

社会調査の Data Life Cycle — DDI と R を利用した統合環境 —

○関西学院大学

中野康人

1. 報告の目的

本報告の目的は、社会調査および社会調査にもとづく研究を、より効率的かつ体系的に実践するための「道具」を提供することにある。具体的には、社会調査の国際規格である DDI(Data Documentation Initiative) とデータ分析環境 R を利用した、社会調査の統合環境を提案する。

2. Data Life Cycle

社会調査の実践は、複数の段階を経ておこなわれる。企画・準備段階、実査段階、分析段階、公表段階など、社会調査全体の過程をここでは Data Life Cycle と呼ぼう。この DLC のそれぞれの段階において、われわれは異なるソフトやファイルを使って作業を行うことが多い。

例えば、「主観的幸福感」の研究のために調査を企画したとする。このとき、先行研究から「主観的幸福感」という概念を操作化した質問項目や変数を照査し、目的にあった質問文（『全体的にいうと、現在、あなたは幸せだと思いますか』）と回答の選択肢（「非常に幸せ」、「やや幸せ」、「あまり幸せではない」、「全く幸せではない」）をワープロソフトなどで書き溜める。実査後、集められた回答を表計算ソフト等でデータ化し、分析ソフトが読み込める形にする。このとき、変数名（主観的幸福感）や値ラベル（1：非常に幸せ、2：やや幸せ、3：あまり幸せではない、4：全く幸せではない）をそれぞれのソフトに入力していく。報告書や分析結果の公表において、質問文や値のラベルが必要になる場合は、統計ソフトが出力した分析結果をワープロソフト等に貼付けて整形していくことになる。

このように、質問項目や回答選択肢など、同じような情報をそれぞれの段階で異なるツールに別々に投入するという作業は、非効率的である。さらに、調査後に報告書やデータを公開する際に、調査企画段階で参照した先行研究や先行調査の情報が不透明になってしまうことも懸念される。

3. DDI

DDI は DDI Alliance (<http://www.ddialliance.org/>) によって提唱されている、社会調査の国際規格である。Data Life Cycle にかかわる情報を、規格化された XML ファイルとして保存し、効率的・統一的にそれらを管理・利用する枠組みである。もともとは、ICPSR などの社会調査データ・アーカイブが、データ収集をする際の規格として整備されたものであるが、社会調査を実践するものにとって、日々の調査・研究活動を効率化する道具となる。

4. package DDIR

報告者は、データ分析環境 R と DDI を接続して、社会調査の統合環境を構築することを提案する。報告者が整備した R の package DDIR は、R で DDI の XML ファイルを読み／書き可能にするものである。DDI の規格に基づいて調査データの情報を蓄積しておけば、そこからデータ分析時に必要な変数名やラベル情報を自動的に抽出できるようになる。さらに、分析結果の出力生成まで視野に入れれば、調査報告書の大半を半自動的に行うことができる。

DDIR は、社会調査を「再現可能な研究」Reproducible Research とする道具となり、社会学における知の蓄積に貢献する。