

## JGSS データダウンロードシステムの構築

### —システムのデザイン・開発・修正—

郭 凱鴻

大阪商業大学 JGSS 研究センター

金 政芸

大阪商業大学 JGSS 研究センター

竹本 圭佑

藍野大学医療保健学部

岩井 紀子

大阪商業大学総合経営学部

### Development of JGSS Data Download System: Design, Construction and Revision

Gaiko KAKU  
JGSS Research Center  
Osaka University of Commerce

Jeongwoon KIM  
JGSS Research Center  
Osaka University of Commerce

Keisuke TAKEMOTO  
Faculty of Health Sciences  
Aino University

Noriko IWAI  
Osaka University of Commerce

Funded by the JSPS Program for Constructing Data Infrastructure for the Humanities and Social Sciences Grant Number JPJS00218077184, JGSS Research Center is developing JGSS Data Download System (JGSSDDS). In Japan, JGSS data were deposited to and distributed by Social Science Japan Data Archive (SSJDA) and JGSS Research Center only handled special use data, geographical data, and data on regional block and prefectures. In order to organize the data usage applications and data management, JGSS Research Center in collaboration with National Institute of Informatics develop JGSSDDS, adding to the WEKO3 that has been developed by the NII. Designing the JGSSDDS began in July 2019, and numerous revisions were needed during the development process. Since the Fall 2020, the three organizations had routine meetings to revise the system. Due to delay in the development of the WEKO3 and merging of the JGSSDDS into WEKO3, the launch of the JGSSDDS is slightly delayed and will be summer 2021. EASS and other data which JGSS Research Center receive from other researchers for data preparation for sharing will also be distributed via JGSSDDS.

Key Words: JGSS Data Download System, JGSS Research Center, Data sharing

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」受託を契機として、JGSS 研究センターは、「JGSS データダウンロードシステム」(JGSSDDS) の構築を進めている。JGSS データは、国内では東京大学社会科学研究所 SSJDA に寄託され、一般公開前のデータと地点情報のみ、本センターで申請を受け付けてきた。国内での利用申請と管理作業を本センターに集約するために、国立情報学研究所 (NII) が更新を進めている情報公開システム WEKO3 に組み込む形で、本センターと NII が JGSSDDS を共同開発している。JGSSDDS のデザインは 2019 年 7 月に開始したが、開発過程で多くの修正を要した。2020 年秋以降は、3 機関合同のオンライン会議を定期的にもち改善を重ねた。WEKO3 開発の遅れと JGSSDDS との機能の組合せから生じる問題から遅れてはいるが、2021 年夏には稼働する予定である。JGSSDDS には、EASS や外部の研究者から整備を依頼されたデータも搭載される。

キーワード: JGSS データダウンロードシステム、JGSS 研究センター、データ公開

#### 1. JGSSDDS 開発の背景

本稿では、JGSS 研究センター (以下、本センター) が NII と共同で進めている「JGSS データダウンロー

ドシステム」(以下、JGSSDDS)の開発の背景と、開発過程のなかで生じた問題とその修正の流れ、今後の展望について紹介する。JGSSDDSでは、本センターが公開するデータの学術利用を希望するユーザが、ウェブ上でデータの利用申請とダウンロード、成果物登録、利用報告を行うことができるデータアーカイブシステムである。同時に、本センターはJGSSDDSを運用することで、利用者の研究成果情報などから本センター公開データの二次分析研究の広がり状況をより容易に把握することができる。

まず、JGSSDDSの開発に至った経緯から説明しよう。本センターは、2018年から日本学術振興会(JSPS)の「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、データインフラ事業)の拠点機関に採択され、その委託業務に取り組んでいる。データインフラ事業は、「人文学・社会科学に属するデータを分野や国を超えて共有・利活用する総合的な基盤を構築することにより、研究者がともにデータを共有しあい、国内外の共同研究等を促進することを」<sup>1</sup>目的としており、本センターの他に、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、東京大学史料編纂所、一橋大学経済研究所、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・解析センターがデータ共有基盤の構築・強化委託業務を遂行する拠点機関として事業に参加している。拠点機関には、機関が保有するデータだけでなく、外部の研究者や研究機関が実施した調査データを整備して保存し、データ共有のために公開することが求められる(「データアーカイブ機能の強化」の取り組み)。

しかし、本センターが実施する日本版総合的社会調査(JGSS)のデータは、これまで外部のデータアーカイブに寄託し、公開してきた。JGSSデータは、日本では東京大学社会科学研究所附属社会調査・データアーカイブセンターのSSJDA(Social Science Japan Data Archive)に、海外ではミシガン大学のICPSR(Inter-university Consortium for Political and Social Research)とドイツのGESIS(Leibniz Institute for the Social Sciences)に寄託してきた。外部のデータアーカイブに寄託することは、データの管理とユーザへの提供を寄託先で担ってくれるなどのメリットがあり、本センターはこれまで独自のデータアーカイブを持たない方針を取ってきた。

ただし、データを外部に寄託することにはデメリットも存在する。寄託先の事情によっては、データを寄託してから公開されるまでかなりの時間がかかり、予定より遅くデータが公開されることがある。データのバージョンアップをする際にも、同様に寄託先の処理を待たなければならない。

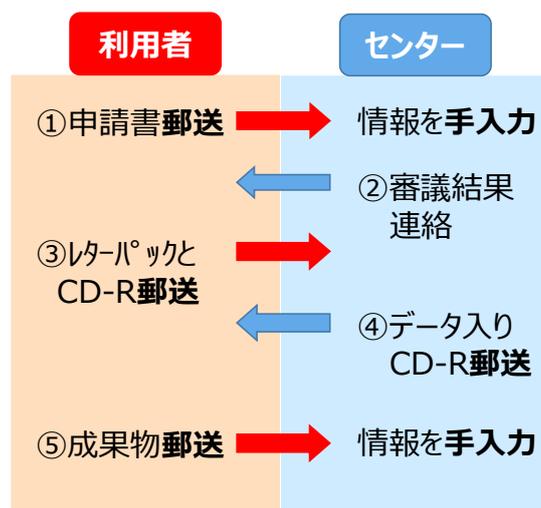


図1 JGSSDDS 開発以前の本センターで直接公開するデータ利用の流れ

<sup>1</sup> <https://www.jspa.go.jp/j-di/torikumi.html>

また、調査対象者個人の特定につながりやすい情報を含むデータは、外部のデータアーカイブへの寄託が難しい。JGSS データも、個人の特定につながる可能性がある調査対象者の地域ブロックと都道府県情報は削除してから外部のデータアーカイブに寄託している。地域ブロックと都道府県情報は、保証人(所属機関長)の署名を求めるといった通常より厳しい制限を設け、本センターが直接に郵送で公開していた。同様に、パネル調査である JGSS ライフコース調査 (LCS) のデータも郵送で提供してきた。しかしこの場合、図 1 で示しているように利用申請からデータの配布 (CD)、利用報告、成果物提出までの本センターと利用者のやりとりのすべてが郵送で行われていたため、効率的とはいえない。

JGSSDDS は、こうした問題の解決を目指して開発しているシステムである。JGSSDDS の運用により、JGSS データなど本センターが実施した調査データは、寄託先の公開を待たず、データ整備が終わったらすぐ JGSSDDS から公開できるようになる。また、公開データの利用に必要な一連の過程を一つのシステム上で行えるため、データの利用の各手続きを簡単に行うことができる。

## 2. JGSSDDS のシステムデザインと共同開発

### 2.1. システムデザイン

独自のデータアーカイブを開発する方針が決まり、本センターはすぐにシステムのデザインを始めた。データインフラ事業の拠点機関のうち、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は、SSJDA (Social Science Japan Data Archive) といったデータアーカイブを 1998 年からいち早く開設している。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・解析センターでも 2017 年から「データ申請・管理システム (DMS)」から複数のパネルデータを公開している。本センターは、5 月からこれらの拠点のデータアーカイブならびに ICPSR、GESIS などの海外のデータアーカイブを参考にしながら JGSSDDS のシステムデザインを進めた。

システムデザインの際に特に重点をおいたのは、データの種類によって利用申請時に複数回の承認をユーザに求める機能である。利用承認の段階は、ユーザの役職とデータの種類とによって異なるように設計した。大学教員や常勤研究員は、JGSS の通常データや LCS データ、EASS データを本センターの承認のみでダウンロードできる<sup>2</sup>。一方で、調査対象者の特定につながりやすい JGSS の地域ブロック・都道府県データと、JGSSDDS では申請のみで実際の利用は本センターでオンサイトになる地点情報データの申請は、2 段階の承認が必要になる。これらのデータの場合は、利用者の所属機関長などが保証人となってユーザの申請を JGSSDDS 上で承認した後、本センターが最終承認をすることでダウンロードが可能になる。また、大学院生のデータ申請は常に指導教員の承認が必要になる。たとえば、大学院生が JGSS の通常データの利用を申請した場合は、指導教員と本センターの承認が求められ、JGSS の地域ブロック・都道府県データを申請した場合は、指導教員、保証人、本センターの 3 段階の承認を経て利用が可能になる。学部生も大学院生と同じく申請に指導教員の承認が必要であるが、JGSS の地域ブロック・都道府県データと地点情報データのオンサイト利用申請は制限される。他のデータアーカイブにはあまり見られないこうした多段階承認機能を持つことで、JGSSDDS では匿名性が十分に確保されていないデータを公開することが可能になる。

ユーザの JGSSDDS 利用の流れは、図 2 のようデザインした。最初に、ユーザは JGSSDDS でアカウントを登録し、氏名、所属、役職などのユーザ情報を記入する<sup>3</sup>。アカウントを登録すると、データの利用申請が可能になる。登録後にログインして JGSSDDS のデータ一覧から利用を希望するデータの掲載ページに

<sup>2</sup> 通常の JGSS データでは、調査対象者が特定されないようにデータの一部の情報を修正・削除している。たとえば、調査地点の情報が含まれているため、対象者の特定につながる可能性がある識別番号 (ID) は通常データではランダムに番号を振りなおしている。また、JGSS-2008 以降のデータでは、地域ブロックと都道府県の情報も削除して公開している。

<sup>3</sup> 今後、全国の大学や研究機関などが連携して構築している「学術認証フェデレーション (GakuNin)」のシングルサインオンにも対応する予定である。GakuNin に参加している所属機関のユーザは所属機関から取得した ID とパスワードを利用して、GakuNin に参加している学外機関のサービスを ID とパスワードの再入力をせずに利用できる。詳細については、GakuNin のホームページ (<https://www.gakunin.jp>) を参照のこと。

ある申請ボタンを押すと、利用申請の画面が表示される。ユーザはデータの利用規約に同意し、研究題目、研究計画などの情報を記入して利用申請を行う。本センターは、ユーザの利用申請内容の審査を行う。ただし、データによっては先述したように、指導教員、保証人の承認が先に求められる場合もある。本センターが最終承認を行うと、ユーザはデータページにあるダウンロードボタンを押してデータをダウンロードできる。ユーザはダウンロードしたデータを分析し、論文や書籍などの成果物を発行した場合に、それを JGSSDDS にアップロードして登録する。なお、ユーザには年度末に研究の結果や進捗状況などについて利用報告を行うことが求められる。利用報告時に、データ利用を次年度にも継続するか終了するかを選択することになる。

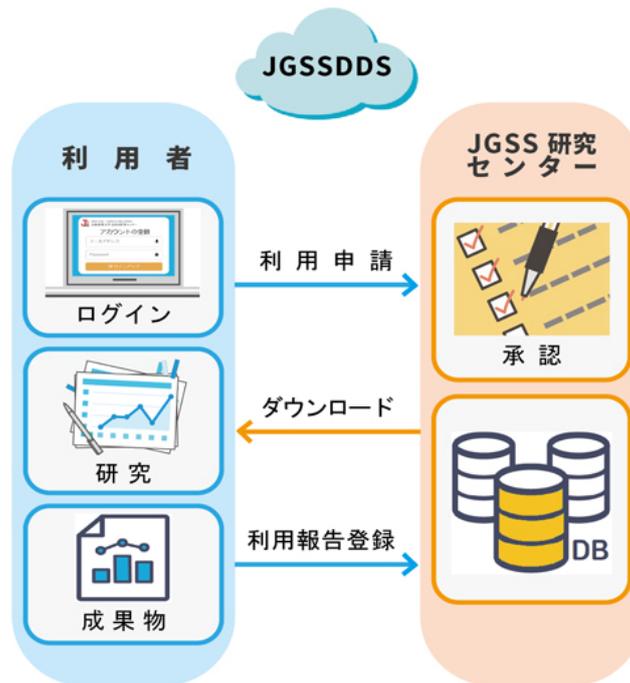


図 2 JGSSDDS 利用の流れ

## 2.2. 国立情報学研究所との共同開発

JGSSDDS のデザインを固め、委託業者の選定を進めていた 6 月末に国立情報学研究所 (NII) から共同開発の提案を受けた。本センターは、6 月に独自のデータアーカイブを構築する意向を、データインフラ事業の委託元である日本学術振興会に伝え、この取り組みがデータインフラ事業の中核機能として学振が NII に委託している業務など他の業務と重なったり競合したりすることはないかについて尋ねた。その際に、NII から JGSSDDS 共同開発の提案を受けた。NII は、データインフラ事業において中核機能の構築の 1 つである人文学・社会科学総合データカタログ (Japan Data Catalog for the Humanities and Social Sciences: JDCat)<sup>4</sup>とオンライン分析システムの開発を担っている。一方、NII は多くの大学で機関リポジトリとして利用しているクラウド型の機関リポジトリサービスである JAIRO Cloud を、2012 年からオープンアクセスリポジトリ推進協会 (JPCOAR) と共同で運営しており、2017 年からは JAIRO Cloud の新しい基盤ソフトウェアとなる WEKO3 の開発を進めていた。WEKO3 はこれまで文献リポジトリとして運用されていた既存の WEKO2 基盤の JAIRO Cloud を、研究データリポジトリとしても運用できるようにすることを開発目的の 1 つとしていた。ただし、WEKO3 は当初から汎用的なリポジトリシステムとして開発が進められていたため、本センターがデザインした複数回承認による制限公開機能や、利用報告、成果物報告などの機能は、

<sup>4</sup> JDCat では、人文学・社会科学分野データのメタデータ (データのタイトル、概要、作成者、配布者などの情報) を一括して検索できる。本センターを含むデータインフラ事業の 5 つの拠点機関が保有するデータのメタデータを公開する予定である。

開発計画に含まれていなかった。

NII の提案を受けてから、本センターと NII は共同開発に向けて協議を行い、WEKO3 の機能を拡張する形でこれらの機能を開発することで JGSSDDS を実現することを合意した。JGSSDDS が WEKO3 をベースにすることになり、JGSSDDS の開発も NII が WEKO3 の開発を委託している業者に依頼することになった。7 月からは NII が WEKO3 の開発を委託している委託業者も加えて協議を重ね、先述した多段階承認などの JGSSDDS の機能を、WEKO3 を基盤にしてどのように実現できるかについて具体的な方針を固めていった。本格的な開発は 10 月から開始した。開発が進む間にも、ウェブ上でタスク管理ソフトウェアを利用して本センターと NII、開発業者の三者でやり取りを続け、毎月 1 回の定例会議を開いて開発の詳細な内容を決めながら開発を進めた。JGSSDDS の委託開発は 1 月末に完了した。

表 1 JGSSDDS 開発のながれ

|                      |  |
|----------------------|--|
| 2019 年 5 月 16 日      | NII が JGSS 研究センターに来学してヒアリング  |
| 2019 年 5 月 21 日      | JGSS の内部会議で JGSS データのダウンロードシステムを持つ方向性を決定<br>他の拠点へのヒアリングを重ねてシステム (JGSSDDS) をデザイン    |
| 2019 年 5 月 29 日      | システム開発の会社と打ち合わせ・見積依頼を開始 (協議 6 社、見積 4 社)  |
| 2019 年 6 月 20 日      | 学振の「データカタログ」との連携を想定し、NII と相談開始   |
| 2019 年 6 月 24 日      | 第 1 回の NII とテレビ会議<br>・ NII との協力について協議  |
| 2019 年 6 月 25 日      | JGSS の内部会議で NII と協力する方針を決定   |
| 2019 年 7 月 1 日       | 第 2 回の NII とテレビ会議<br>・ NII のサポートを受けつつ、<br>WEKO3 の機能を拡張する形で JGSSDDS を開発することに合意      |
| 2019 年 7 月 12 日      | 委託業者とテレビ会議   |
| 2019 年 9 月 11 日      | 委託業者と契約締結  |
| 2019 年 9 月 26 日      | 学振の「データカタログ」のメタデータに関する説明及び意見交換会  |
| 2019 年 10 月 11 日     | 第 1 回の JGSSDDS 開発のテレビ会議 (JGSS、NII、委託業者)<br>・ 開発のキックオフとスケジュールの確認                    |
| 2019 年 10 月 23 日     | 委託業者から担当のエンジニア 2 名が来学<br>・ JGSS データや現行運用について説明<br>・ WEKO3 に基づく JGSSDDS のデモとイメージの共有 |
| 2019 年 11 月 8 日      | 第 2 回の JGSSDDS 開発のテレビ会議 (JGSS、NII、委託業者)<br>・ 現在までの実績確認                             |
| 2019 年 12 月 6 日      | 第 3 回の JGSSDDS 開発のテレビ会議 (JGSS、NII、委託業者)  |
| 2019 年 12 月 18 日     | 連絡協議会で JGSSDDS を含む取組状況報告   |
| 2020 年 1 月 11 日～12 日 | JGSS 国際シンポジウム 2020 にてデータインフラ事業の取組を報告   |
| 2020 年 1 月 17 日      | 第 4 回の JGSSDDS 開発のテレビ会議 (JGSS、NII、委託業者)  |
| 2020 年 1 月 31 日      | JGSSDDS の開発完了、納品   |
| 2020 年 2 月～4 月       | NII の実験環境の構築の待ち  |
| 2020 年 5 月～8 月       | 実験環境での JGSSDDS のテスト<br>・ 49 件の修正必要箇所   |

JGSSDDS が WEKO3 を基盤とする JAIRO Cloud 上で運用されることによる利点は大きい。第一に、JAIRO Cloud サービス上で運用することで、低コストでシステムを維持管理することができる。JGSSDDS のようなウェブ基盤のサービスは、通常は専用のサーバを構築して運用されることが多い。しかし、独自サーバの構築には、サーバ等のハードウェアの購入などの多額の初期費用が発生し、そのメンテナンスやアップグレードなどの維持管理費も年々重なっていく。一方で、クラウド型サーバをレンタルする場合は、物理サーバを構築する必要がないため、初期費用をかなり抑えることができるが、高いセキュリティ対策を維持するための OS やソフトウェアの更新などの維持管理は依然として必要になる。JAIRO Cloud では、こうした維持管理を国立情報学研究所が担うことになり、JAIRO Cloud サービスの利用料金だけで運用が可能になり、維持管理の負担をかなり抑えることができる。JAIRO Cloud を利用することで、本センターは JGSSDDS のコンテンツ管理の部分だけに集中しながら運用を続けることができるようになるのである。

### 3. JGSSDDS のシステムテストと改修

#### 3.1 システムテストの実施

JGSSDDS のための WEKO3 拡張機能の開発は 2020 年 1 月末に一旦完了したが、WEKO3 のメイン機能開発が継続していたため、システムテストは同年 5 月から開始された。テストは複数の研究員が分担して行い、各機能を相互にクロスチェックする形で進められた。テストにより、計 49 件の課題が抽出された。

抽出した課題は、まず「I. システム全般・管理機能」、「II. 利用者関連」、「III. 英語対応」の 3 つにカテゴリに分類し、さらに「追加」「修正」「バグ」「その他」の 4 類型に課題に分けて整理した（表 2）。「追加」とは、開発期間中に新たに必要となった機能や、当初の仕様で不足していた点を補う必要がある課題である。「修正」は、仕様の解釈の違いや説明不足により、実装された機能が想定されたものと異なる課題である。「バグ」はプログラム上の単なる不具合と判断した課題であり、「その他」WEKO3 メイン機能側の未解決課題など、JGSSDDS 単独では対応できない問題である。

課題類型のうち「追加」の項目が多いのは、開発委託当初とは違って、公開するデータの種類が増えたことが主な原因である。地域ブロック・地点データ、EASS データ、外部研究者から寄託されたデータも公開対象となり、それに伴い新たなカテゴリや申請プロセスに対応できるシステムにする必要があったからである。

表 2 システムテストにより抽出された課題

|                      | 追加 | 修正 | バグ | その他 |
|----------------------|----|----|----|-----|
| I システム全般・管理機能        |    |    |    |     |
| I.1 システム関連           | 6  | 1  | 0  | 0   |
| I.2 自動送信メール関連        | 0  | 1  | 1  | 0   |
| I.3 利用統計ダウンロード関連     | 0  | 2  | 1  | 1   |
| II 利用者               |    |    |    |     |
| II.1 利用者画面の全体にかかわる問題 | 0  | 1  | 0  | 3   |
| II.2 利用申請関連          | 4  | 1  | 1  | 2   |
| II.3 利用報告関連          | 1  | 1  | 1  | 0   |
| II.4 成果物報告関連         | 0  | 1  | 3  | 2   |
| III 英語問題             |    |    |    |     |
| III.1 サイト表示関連        | 2  | 1  | 0  | 1   |
| III.2 メール関連          | 6  | 0  | 0  | 0   |
| III.3 サイトの英語表現関連     | 0  | 0  | 0  | 5   |
| 合計                   | 19 | 9  | 7  | 14  |

### 3.2 改修の実施

抽出された49件の課題については、2020年8月に本センター・NII・委託業者の三者で2回の定例会議を開催し、個別に内容を精査した。最終的には、「追加」課題に関しては、委託業者に改修依頼して開発し、「バグ」や「その他」の課題に関してはNIIが対応することとなった。

2020年11月からは、委託業者による改修作業が本格的に開始され、地域ブロック・都道府県データやEASSデータへの対応、外部データの受け入れ、申請プロセスの新設、プログラム修正などが集中的に行われた。進捗はタスク管理ツールを通じて三者間で随時共有され、定例会議で機能の確認を行った。2021年1月末には委託業者による改修作業が完了し、翌2月に改修部分のテストが実施され、「追加」課題はほぼすべて解決された。NIIによるWEKO3のメイン機能の改修作業も開始されており、WEKO3メイン機能とJGSSDDSの拡張機能との円滑な連携を目的とした対応も着々と進められている。

## 4. おわりに

本稿では、JGSSデータダウンロードシステム(JGSSDDS)の構築に至る背景から、システムデザイン、開発過程で生じた課題とその対応について報告した。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」の受託を契機に、JGSS研究センターは、JGSSデータに加えて外部の研究者や研究機関が実施した調査データも公開できるよう、JGSSDDSの構築を進めている。

本センターは、国内外のデータアーカイブ(ICPSRやGESISなど)を参考にしつつ、JGSSDDSのシステムを設計した。特に重視したのは、JGSS単年度・累積データ、地域ブロック・都道府県データ、EASSデータ、外部データといったデータの種類や、利用者の役職に応じて、多段階の承認を求める機能である。本センターは国立情報学研究所(NII)と連携し、汎用的なりポジトリシステムであるWEKO3を拡張する形でJGSSDDSの開発を進めた。

しかし、データ整備の加速や外部データ公開方針の確立により、公開予定のデータの種類と量は当初の設計を大きく上回った。その一方で、開発当初の要件をすべて満たすことはできず、一部機能の追加や修正が必要となった。開発段階のテストでは、3つのカテゴリと4つの課題類型に分類された計49件の課題が抽出された。これらの課題をもとにNIIと分担してシステムテスト改修が進められた。

今後は、JGSSDDSにJGSSやEASSのデータ、さらに本センターが整備を支援する他機関・研究者のデータを掲載し、データの利活用を一層促進していく。また、「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」に参加する他拠点とも連携し、総合データカタログの整備やデータ共有の推進に取り組んでいく。