

58 日本における法定伝染病統計の分析

一九〇〇—一九六〇 (四)

平山 勉

本報告では、「日本における法定伝染病統計の分析(二)」「(三)」で使われた、『衛生局年報』『府県統計書』などから得られる豊富なデータを、より有機的なデータベースにする方法を具体的に提示するとともに、歴史学研究におけるデータベースのあり方について展望を試みたい。

近年、データベースという用語は広く使われており、歴史学の研究でも積極的に活用する傾向が生まれている。最も広い定義は、インタビュ(オーラル・エビデンス)を文字に起こしてワープロに打ち込んだものや、スキャナーで画像保存された公文書などを含む、「デジタル化された歴史資料」というものであろう。しかし、数値データからなるデータベースについては少し誤解が

あるように、表計算ソフト (Microsoft Excel, Lotus 1-2-3 など) のシートに数値を打ち込んだだけのものを、データベースとする人が少なくない。今回、改めて提起したいのは、誰が使っても「同じ結果」が得られる表形式のデジタルデータは、実証の根拠となるデータファイルではあっても、決してデータベースとは言えないということである。つまり、どんなに正確なデータが大量に打ち込んであっても、データを自分なりにアレンジすることによって、何か新しい発見をする「可能性」が担保されていないければ、歴史学の研究に寄与しうるデータベースとは言えないのである。

このような考え方をすると、これまで苦勞して入力してきたデータが、あまり価値のないもののように感じられるかもしれない。ある意味で、それは、新しい史料の発見によってデータが追加されるという、歴史学研究の特性による制約とも言える。しかし、新しいデータに合わせてフォーマットを作り変えていく「柔軟性」と、そうしたデータを付け加えていく「拡張性」を備えることで、有機的なデータベースを構築することは可能であ

る。それどころか、そもそも「完成したデータベース」というものが存在するはずがないことを考えれば、史的資料の発掘状況に合わせて築き上げていくスタイルこそが、データベースとしての本質を表現しているのである。

市販のデータベース・ソフトを使って、複数の種類からなるデータファイルをデータベースとして構築する場合、リレーショナル機能を使って連結していく方法と、一定の形式で一括管理する方法とがある。リレーショナル機能を使ったデータベースは、個票を扱う場合に優れており、一括管理型のデータベースは、数値表を処理するのに適している。今回の報告では後者について取り上げ、それぞれの数値をカード形式で整えて一括管理した上で、クロス集計を通じて様々な数値表を作成する方法を具体的に提示する。というのも、この「クロス集計」と「検索」とを組み合わせることで「自分なりのアレンジ」を出すことができ、新しい発見の可能性が担保されると考えられるからである。なお、使用するソフトは、

Lotus Approach による。

言うまでもなく、個人が入力できるデータには物理的な限界があり、また、問題関心から外れたために入力されないデータもあるだろう。しかし、多くの研究者が、個々にデータを入力していることは、誰もが認める最近の研究状況であると思われる。こうしたデータがひとつのデータベースとなつて、デジタルデータにもとづいて研究を進める人々の間で共有され、新しい知見を共感するようなかたちで、歴史学の研究がより豊かなものになればと願う次第である。

(慶應義塾大学グローバル・セキュリティ・リサーチ・センター)