

社会科学分野における 研究データ公開

国立情報学研究所
朝岡 誠

J-STAGEセミナー
2020年8月28日



発表の流れ

- 社会科学データの公開
 - 社会科学データの特徴
 - 海外と日本の現状
- 人社データインフラ事業による研究データ環境の整備
 - 事業の目的と内容
 - ✓ 総合的データカタログ
 - ✓ オンライン分析システム
 - ✓ 拠点機関のアーカイブ機能強化
- 制限公開対応の必要性
 - 機関リポジトリへの応用

社会科学データの公開

•社会科学データの特徴

- 人間の営みや社会事象を把握するために人々の意識、行動を調査
 - 個票データとして管理される
 - プライバシー情報が含まれるデータのため、慎重な取り扱いが必要

•社会調査データアーカイブを通じたデータ共有

- 1960年代から、調査データは公共財として認識
 - 二次分析が可能な形でデータを提供
 - プライバシー情報の秘匿処理
 - 分野独特のデータ共有方法
- 利用申請を通じて、条件を満たした利用者にデータ提供

社会科学データの共有

• 海外の最新動向

• オープンサイエンスの進展

- ジャーナル投稿や研究資金獲得時に推奨
- 共有されたデータを用いた「二次分析」が一般化
→ 研究データ提供・利用共有の活発化

• デジタル情報インフラの整備

- データアーカイブのデジタル化
 - 利用申請システムの導入
 - 紙申請、郵送提供からオンライン申請システムへ
 - 利用者判断のために認証フェデレーションの利用
- 研究データの再利用促進のためのインフラ整備
 - データカタログ
 - オンライン分析システム

社会科学データの共有

•日本の現状

- 研究機関を通じて研究データの共有が行われているが、情報インフラの整備が遅れている
 - データアーカイブのデジタル化対応の遅れ
 - 情報発信の遅れ、海外研究者との共同研究機会の喪失
→人社系データインフラの遅れに対する危機感



人文学・社会科学データインフラストラクチャー構築推進事業(人社データインフラ事業)

- 日本学術振興会による事業
- 2018年度から2022年度の5年間で
 - ✓ 研究機関に研究データの整備を委託
 - ✓ 研究データ再利用促進のためのインフラ整備

発表の流れ

- 社会科学データの公開
 - 社会科学データの特徴
 - 海外と日本の現状
- 人社データインフラ事業による研究データ環境の整備
 - 事業の目的と内容
 - ✓ 総合的データカタログ
 - ✓ オンライン分析システム
 - ✓ 拠点機関のアーカイブ機能強化
- 制限公開対応の必要性
 - 機関リポジトリへの応用

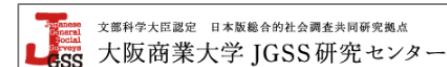
人文学・社会科学 データインフラストラクチャー構築推進事業

中核

- 研究データ再利用促進のためのインフラの開発・整備
 - 人社データカタログ
 - オンライン分析システム
- 研究データ公開ガイドラインの整備・周知
→ 研究データ共有のためのシステム整備

拠点機関

- 研究データの収集、アーカイブ機能強化
- メタデータの作成、多言語化
- 研究データ公開支援
→ 共有される研究データの整備



人社データカタログ

2021年4月 一般公開

CiNii Research



人社データカタログ

日本語 General Social Surveys <JGSS-2008>

東大社研・若年/若年未満調査 (JLPS-Y) Wave1-2, 2008

海外のデータ
検索サイト

国際標準メタデータ
DDIによるメタデータ
自動収集機能



- 社会調査・アンケート・インタビュー・共同研究拠点
東京大学 社会科学研究所
附属社会調査・データアーカイブ研究センター
- WEKO3
文部科学大臣官邸 日本版総合的社会調査共同研究拠点
大阪商業大学 JGSS 研究センター
- WEKO3
パネルデータ設計・解析センター
Panel Data Research Center at Keio University
- HISTORIOGRAPHICAL INSTITUTE THE UNIVERSITY OF TOKYO
東京大学史料編纂所
- WEKO3
一橋大学経済研究所
Institute of Economic Research, Hitotsubashi University
- WEKO3

日本の人文社会科学データの公開サイト

人社データカタログ

The screenshot shows the NII RCOS Human Resources Data Catalog interface. At the top right, there is a language selector set to 'English' and a user profile 'wekosoftware@nii.ac.jp'. Below the header, there are navigation tabs for 'Top', 'WorkFlow', and 'Communities'. The main content area is titled 'ファセット検索' (Faceted Search). On the left, there are three filter sections: 'トピック' (Topic) with options like '労働と雇用(693)', '人口統計学と人口(248)', '経済(723)', and '健康(524)'; 'アクセス制限' (Access Restriction) with '一般公開(744)' and '制限公開(1256)'; and 'データタイプ' (Data Type) with '量的調査(1283)' and '政府統計(754)'. The '経済(723)' option is selected. The search bar contains the text 'Type and press enter to search'. Below the search bar, there are radio buttons for 'Full text' (selected) and 'Keyword'. The search results section shows '1 - 2 of 2 results.' and includes an 'Export' button. The first result is '日本版General Social Surveys <JGSS-2008>' with a description and a 'Show source' link. The second result is '東大社研・若年パネル調査 (JLPS-Y) Wave1-2, 2008' with a description and a 'Show source' link. Red boxes highlight the filter sections and the search results.

トピック、アクセス制限、
データタイプ等で絞り込み可能

検索内容に応じて、各機関の最新情報をもとに研究データが検索できる
→詳細情報から公開サイトにアクセス可

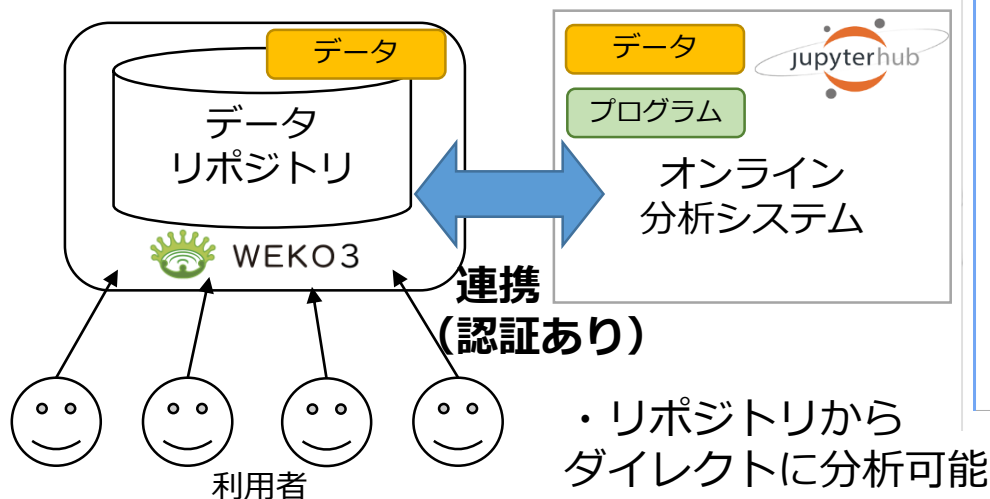
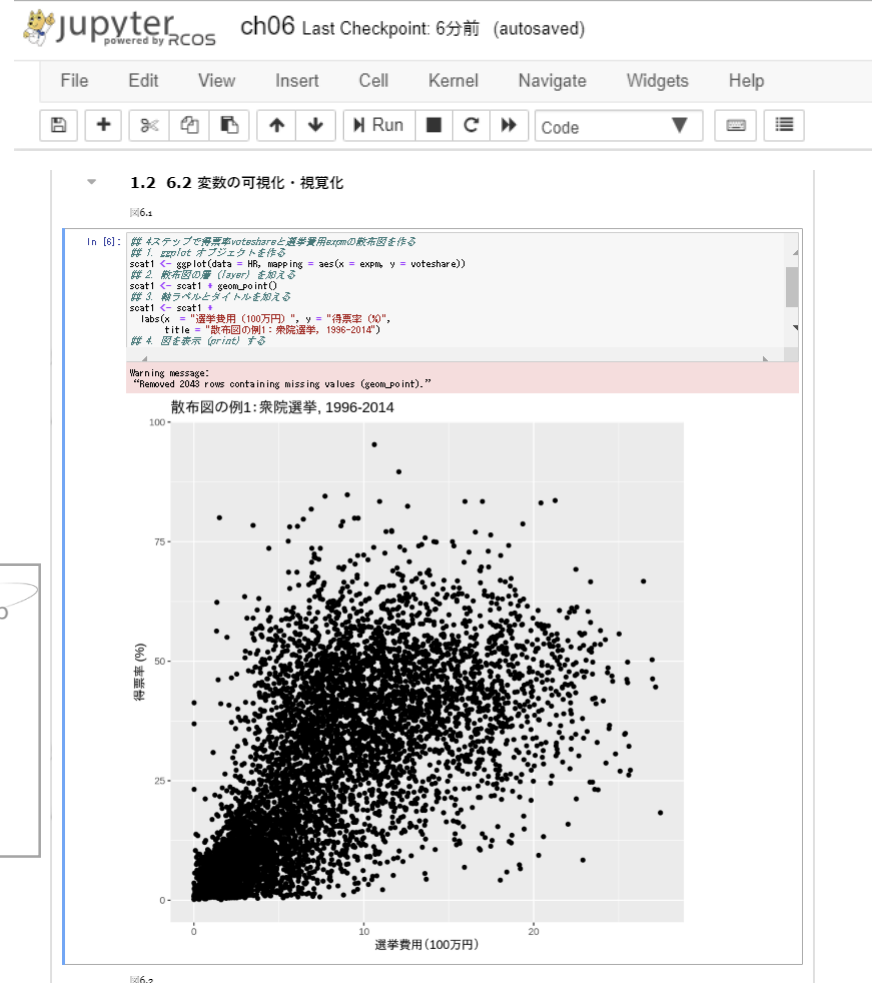
オンライン分析システム

2022年一般公開

JupyterHub

Webブラウザ上から、学認を使ってアクセス

- ・研究データ、分析プログラムをアップロードが可能
- ・R、Pythonを用いたデータ分析が可能



拠点機関のアーカイブ機能強化

社会科学拠点機関の状況

- 独自のインフラを持たずに運用したい
 - 少人数による運営、インフラを整備する人員がない
- 外部システムと連携して利用促進を行いたい
 - データカタログ、オンライン分析システムとの連携
- 利用者審査を行った上でデータを提供したい
 - 利用資格、利用目的の制限



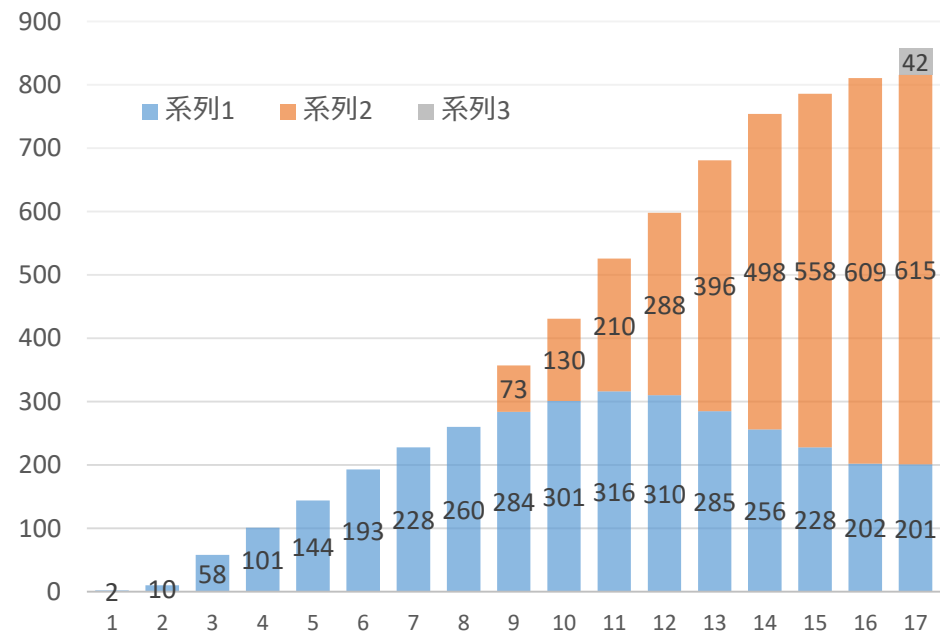
機関リポジトリサービスJAIRO Cloudの利用

クラウド型サービスによる運用



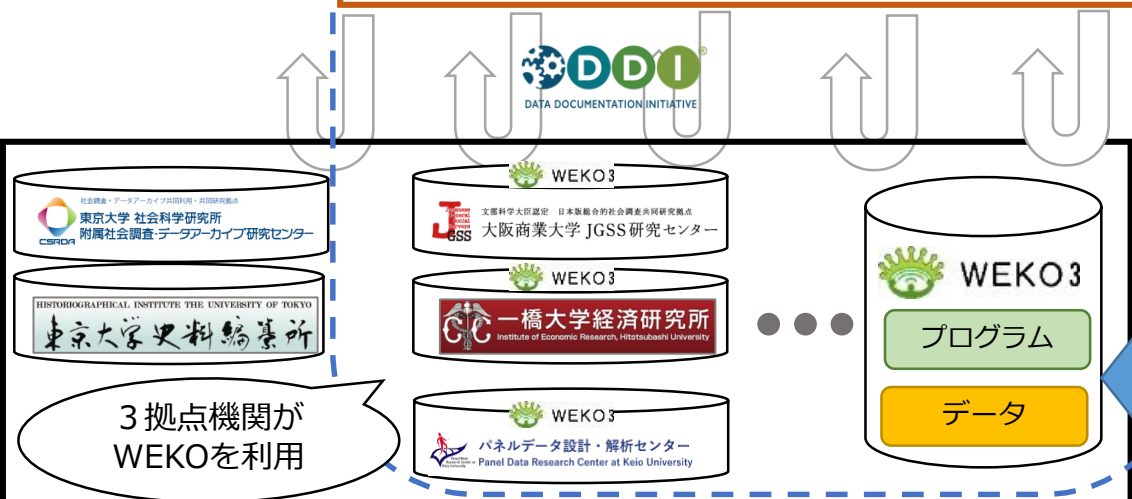
- 研究成果を公開するデータベース機能を提供するクラウドサービス
- 主に大学図書館が博士論文や研究紀要を公開するために利用
- 今年10月から論文だけでなく研究データを取り扱うことが可能

日本の研究教育機関におけるJAIRO Cloud 利用



研究データインフラとの連携

JAIRO Cloud
の仕組みを
使って構築



3 拠点機関が
WEKOを利用

連携
(認証あり)

利用申請システムの実装

- 拠点機関 大阪商業大学 JGSS研究センターとの共同開発で、利用申請システムを実装

JGSS

ワークフロー Communities

Type and press enter to search

Full text キーワード

インデックスツリー

JGSSデータ

アイテム

JGSSデータ

C. 機関への利用申請 データ

<input type="checkbox"/>	名前 / ファイル
<input type="checkbox"/>	datafile.zip (879.6 kB)
	🔒 アクセス制限

データ概要画面

JGSS

ワークフロー Communities

フロー	C. 機関への利用申請	作成日	2020-09-22	更新日	2020-09-22
アクティビティ	A-20200522-00001	作成者	jgssuser1@gmail.com	更新者	jgssuser1@gmail.com
ステータス	ToDo				

アイテム

ステップ

```

graph LR
    Start[Start] --> Item[Item Registration for Usage Application]
    Item --> Approve[Approval by Administrator]
    Approve --> End[End]
  
```

利用規約

日本版組合社会員等の積極的な実践とデータの公開は、調査に協力して下さる回答者と、大阪商業大学JGSS研究センターと、調査データの利用者との間に、信頼関係が成立していることで、実現しているものです。そのため、次の利用規約に同意をさせていただきます。

- 提供された調査データは、学術目的での二次分析にのみ利用します。
- 調査対象者の個人情報を含む、個々の対象者の特定に結びつく可逆性のある形式のデータの事例研究やモニタリングなどの研究発表を行います。調査対象者の個人情報は、本利用規約に賛同した者に対してのみ、第三者には開示しません。共同研究者データを利用する場合、共同研究者全員が個別に利用を申請します。
- 二次分析の結果を公表する際には、データに提供されている謝辞 (Acknowledgement) をそのまま論文、報告書などに明記します (英語・日本語のどちらか一方)。
- 報告書等の論文等を発表した場合、大阪商業大学JGSS研究センターのデータダウンロードシステムから登録します。その際、当該論文等の電子ファイルをアップロードして登録する場合、大阪商業大学JGSS研究センターに登録します。
- データを利用している間は、任意ごとに、JGSSデータダウンロードシステムから研究の進捗に関して利用報告を行います。その際、データ利用を次年度も継続するが終了するシステム上で選択します。
- 利用期間終了後は、調査データを削除します。
- その他、データの利用に関して、大阪商業大学JGSS研究センターの指示に従います。
- 提供された調査データ等の利用により何らかの不利を蒙ったとしても、大阪商業大学JGSS研究センターの責任は一切ありません。
- 本調査の個人情報に関して、以下の事項について、大阪商業大学JGSS研究センターが各者への情報の提供、開示を行うことを了承します。(調査・提供される情報) 氏名、所属、職名、研究会等での写真、研究上の役割、研究題目と概要、JGSS関連文庫の裏表情報 (調査・提供先) JGSS研究センターのホームページ及び編集・刊行物 (基礎集計表・コードブック、研究論文集等)、運営委員会が認めた第三者機関 (データ・アーカイブ、共同研究拠点の報告書など)
- 申請者が大阪学生 (または大学の学部学生) の場合、学務課長の承認が必要です。

※調査や論文、学会の大会報告、授業の実習報告書など印刷されたものを除きます。「論文等」が、印刷物の一部であり全体の提出が困難な場合、該当する「論文等」を写真したもので可也します。論文等のタイプ利用規約を確認の上、スクロール最下部にある「利用規約に同意する」にチェックを入れてください

ヒストリー

アクション	更新日	更新者	結果	コメント
Start	2020-05-22 17:58	jgssuser1@gmail.com	作業済	Begin Action

利用申請画面

人社データインフラ事業では…

- 研究データ再利用促進のためのインフラの整備
 - 人社データカタログ
 - オンライン分析システム
- 拠点機関のアーカイブ機能の強化
 - 機関リポジトリサービスJAIRO Cloudの利用
 - クラウド型サービス利用によるシステム運用コストの軽減
 - 人社データカタログ、オンライン分析システムとの連携
 - 利用申請システムによる条件付きデータ提供への対応



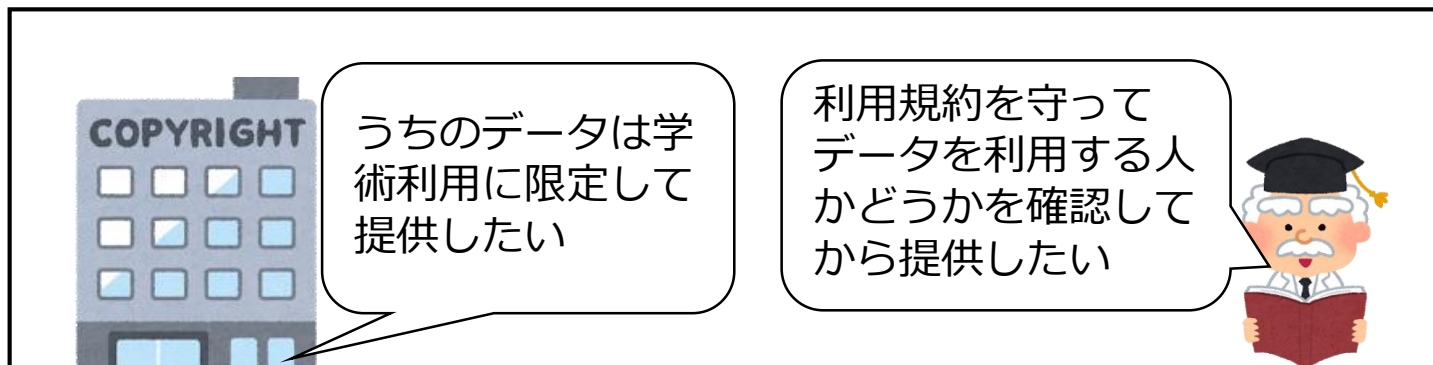
社会科学分野だけでなく、他分野の研究データの公開のためには必要な機能

発表の流れ

- 社会科学データの公開
 - 社会科学データの特徴
 - 海外と日本の現状
- 人社データインフラ事業による研究データ環境の整備
 - 事業の目的と内容
 - ✓ 総合的データカタログ
 - ✓ オンライン分析システム
 - ✓ 拠点機関のアーカイブ機能強化
- 制限公開対応の必要性
 - 機関リポジトリへの応用

制限公開対応の必要性

- 制限公開：条件を満たした利用者へのみデータを提供
 - データ公開が促進される一方で、機密保持の観点から公開になじまないデータもある
 - 研究データ公開ポリシーが制定されていく中で、制限公開によるデータ提供が増えることが予想される



制限公開の例

制限公開の必要性

科学技術専門家ネットワーク構成員への調査

■あり ■なし ■わからない ■無回答

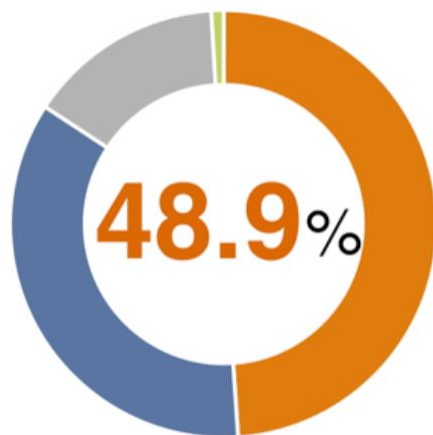


図 45 機密情報の有無 (n=1,396)

■問題である ■やや問題である ■あまり問題ではない ■問題ではない ■わからない

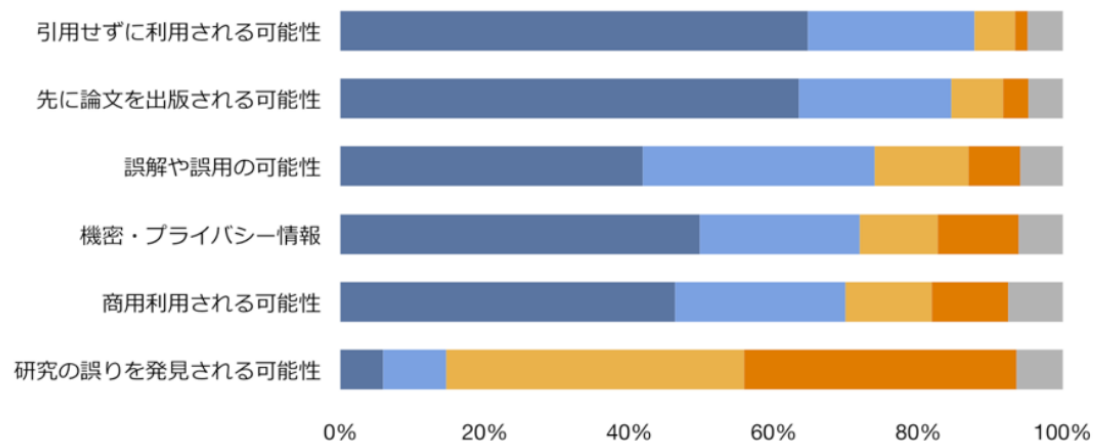
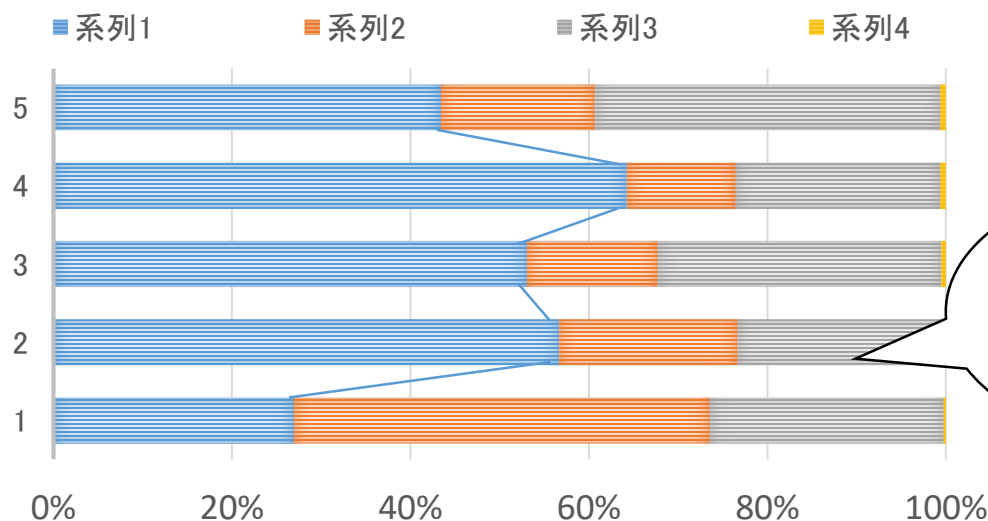


図 53 カレントデータの公開に関する懸念 (n=1,396)

池内有為, 林和弘, 赤池伸一「研究データ公開と論文のオープンアクセスに関する実態調査」, NISTEP RESEARCH MATERIAL, No.268, 文部科学省科学技術・学術政策研究所.DOI: <http://doi.org/10.15108/rm268> より引用

制限公開対応の必要性



自然科学以外の分野は、約半数のリポジトリが制限公開に対応

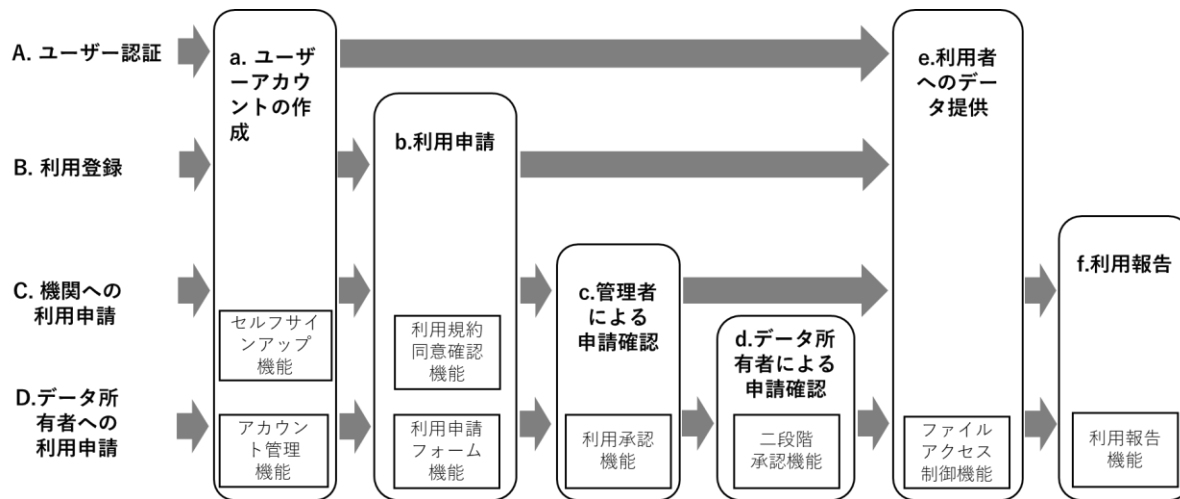
Re3data(<https://www.re3data.org/>) 2019年12月15日現在

Re3dataに登録されたリポジトリにおけるデータへのアクセス方法

→社会科学分野だけでなく、他の分野でも研究データ公開のためには制限公開に対するニーズは大きい

多様な制限公開方法への対応

- 国内の研究機関の制限公開データの提供フローを類型化し、4つのフローを見出した。



- | | |
|------------------|---|
| A. ユーザー認証： | アカウントによるログインで可否を判断 |
| B. データの利用登録： | 利用規約を遵守させることを目的に利用 |
| C. 機関への利用申請： | 利用者を厳密に審査するために用いられる |
| D. データ所有者への利用申請： | データ所有者が自分たちのデータを知的財産として管理する必要がある場合に行われる |

機関リポジトリへの応用

機関リポジトリで制限公開機能の運用

- エビデンスデータの運用
 - 誤用を防ぐため、利用目的を制限したいデータ
 - 一部機密情報を含むエビデンスデータ
- 制限公開機能を利用したデータ提供
 - 学内研究プロジェクトによる利用
 - 紙申請による利用申請フローの電子化

→機関リポジトリで制限公開によるデータ提供を可能とすることで、一般公開が難しい研究データの公開を促進する

まとめ

- 人社データインフラ事業で社会科学研究データの再利用を促進するインフラが整備されている
- 機関リポジトリサービスJAIRO Cloudは社会科学データのリポジトリとして活用される
- JAIRO Cloudに制限公開機能を実装することで、一般公開が難しい研究データの再利用を促進する

研究データの公開で相談したいことがあれば
NII RCOSにご連絡ください！

RCOS

<https://rcos.nii.ac.jp/>

asaoka@nii.ac.jp

研究データの公開で相談したいことがあれば
NII RCOSにご連絡ください！

本発表の一部は、JSPS 人文学・社会科学データインフラストラクチャー構築推進事業 JPJS00220180001, JPJS00218077184の成果です