



大学共同利用機関法人

人間文化研究機構

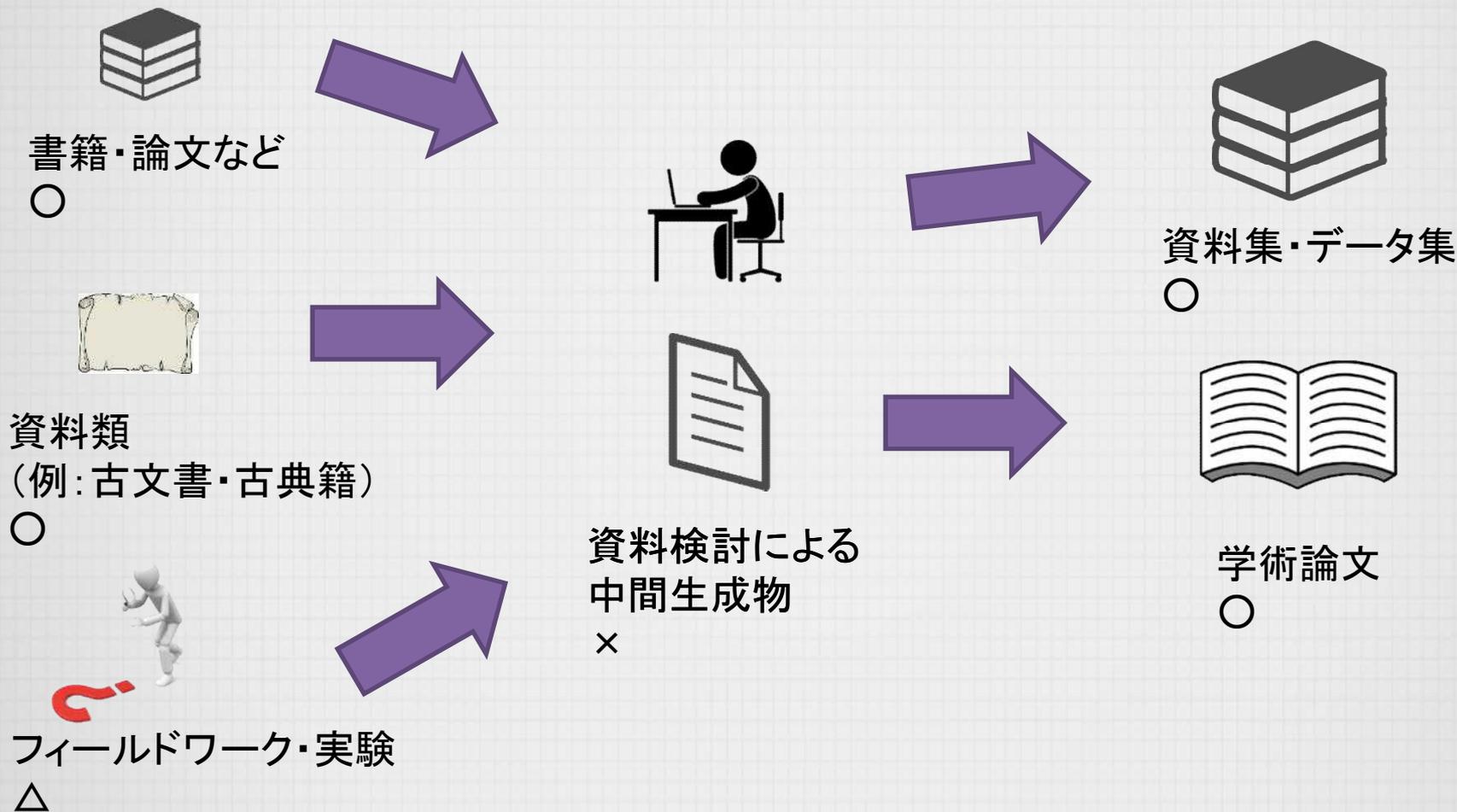
資料3

科学技術・学術審議会学術分科会  
学術情報委員会(第12回)  
平成30年9月18日(火)

# 人文学における オープンサイエンスの現状

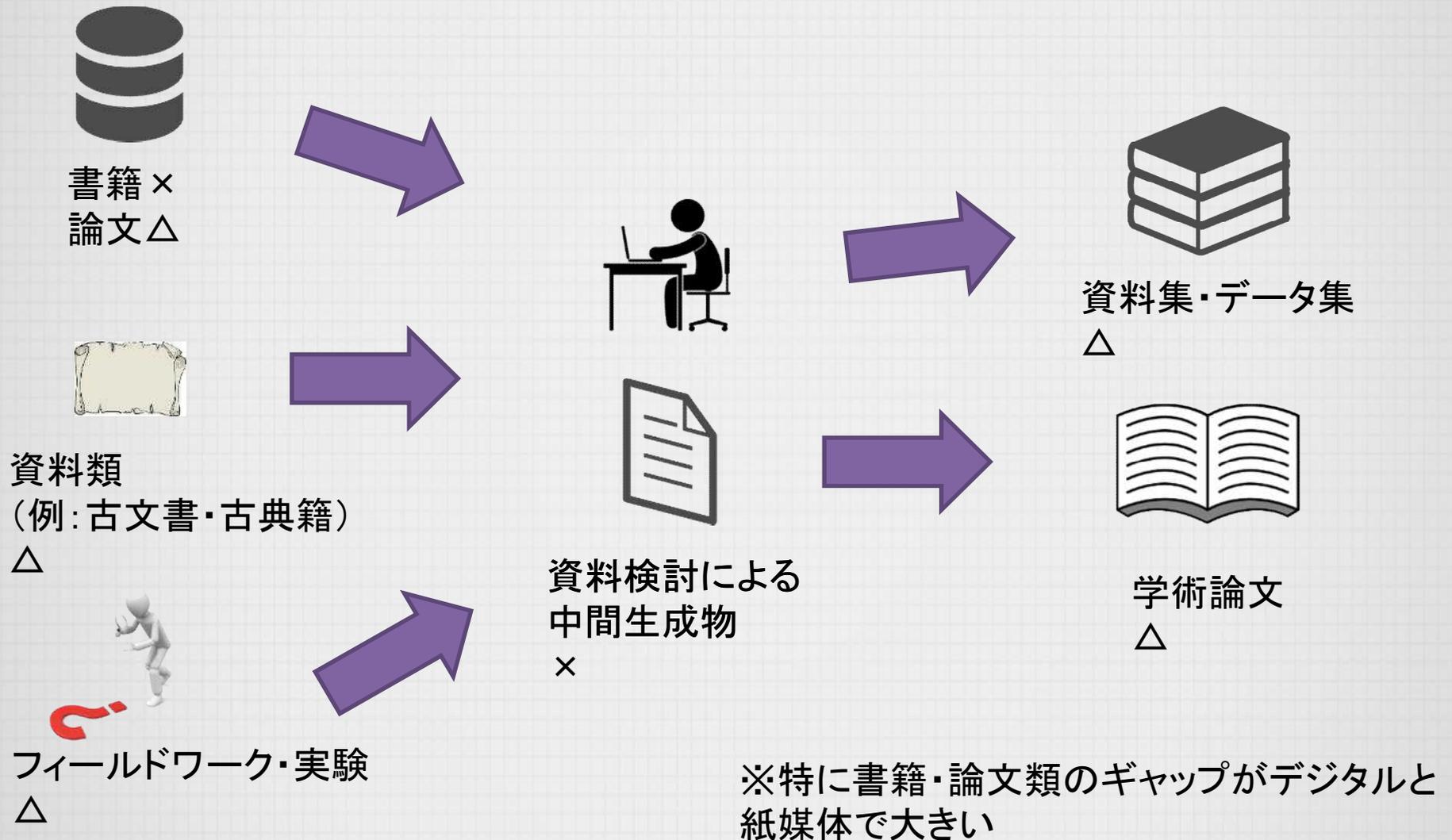
大学共同利用機関法人 人間文化研究機構  
国立歴史民俗博物館  
後藤真

# 研究の流れとアクセス状況概観 (非デジタル)



# 研究の流れとオープン化状況概観

3



# 人文学におけるアクセス原則

- 原則として資料については、公平なアクセスを確保
  - 論文として使った資料は第三者もアクセスできる
  - 資料調査を行なった研究者は資料を公開したのちに論文を書くという倫理
  - 資料公開自体が一つの業績になる分野も
- フィールドワーク等については若干状況が異なることも

# 人文学におけるデジタルデータ蓄積

- 全体に立ち遅れ（他分野比較 国際比較 両面）
- 論文・資料・中間生成物のいずれも
- 理由
  - ステークホルダー
  - 言語的課題
  - 分野の大きさと分散化

# 資料のデータ化

- 基盤データとして、いくつかのものがある

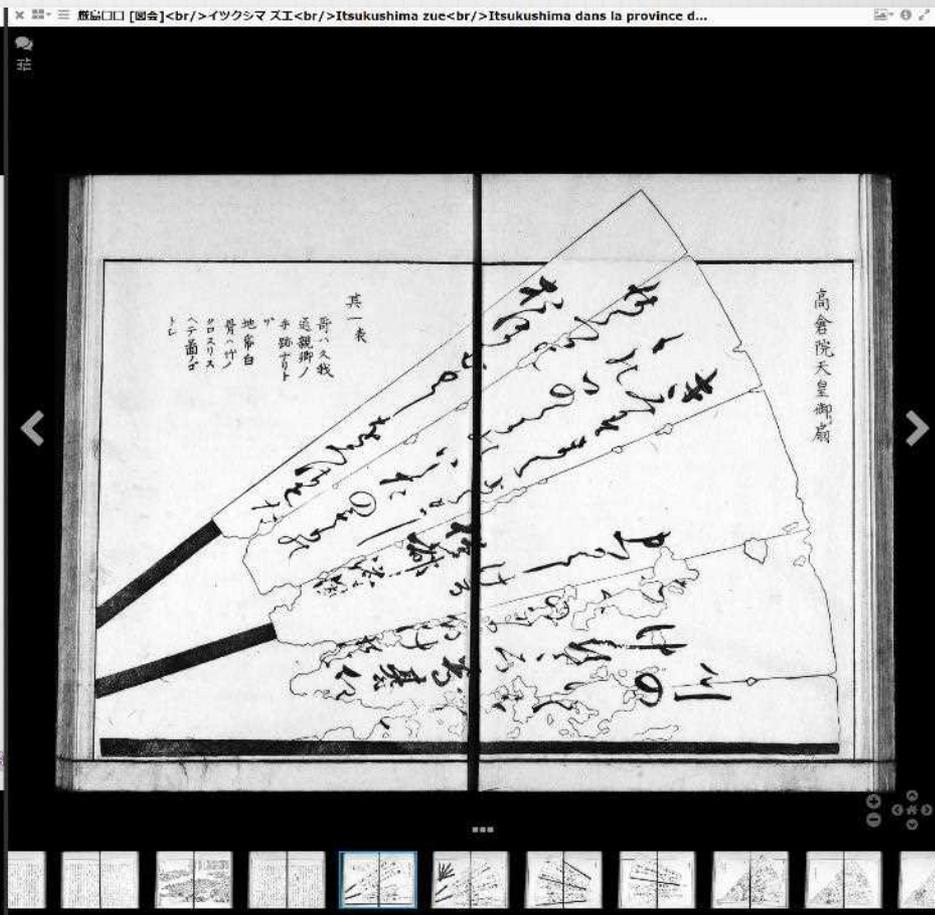
文学	国文学研究資料館・古典籍総合目録および画像、国会図書館デジタルライブラリー
言語学	国立国語研究所のコーパスなど
史学	東京大学史料編纂所・国立公文書館デジタルアーカイブ・文化財機構・歴博など
文化人類学	国立民族学博物館(標本・映像)
地理学	東大空間情報科学研究センター
社会学系	大原社会問題研究所
考古学	奈良文化財研究所(発掘調査報告書リポジトリ)

- ほかに画像データの国際標準規格(IIF)への対応と充実(東京大学・千葉大学・京都大学・歴博・国文研など)

# 国際標準画像資料データ提供例 (いずれもCC BY 4.0(相当)提供)

国立歴史民俗博物館所蔵

ルーヴァン・カトリック大学所蔵



# 国際標準画像資料データ提供例

## 千葉大学学術リソースコレクションとは

千葉大学学術リソースコレクション (c-arc) は、千葉大学附属図書館がウェブ上で公開・提供するコンテンツを学術リソースとして広く使ってもらうためにコレクションとしてまとめたものです。画像データについては、International Image Interoperability Framework (IIIF) という技術を用いて、学術リソースとして簡単に活用できる環境を提供しています。本サイトは、デジタルコンテンツを活用した研究・教育・学習を実現するための新しい教育研究基盤である「デジタル・スカラシップ」構築の一環として、アカデミック・リンク・センターが制作しました。

> IIIFの使い方

## コレクション

### 古画コレクション



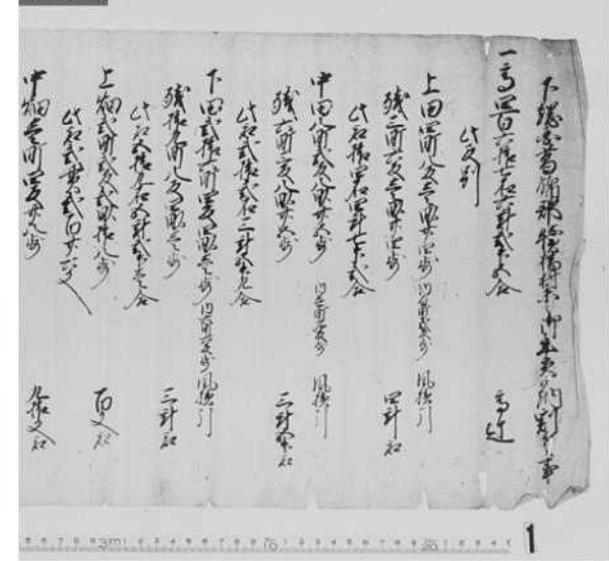
千葉大学工学部工業意匠学科・デザイン工学科卒業生デザイン集

### 江戸・明治期園芸書コレクション



萩庭植物標本データベース

### 野家文書



真菌・放線菌ギャラリー

千葉大学学術リソースコレクション (<https://iiif.ll.chiba-u.jp/>) より



# 資料データの課題

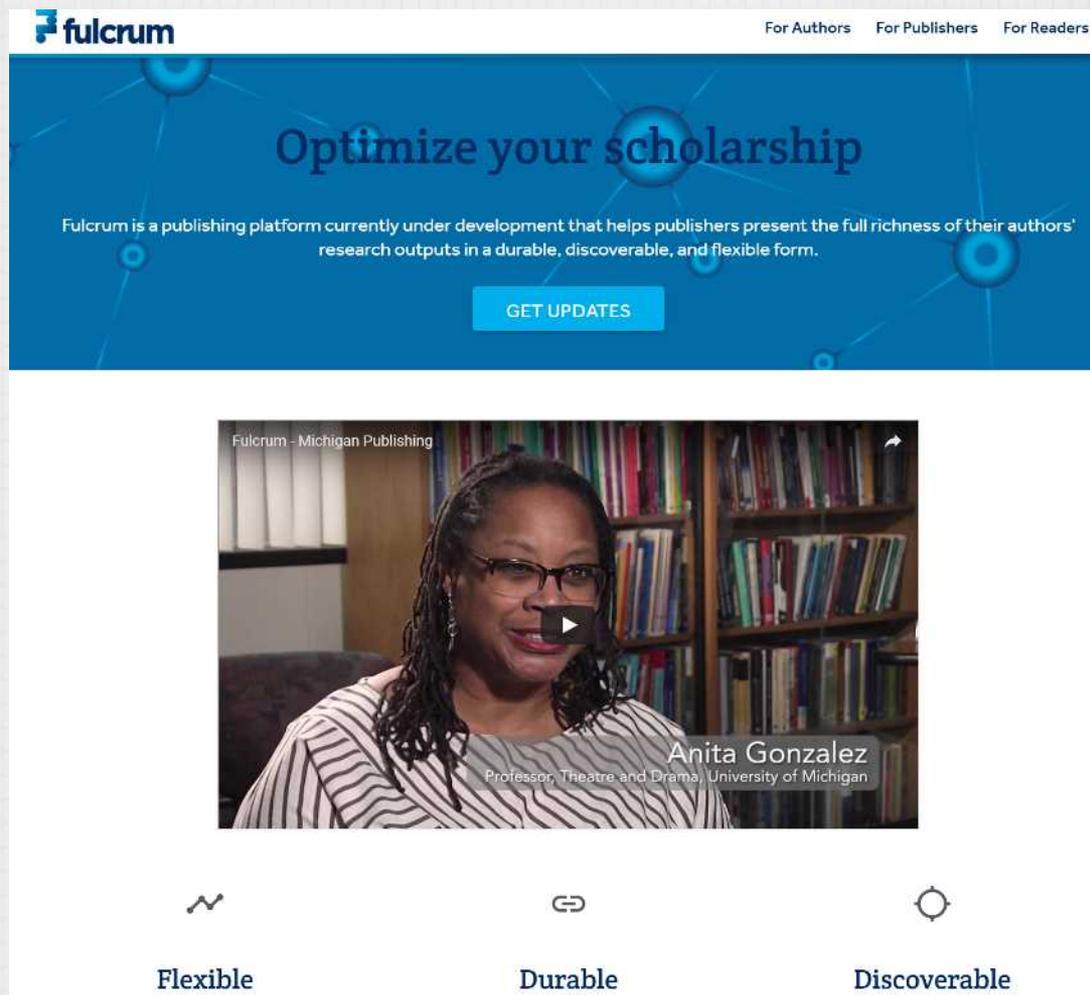
- 機械処理できるデータの不足
  - ほとんどのものが画像、もしくは目録まで
  - 本来、文字資料が多いのでテキストデータがあることが望ましい
  - アメリカ・ヨーロッパなどでは人文学テキストデータの標準化が進む(Text Encoding Initiative)→人文情報学(Digital Humanities)への発展など
  - 日本の場合、文字の障壁も存在する
- 個人所有資料の問題
- ワンストップサービスの不在／困難さ
  - データの質を整えにくい

# 論文のデータ化

- リポジトリサービスの一般的充実
- 機関レベルでのデータ提供は整いつつある
- 一方で、学協会が出版するもののデジタル化は立ち遅れ気味
  - 史学・文学などが特に遅れる 日本におけるクオリティ・ペーパーのデジタル化の遅れ
- 書籍の電子化という課題
  - 海外(特に英米圏)の場合、Top Publisherは電子化と論文DBなどへの登録は進む
  - 日本の場合、学術書籍の電子化もだが、中にある論文や引用のデータベースがない



# アメリカにおける大学出版の応用事例



**fulcrum** For Authors For Publishers For Readers

## Optimize your scholarship

Fulcrum is a publishing platform currently under development that helps publishers present the full richness of their authors' research outputs in a durable, discoverable, and flexible form.

[GET UPDATES](#)

Fulcrum - Michigan Publishing

Anita Gonzalez  
Professor, Theatre and Drama, University of Michigan

**Flexible**

**Durable**

**Discoverable**

By adopting an agile development approach and working in partnership with the Samvera

Built on research university library infrastructure specifically designed to curate

Interoperable with other publishing tools and integrated into the information supply chain.

# 論文データの課題

- クオリティペーパーの電子化
  - 日本史におけるもっともレベルの高い研究は日本語で行われる(ドイツ史ならドイツ語)→しかしそれらの電子化が進んでいない→水準の高い日本研究が国際的に可視化されず、日本研究そのものの低調化につながるおそれ
- 言語問題
  - 一方で、検索によるアクセスという観点からは英語の情報は重要→最低限、コンテンツにたどり着く前のメタな部分までは英語の情報が求められる(資料も同様)
- 書籍の電子化
- 論文情報の管理(引用など)
  - 大学における人文学のプレゼンスの可視化にも重要

# 中間生成物

- 全体に厳しい
- 個人研究に依存しがちな人文学
  - 例外として
    - 考古学の調査ノート
    - 古文書の調査
    - フィールドワークノートのうち、チームで行われたもの
      - ※このうち、デジタルデータとして公開が進むのは考古学の調査
- 資料的価値を含むものは電子化の例もある
- 論文に関係するファクトデータとしてのデータベース
  - 大学・共同利用機関含めて量が多い
  - 長期保存と分散の状況において課題が残る

# データの長期保存

- 注目されつつはある
  - iPresというデジタル保存の学会が2017年度に京都で開催
  - データ保存のための運用(標準化・データ管理モデルの検討)に関する研究も
  - 研究資源としての安定性という意味では人文学特有ではないが
  - 図書館・博物館という観点からは検討が進みつつある
  - とりわけ、いかにシンプルに、共有可能な形でデータを提供できるかという視点から



# 課題解決のために(長期的)

- 資料テキストデータの整備
  - 特に自然言語処理解析等、「新たな人文学」のためのデータ整備の重要性
  - そもそも本文が検索できないと...
- 日本語書籍・論文のデータベース管理
  - 引用の関係を明確に
  - 研究がどのようなサイクルで進んでいるのか 社会にどのように展開しているのかの可視化を
  - 日本研究の国際展開のために
- 中間生成物
  - Digital Humanities の発想からは可能
  - データ処理をどのように行ったかの記録や、資料へのアノテーションなどのノートを保管・(適切な)共有



# 課題解決のために(短・中期的)

- ワンストップサービス
  - 場合によっては検索できなくてもよい
  - どのような資料にどこまでアクセスできるかのポータル(日英)
  - 長期的な維持という観点を含めつつ
  - いくつかの事例をもとに
- 大学等の研究データベースの長期的保存策
  - データを預かる場所などの検討
  - 人間文化研究機構等でも検討はされているが
- 学協会の論文デジタル化
  - せめて雑誌論文だけでも引用のデータベース構築を