

連載

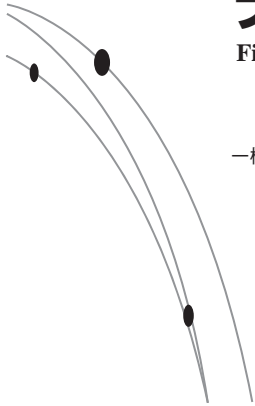
## フィールド・アイ

Field Eye

パリから——①

一橋大学准教授 神林 龍

Ryo Kambayashi



### データ・アクセスの行方

少々の紆余曲折の末、筆者は2010年4月から経済協力開発機構（OECD）雇用政策局に滞在する機会を得た。突然決定が下された時にはすでに今年度の予定も決まりつつあったものの、諸方面のご迷惑を顧みずすべてにご破算を願って、現在パリで過ごしている。そのひとつの余波ではないが、このような連載を担当する機会をいただいた。とても光栄に思うのだが、目下のところ私の語学力が記事を作る上で最大の障害となっている。

確かにフランス語なるものを勉強したことはあったが、それは今となっては20年ほど前にさかのぼるし、そもそもとりあえず単位さえとればよいという（当時の）典型的大学生の態度だった。ということは、講義には必要ときだけ出席し、試験は大学受験で培った手練手管でぎりぎり乗り切ってきたわけである。身につくはずがない。おまけに、筆者が滞在しているOECDは、フランスにあるのだけれども、国際機関という性格上英語が仏語と並ぶ公用語となっている。筆者は学部大学院を通じて日本国内で過ごしているので英語もどうにか意思疎通ができる（と自分に言い聞かせる）水準であるが、仏語よりはるかにましで、今のところ仕事でこの国の生粋の言語を使うことは一切ない。したがって、この国で何が起っているかも、おぼろげながらしかわからない。

というわけで、筆者は今、完全にエトランジェなのである。この連載も、こういう立場から見聞きしたこと、考えたことを書きつけようと思う。ゆえに、話題はフランス国内のみならず、筆者が興味深いと思ったことを取り上げよう。

いきなり自己中心的な言い訳で始め恐縮しているのだが、今回は米国で開かれた国際会議のことをお話ししよう。この会議はWorkshop on Data Accessと通称され、米加英独蘭など欧米諸国の統計部局の人々を中心に、政府統計の個票の利用方法などに関する情報共有を目的として開催されている。今年5月に3回目がミシガン大学 Interuniversity Consortium for Political and Social Research (ICPSR) 主催で、1日半程度の日程で行われた。

今となってはご存じの方は少なくないと思われるが、現在、欧米での政府統計の個票の利用方法は、public-use fileと呼ばれる匿名化された公開データ以外に、俗にオンサイト施設と呼ばれる、外界から隔離・監視された無菌室のような場所で利用する方法が普及してきている。当初、すでにデータ・アーカイブの伝統を築いていたミシガン大学ICPSRやエセックス大学U.K. Data Archive (UKDA)などは、データは公開され共有されることが原則であると主張し、この方法が新たな利権を生み出すことに批判的だった。ところが、彼の地もプライバシー保護の波にさらされたことに他国と違いはなく、データの寄託者などによって重度の匿名化が要請された結果実用に耐えられなくなるデータセットが続出した。そのため両者とも近年オンサイト施設の構築を決定し、英国ではそれに伴い統計部局からも公開データが提供されることとなった。この会議で報告あるいは討論された話題のひとつはこういった組織の方向転換についてだったが、他の出席者には周知の事実らしく、それほどつつこんだ話にはならなかった。結局のところ、ICPSRとUKDAの方針転換が象徴するように、データ・アクセスは公開データとオンサイトの二本柱で構成される方向に収斂しつつあるようである。

この会議で熱心に議論されたことは、データ・アクセスの国際化についてであった。匿名化されたデータでさえ、外国での利用が想定されていない日本では、データ・アクセスを外国に居ながら許すという事態はおよそ考えられない。しかし、欧米諸国ではすでに現実のものとして考えられ、話し合われているのである。もちろん、こうした一見するとラディカルな施策には伏線がある。

ひとつはリモート・アクセスと呼ばれる技術の飛躍的進歩である。これは、シン・クライアントと呼ばれる構造で古くから知られていたネットワーク構築方法

の引き写しで、個別端末には画面とキーボードの出入力しか許さず、すべてを中央のサーバーで管理・運用するというシステム構想である。サーバーの計算能力と管理能力、そして通信速度の発展を前提に、近年ではパッケージでシステムが販売され、比較的簡単に組織全体に導入出来るようになってきた。このシステムを通じてデータにアクセスすれば、すべてのユーザーの行状はサーバーで管理・制限・記録される。現在のオンサイト施設が持つ唯一の利点は、目視による監視だけになる。目視による監視の目的は、要するに、悪意あるユーザーの行為を掣肘することにあったが、オランダやスウェーデンなどいくつかの国の統計部局は、こういった悪意あるユーザーを前提とすること自体、現実的ではないと考えた。その結論として、すでにリモート・アクセスによるデータ・アクセスを許容している。

もしもインターネットを介してリモート・アクセスによるデータ利用が認められたとしたら、どこからアクセスしているかという現実世界の地理的情報は意味をほとんどなさない。つまり、スウェーデン統計局のサーバーにスウェーデン国内のどこかの大学からアクセスするか、日本のどこかの大学からアクセスするかは、インターネットを介している以上ネットワーク的には等価である。この方法が十分実用的であれば、個票の国際共有には技術的には障害が小さくなる。

もうひとつの伏線は実際に共有が進められていることである。公開データの収集に関しては米国ミネソタ大学人口研究所 Integrated Public Use Microdata Series International (IPUMS-I) の試みが有名である。IPUMS-I は各国の国勢調査を中心とした公開データを収集し、一手にアクセス可能にするアーカイブで、現在 55 カ国のべ 150 回以上の国勢調査、総計 3 億人以上のデータが収蔵されている。国勢調査という性格上調査項目は限られるものの、サンプルサイズは大きく、一部のデータは職業や産業分類について国際分類に調整し直されているなど、国際比較をするうえで非常に便利なプラットフォームが形成されている。また

EU ではすでに Eurostat と呼ばれる統一統計局が設立され、European Community Household Panel (ECHP) など統一統計が実施されて久しい。これらの統計はもちろん EU 内で共有されている。さらに、Eurostat では EU 各国の労働力調査をも共有する計画が進行中である。EU という統一法規に根拠をもつ特殊な事例とはいえるが、データ・アクセスの国際共有に際して起こりえる問題点などを考えるにはうってつけの先事例となっている。

このように、データ・アクセスの国際化はすでに夢のような話ではなくなってきており、久しぶりに高揚感のある会議だった。とはいうものの、ようやく匿名データの制限的リリースにこぎ着けた日本の現状を考えると少々寂しい気もした。この寂しさの原因は、単に日本が各国の後塵を拝しているからというよりも、日本の統計の論理が内向きに過ぎることにあるのではないかと思う。自国の個票データを外国にしながら利用可能にするというのは、現在統計に携わる人々からするとおよそ理解しがたい状況だろう。しかし一方で、日本人研究者は各国統計の多大な恩恵にも与かっている。実際筆者も米国 CPS を扱っているし、IPUMS-I では日本人利用者数が堂々 20 位以内に入っている（しかもデータを拠出していない国で唯一の大口利用国であることは会議でも指摘された）。もちろん、各国がデータを提供する理由は単なる国際協力の精神だけではなく、それが自国（あるいは自分の組織）の利益になると考えているからであろう。また、日本も 30 年越しにルクセンブルク所得研究にデータを拠出可能になるなど、法律の範囲でできる努力をしていると思う。とはいえ、統計の作成・利用に関して、もう少し国際的な視点も必要なのではないかと感じた会議だった。

かんばやし・りょう 一橋大学経済研究所准教授。最近の主な著作に『日本の外国人労働力——経済学からの検証』（共著、日本経済新聞出版社、2009年）。労働経済学専攻。