

2014年6月
海外研究員(デリー)
坂井 華奈子

ジャワハールラーネルー大学センサス・データ・センターと 中央図書館政府刊行物部門を訪問して —インド政府情報の保存と利用の観点から—

ジャワハールラーネルー大学 (JNU) にインドの人口センサスのマイクロ (ミクロ) データが利用できる場所があると聞いて、4月にデリーで開かれた図書館情報学関係の国際会議で知り合った JNU の図書館員に紹介を頼み、訪ねてみた。

センサス・データ・センターのウェブページ¹によると、センターはインドの大学や研究機関に所属するすべての学生・研究者に開かれている、と書かれていたが、外国の機関に所属するインド研究者にとっても有用なデータであろう。最近ではビッグデータやオープンデータの有用性が注目されているが、12億人の国民を抱えるインドの人口センサスのデータを詳細な分析ができる形で公開することにより、社会経済的状況について、これまで以上に深く分析できる可能性が広がる。

今回は、ウェブサイトの情報になかった海外の研究者の利用の可否や、実際の利用環境などを確認するために訪問することにした。

データ・センターを訪問する前に、仲介をお願いしたライブラリアンに面会するために、まず JNU の中央図書館²を訪ねた。JNU 中央図書館の9階には政府刊行物部門があり、インドの政府刊行物をはじめ、海外の政府刊行物、国連などの国際機関の刊行物が所蔵されている。以前も出張で訪問したことがあったが、新しい担当者を紹介してもらい、政府情報の図書館における提供について情報交換を行った。JNU 図書館での政府刊行物の収集方針について聞いてみたところ、センサスは非常に重要なので冊子体で出たものはすべて購入しているが、最近は冊子体では出していない (CDしかない) ものがあり、それらは網羅的に入手できていないことが課題らしい。その点については筆者も同様に認識しており、経験的にも書店・代理店への依頼の仕方によって、CDは対象からもれているのではないかと感じていた。また、書店への発注だけでなく、出張で政府刊行物の販売所を訪ねた際もセンサスの冊子体の取り扱いがあったが、CDについてはセンサス・オフィスの Data Dissemination Unit³に行くように、と言われた経験がある。

冊子と CD の両方で同じタイトルが出ている場合に、限られた予算の中でどちらを購入すべきかは判断の難しいところである。長期的な保存にとっては冊子体の方が安心でき、CD などのようにデジタルの場合は媒体の安定性や、ファイル形式や扱えるソフトウェアなどが将来的に OS やバージョンが変わっても利用できるかという点で不安が残るが、集計・分析を行うには圧倒的にデジタルの方が利便性が高い。また、冊子体での発行がなく CD のみのタイトルについては特にその把握と収集が今後の課題である。

¹ JNU Census Data Centre <http://www.jnu.ac.in/CDC.htm>

² JNU Central Library <http://www.jnu.ac.in/Library/>

³ Data Dissemination Unit, Census of India <http://censusindia.gov.in/2011-common/ddU.html>

しかし、JNUの図書館ではセンサス以外の政府刊行物は寄贈に頼っていて、継続して集めるべき年刊類でも90年代後半や2000年代前半あたりで止まってしまっているタイトルがかなりあるとのことであった。おそらく、政府情報のウェブサイトへの掲載が進んで以降、省庁での冊子体での発行部数が減少し、寄贈の配布部数も削減されたのではないかと推測される。これまで寄贈に頼る受け身の体制であったため、送付が止まってもメンテナンスがされておらず、届かなくなってからの督促などもあまりしていなかったようであるが、今後、欠号部分についても何らかの方法で入手を図りたいとのことであった。

筆者が、政府のウェブサイト上に掲載されている政府刊行物について、古くなったバックナンバーが更新時に削除される可能性など、安定的な提供についての懸念を述べたところ、共感が得られた。現在所蔵している省庁のAnnual Reportなどの資料については、冊子体の保存・提供に加え、内部向けにデジタル化し、保存・提供する方向を検討しているそうである。政府刊行物を利用する学生や研究者は、同じシリーズから特定の数ページのみを複写する機会が多いため、そのような利用方法にはデジタルの方が利用者にとっても効率的だと考えているそうである。

インドの図書館の書庫や書架スペースには空調が天井扇しかないことが多く、気温が40度～45度を超える環境下に保存されてきた古い冊子体については劣化が目立つ。同様の理由で、筆者が現在所属している経済成長研究所(IEG)図書館が所蔵しているセンサスのマイクロフィッシュは、劣化が進みすぎてケースに張り付いてしまい取り出すことができなくなっていた。冊子体資料のデジタル化と図書館での提供においては著作権処理などの課題はあるかもしれないが、保存環境の点からも、特にインドにおいては、デジタル化は重要なのではないかと感じた。

現在の新しい政府刊行物部門の担当者は着任して半年ほどのことであったが、学位論文の電子化や機関リポジトリなどの担当も兼ねており、デジタル化やデータリポジトリについても意欲を持っていた。しかし、部門内では専任は彼のみであり、数名の部下は十分な知識を持たないスタッフのみで、実際の業務にあたっては常に指導・確認が必要なため進捗ははかばかしくなく、新しい事業は構想はあるが実現はすぐにはできないだろうとのことであった。ただし、過去のセンサスのデジタル化については、現在インド政府とUNFPA(国際連合人口基金)がプロジェクトを立ち上げているそうである。古いCensus of Indiaの冊子体は本国であるインドにおいても所蔵状況がまちまちであり、一箇所ですらえることが難しいため、各図書館を回って欠号を埋めるために協力を依頼しているそうである。筆者の所属するIEG図書館でも同様の話を聞いており、今後の動きが注目される。

センサス・データ・センターについては、同じ大学内でも図書館とは直接かかわりのない部署であるためこれまで訪問したことはなかったが、政府刊行物部門とは利用者層が重なっており協力したいと考えていたそうで政府刊行物部門の担当者とも一緒に訪問することにした。

センサス・データ・センターはJNUのコンベンションセンターの1階(日本でいうところの2階)の一室に設置されており、中央図書館からは徒歩で5分ほどの距離にあった。インド政府のOffice of the Registrar GeneralとCensus Commissioner of India、そしてJNUの共同で運営されている。施設はJNUに属し、JNUのCentre for the Study of Regional Developmentが監督しているが、データについてはセンサス・オフィスから順次提供されている。

センサス局は収集したデータから重要な指標や集計表のサマリーを公開しているが、インドの社会経済的状況の人口統計学的な深層分析において大きな可能性を秘めているユニットレベルのマイクロデータを公開するために、JNUをはじめインド国内の大学や研究機関にこのようなマイクロデータを公開する拠点を設けている。2012年5月にオープンしたJNU Census Data Centreが国内初の拠点であり、現在までにPunjabi University、Kerala University、Goa University、Gokhale Institute、Nabakrushna Chaudhury Centre for Development Studies等に同様の施設が設けられている。⁴⁵⁶⁷⁸

JNUのセンサス・データ・センターにはセンサス・オフィスから提供されたユニットレベルのマイクロ・データが保存されたサーバと、そこからコピーしたデータ・テーブルが保存された利用者用の端末4台、およびプリンター1台が提供されている。ただし、元データには住所など個人を特定できるような情報が含まれているが、その部分は削除して、問題のない部分のみを提供しているようだ。また、データそのものの複製は許可しておらず、ノートPCの持込やCD/USBメモリ等へのコピーは禁じられている。利用者は、センターの端末で集計・分析を行い、結果の表をプリントアウトしたものを入手できる。利用者用の端末は、ネットワークから遮断されており、USBスロットはふさがれていた。

現在マイクロデータが利用可能なのは1991年、2001年に加え、2011年についても利用可能になり次第、順次提供されている。筆者の訪問時点で利用可能となっていた2011年のデータ・テーブルは以下のとおりである。

- Primary Census Abstract,
- General Population Tables,
- Socio Cultural Tables(現時点では Age のみ)
- Housing Tables

また、販売されているCD版やウェブサイトで購入可能な公開版のデジタルデータについても、同様に端末に保存されており分析に利用できるが、こちらについてもセンサス・データ・センターではコピーは許可していない。センターはあくまでも分析のための拠点であり、データの配布はData Dissemination Unitの役割ということであった。データのコピーを持ち帰ることができないと知ると、かなり憤慨して帰る利用者も多いそうである。マイクロデータのコピーについては情報の特性上許可が難しく、公開データのコピーについては販売されているデータであり、図書館で著作権の問題で資料の全文コピーができないのと同じようなことであると説明された。

⁴ “Workstation for Research on Census Data at Jawaharlal Nehru University, Delhi” Census of India <http://www.censusindia.gov.in/2011-common/Workstation-Research.html>

⁵ “CENSUS STUDIES WORKSTATION OPENED IN BHUBANESWAR” The Pioneer, 3 June 2014 <http://www.dailypioneer.com/state-editions/bhubaneswar/census-studies-workstation-opened-in-bhubaneswar.html>

⁶ “South India's first census research centre at Goa University” Times of India, 4 April 2013 <http://timesofindia.indiatimes.com/city/goa/South-Indias-first-census-research-centre-at-Goa-University/articleshow/19369983.cms>

⁷ “Pune gets Census Data Centre at Gokhale Institute” Sakal Times, 8 February 2014 <http://www.sakaaltimes.com/Tiny.aspx?K=O1IYG>

⁸ “Census research work station at Kerala University” The Hindu, 19, August 2013 <http://www.thehindu.com/news/cities/Thiruvananthapuram/census-research-work-station-at-kerala-university/article5037877.ece>

センターは大学の営業日の 10:00~13:00、14:00~17:00 に利用できる。一日の利用者数を尋ねたところ、あまり多くなく、3人~5人程度とのことであったので利用者用端末の数(4台)は妥当と思われる。訪問時は外気温が 46 度を超える猛暑であったが、センターの室内はエアコンもきいており、滞在して分析するには快適な環境だと思われる。

サーバには統計分析用のソフトウェアである Stata がインストールされているとのことだったが、利用者用端末には Excel や Notepad など基本的なソフトのほかに CSPro

(<https://www.census.gov/population/international/software/cspro/>) というソフトウェアがインストールされていた。

マイクロデータのファイルはテキスト形式で提供されていた。必要なデータを探すには、Excel の表が用意されており、スタッフも支援してくれる。

外国人の利用も可能かどうかを尋ねたところ、ほかの利用者と同様に記帳(住所、氏名、訪問日時、サイン)すれば利用可能との回答を得た。

インドのセンサスは他の国の国勢調査と比較して調査項目が幅広く、その結果は複数年にわたって数百冊を超える報告書として刊行される。これまでは収集された膨大なデータのうち、すでに集計され、公表された形での情報しか得ることができなかったが、この元データをデジタルな形で提供することによって、項目を詳細に絞り込んで掛け合わせたり、自由な切り口からの分析が可能となり、様々な分野に有用な知見が得られることが期待できる。

データのコピーは許可されていないが、マイクロデータがデジタル形式で分析用に提供されているセンサス・データ・センターはインドの社会経済的な調査・研究にとって重要な施設であると感じた。

Census Data Centre , Jawaharlal Nehru University

<http://www.jnu.ac.in/CDC.htm>

利用時間

大学営業日の 10:00~13:00、14:00~17:00

利用のための手続き・条件

記帳(住所、氏名、訪問日時、サイン)

ただし、ノート PC の持込や USB メモリの利用は禁じられており、データのコピーは不可。利用者用の PC から分析結果をプリントアウトすることは可能。



