

一橋大学経済研究所リポジトリ における 電子データの整備等について

2022年5月31日

学術情報基盤オープンフォーラム

伊原 一（一橋大学経済研究所准教授）

1. データリポジトリの概要

一橋大学経済研究所は、日本学術振興会「人文・社会科学データインフラストラクチャー構築推進事業」の拠点事業の一環としてデータリポジトリ構築の取り組みを行っている。

令和3年に、オープンアクセスリポジトリ推進協会と国立情報学研究所が共同運営するクラウドサービス新JAIRO-Cloudに構築環境を得て、同年6月からIPアドレスによるアクセス制限を解除し、試験公開を開始した。

一橋大学経済研究所データリポジトリでは、下記のデータ等を提供している。

- (1) 公的統計メタデータ
- (2) 長期経済統計
- (3) 日本統計年鑑
- (4) 公的統計匿名データ
- (5) オーダーメイド集計補助
- (6) JIPデータベース

(1) 公的統計メタデータ

いくつかの主要な公的統計について、統計年鑑に掲載された統計表などの数字を理解するために必要となる、調査に使用された調査票の画像イメージや産業分類など、統計調査のメタデータを収集、公開している。

(2) 長期経済統計

近代日本経済の歴史統計を、国民経済の計算体系に即しつつ経済活動の諸分野にわたって推計、加工して体系的に集積した『長期経済統計』（大川一司・篠原三代平・梅村又次監修、東洋経済新報社、1965-1988年）のデータベース。

(4) 公的統計匿名データ

公的統計の匿名データを利用して作成した統計表等（未収録）。

(5) オーダーメイド集計補助

一橋大学経済研究所は、日本学術振興会の「人文・社会科学データインフラストラクチャー構築推進事業」の一環として、オーダーメイド集計補助プロジェクトの公募を行っている。本プロジェクトは、一定の条件の下で、この作成料を補助しており、補助を受けたオーダーメイド集計により作成された統計表は、定められた期間後公開される。

(6) JIP データベース

本研究所が独立行政法人経済産業研究所と共同で作成・更新している、日本の産業構造や生産性動向、労働・資本投入等の長期分析を可能にする基礎データ。

各部門別に全要素生産性（TFP）を推計するために必要な資本サービス投入指数と資本コスト、質を考慮した労働投入指数と労働コスト、名目及び実質の生産・中間投入、TFPの上昇率を計算した成長会計の結果等の年次データから構成されている。

2. データリポジトリの提供サイト

(1) 一橋大学経済研究所データリポジトリ <<https://d-repo.ier.hit-u.ac.jp>>



The screenshot shows the homepage of the Hitotsubashi University Data Repository. The browser address bar displays the URL <https://d-repo.ier.hit-u.ac.jp/?page=1&size=20&sort=controlnumber>. The page features a navigation menu with the following items: [メイン](#), [このリポジトリについて](#), [日本統計年表データベース](#), and [関連リンク](#). A notice section titled "お知らせ" contains two announcements: one regarding service interruptions on August 27, 28, and 30, 2021, and another regarding the start of a trial public release on August 29, 2021. The "Main" section includes a search bar with the placeholder text "入力欄、Enterキーで検索してください" and buttons for "検索" and "詳細検索". At the bottom, there are links for "全文" and "キーワード".

(2) 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 <<https://d-infra.ier.hit-u.ac.jp/Japanese/overview/>>



The screenshot shows a web browser window displaying the Japanese overview page for the project. The page title is "日本学術振興会 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 一橋大学経済研究所". The URL is <https://d-infra.ier.hit-u.ac.jp/Japanese/overview/>. The page features a navigation menu with "業務概要" (Business Overview) selected. The main content area is titled "業務概要" and contains a paragraph of text describing the project's goals and timeline. Below the text, there are two buttons: "一橋大学経済研究所長" (Hitotsubashi University Institute of Economic Research Director) and "運営委員会" (Steering Committee).

日本学術振興会
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
一橋大学経済研究所

English
トップページ
JSPSデータインフラ事業

業務概要 日本統計年鑑・帝国統計年鑑 公的統計調査票様式 公開データベース集 オーダーメイド集計

業務概要

一橋大学経済研究所では、日本学術振興会の「人文学・社会科学データインフラストラクチャー構築プログラム」の委託業務として、2018年10月からの5か年間、政府統計に関するプロジェクトを実施しています。このプロジェクトでは、民間問をカバーする政府統計データのデータベース化、調査票情報等メタデータの整備、政府統計標準の集計データや加丁統計（オーダーメイド集計、歴史統計、産業構造データベース及び国際比較データベース等）の整備、及びこれらデータの英語化や国内外への発信、個票データの利活用可能性の向上、等を通じて、国内外研究者コミュニティによる政府統計データやその集計・加丁データの利活用を促進する総合的なシステムを構築します。

実施体制

一橋大学経済研究所長
運営委員会

(3) 人文学・社会科学総合データカタログ (JDCat)

データリポジトリは、令和3年5月に公開を開始したJDCatとのデータ連携を行っている。

JDCatとは、「人文学・社会科学総合データカタログ (Japan Data Catalog for the Humanities and Social Sciences)」の略称で、独立行政法人日本学術振興会が実施している「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」(以下「本事業」という。)の一環として構築された、人文学・社会科学分野データのメタデータ(データセットのコンテンツ、コンテキスト、出所などを表現する手段)の一括検索が可能なデータカタログ。

(3) 人文学・社会科学総合データカタログ (JDCat) <<https://jdcats.jps.go.jp/>>



人文学・社会科学総合データカタログ
JDCat Japan Data Catalog for the Humanities and Social Sciences

システムメンテナンスのため、以下の日程でJDCatを停止いたします。
2022年5月26日（金）14:00～17:00
上記時間帯にてJDCatにアクセスできない時間帯が発生いたします。
ご不便をおかけいたしますが、ご理解いただけますようお願い申し上げます。

[トップ](#) [お知らせ（更新情報等）](#) [JDCatとは](#) [JDCat利用マニュアル](#) [拠点機関のリンク](#) [データ共有のための手引](#) [JDCat分析ツール](#)

拠点機関へのリンク

本事業では、公募により採択された以下の研究所等との連携協同により、取組を実施しています。

<社会科学分野>

一橋大学 経済研究所 （日本学術振興会の外部のサイトへリンクしています）

社会科学における公約統計の利用促進のため、長期関をカバーする政府統計データ（日本統計年鑑）、調査票様式等のメタデータ（国勢調査、事業所・企業統計調査、毎月勤労統計調査）、政府統計標準の集計データ、および加工統計（長期経済統計データベース（LTES）、アジア長期経済統計データベース）などを受け入れています。
今後、政府統計データや歴史統計を中心に委託受入を行うか否か、どのように行うかについては現在検討中です。

収集したデータ

日本統計年鑑／調査票様式等／長期経済統計：（日本）長期経済統計データベース（LTES）、アジア長期経済統計データベース／オーダーメイド集計／JIPR-JIP-CIPデータベース／くらしと健康の調査（JSTAR: Japanese Study of Aging and Retirement）／大株主・役員データベース2012／POSデータに基づく独自統計／近世経済データベース／新聞記事テキストデータ（明治末～昭和45年）／個別企業資料（採訪資料）

慶應義塾大学 経済学部附属経済研究所パネルデータ設計・解析センター （日本学術振興会の外部のサイトへリンクしています）

慶應義塾大学パネルデータ設計・解析センターは、パネルデータに関するデータアーカイブを構築し、外部の研究者や調査機関より社会科学関連のパネルデータの委託を受け入れています。
・ 委託されるデータは匿名処理後のものとし、センターの提示するフォーマットに沿ってメタデータや変数説明も合わせて提供いたします。

(4) 学術機関リポジトリデータベースへの登録

データリポジトリの書誌情報等は、学術機関リポジトリデータベース (IRDB: Institutional Repositories DataBase) に登録されている。

(4) 学術機関リポジトリデータベースへの登録 <https://irdb.nii.ac.jp>



The screenshot shows the IRDB search results page. The search criteria are '一橋大学 経済研究所'. The results are displayed in a list format, showing the first 20 items out of 21,628 total items. The results are filtered by 'すべて' (All) and '本文あり' (Full text available). The search results are sorted by 'レコード番号(新しい順)' (Record number (newest first)).

IRDB 検索 コンテンツ統計(全体) ログイン English

一橋大学 経済研究所

すべて 本文あり 詳細検索

21628 件中の 1 件目～ 20 件目を表示

20件ずつ表示 レコード番号(新しい順) ページCSV出力

平成7年～平成15年農業経営統計調査：調査票様式ほか調査概要
農林水産省
一橋大学経済研究所データリポジトリ；一橋大学

平成2年度～平成6年度農家経済調査：調査票様式ほか調査概要
農林水産省
一橋大学経済研究所データリポジトリ；一橋大学

昭和52年度～平成元年度農家経済調査：調査票様式ほか調査概要
農林省 | 農林水産省
一橋大学経済研究所データリポジトリ；一橋大学

昭和46年度～昭和51年度農家経済調査：調査票様式ほか調査概要

資源タイプ
other (11584)
dataset (6031)
journal article (2137)
periodical (1019)
technical report (452)
book (191)
report (117)
still image (79)
departmental bulletin paper (31)
conference object (2)
もっと見る

出版タイプ
NA (2137)
VoR (1728)
AM (2)

言語
jpn (18167)
eng (3388)
fra (80)

3. データ整備の手順

(1) リポジトリ構築

JAIRO Cloud環境の確保（有料）

テストサイト構築

コンテンツ登録

テスト運用（IPアドレス制限）

コンテンツ追加更新

運用開始（IPアドレス制限解除）

(2) コンテンツ登録

統計年鑑の例

原資料入手（大学図書館等）

著作権の確認（総務省統計局）

使用許諾（取り交わし文書）

スキャナ入力

OCR認識

テキストデータ校正

書誌情報作成

（凡例、JDCat統制語彙等の情報付与）

コンテンツ登録（リポジトリ）

4. データ利用の留意点

(1) 公的統計メタデータ

- ・ 調査票のスキャン画像は「Sample」のすかしが入っている。
これは、調査票イメージをかたり調査（偽調査）などに悪用されないようにするためである。
- ・ 調査項目リストと産業分類リストは、
フォントがUTF-8になっていてエクセルで開くと文字化けするためメモ帳等で開く必要がある。

4. データ利用の留意点

(2) 長期経済統計

- ・収録データは農林業と鉱工業のみ

(3) 日本統計年鑑

- ・昭和3～15年版の発行者は東京統計協会となっている
- ・昭和2年以前と昭和16年以降は未収録

(4) 公的統計匿名データ

- ・未収録となっている

4. データ利用の留意点

(5) オーダーメイド集計補助

- ・一定期間後に公表するため、現時点で2018年度のみ結果表を収録

(6) JIPデータベース

- ・2015と2021のみを収録。

5. データ整備の課題等

(1) 歴史資料等の電子化の課題

- ・ 統計表は通常の記事と異なり、表形式のOCR認識が必要
- ・ 統計表の罫線は誤認識の原因となる
- ・ 貴重な歴史資料は、非破壊で画像データ入力が必要
- ・ 著作権法上は70年以上経過していれば公表可能
- ・ 旧漢字を含むためOCR認識が困難

5. データ整備の課題等

(1) 歴史資料等の電子化の課題

- ・ 時系列接続は、収録表が変化するため困難
- ・ 単位が変わることがある。

(尺貫法等)

- ・ 対象地域が変わることがある。

(外地等)

- ・ 品目名称が変わることがある。

(サツマイモ／甘藷等)

5. データ整備の課題等

(2) オープンデータ化の留意点

- ・ 著作権の制約に注意が必要
- ・ 電子図書館では1冊ごとに1利用者
- ・ データだけでは利用できない
- ・ 基データの書誌情報が必要
- ・ メタデータ（周辺資料等）も必要
- ・ 文字化けや誤植等への対応
- ・ 問い合わせ先の明示
- ・ 引用方法の明示
- ・ 原資料の所在明示

5. データ整備の課題等

(3) データ提供の課題

- ・ 事業の継続性が必要
- ・ 構築後も事業維持の予算が必要
- ・ 機材の更新が必要
- ・ データベース技術が必要
- ・ 技術進歩への対応が必要
- ・ 高解像度化への対応
- ・ ファイル形式の多様化
(テキスト、画像、エクセル、DB等)
- ・ アプリケーションの進歩

5. データ整備の課題等

(4) データリポジトリ改善の課題

- ・コンテンツの充実
- ・検索機能の強化
- ・事業の継続性

(データインフラ事業は令和4年度まで)

- ・機関リポジトリ等への登録

6. 機関リポジトリの紹介

一橋大学では、大学附属図書館が機関リポジトリとして「HERMES-IR」を運用している。

機関リポジトリ（IR：Institutional Repository）とは、機関（大学等）で生産された研究成果（全文）を電子的に保存し、発信するインターネット上の集積庫のこと。

一橋大学機関リポジトリ「HERMES-IR」 <<https://hermes-ir.lib.hit-u.ac.jp>>



The screenshot shows the HERMES-IR website interface. At the top, there is a navigation bar with the university logo and the text "一橋大学機関リポジトリ HERMES-IR Research & Education Resources". A search bar is prominently displayed in the center, with a "検索" (Search) button and a "クリア" (Clear) button. Below the search bar, there are several news items, including one dated 2022/3/25 regarding service suspension and another dated 2022年1月19日 regarding document information. The interface is clean and professional, with a dark red header and a white main content area.

(1) 主な収録内容

① Research & Education Resources

学術雑誌論文・研究報告書・博士論文・紀要・
会議資料・教材・図書・その他

収録件数 26497件 (2022年5月現在)

14946件	本学紀要論文	221件	研究報告書等
4800件	雑誌掲載論文	159件	会議・学会発表資料
3650件	ワーキングペーパー	1件	教材
1553件	研究会刊行物	0件	学園史関係資料
819件	学位論文	98件	その他
250件	図書		

② Special Collections

フランクリン文庫・井藤半彌旧蔵ドイツ紙幣等
コレクション・戦前期アジア諸国写真コレクション
ン・附属図書館所在肖像画等コレクション・福田
徳三関連資料など

収録件数 23426件（2022年5月現在）

8734件 都留重人名誉教授寄贈資料（経済研究所資料室所蔵）

4036件 戦前期アジア諸国写真コレクション

3476件 日本・旧満州鉄鋼業資料：水津利輔氏旧蔵資料

2605件 学園史関係資料 = University Archives

1238件 札差関係資料

695件 井藤半彌旧蔵ドイツ紙幣等コレクション

447件 日本郵船株式会社社会計帳簿類

2195件 その他

(2) 研究成果の登録方法

- ①論文（全文）の電子ファイルを用意
- ②利用許諾（同意）書を記入
- ③電子ファイルと利用許諾書を送付
- ④図書館で論文の検索を可能にするためのキーワードを付与して公開

研究成果を登録・公開するには

対象とする研究成果

- ・学術雑誌論文
- ・研究報告書
- ・博士論文
- ・紀要
- ・会議資料
- ・教材
- ・図書 等

研究者データベース

HRI: Hitotsubashi Researchers Information

に研究成果を登録する。
※教員の方のみ

著作物の
電子ファイル

+

著作物利用許諾
(同意)書

★許諾要件を確認の
上、ご提出ください

メール添付または電子メディアを提出

HERMES-IR
Hitotsubashi University Repository

に登録します！



登録にあたり、出版社や学会等との間の著作権の確認を行ってください。難しい場合はご相談ください。
また、共著者が存在する場合には登録前に同意を得てください。

(3) 機関リポジトリの特徴

- 大学附属図書館が運営
- 主に刊行物を収録（図書館業務）
- 大学紀要に掲載した論文は全て収録（義務化）
紀要論文の数が多い
- 学位論文については出版の制約で掲載できないものがある
(著者最終稿を掲載できる場合がある)
- 著名な歴史資料等を電子化して収録
- 各種研究データの収録も可能

(4) 機関リポジトリのメリット

①研究者のメリット

研究成果（全文）がインターネットで公開される
研究成果がデジタル保存される
研究成果のアクセス及び引用が向上する
読者層が拡大する

②大学のメリット

知名度が向上する
研究・教育活動を公表できる
国際的学術交流に寄与する

(4) 機関リポジトリのメリット

③利用者のメリット

論文を閲覧できる

検索が容易

利便性が向上

7. 研究データ整備の課題等

(1) 提供上の課題

- 研究データは10年保存（義務）
- 研究データは公開向けに編集されていないものも多い
- 研究者自身による管理が必要
- データ利用には専門知識が必要
専門用語、旧漢字（漢文）、手書き文字、尺貫法、法律等

7. 研究データ整備の課題等

(1) 提供上の課題

- 画像データの利用はOCR処理が必要
- 古いOSの分析アプリ等は動かない（再現性）
- プログラムのソースコード等も必要
- 中間生成物の必要性

7. 研究データ整備の課題等

(2) 収録上の課題

- 複数のバージョンが存在
(試行錯誤の研究プロセスの場合)
- 研究データは膨大
(量が多く、ビッグデータ等も含まれる)
- IoTの進展に伴うデータの大量発生
- 全ての研究データを収録することは物理的に不可能(?)

7. 研究データ整備の課題等

(2) 収録上の課題

- ・ 誰がどこに何を登録するか
(受皿と標準化が必要)
- ・ 研究者DB、機関リポジトリ等と相互リンクが必要 (研究者<論文<研究データ)

7. 研究データ整備の課題等

(3) 個人情報を含むデータの制約

- ・ 公開できるのはオープンデータのみ
- ・ 個人情報を含むデータ（クローズデータ）は公開不可
- ・ 手書き資料はOCR処理が困難
- ・ 入力を外注する場合は国外持ち出し不可（漏洩対策）

7. 研究データ整備の課題等

(3) 個人情報を含むデータの制約

- ・ データ入力後にデータクリーニングが必要
(家計簿等)

合計と内訳が合わないことがある

出生年と年齢が合わないことがある

項目間に矛盾が含まれることがある

不詳が多数発生する場合がある

- ・ 標本バイアス等に注意が必要

8. 研究データリポジトリの方向性

研究データリポジトリ整備・運用ガイドライン（平成31年3月内閣府）では、研究データリポジトリについて、「国研・大学等により整備されるものであって、とくに公的資金により得られた研究データを再利用できる形で保存・管理し、当該機関内または機関外へのデータ利用サービスに供するもの」と定義している。

8. 研究データリポジトリの方向性

広義の研究データを全てリポジトリに収録することは困難であるが、当面の方向性として、公的資金（科研費等）による研究成果について研究データを収録して提供する研究データリポジトリの構築と運用が求められている。