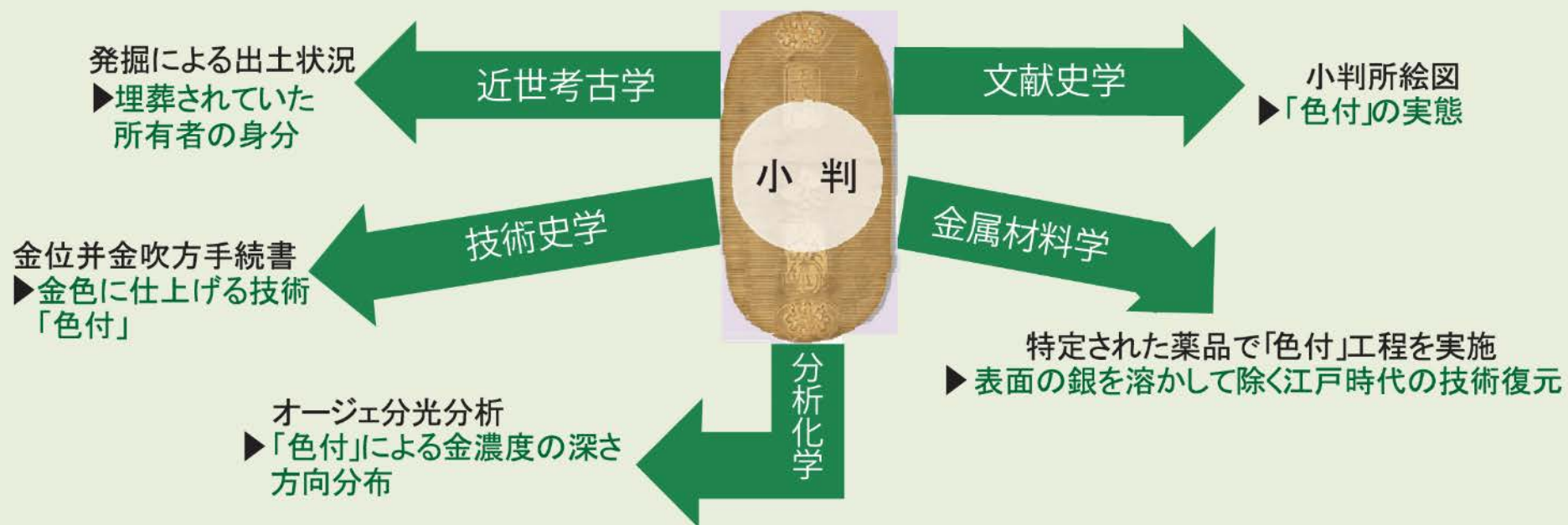
色付前(白色)



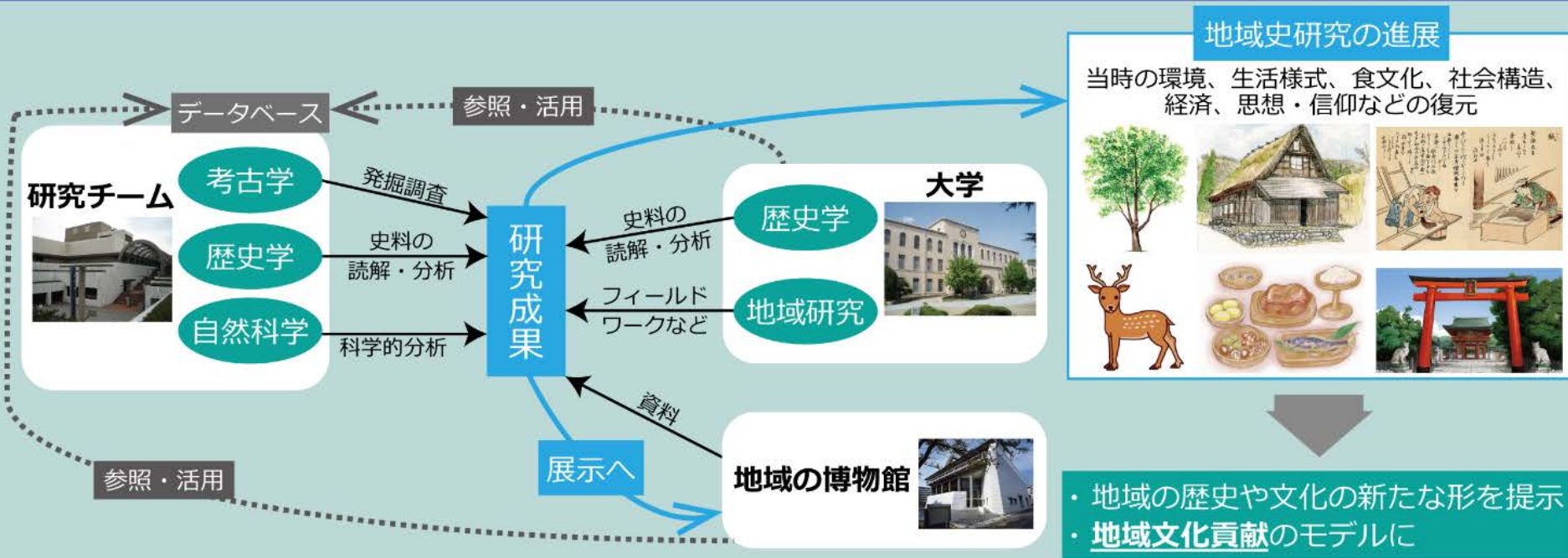
色付後(金色)

特定された薬品で「色付」工程を実施
▶表面の銀を溶かして除く江戸時代の技術復元

② 構築したデータの研究活用ワークショップ



③ 研究成果の社会的活用ワークショップ



詳細は

<http://www.metaresource.jp>



博物館や研究モノ資源の メタデータの特徴と課題

博物館の特徴と課題

- 総論として：横断検索に反対する組織はない
(歴民協)
 - 各論になるとつまづく
- 資料ごとに性質が異なるのでメタデータが異なる
 - 文書・考古資料・美術・地図・植物・鉱物・動物…



- 担当ごとに持っているデータ形式が違う館も
- データのセンシティブ度合いが異なる
 - 取得（寄贈・購入など）・寄託など
 - 展示
 - 登録されているものが「すぐに見られるか」どうか

メタデータをいかに統合するか

- niहुINTの経験をもとに
 - 複雑なものは困難→なるべくシンプルに（6個）
 - もとのデータは各組織が持つ
 - 横断検索は情報発見のツール
 - データは各自を原則
 - 組織で持ちきれないものだけを横断検索側で持つ
- （ただし、更新や組織改編の問題は残るので対応が必要。URIの維持などについては課題を整理する必要がある）

横断検索のための4つの層

横断検索層

変換A

個別目録層

変換B

原データ層

変換C

現物層

全体像を見る

データの階層
などを見る

(公開)
データを
取得する

現物に
アクセスする



横断検索に向けて それぞれで異なる課題

- 横断検索層

- 速度・何が見つかるか
 - 資料そのもの
 - データのリーチ度合（Webで全部見られる・近くで複製を見られる・世界に一つのモノへのアクセス情報だけ）

- 変換A

- メタデータマッピング（誰がやるかも含めて）
- データを個別層と横断層の何をどちらで保持するか
- クラウドなどの活用の場合の課題（目録DBは著作権よりセキュリティポリシーでひっかかる）
- APIコスト（個別目録側でどのぐらいの負担があるか）

横断検索に向けて それぞれで異なる課題

- 個別目録層

- そのデータは何を対象としているか
- 公開レベルはどのぐらいか（その館だから公開してよいのか、それとも通常の公開情報か）
- データは目録のみかコンテンツがあるのか

- 変換B

- 館のデータを公開データに変換する際の問題
 - 項目を一括で削ればよい場合とそうではない場合
 - 予算より「資料がわかる人」の確保→選り分けには人が必要

横断検索に向けて それぞれで異なる課題

- 原データ層

- バックアップなどを持つフォローをするか
- リポジトリなどでの公開への支援（ここでのオープンデータが重要 cf:国文研）
- 横断検索から原データへのアクセス経路確保
- 紙も公開されていれば重要（後述）

- 変換C

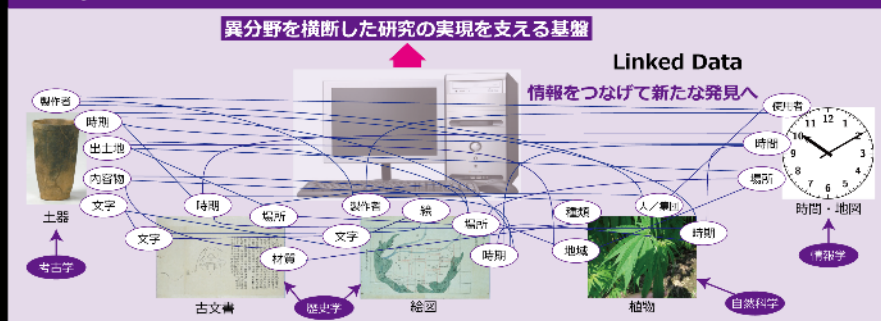
- 通常は横断検索側では関係ないが障害要因はある…

前に進めるために

- （繰り返しますが）博物館としても**総論は賛成**
- 各論の障害要因がどこにあるのかを見極めることが大事
- どの層、どの変換部に課題があるのかの状況に即した対応策提示の必要性
 - 例えば紙の情報でも公開されているものか、そうでないものかは、公開のコストを左右する
 - デジタルデータの非公開と紙の公開ではどちらが作業障壁が低いかはわからない
- 各論の課題を軽視すると総論で反対されるようになる

横断検索が粒度の細かい文化にリーチする可能性

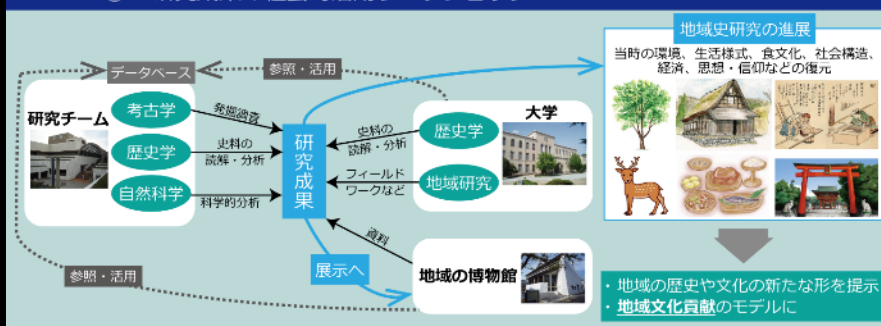
① アクセスデータ構築のワークショップ



歴博の総合資料学は研究中心ではあるが

横断的な情報基盤—それもGoogleとは異なる—の応用で、より粒度の細かいかつ多分やな歴史研究資料へとアクセスできるモデルを検討している

③ 研究成果の社会的活用ワークショップ



横断検索は

- スケールのメリット (今までとは違う人が検索する)

- 同じ地域の別の館を探せる

という観点から、今までとは異なる導線が期待できる→地域内のデータであれば、誘導がうまければ、アクセス情報だけあればよいものも

バラバラな状況をうすくカバー
※地域の博物館の「背中をおす」
きめ細かいしくみとともに

ワンストップサービスとしての
横断検索→まずはアクセス経路を！

ありがとうございました