



# Understanding selection bias, time-lags and measurement bias in secondary data sources: Putting the *Encyclopedia of Associations* database in broader context



Shaun Bevan<sup>a,\*</sup>, Frank R. Baumgartner<sup>b</sup>, Erik W. Johnson<sup>c</sup>, John D. McCarthy<sup>d</sup>

<sup>a</sup> Mannheim Centre for European Social Research (MZES), University of Mannheim, Germany

<sup>b</sup> Department of Political Science, University of North Carolina, Chapel Hill, United States

<sup>c</sup> Department of Sociology, Washington State University, United States

<sup>d</sup> Department of Sociology and Criminology, Pennsylvania State University, United States

## ARTICLE INFO

### Article history:

Received 1 January 2013

Revised 9 April 2013

Accepted 12 August 2013

Available online 22 August 2013

### Keywords:

Secondary data

Measurement

Associations

Interest groups

Social movements

## ABSTRACT

Secondary data gathered for purposes other than research play an important role in the social sciences. A recent data release has made an important source of publicly available data on associational interests, the *Encyclopedia of Associations* (EA), readily accessible to scholars ([www.policyagendas.org](http://www.policyagendas.org)). In this paper we introduce these new data and systematically investigate issues of lag between events and subsequent reporting in the EA, as these have important but under-appreciated effects on time-series statistical models. We further analyze the accuracy and coverage of the database in numerous ways. Our study serves as a guide to potential users of this database, but we also reflect upon a number of issues that should concern all researchers who use secondary data such as newspaper records, IRS reports and FBI Uniform Crime Reports.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

Our aims in this paper are twofold. The first is to describe a systematic data set we have assembled that will allow researchers to track, over an extended period, the national associations in the US that make up what political scientists call the interest-group system or the associational universe. The source of that data is the *Encyclopedia of Associations*, a publication that was not originally assembled for research, but which has extensive research value nonetheless. The second is to situate the potential biases that we enumerate and assess in the EA into a broader context of problems of bias that have been investigated in the use of other widely used data sets, including FBI Uniform Crime Reports, IRS 990 filings, and newspaper reports of protest. We call such data secondary, in that it is not initially collected by researchers for their own purposes, but distinguish it from the typical use of the term to describe survey data that is used by researchers who did not collect the survey data themselves. We begin by introducing the EA data set, then describe our assessment of its biases, and conclude with a discussion of the lessons we believe our efforts teach, comparing and contrasting our assessments of the EA with similar efforts to assess biases in newspaper, IRS and FBI crime data.

\* Corresponding author. Address: University of Mannheim, MZES, D-68131 Mannheim, Germany. Fax: +49 621 181 2845.

E-mail address: [shaun.bevan@gmail.com](mailto:shaun.bevan@gmail.com) (S. Bevan).

## 2. The encyclopedia of associations

Since 1956 Gale Research (which became part of Thomson Learning in 1984, and a part of Cengage Learning in 2007) has published an annual volume listing US voluntary associations active at the national level: the *Encyclopedia of Associations* (EA).<sup>1</sup> The EA is intended to act much like a trade directory that potential members and other interested parties can use to locate information about groups of interest. This directory is the publisher's best attempt in any given year to capture the total number of national-level voluntary associations operating in the United States. The annual nature and consistently meticulous search process used by the compilers of the EA make it an excellent resource for the creation of a time-series dataset on national-level voluntary associations.

For each organization listed, the EA provides contact information, an abstract describing its activities, information about its membership size, staff, budget, and details such as whether it publishes magazines, holds annual conventions, etc. The EA is organized into 16 sections including Trade and Business, Health, Agriculture, Environment, Public Affairs, Social Welfare, among others, and each group is associated with a keyword that further highlights the main focus of its activities. Hundreds of highly detailed keywords are used. The 16 sections used to organize the Sections of the EA have been used in a consistent manner since 1959, with only minor changes to the subject names over time.<sup>2</sup>

The EA's empirical value has been demonstrated previously by other research, generally<sup>3</sup> within specific areas of interest, e.g., the women's and civil rights movements (Minkoff, 1995, 1997), gay/lesbian movement (Nownes, 2004) labor unions (Martin et al., 2006), human rights (Walker et al., 2010) and the environment (Johnson et al., 2010; Johnson and Frickel, 2011). Many scholars are interested in tracking populations of associations within policy sectors or issue-domains (Aldrich and Ruef, 2006; Davis et al., 2005; Hannan and Freeman, 1989; McCarthy and Zald, 1977; Minkoff and McCarthy, 2009; Perrow, 1992; Walker, 1983, 1991), and they use a variety of methods to do so. Important sources of data include general organizational directories like the EA and *Yearbook of International Associations* (see e.g. Boli and Thomas, 1997; Smith and Wiest, 2012), as well as domain specific directories (e.g. Edwards, 1994), IRS 990 forms (Brulle et al., 2007), activist biographies (Rosenthal et al., 1985), newspaper reports (Soule and King, 2008), and combinations of secondary sources and qualitative methods to build systematic enumerations of organizations within a defined geographical area (e.g. Andrews and Edwards, 2005; Kempton et al., 2001).

Despite the variety of strategies employed to identify relevant populations of social change organizations, limitations of data access have largely restricted previous analyses to single topics or to a single population or community of organizations. Building comprehensive knowledge from case studies one issue-domain at a time is a limited route to generalizable theory. Whereas information on commercial establishments, trade unions, and registered political parties, for example, is relatively plentiful and readily accessible, the lack of available data for systematic research on the characteristics and dynamics of broad associational populations has been the subject of repeated lament in social movements and interest group literatures (Gray and Lowery, 2000; Knoke, 1986; Lowery, 2012; McCarthy and Castelli, 2002; Schlozman and Tierney, 1986; Walker 1991). By aggregating and making available the full data associated with the entire set of organizations listed in the EA (Gale Research, 1970–2005), we aim to spur more comprehensive assessments of the population dynamics of US voluntary associations.

In 2012 the Policy Agendas Project (PAP; see [www.policyagendas.org](http://www.policyagendas.org)) made available annual counts of associations present per year in each of the PAP topic categories. The PAP also makes available time-series databases relating to national government activities of all types from 1948 to present. By linking the EA database with those of the PAP, we allow scholars to study the linkages between growth and development of the voluntary sector with that of government. As government has become more active in education policy, have more education-related associations formed? Or, were associations leading indicators, preceding government activities in various fields? The PAP database also provides a cross-walk through its issue codes to a database composed of all reports of protest in the *New York Times* between 1960 and 1995, inclusive (<http://www.stanford.edu/group/collectiveaction/cgi-bin/drupal/>), allowing for the first time systematic analyses of public protest, the formation of voluntary associations, newspaper coverage, and government activity not just for a single issue-domain, but in all areas of public policy.

Our intent here is to assess the quality of the database and to bring front-and-center what we have learned about the apparent lag between the time when a group forms and the time when it is reported in these published volumes. The lessons we have learned throughout our efforts to create a high quality database from an existing published source, we believe, should have broad resonance for any scholars interested in using secondary data of any type.

### 2.1. Creating a database

We began the daunting project of compiling a time-series database covering 35 years and more than 20,000 records per year by looking at two sections of the database: Labor Unions, and Public Affairs. The Labor Union section had fewer than 300 records and focusing on it allowed us to understand the logistics and difficulties of the large task we were undertaking.

<sup>1</sup> Before 1974, volumes were published less regularly. Today, the printed version still appears each year, but many university libraries offer an electronically searchable database that is continuously updated: *Associations Unlimited*. The EA can be thought of as an annual snap shot of this database.

<sup>2</sup> The only major change in the 1970–2005 time period is the addition of a section on Fan Clubs in 1987.

<sup>3</sup> A bibliography including all social science journal articles using the source in some way generated from an exhaustive search of the JSTOR database is available (see Baumgartner and McCarthy, 2009).

Download English Version:

<https://daneshyari.com/en/article/956002>

Download Persian Version:

<https://daneshyari.com/article/956002>

[Daneshyari.com](https://daneshyari.com)