

# データ・アーカイブと 個人データの保護

高千穂大学 経営学部 教授 渋 谷 政 昭

## 文書館

成功した支配者は栄光を誇示するために歴史書を作り施政の記録を残す。その中には粉飾も多いが、すべての事象を綿密に記録させる人も現れ、制度となって長く継続することがある。逆に支配者を追放すると記録一切を灰にして過去の権威を抹殺する。

現代国家も多かれ少なかれ「古文書館」を作り、行政記録を保存している。これら古文書がアーカイブ、古文書館を運営する人がアーキビストである。コンピュータの世界にこの言葉が入って、処理中でないファイルを圧縮して保存したものアーカイブと呼ぶ。

US ArchiveがワシントンD.C.にあり、その資料がしばしば話題となる。ここを題材にしたJ.アーチャーの小説「盗まれた独立宣言」はフセインのきわめて平和的な謀略物語である。日本でも多くの自治体が中央官庁に先駆けて「文書館」を作り、学芸員を養成し、情報公開に努力されている。

## データ・アーカイブ

当然とも言えるが、経済・社会調査データを蓄積し、これを広く研究利用に開放するデータ・アーカイブが作られ、ウェップのアクセスが便利となっている。「R」という便利なデータ分析用ソフトが無料利用可能となり、実証的研究が盛んである。

イギリスのData Archive, University of EssexおよびNational Digital Archive of Datasets、ルクセンブルグのLuxemburg Income Study、アメリカのInter-university Consortium for Political and Social Research, University of Michiganなどが先駆的先進的な施設である。データ・アーキビストが重要な専門職として認知されている。研究機関だけではなく、ヨーロッパ連合諸国、アメリカなどの統計局も最近オンライン・データ・アーカイブを充実し、すべての人にサービスしている。例えば、US Bureau of Censusでは2000年国勢調査のかなり詳しいデータを公表している。

調査に当たって調査機関は、その目的を明記し、それ以外にデータを利用せず、調査対象の個人・組織名を秘匿することを約束する。調査結果は、通常多重分類表という形式で公表される。これに賛同し協力してくれる人々の善意によって調査の信頼性が保障される。しかし収集された個票には要約された集計表以外の情報も潜在しており、データ・アーカイブはできるだけ加工しない調査個票を含める。



### 個人データの漏洩

調査個票を公開したとき問題になるのは個人・企業データの漏洩、プライバシーの侵害である。個人を特定できるデータは抹消されているが、年齢、性、職業、収入など分析にあたって必要な補助情報の項目が多く、分類が詳しいと個人を識別できてしまう。その結果個人に被害が生じることはあまりないが、サンプルの中に含まれていたこと自体公開されないはずであり、調査者と被調査者の間の信頼を壊す。これは本質的な矛盾で、公開するデータの情報を増すと、個人を識別できる可能性はどうしても高くなる。個人保護を優先すればデータの詳しさを制限せざるをえない。特に問題は職業分類と地域の分類で、これらを細かくすれば個人が浮かび上がってしまう。個人の秘密とは何かも重い課題である。難病患者はそれを他人に知らせたくない。しかし、多くの同病者が苦しんでいるならば、対策のためにその事実の公開を辞さない。

どこに線を引くか、数理統計学の一つの課題として研究しているが、すばらしい名案はあまり期待できない。任に当たる人の高度な知性と判断にたよる面が強い。

### 調査データの公有

中央官庁のことは別の資料に譲り<sup>1</sup>、地方自治努力の一環としてデータ・アーカイブ作成を推進していただきたい。例えば、介護保険が市町村主体で実施され、独自の介護サービス提供が可能であり、各市町村の介護市場での競争が容認されている。保険料負担者、介護受益者、それらの家族、介護サービスにあたる人、団体・企業、すべての努力を集めて業務が進行している。福祉関連予算とも連結する必要がある。その結果すべての人の満足を得るためにには、できるだけ具体的なデータを示して、関係者の知恵を得なければならない。現在進行の業務データは自然と多くの人の目に触れ、プライバシー侵害、秘密漏洩の危険が非常に高い。過去の記録にも危険があるが、十分配慮し関係者の同意を得て、公開することが重要である。

種類が違うが、岡山県と岡山大学が協力して、県内商圈での購買調査を公開して、経済活動に役立てている。「数値目標」を政治家が選挙公約として取り上げているが、それを空虚な内容にしないためには、計画、実施、評価、を公開で進める基盤が欠かせない。

基本的には役に立つデータを協力して作り、皆で利用して役立てる。調査対象者にも法と規則ではなく、納得して合意していただく。調査データを官有ではなく、公有にしなければ基本的矛盾を解決できない。

(慶應義塾大学名誉教授)

1 松田、濱砂、森、編著、2000、『ミクロ統計分析』日本評論社