



「魅力ある新国立公文書館」に向けて ～主にデジタルの立場から～

立命館大学情報理工学部 上原哲太郎

R 国立公文書館の機能

「パブリック・アーカイブ宣言」より

1. 歴史的に重要な公文書の収集・保存を、一層進めます。
2. インターネットを通じたご利用のため、公文書のデジタル化を一層推進します。
3. 生涯学習、社会教育にも活用できるデジタルアーカイブの更なる拡充を進めます。
4. 積極的な展示及び学習活動を通じて、公文書に触れ、興味を持っていただく機会を広げます。
5. 地方公共団体や研究機関、学会など、関連する団体との連携を強めます。
6. 海外の公文書館などとの国際交流の輪をひろげます。
7. 文書管理の専門家を育成します。
8. 職員全てが、利用者本位のサービス意識で活動します。

「研究者のためのもの」から「国民みんなのもの」へ
そのために必要な変化は何か？ ITは何ができるか？

「デジタル」の特性と公文書管理

- 複製が容易・そもそも原本と複製の区別がない
- 歴史的公文書がそもそもデジタルであれば積極的に複製することで受入処理が効率化できるはず
- 利用者にとって閲覧と「写しの交付」の区別はあいまいに
- 検索性が大幅に増す
 - 特に文字・語句の検索は超高速になる
 - AI技術の発達により「あいまい検索」の精度も今後向上
 - 前提として「正規化された文字データ」が必要
- 大量文書の詳細分析が容易になる
 - 統計的処理やデータマイニング技術の活用が容易に
- メディアが文書 = 文字に閉じなくなる
 - 写真・音声・動画・データベースなど
 - スプレッドシートは数式ごと保存可能に



公文書館における受入から公開までの タイムラグの低減

- 現在は「紙」が原則なので
利用可能になるまで時間がかかる
- 現在「1年以内」→最悪1年近い空白期間
- デジタルで受入をすれば公開までの時間を短縮できる
- 少なくとも長期保存のための処理が不要に
- 目録作成の作業も見直し可能
 - 公文書ファイル管理簿で代用など

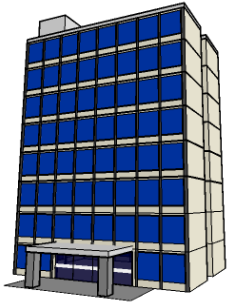
R デジタルでもタイムラグは0にならないが…

- 現在、歴史公文書は
国立公文書館で受入られてから利用可能になるまで
非公開状態に→情報公開制度との整合
- しかしデジタル化によって解決可能に
 - 複製が容易→受入処理は複製で行ってよい
 - 国立公文書館での受入処理が完了するまで
行政機関側で行政文書原本データの保管を継続すれば
情報公開制度の空白期間が生じるのを避けられる

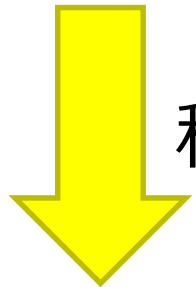
クラウド化が進めば



国民にとっての情報公開がシームレスに？



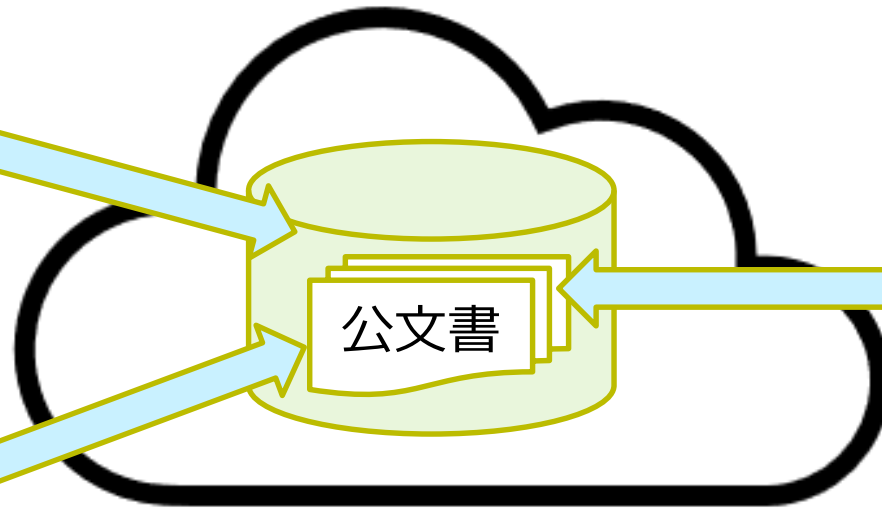
行政機関



移管



公文書館



国民



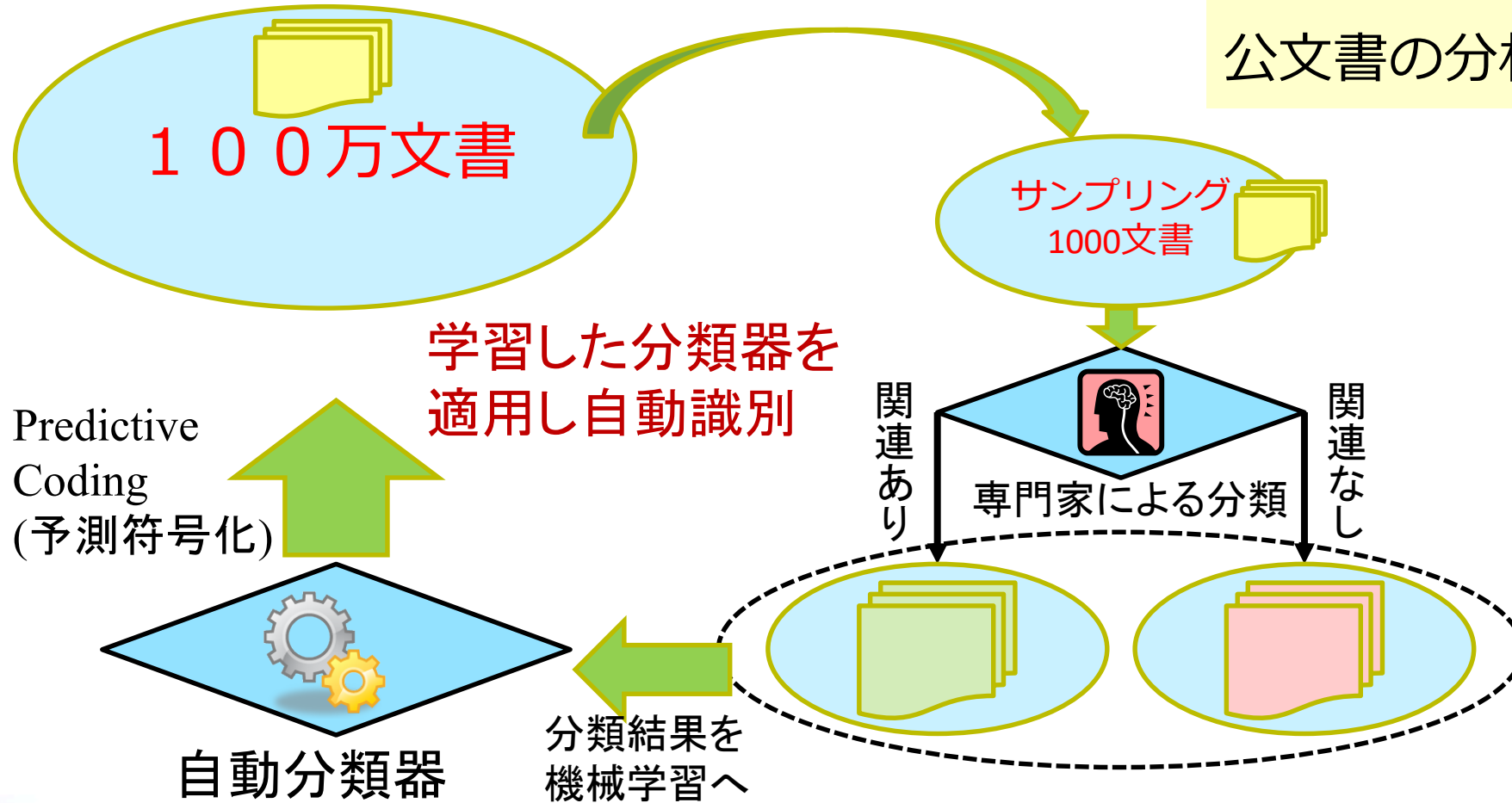
各データ（文書）への
アクセス権の問題

歴史的公文書はビッグデータ

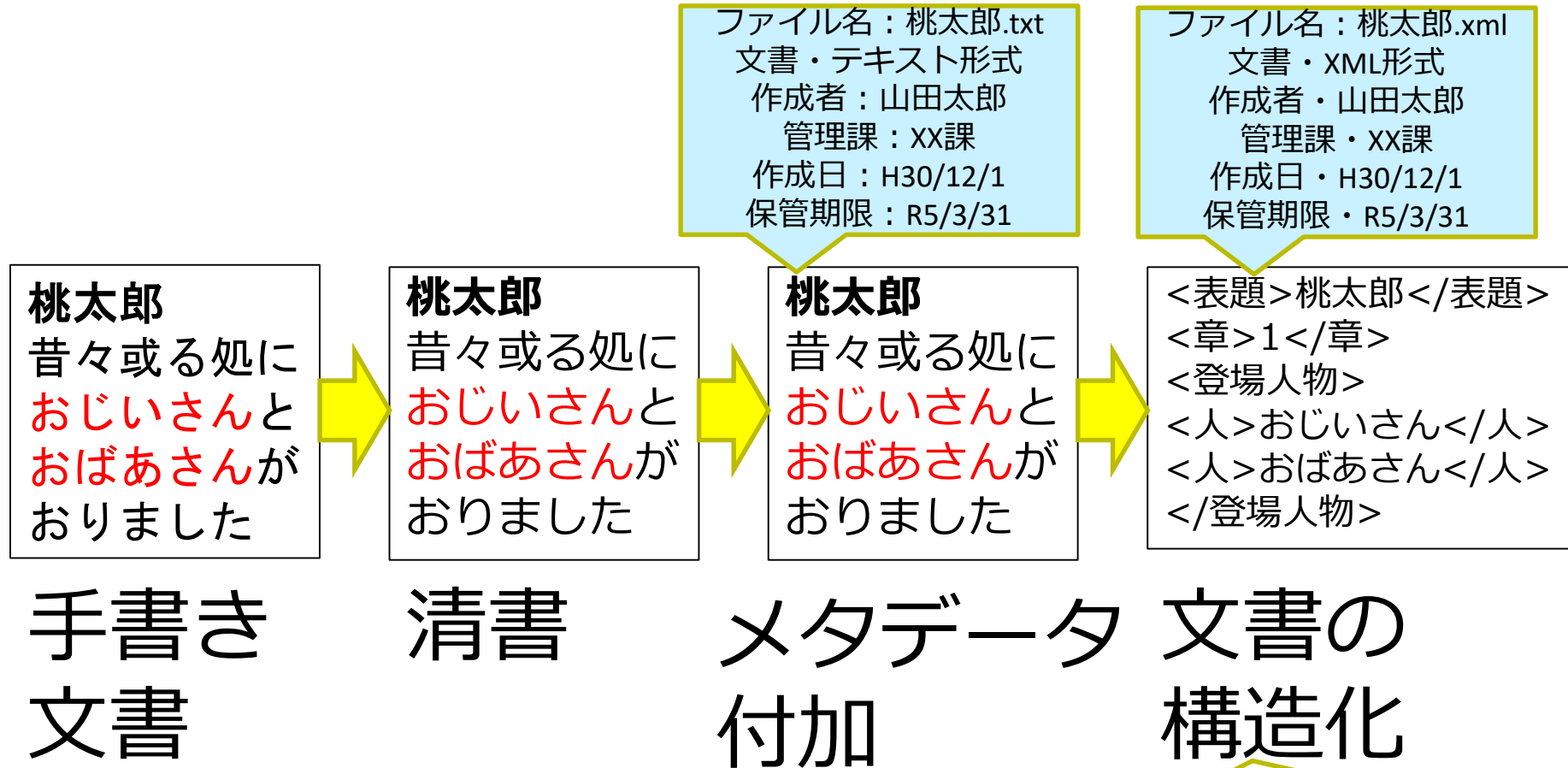
- **ビッグデータとは…**
 - 「非構造化データを含む」多様で巨大なデータの集合
 - 計算機処理技術の向上により検索・統計処理が可能に
- **ビッグデータを様々な角度から処理することで新しい発見が容易になる**
 - 用語の変遷、言葉の用法の変遷、形式の変遷…
 - 「歴史的経緯」の分析
 - 各種分析は大幅に効率化し、非専門家にも可能になる
 - 「趣味」で色々な分析結果を披露してくれる人が産まれるかも

機械学習による文書分類

Discovery制度がある国では裁判活用が盛ん
公文書の分析にも使える



公文書の構造化が進めばAI処理も容易に



文書の「内容」と「表示形式」の分離
文書に対する「機械可読な意味づけ」

R 大量の文書を「そのまま」利用に供する可能性

- ビッグデータとなった大量の公文書データを複製して利用者に提供するサービスの可能性
- 利点：
 - AI（機械学習）の元データとして有益
 - 手元にデータを持つことで分析の自由度が上がり何らかの新たな発見に繋がる可能性
- 欠点：
 - 一度複製されると流通は止まらない
 - 「私的公文書館」登場の可能性
 - 恣意的に一部データを切り取ったものなど

R デジタル専門職の必要性

- 今後デジタルで受け入れられるデータが増えると行政機関にも公文書館にもデジタル専門職が必要に
- 求められる知識や技能
 - 特に「自然言語処理」に関する知識
 - データベース
 - 統計処理・データマイニング・データサイエンス…
 - デジタル文書の長期保存に関する知識
 - メディア・データ形式…
 - プログラミング

R まとめ

- デジタル・ネイティブが増えるにつれて
ITを駆使してデジタルデータの利用・分析が行える
人材は増えてくるはず
- デジタルネイティブにとって「検索」は当たり前
- 分析に長けた人材は一般国民に増えてゆく

- 新公文書館は国民の公文書利用ニーズの変化
= ビッグデータとしての直接利用ニーズに
応えてゆく必要