



CONSIRT

Cross-National Studies: Interdisciplinary Research and Training Program

Working Papers Series

CONSIRT Labs: Methodology of Survey Data Harmonization

A History of Cross-national Survey Data Harmonization Projects in the Social Sciences: Emergence of an Interdisciplinary Methodological Field

Joshua Kjerulf Dubrow, CONSIRT and the Institute of Philosophy and Sociology, Polish
Academy of Sciences

dubrow.2@osu.edu

Irina Tomescu-Dubrow, CONSIRT and the Institute of Philosophy and Sociology, Polish
Academy of Sciences

tomescu.1@osu.edu

October 2014

About CONSIRT Working Papers

Papers in the CONSIRT Working Papers Series are pre-publication versions of the author(s)' academic work. Working Papers posted on consirt.osu.edu are in progress, under submission, or forthcoming. Working Papers are produced on this site by the author(s) for the consumption of the international scientific community and others interested in this work. CONSIRT Working Papers are not subject to double-blind peer review. CONSIRT reviewed the Working Paper submission for general suitability, but the quality of the Working Paper itself is the responsibility of the author(s). The form and content of papers are the responsibility of individual authors. The cover sheet is standard across all papers (subject to change), but the format of the rest of the work is not standardized. Comments on papers or questions about their content should be sent directly to the author, at his or her email address.

Citation of CONSIRT Working Papers

Working Papers may be cited without seeking prior permission from the author. The proper form for citing CONSIRT Working Papers is:

Author(s). Year. "Title." *CONSIRT Working Papers Series* at consirt.osu.edu.

Once a CONSIRT Working Paper has been published elsewhere, it is customary that it be cited in its final, published version, rather than in its Working Paper version.

Copyright Statement

Copyright to CONSIRT Working Papers remains with the authors. Posting a Working Paper on consirt.osu.edu does not preclude simultaneous or subsequent publication elsewhere, including other Working Papers series. It is the current policy of CONSIRT to maintain all Working Papers posted on the CONSIRT website unless otherwise notified by the author(s). Users of this website may consume papers for their own personal use, but downloading of papers for any other activity, including but not limited to electronic reposting to the internet in any way, may only be done with the written consent of the authors. It is the author's responsibility to know if copyright has been transferred upon publication elsewhere and to ask CONSIRT to have it removed from the site, if necessary.

A History of Cross-national Survey Data Harmonization Projects in the Social Sciences: Emergence of an Interdisciplinary Methodological Field ¹

Joshua Kjerulf Dubrow, CONSIRT and the Institute of Philosophy and Sociology, Polish Academy of Sciences

Irina Tomescu-Dubrow, CONSIRT and the Institute of Philosophy and Sociology, Polish Academy of Sciences

Abstract

We examine the history of survey data harmonization (SDH) projects in the social sciences, focusing on *ex post* cross-national survey harmonization. The potential for interdisciplinarity -- the integration of two or more disciplines that produces emergent, original, interesting and useful knowledge -- is not often realized in practice. Cross-national survey data harmonization as a scientific/intellectual project is an example of when the potential for interdisciplinarity does actually lead to interdisciplinary knowledge, in this case the synthesis of knowledge from economists, sociologists, political scientists, survey specialists and data archivists. SDH's main characteristics are: (a) it is part of a new intellectual movement in science toward Big Data, (b) it is such an unusually complex method that it can be considered an emerging interdisciplinary field, or perhaps a "supra-method," and (c) despite of -- or perhaps because of -- the other two reasons, there is no coherent set of guidelines on how best to solve the severe methodological challenges SDH imposes.

Acknowledgements

This work is funded in part by the project "Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective", and is funded by the National Science Centre, Poland, under the grant number 2012/06/M/HS6/00322. We thank Kazimierz M. Slomczynski and Marta Kolczynska for their comments on an earlier draft, and Dean Lillard for advice on the citation of harmonized data.

¹ We are grateful for comments by Kazimierz M. Slomczynski and Marta Kolczynska.

The potential for interdisciplinarity -- the integration of two or more disciplines that produces emergent, original, interesting and useful knowledge -- comes to fruition when the relationship of the integrated disciplines is strong and symmetrical (Wagner et al. 2011; National Academies 2004; Dubrow 2011; Dubrow and Kolczynska 2014). The strength and symmetry of disciplinary relationships depends on the relative similarity in objects of study, methods of inquiry, the size of the profession and its disciplinary apparatus, and evolutionary history (Dubrow 2011; Dubrow and Kolczynska 2014). There are plenty of examples of when this potential does not lead to interdisciplinary knowledge, such as that of sociology and American Studies (Dubrow 2011; Jacobs 2014) or in political sociology (Dubrow and Kolczynska 2014). Cross-national survey data harmonization (SDH) as a scientific/intellectual project is an example of when the potential does lead to interdisciplinary knowledge, in this case the synthesis of knowledge from economists, sociologists, political scientists, survey specialists and data archivists.

We examine the history of survey data harmonization (SDH) projects in the social sciences, focusing on *ex post* cross-national survey harmonization. Some recent works have presented a history of data harmonization in the social sciences. Burkhauser and Lillard (2005) present a very useful 25 year overview of *ex ante* and *ex post* harmonization attempts, with a focus on its uses for the development of social policy². A more limited history -- because it was designed to present principles of SDH in general -- is by Granda et al. (2010).

SDH in the social sciences is worth a closer look because it is an unusual scientific project that challenges what we know about the evolution of such projects. SDH's main

² <http://www.econstor.eu/bitstream/10419/18337/1/dp486.pdf> Accessed February 12, 2014. See especially pp. 9 - 18.

characteristics are: (a) it is part of a new intellectual movement in science toward Big Data, (b) it is such an unusually complex method that it can be considered an emerging field, or perhaps a “supra-method,” and (c) despite of -- or perhaps because of -- the other two reasons, there is no coherent set of guidelines on how best to solve the severe methodological challenges SDH imposes.

SDH is a rare phenomenon, and we start by defining what it is not. It does not conform to the popular notion of a scientific/intellectual movement (SIM). SIM is defined as “collective efforts to pursue research programs or projects for thought in the face of resistance from others in the scientific or intellectual community” (Frickel and Gross 2006: 206). While SDH is a collective effort, it is not a SIM, or at least not yet, because (a) the collective process has not adequately scrutinized the methodology and (b) the project it has not faced any opposition. The scientific community willing to take on such a monumentally difficult task as SDH is small, but small as it is, SDH as a method has not emerged from a process of intense scrutiny. There is little debate about SDH because few have taken the trouble to develop a coherent set of methodological guidelines that can be debated. It is a method without opposition: There is no evidence whatsoever that there is a competing methodological paradigm to it, or any stated resistance to the existence of SDH as a way to understand human thoughts and behaviors. In fact, SDH projects, as part of the Big Data wave sweeping the sciences in general, are growing in popularity among social scientists and the institutions that fund them.

SDH is a popular form of data harmonization that both presaged, and is now a part of, the Big Data movement sweeping the social sciences. Big Data, According to Peter Elias of the UK’s Economic and Social Research Council, means:

“Electronic data generated from research infrastructures (e.g. astronomy, particle physics, microbiology, etc.) and Electronic data not designed for research but with potential research value which records transactions, communications, physical movements (e.g. customer databases, service delivery records, internet search activity, etc.)³.”

SDH, which contains an unusually large number of cases, variables or countries, culled from an impressively long list of surveys, is Big Data.

Second, SDH is an unusually complex process with a host of methodological challenges involved in competently harmonizing survey data projects that were not specifically designed to be harmonized. Methods in the social sciences develop because they are a problem-solving enterprise and they diffuse because other social scientists come to see their usefulness in solving their own problems, such as with focus group interviews (Lee 2010), the participant observation method (Platt 1983), or the invention of the scatterplot graph (2005). Development takes time, and in some cases decades of time, and not all methods survive the development stage (such as the life history approach of the 1930s, chronicled by Thomas 1978).

SDH is close to the concept of supradiscipline, in which the integration of disciplines produces emergent knowledge in such a way that the contributing disciplines are transgressed, or superseded (Balsiger 2004: 410). SDH may be a supra-methodological field, as it is an endeavor that combines methodologies in such a way as to transcend them. A supra-methodological field is a creature of necessity; the methodological problems involved are so specific, dire and acute that innovative methodological advances are needed to solve them.

3

http://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/events/conferences/peuk/peter_elias_big_data_and_the_social_sciences_pe_final.pdf. Accessed February 6, 2014.

Unlike scientific advances elsewhere, SDH projects in the social sciences have proceeded without a coordinated effort (for the example of coordinated biomedical advance, see Marcum (2008)). Our review of SDH projects since the 1980s reveals that while successive social science SDH projects have been aware of previous attempts, i.e. there is an observable learning process, they have done so without generating a set of general methodological guidelines. SDH in the social sciences occurs without an apparatus: no journals, no professional association, no academic department; SDH does not even have a handbook. It is only in the last decade that there is some attempt at a theory of SDH methodology in the social sciences (Hoffmeyer-Zlotnik and Wolf 2003; Granda et al 2010). Thus, a methodological field emerges in the social sciences, but without the coherent scientific advance one would expect of a 30 year effort.

We focus on the institutional factors that created SDH projects, and this necessarily includes some analysis of the intellectual processes that guided how each project chose its methodology. Our history is presented chronologically to show how cross-national, ex post SDH in the social sciences evolved, and how projects influenced one after the other. As with other histories of methodologies (e.g. Platt 1983, Friendly and Denis 2005), ours is based on archival research: in addition to journal articles, much of the archives we examined are available on the internet, and as such we use websites and reports of research projects and workshops reported there.

The authors of this article are involved with a current SDH project, the Harmonization Project based in the Polish Academy of Sciences. Inasmuch as history is designed to be learned from, in this article we attempt to draw lessons from the choices made by SDH projects for those who wish to initiate their own, large scale SDH project.

The Methodological Challenges of SDH

SDH dates back to the early 1980s, but it is safe to say that most social scientists are unfamiliar with the literature on it. As Matsumoto & Van de Vijver (2010: 6) put it, “Understanding why any study is to be conducted in the first place leads to questions about how to conduct it, which is a discussion in the realm of research methodology.” Thus, we begin with a brief outline of the basic research problem all cross-national survey research faces, and how SDH attempts to deal with it. Then, we go in-depth on the definition of SDH, and outline the methodological problems in actually conducting an SDH project.

Substantive knowledge of the relationship between any two major concepts (inequality and democracy, for example) covered in cross-national survey projects depends on a methodological issue: country variation. If we examine contemporary Europe with the European Social Survey, for example, we have too little variation in level of democracy, for example. To get more variation, we typically turn to multiple international survey projects. Some are regional focused, such as the Afrobarometer. Others are worldwide, but with too few countries, such as the International Social Survey Programme. If we turn to the World Values Survey, the one international survey project that includes all major regions of the world and has wide country variation, we incur two huge costs: (a) limits on both the measurement of the concepts and the countries that are a part of that particular international survey project, and (b) an inheritance of the methodological shortcomings inherent to any single international survey project. Cross-national SDH projects are created because researchers think there is a better way to deal with better country coverage: the harmonization of multiple survey projects. Harmonization has many benefits, and chief among them is the potential to increase cross-national variation and thus lead to new substantive insights into human thoughts and behaviors.

According to Granda and Blaszcyk (2010) and Granda, Wolf and Hadron (2010), harmonization is a generic term for procedures aimed at achieving or at least improving the comparability of surveys over time and/or surveys from different countries. Depending on whether scholars intend to (a) design a study to collect new, comparable, data, or (b) use existing data that were not a priori designed as comparative, the literature points to the strategies of input and ex-ante output harmonization for the first situation, and the strategy of ex-post output harmonization (or, simply ex-post harmonization) for the latter case (CHINTEX Synopsis p.1-2, Granda et al 2010). While ex ante harmonization means that surveys are fielded in different countries with the design intent of an easier harmonization after the data are collected, ex post means that the surveys were not designed specifically for harmonization. There is no common definition of ex post, but from the literature, we can generically say that it is a process (a) in which different survey datasets that were not specifically designed to be compared are pooled and adjusted (i.e. recoded, rescaled, or transformed) to create a new integrated dataset that could be analyzed as a typical single-source dataset; and (b) that is based on clear criteria that specifies which datasets are included into the new dataset and clear methods for how variables in the new dataset are created.

Thus, if one wants to identify all of the methodological challenges inherent in SDH, one should start with the overarching methodological challenge in data comparability, and then recognize that there are nefarious methodological problems replete with error in each step in the entire harmonization process, i.e. moving from source variables -- those of the data set prior to harmonization -- to target variables -- the variables created based on source variables. The challenge of cross-national SDH is to produce meaningful data that has accounted for all of the error produced in what Granda and Blaszcyk (2012) refer to as the data lifecycle, a lifecycle that

stretches from the initial data source (e.g. each country involved in the international survey research project, such as Poland from the European Social Survey) to the harmonization decisions undertaken by the SDH project (creation of the target variables), to data cleaning of the final master file (i.e. the harmonized data). Thus, not only do SDH projects inherit the errors of the initial data source, but they may create their own in the harmonization process. In short, SDH is an almost incredibly daunting endeavor.

In general, there are two types of SDH projects. We concentrate on large scale projects designed to produce data on a range of research topics rather than on overly specific research questions. We refer to these efforts as “Massively Coordinated SDH,” because the project typically involves multiple institutions -- including governments, and especially their money -- and large numbers of researchers and assistants. Typically, these projects produce the data itself and user manuals for the data, and some articles that show that these data can address substantive issues. The second type is those designed by small research teams to answer specific research questions -- and not to produce data on a range of research topics beyond the specific research question. In this second type, harmonization procedures are less transparent, and there is little attempt to disseminate the data. This type can be called, “Lone Wolf Harmonization,” but it is not the focus of this paper.

The Beginning of Massively Coordinated SDH: The 1980s

One of the earliest attempts at Massively Coordinated SDH, and perhaps its most successful, was the Luxembourg Income Study, now simply called LIS. The idea of LIS was generated by a conference on the topic of poverty in cross-national perspective, held in Luxembourg in 1982

(for a detailed history, see Smeeding et al. 1985: 2-4⁴). Smeeding et al. (1985: 2) describe how the idea of LIS took shape:

“Some participants in that conference were highly experienced in the microanalysis of income distribution data sets for their own countries. It became apparent during the conference deliberations that it would be possible to pool the knowledge and experience in these various countries to create internally and externally consistent data sets for comparative studies which are far superior to those currently in existence.”

The government of Luxembourg agreed, and from their sponsorship, LIS was born. From 1982 until 2006, Timothy M. Smeeding, a PhD in Economics from the University of Wisconsin-Madison, directed the project⁵. At the outset, LIS intended to include social indicators other than income, and to be useful in both basic and applied “social economic research” (with regard to some specific issues; see Smeeding et al. 1985: 2-3). At that point, LIS was in its “experimental stage,” and was envisioned to update and expand its data over time (Smeeding et al. 1985: 4). At the beginning, the budget was not enough to cover their wider ambitions: “There was only enough money to conduct a feasibility study on the seven datasets currently included” (Smeeding et al. 1985: 5). Two years later, LIS claims to have “moved beyond the initial experimental stage” (Rainwater and Smeeding 1987: 5)⁶.

It was 1983, and LIS attempted to codify the procedures designed “to produce the final harmonized LIS data set” (Smeeding et al. 1985: 4). For each dataset under consideration, LIS

⁴ <http://www.lisdatacenter.org/wps/liswps/1.pdf> Accessed February 7, 2014.

⁵ <http://www.lafollette.wisc.edu/facultystaff/smeeding/smeeding-timothy-cv.pdf> Accessed February 7, 2014.

⁶ <http://www.lisdatacenter.org/wps/liswps/12.pdf> Accessed February 7, 2014. Harmonization speeds forward via technological advance. Since its inception, LIS faced a major problem in supplying its data to interested researchers. By 1987, LIS began to solve the problem of data access through the then new system called BITNET, described as “an electronic mail and file transfer network” that linked about 400 academic and research institutions around the world (Rainwater and Smeeding 1987: 9). It was the early internet.

identified an expert who knew that data well and asked them about the strengths and weaknesses of that data “for the type of comparable social policy analyses” they hoped to do (Smeeding et al. 1985: 4). As Smeeding et al (1985: 4-5) describe it, LIS tried to obtain a representative sample of suitable datasets, they developed criteria for inclusion: timeliness⁷, data quality⁸, sample size, income accounting unit⁹, and geographical location. These data are collected by government agencies or in close collaboration with them and their national statistical agencies. Some governments were suspicious of the LIS project, and refused to share some of their datasets (Smeeding et al. 1985: 5). Datasets that were available but did not meet the criteria were excluded. The first version of the LIS database contained data from a handful of Western countries -- USA, Norway, Canada, UK, Germany, and Sweden -- and Israel.

Ex-post harmonization was the goal, and nothing was more troubling than income, their main variable of interest. “It should be stressed,” Smeeding et al. (1985: 11) write of the income harmonization, “that international comparability and consistency rather than perfection is our goal.” The point was to make a generally comparable dataset that also allows individual researchers to decide on what to do with certain anomalies.

Being aware of imperfections of existing datasets, the LIS team put a lot of effort in understanding ways in which “relative data quality” (p. 18), as they called it, could affect the

⁷ Data had to be available for 1979, the baseline for all datasets. They transgressed that rule by including 1981 data from Germany and Canada, which they justified on practical grounds (see Smeeding 1985: Footnote 3, p. F-1).

⁸ “...(as measured by response rates and other indicators of nonsampling error)” Smeeding et al 1985:4. They begin to discuss data quality in detail on p. 18.

⁹ Household, family or both. This problem required some harmonization of what “household” vs. “family” means.

comparability of income measures between countries. This included not only constructing a conceptual model of levels of income reporting, but also an analysis of the quality of administrative data -- that was used for comparison -- to make necessary adjustments for income definitions and populations.

At the outset of the project, with almost no methodological literature on data harmonization to draw upon, it was difficult to anticipate all the problems that seemingly simple variables would pose to harmonization. In the early years, LIS spent a lot of time thinking how to get at comparability, and considered many elements of harmonization. As the data developed over the decades, their harmonization procedures have apparently satisfied its many users. In the early 1980s, though, LIS did not look as deeply at the problems in harmonizing demographics as they had with income. Of demographics, LIS claimed then that “most of these variables are self-explanatory” (Smeeding et al 1985: 16). The Appendix in which they listed these demographics did not contain much more information. Education, for example, has the following comment: “If not coded in years, recoded as minimum years needed to attain given level of education” (Smeeding et al. 1985: 40). It is not clear why, at that time, they did not use the International Standard Classification of Education (ISCED), which was created by UNESCO in 1976¹⁰. The lesson to draw, here, is that in SDH, the best approach is to assume that all variables will pose major difficulties¹¹.

¹⁰ Classifications Newsletter, Number 27, August 2011, p. 5. ISCED was not updated again until 1997, and then again in 2011.

http://www.uis.unesco.org/Education/Documents/UNSD_newsletter_27e_ISCED.pdf Accessed February 7, 2014. Sometime afterward, LIS began to also use ISCED <http://www.lisdatacenter.org/wp-content/uploads/standardisation-of-education-levels.pdf> Accessed February 7, 2014.

¹¹ The LIS harmonization process is described here: <http://www.lisdatacenter.org/wp-content/uploads/our-lis-documentation-harmonisation-guidelines.pdf>. According to the LIS Guidelines (p. 12): “Socio-demographic variables... are all individual level variables and report the major socio-demographic

While LIS was getting off the ground, scholars interested in the concept of “time use” also started to consider how to compare all of the Time Use Studies (TUS) conducted in various countries, past and present¹². The resulting project, named the Multinational Time Use Study (MTUS), has its roots in the 1970s, but only took shape as a harmonized time use study in the 1980s (for a detailed history, see MTUS User’s Guide 2013: Chapter 2). Jonathan Gershuny, PhD in History and Social Studies of Science from the University of Sussex, and now Professor of Sociology at Oxford University, led MTUS from the beginning. The European Foundation for the Improvement of Living and Working Conditions (EFILWC), an agency of the European Union, paid for the initial release of MTUS; the collaboration between MTUS researchers and the EU led to the Harmonized European Time Use Study, or HETUS (see below).

MTUS is based on time use diaries, though they did collect sociodemographics, as well. As MTUS (MTUS User’s Guide 2013 Chapter One: 5) explains:

“The time use diary is a narrative account and not a series of quantitative questionnaire answers. People can give a full account of their day without necessarily completing all columns of the diary for all potential time slots...”

The early MTUS lacked some basic methodological rigors -- “in early versions, limited or no efforts were made to clean data,” and “some diaries originally included as good quality diaries now are classified as low quality diaries not suitable for more analysis” (MTUS User’s Guide 2013: 25) -- that MTUS now does thoroughly. They follow their own harmonization rules which

characteristics of the household members.” They are: Living arrangements, demographics, immigration, health, and education. LIS’ wealth data also has harmonization guidelines, but as of this writing, they are not available on the website. Instead, they have a document on “behavioral variable mapping” <http://www.lisdatacenter.org/wp-content/uploads/2011/02/behavioural-variable-mapping-2011-03.pdf>

¹² MTUS has time use data from the 1960s to the present.

they consider as appropriate to TUS. For example, MTUS refuses to impute data: “We also do not impute data based on estimations from what similar people do on similar days. All adjustments to diaries work from information that the participant supplies” (MTUS User’s Guide Chapter One: 6).

The Rise of SDH: The 1990s

There were three significant SDH projects of the 1990s. One was the Cross-national Equivalent File (CNEF) (<http://cnef.ehe.osu.edu/>). CNEF is simultaneously based on the successful LIS model¹³ and designed to overcome some of the problems of LIS. As Frick et al. (2007: 628) write,

“While the standardized LIS data are impressive, they cannot meet some goals of the cross-national research community. For example, the LIS allows researchers only indirect access to the underlying confidential microdata which in several cases is official data. Further, researchers cannot easily get access to the original data sources. This limitation means that most researchers must accept the LIS standardization rules. Finally, and perhaps most importantly, the LIS data are cross-sectional, and so do not serve researchers interested in longitudinal analyses.”

CNEF harmonizes household panel studies (such as the US Panel Study of Income Dynamics PSID and British Household Panel Study BHPS). It began in 1990 as a project funded by the U.S. National Institute on Aging and administered by Cornell University. The idea was to create a cross-national harmonized panel dataset dating back to 1980¹⁴. Initially, it included only Germany and the U.S., but by 2007 it expanded to six countries. In the early years, variables

¹³ Burkhauser and Lillard (2005: 10).

¹⁴ It has since been dated back to 1970 <http://www.human.cornell.edu/pam/research/centers-programs/german-panel/cnef.cfm> Accessed February 11, 2014.

were limited to income and demographics¹⁵. The CNEF project was initially directed by Richard Burkhauser, a PhD in Economics from the University of Chicago and Gert G. Wagner, a PhD in Economics from Berlin University of Technology.

Unlike LIS, CNEF was designed to be developed and enhanced by its user community. As Frick et al. (2007: 629) write, “Equivalently defined variables are added when researchers develop cross-nationally comparable measures as part of a particular research project. ... Consequently, the harmonized data included in the CNEF are an amalgam of the knowledge of many researchers answering a diverse set of questions. Just as importantly, the CNEF continuously evolves as researchers refine and add to the set of harmonized variables.”

CNEF can be called a bottom-up approach, with users having strong say in the direction of CNEF’s target variables, as opposed to LIS’ top-down approach. When it comes to top-down or bottom-up in SDH, there are no ideal types, as LIS uses its working papers to understand how users use the data (with the assumption that LIS makes adjustments to its dataset based on use patterns). According to Burkhauser and Lillard (2005: 12), successful SDH of panel studies have been driven and implemented by active researchers, not by government bureaucrats¹⁶, and they

¹⁵ Frick et al (2007: 630) write: “The original core variables to be harmonized were income and demographic characteristics of respondents to the PSID and the German Socio-Economic Panel - SOEP, and reflects the objectives of the original project that motivated the creation of the CNEF ± to compare and understand income-based inequality and income mobility in the US and Germany.”

¹⁶ “The effort to harmonize existing panel studies share one significant organizational feature: active researchers conceived, planned, and carried out how the data would be harmonized. While data managers, some in government statistical agencies, were often involved in the process, it was researchers who decided how to define equivalently the variables of interest.”

do so by identifying theories and research questions before they harmonize¹⁷. All of these efforts, however, are dependent on government funds.

Lessons from a Troubled Project: The European Community Household Panel (ECHP)

The European Community Household Panel (ECHP), a Massively Coordinated SDH project run by the European Union's Eurostat, provides lessons for present and future SDH projects. Though an ex-ante project, ECHP's unanticipated end had such far reaching and long lasting implications that it is worth discussion in this history¹⁸. ECHP lasted from the mid-1990s to 2001. As Burkhauser and Lillard (2005: 14) explain,

“Led by Eurostat, the ECHP attempted, by using a common survey instrument, to create a set of country based data sets that were comparable across countries. The ECHP goal was to create comparable panel data for all European Union (EU) countries.”

Burkhauser and Lillard clearly see ECHP as a failure, and describe the reasons they think it failed (2005: 14):

“The ECHP was plagued by problems from the outset. In part these problems may have arisen because the ECHP was developed by Eurostat and implemented by each country's statistical agency with little or no consultation with the research community. ... end users played a minor role in the creation and implementation of the survey instrument. Most troubling, the ECHP project failed to utilize the long experience of researchers who were running mature panel surveys in EU countries.”

¹⁷ Burkhauser and Lillard (2005: 12): “Researchers guided by theory and concepts flowing from the research pertinent to the object of their studies are best able to make the assumptions necessary to harmonize data across countries.”

¹⁸ For more discussion on the shortcomings of ECHP, see Burkhauser and Lillard (2005: 14-15); CHER 2003:6-7; and undated CHINTEX document, p. 5; for more information on ECHP, see the EuroPanel Users Network. <http://epunet.essex.ac.uk/echp.php.html> Accessed February 13, 2014.

The listed ECHP's main failures: "Long delays in processing, Problems with initial responses, Problems with attrition rates, Non-uniform implementation, Lack of input from the research community in design and response to users over time, Initial failure to take advantage of existing panels, Poor dissemination strategy to get the data to the international research community and High costs of use for individual researchers" (14-15).

CHER, an EU funded project (see below), also had many complaints about ECHP data, mostly about the lack of comparability with existing datasets. As CHER (2003: 7) described it:

"The lack of longitudinal data that are at the same time comparable, well documented and closely related to relevant macro and meso information and of user-friendly access, has truly regrettable consequences. The potential for a cross-national database to compare the situation in one country with those in other countries is not sufficiently used, and comparative analysis of European issues is still underdeveloped."

Though referred to as a failure and maligned by subsequent projects, Eurostat -- who was in charge of ECHP -- has a different spin on how ECHP ended. Eurostat (2012: 11) sees ECHP in somewhat heroic terms, retired only because the nature of its agreement had ended (in a gentlemanly manner) and because of the changes to the EU's political circumstances:

"The EU commitment to fighting social exclusion was confirmed with the formation of 'The European Community Household Panel (ECHP)' a pioneering data collection instrument. Launched on a gentleman's agreement basis in 1994 it expired in 2001. However the political scene has changed, notably with the introduction of open method of coordination in the fields of social inclusion and pensions reform. Other important changes included enlargement of the EU from 15 to 25 states (and demands for coverage of other neighbouring countries), and the publication by the United Nations expert group on household income statistics of a detailed report and recommendations. In recognition of these changes, the ECHP is being progressively replaced with data collection under the EU-SILC regulations..."

The end of ECHP caused difficulties for Eurostat and the users of ECHP data. The demise of ECHP was also the end of EPAG, the European Panel Analysis Group built to use ECHP data,

and the EuroPanel Users' Network (EPUNet), created by EPAG to disseminate information and coordinate analyses of ECHP data.

ECHP's end also caused an existential crisis for the EU's harmonization efforts, ex ante in particular. Is ex ante possible? Is it worth the expense?

To answer these questions, the EU created CHINTEX. CHINTEX is an improbable acronym that somehow stands for, "The Change from Input Harmonization to Ex-post Harmonization in Samples of ECHP -- Implications on Data Quality." CHINTEX began in 2000 and was funded for three and half years by the European Commission, and was coordinated by Germany's Federal Statistical Office¹⁹. It arose out of a problem posed by the end of ECHP, when a few countries pulled out of the ex-ante collaboration on which ECHP depended. In short, the point of CHINTEX was to wonder whether ex ante efforts were worth the time and expense. According to a document that summarizes CHINTEX (date unknown, p. 1):

"It is the overall objective of CHINTEX by means of this unique data situation to clarify if it is necessary to have centralised, standardised survey instruments to achieve harmonisation and comparability or if this objective can also be achieved by ex-post harmonisation, by which independent national sources are satisfactorily converted to common concepts, definitions, survey questions etc."

Secondarily, they sought to better understand the issue of data quality²⁰.

¹⁹

https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/FinalConference/Downloads/Hahlen.pdf;jsessionid=7E7753BEE731A012B3D321FA73D189F0.cae3?__blob=publicationFile
Accessed February 13, 2014.

²⁰ "Furthermore, the project investigates important hypotheses about the data quality of panel surveys (non-response, reporting errors and panel effects) which are of general interest for survey statisticians." (CHINEX data unknown, p. 1).

In their final conference in May 2003, CHINTEX attempted to explain what they learned from studying ECHP's methods and how it can be applied to its successor project, EU-SILC (and, presumably, other such projects)²¹. CHINTEX's main contribution was to offer a wide range of methodological guidelines for SDH, in topics such as total survey error, panel effects, weighting and imputation, in the form of working papers and powerpoint slides. CHINTEX cost 1.3 million Euros.²²

CHINTEX's post-mortem methodological analysis of ECHP is the rare trove of methodological literature on SDH. In analyzing the feasibility of ECHP years after it was created, CHINTEX contributed to what is becoming a theory of data harmonization²³. Separately, Granda et al. (2010) also contribute to this on-going effort.

Was ECHP a Failure?

The claim that ECHP was a failure is debatable. Those suggesting or outright declaring that it was a failure – CNEF, CHER, and CHINEX -- present a litany of complaints about its administration. Further evidence is that two countries pulled out of ECHP prematurely, which likely encouraged its end. Eurostat counters by arguing that ECHP would have ended, anyway,

²¹

<https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/FinalConference/Einfuehrung.html> Accessed February 14, 2014.

²² See CROS, a website maintained by the European Commission <http://www.cros-portal.eu/content/chintex>. "The CROS Portal is a content management system based on Drupal and stands for "Portal on **C**ollaboration in **R**esearch and Methodology for **O**fficial **S**tatistics". <http://www.cros-portal.eu/page/about-cros-portal> "The European Commission maintains this website to enhance public access to information about its initiatives and European Union policies in general." <http://www.cros-portal.eu/page/legal-notice>.

²³ "One objective of this chapter is to establish a simple framework of harmonisation which is useful to pinpoint the research issues of CHINTEX and to highlight the differences to other issues of harmonisation."

because the gentleman's agreement was due to expire at that time. Of course, since ECHP was allegedly the product of a gentleman's agreement, there can be no definitive paper-trail that would prove it was to end when it did. Perhaps it was, as Eurostat seems to suggest, a pilot study designed to create its successor project (EU-SILC), but I do not know of documentation from ECHP's beginning that suggests that it was just a big pilot study.

A counter argument – that ECHP was useful, though limited – can be based on how many academic products were produced out of it. Setting aside the idea of whether the money spent for ECHP was worth it, to establish this argument we would need comparable information on academic products of each major SDH project, e.g. the number of articles that are based on LIS, CNEF and the like. Such data does not exist, mainly and unhappily because there is no requirement of data users to actually cite the data that they are using. SDH projects encourage users to cite the data, but there is no enforcement. Still, we can make a rough sketch of popularity or, at least, name recognition, based on mentions of the SDH data in major research article databases.

We did this for ECHP. Using the ISI Citation database, under “topic”, there were 245 mentions in articles of the search term, “European Community Household Panel.” This number is similar when searching for in EBSCO and SCOPUS. It would be unreasonable to assume that these are all negative citations of the data, considering that 13 of these articles were published in 2013 alone, and all of them are empirical articles that faithfully use the dataset. To call ECHP a failure, then, is to ignore a very important aspect of any SDH project: That scholars use the dataset to produce scientific knowledge. Clearly, ECHP has been successful in this regard.

While ECHP was signaling the administrative difficulties in ex ante harmonization across national statistical agencies, another European project, HETUS, the Harmonized European Time Use Surveys, embraced it. HETUS, however, did not go all in on ex ante, but rather combined both ex ante and ex post harmonization. HETUS began in the 1990s when “the need for increased comparability became recognised” (Eurostat HETUS 2000: Preface). In 1996-7, Eurostat funded a series of pilot-level time use studies in 18 countries designed to test the feasibility of harmonization. In 1998, the Statistical Programme Committee (SPC) recommended that “harmonisation of time use data was feasible despite recognised national differences” (Eurostat HETUS 2000: Preface). In 2000, around the time that the ECHP project was drawing to a close, Eurostat drew up harmonization guidelines for the Time Use Studies (TUS). Of TUS harmonization, Eurostat lamented its cost: “The chosen survey design is rather expensive, and in some cases it might be somewhat more expensive than a non-harmonised national design would have been” (Eurostat HETUS 2000: 8). To date, HETUS has 15 “comparable countries.”²⁴

SDH Booms: The Early 2000s

The early 2000s saw the maturation of LIS, CNEF and HETUS, the end of ECHP, and the creation of new SDH projects. An early project was the Consortium of Household Panels for European Socio-economic Research (CHER). CHER was initially funded by the European Commission for over 1 million Euros between 2000 and 2003, and coordinated by CEPS, a

²⁴ <https://www.h2.scb.se/tus/tus/introduction1.html> Accessed February 14, 2014.

research bureau in Luxembourg²⁵. CHER is substantively similar in its harmonization aims as CNEF, namely the harmonization of panel data that has already been collected.

Unlike the beginning of CNEF, the aims of CHER at the outset were much more ambitious. In 2003, CHER reported to the European Commission that its aims were to collect and harmonize panel data on “demography, health, education and training, employment and activity, income and expenditure, housing and household durables, subjective information and social relations” (CHER 2003: 6). CHER grew out of frustration with the ECHP data, and was built to overcome its shortcomings. By 2003, CHER expanded into many more countries than CNEF, and had data dating back to the 1980s. CHER had eclipsed CNEF in terms of country and topic coverage, but the project ended in 2003, and was not updated²⁶.

Although CHER was designed to overcome the problems of ECHP, EU-SILC has been named (by Eurostat) as its true successor²⁷. Like CNEF and CHER, EU-SILC is ex-post

²⁵ http://ec.europa.eu/research/social-sciences/projects/010_en.html Accessed February 13, 2014. According to CEPS’ website, “CEPS/INSTEAD is a centre of reference for research in the social and economic sciences in the Grand Duchy of Luxembourg, a public institution under the jurisdiction of the Ministry of Higher Education and Research.” <http://www.ceps.lu/?type=module&id=53> Accessed February 13, 2014.

²⁶ There is no official information about the fate of CHER. From existing documents, it seems to have ended in 2003 when they submitted their final report to the European Commission. To get more information, and since CHER was a CEPS project, I emailed a colleague to ask when CHER ended. They wrote, “The project has been on for a while as an experiment to test whether various national level panel studies could be homogenized and merged in one big data-set. I don't know exactly the date when the CHER was closed probably it was 2003, but the data it contains should be for sure older, probably 1999. As a matter of fact, it was a nice data-set but I've never used it because it was too old. Not by chance only 16 papers have been produced with CHER.” Personal Communication, February 13, 2014.

²⁷ As of April 2006, EPUnet “is now finished,” according to its website. “Jean-Marc Museux from EUROSTAT has produced three papers on EU-SILC, the successor to the ECHP” http://epunet.essex.ac.uk/view_news.php%3FID=36.html. Accessed February 13, 2014.

harmonized panel data. It was formally created in 2004 and run by Eurostat²⁸. It has data dating back to the start of its data collection process, in 2005. Unlike any other harmonization project, however, EU-SILC “is organised under a framework regulation and is thus compulsory for all EU Member States” (Eurostat 2012: 13), and as such is expressly designed to inform European Union economic and social policy. Thus, it has data for all EU countries, plus others in or near the European continent.

The only ex-post, cross-national SDH project not to use panel data, and the only major project from 1980s to the 2000s not to be run by economists or statisticians, was the International Stratification and Mobility File [ISMF]²⁹. The goal of ISMF is to be able to compare social stratification and social mobility patterns across countries and time. By 2009, the data expanded to over 250 surveys from 52 nations, with some surveys dating back to the 1940s. Its focus is on educational and occupational status of both respondents and their parents, and has harmonized demographics, education, employment status, occupation and income.

The ISMF project is a significant advance in SDH for many reasons. First, it is the rare attempt to create a harmonized datafile of cross-national, cross-sectional surveys stretching far back in time. Second, ISMF is run by Ganzeboom and Treiman, two scholars who have been at

²⁸ “The EU-SILC project was launched in 2003 on the basis of a “gentlemen’s agreement” in six Member States (Belgium, Denmark, Greece, Ireland, Luxembourg and Austria) and Norway. The start of the EU-SILC instrument was in 2004 for the EU-15 (except Germany, the Netherlands, the United Kingdom) and Estonia, Norway and Iceland.” http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc Accessed February 13, 2014.

²⁹ Most SDH is about economics and was run by economists. Only the TUS, ISMF and now, the Harmonization Project, are not explicitly about economics; they are run by sociologists.

the forefront of measurement standardization of stratification variables -- occupation, and now education³⁰ -- for decades.

The ISMF does not have a public official history and it is not clear when it began. Most current information on ISMF is from the personal website of its co-creator, Harry Ganzeboom, PhD in Social Sciences from the University of Utrecht (Netherlands)³¹.

The ISMF seems to have grown out of the needs of the social stratification and mobility research community, particularly those from the International Sociological Association's Research Committee (ISA RC) 28, "Social Stratification," for comparable cross-national and cross-time data. In 1990, Treiman and Ganzeboom complained about the poor attention given to comparability issues between empirical studies of status attainment based on surveys conducted in different nations (pp. 109, 116-117). In a sharply worded rebuke, Treiman and Ganzeboom (1990) wrote of the consequence of poor comparability: "Results obtained by analysts who have not troubled to standardize their samples, their measurement instruments, and their models must be dismissed as simply meaningless" (116). In 1990, no such data existed, and their solution was to standardize analyses with common variables and methodological approaches. "Because no one analyst will command the data or the resources to carry out the definitive cross-national comparison of status-attainment processes," they wrote, "our only alternative ... is to generate

³⁰ Schröder, Heike & Harry B.G. Ganzeboom (2013), "[Measuring and Modeling Levels of Education in European Societies](#)". European Sociological Review.

³¹ We have requested from ISMF information on the organizational and intellectual history of ISMF, but as of this writing have not received a response. The data set is created and maintained by Harry B.G. Ganzeboom at VU University Amsterdam, in collaboration with Donald J. Treiman and Elizabeth Stephenson, University of California-Los Angeles. A note to avoid confusion: Ganzeboom is also the director of SILC, an acronym for Social Inequality and the Life Course, with no relation to EU-SILC whatsoever.

the necessary information collectively” (116). Sometime between 1990 and 1993, they created the ISMF, but the only evidence I’ve found is that, in 1993, Ganzeboom and Treiman published a paper in which they reference the ISMF (Ganzeboom and Treiman 1993: 470)³².

From publicly available information, it is not clear whether ISMF emerged as an ad hoc way of addressing the data comparability problem via SDH or was, at the outset, a Massively Coordinated SDH effort. It is also not clear what harmonization procedures they employed, or the extent to which they have drawn on the growing methodological research on SDH. They do not provide any document that discusses their theory of SDH. In fact, most of what they say is this³³: “...we do provide researchers with some important tools for harmonization efforts, in particular maps to recode occupation and education categories into internationally and historically comparable metrics.” After, they provide a long list of surveys with links to syntax for recoding in statistical software packages.

It does not seem as if the ISMF has been publicized in the same degree as LIS and all such SDH studies. JSTOR lists only 11 articles containing the phrase, “International Stratification and Mobility File”. ISMF may not even be well publicized in RC28. In 2006, Hout and diPrete offer a history of accomplishments of the International Sociological Association’s RC28 committee, in which they only briefly mention that the ISMF confirms a finding on the topic of welfare states, labor markets and occupational mobility (12). Hout and

³² In a book chapter in the International Handbook of Sociology, Treiman and Ganzeboom (2000) describe the history of social stratification and mobility research. While they present a clear trend toward cross-national surveys and attempts at standardizing measurement of stratification variables (see also Ganzeboom and Treiman 2003), SDH of the type attempted by the Harmonia project is not mentioned. Note that these articles are listed under the rubric, “Harmonization Tools and Review Articles” in selected readings of ISMF.

³³ <http://www.harryganzeboom.nl/ISMF/index.htm> Accessed February 14, 2014.

DiPrete cite a paper by Ganzeboom and Treiman from 2000³⁴. The data seems a treasure, but it is not clear why this treasure is not more widely recognized.

SDH Expands: The 2010s

The 2010s has thus far seen the continuation of CNEF, EU-SILC, and ISMF, as well as MTUS and HETUS, and, now, the Harmonization Project, which began in 2013. The project is led by Kazimierz M. Slomczynski, an emeritus Professor of Sociology and Political Science at The Ohio State University and the Polish Academy of Sciences. At its core, the Harmonization Project seeks to harmonize variables relevant to basic models explaining democratic values and political protest. Thus, it includes demographics and social structural variables, along with the political ones. Equally important to the project are the contextual variables that influence individual political attitudes and behavior. Thus, part of the proposed harmonized datafile is contextual data for each country. This is the only current harmonization project to focus on these topics.

The Harmonization Project argues that insufficient harmonization of extant survey data has led to a fragmentation of knowledge concerning democratic values and political protest. There are many countries, and cross-national surveys – even the World Values Survey – cover some, but not enough. Since information on democratic values and political protest are contained in the various survey datasets, combining (harmonizing) all of the major ones will yield a single dataset that covers most of the world. All SDH projects are ambitious, by

³⁴ The cite is: Ganzeboom, H., & Treiman, D. J. (2000).Ascription and achievement after career entry. Paper presented at the 2000 meeting of the RC28.

definition. The Harmonization Project's ambition is in the size of its hoped-for datafile: "The [Harmonization] Program aims at providing researchers with a new empirical base: harmonized data on democratic values and protest behavior for from 1344 surveys, all together including around 1,5 million respondents from over 100 countries" (Harmonization Project proposal 2013: 5). Currently, it is much bigger: it uses 22 international survey projects to create an integrated data file and relational database of 1,720 country-survey-waves that cover 140 countries or territories and about 2.3 million respondents from 1966 to 2014.

Some Lessons of SDH History

This is a history of survey data harmonization projects in the social sciences (SDH) and how they developed from the 1980s to now. In this history we traced an emergent methodological field on cross-national survey data harmonization in the social sciences. Since the 1980s, there has been substantial thought, money and time spent on SDH in the social sciences. In this section we try to draw some lessons of this history.

Why Does SDH Exist?

It seems that the idea of SDH developed as the technological, statistical and other knowledge developed to a point where (a) there came to be a wealth of cross-national survey projects and (b) people think it would be useful if these data could be compared but wondered whether these disparate datasets are, in reality, comparable. Its recent popularity is likely owed to the fact that (c) we finally have the computing technologies to start making such an idea a reality.

Yet, just because we can, does not necessarily mean that it is useful for the social sciences. On the benefits to science in general, Doiron et al. (2013: 1) stated it quite well:

“The benefits of harmonizing and pooling research databases are numerous. Integrating harmonized data from different populations allows achieving sample sizes that could not be obtained with individual studies, improves the generalizability of results, helps ensure the validity of comparative research, encourages more efficient secondary usage of existing data, and provides opportunities for collaborative and multi-centre research.”

These benefits seem to reasonably apply to SDH in the social sciences. SDH is something akin to interdisciplinarity, the combining of different things that were not designed to be combined in order to produce emergent, original and interesting knowledge. The institutional rationale of SDH is, supposedly, the history of the quest for this knowledge. By its nature, SDH is ambitious. Millions and millions and millions of Euros have been spent on it. But there is too

little research that answers the questions: To what extent does SDH provide superior knowledge? Is harmonized data better than non-harmonized? Is ex ante better than ex post?

One thing is clear: the attempt at SDH has produced innovations in cross-national methodology, and has great potential to generate more innovations.

Where Is the Methodological Literature on SDH in the Social Sciences?

In compiling this history, it is clear that there is now a very large methodological literature on SDH in the social sciences but there is no coherence across projects. The problem is that this literature exists in various documents scattered over time, but little of it has been synthesized into a manageable and accessible format. In short, there is a need to pull this literature together to create a handbook of SDH in the social sciences based on the many efforts that have been done. This handbook should start with a theory of SDH in which we can apply general principles, and should continue with chapters on specific aspects of SDH – imputation of data, weighting, the relevance of question wording and ordering, cross-checks with other data sources, and so on – that synthesizes what we know.

There is also a clear need for definitions of SDH and its processes, but such a glossary is only now in the making. Some have written guidelines for data harmonization³⁵. An attempt at a glossary of survey data harmonization is available by Granda and Blasczyk (2010)³⁶. The chapter by Granda et al. (2010) is also relevant. The term “clear” is in development, and existing

³⁵ <http://ccsg.isr.umich.edu/harmonization.cfm> Accessed February 7, 2014.

³⁶ <http://ccsg.isr.umich.edu/pdf/13DataHarmonizationNov2010.pdf>. Accessed February 7, 2014.

SDH projects are not so transparent in explaining how they did what they did³⁷. While there is a call -- and an effort at heeding this call -- for better documentation (Granda et al. 2010: 326), there remain few comprehensive guidelines of the SDH process, especially ones that include details on each step, including the ideas that were abandoned. Granda et al. (2010: 326-328) call for the development of software that standardizes the documentation process. We seem to have a long way to go in this regard.

³⁷ This is a problem with social science in general: Social researchers rarely keep good records on the research process and are reluctant -- for whatever reasons -- to share enough the ups-and-downs of their scientific pursuit.

Works Cited

- Behavioral variable mapping. Undated. <http://www.lisdatacenter.org/wp-content/uploads/2011/02/behavioural-variable-mapping-2011-03.pdf> Accessed February 7, 2014.
- Burkhauser, Richard V. and Lillard, Dean R. 2005. The Contribution and Potential of Data Harmonization for Cross-National Comparative Research, DIW Diskussionspapiere, No. 486. <http://www.econstor.eu/bitstream/10419/18337/1/dp486.pdf> Accessed February 12, 2014
- CEPS/INSTEAD: <http://www.ceps.lu/?type=module&id=53> Accessed February 13, 2014.
- CHINTEX 2003. Final Conference: Harmonisation of Surveys and Data Quality. <https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/FinalConference/Einfuehrung.html> Accessed February 14, 2014.
- Classifications Newsletter. 2011. United Nations Statistics Division (UNSD) Number 27 (August). http://www.uis.unesco.org/Education/Documents/UNSD_newsletter_27e_ISCED.pdf Accessed February 7, 2014.
- Cornell University User Package for the Cross-National Equivalent File (CNEF), 1970-2009. <http://www.human.cornell.edu/pam/research/centers-programs/german-panel/cnef.cfm> Accessed February 11, 2014
- Cross-National Equivalent File (CNEF): <http://cnef.ehe.osu.edu/> Accessed February 22, 2014.
- ČSDA. 2010. Program of Workshop on Harmonisation of Social Survey Data for Cross-National Comparison will be held in Prague on Tuesday 19th October <http://archiv.soc.cas.cz/articles/cz/84/workshop.html> Accessed February 22, 2014.
- Doiron, Dany, Parminder Raina², Vincent Ferretti³, François L'Heureux¹ and Isabel Fortier. 2012 Facilitating collaborative research - Implementing a platform supporting data harmonization and pooling *Norsk Epidemiologi* 2012; 21 (2): 221-224.
- Dubrow, Joshua Kjerulf and Marta Kolczynska. "Who Owns the Study of Democracy? Sociology, Political Science and the Interdisciplinary Promise of Political Sociology since 1945." Forthcoming in *Sociologias* [Brazil, in Portuguese].
- Dubrow, Joshua Kjerulf. 2011. "Sociology and American Studies: A Case Study in the Limits of Interdisciplinarity." *The American Sociologist* 42(4): 303-315.
- Elias, Peter. 'Big Data' and the social sciences – a perspective from the ESRC. http://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/events/conferences/peuk/peter_elias_big_data_and_the_social_sciences_pe_final.pdf. Accessed February 6, 2014.
- EPUNET: http://epunet.essex.ac.uk/view_news.php%3FID=36.html. Accessed February 13, 2014.
- EuroPanel Users Network. <http://epunet.essex.ac.uk/echp.php.html> Accessed February 13, 2014

European Commission website. Consortium of Household Panels for European Socio-Economic Research (CHER) http://ec.europa.eu/research/social-sciences/projects/010_en.html Accessed February 13, 2014.

European Community Household Panel (ECHP):
<http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/echp>

European Union Statistics on Income and Living Conditions (EU-SILC):
http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc Accessed February 13, 2014.

EUROSTAT. 2009. Harmonised European time use surveys. Methodologies and working papers.
http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-08-014 Accessed February 22, 2014.

Fisher, Kimberly and Jonathan Gershuny. 2013. Multinational Time Use Study. User's Guide And Documentation. Version 6 (16 July).

Frick, Joachim R., Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps, and Mark Wooden. 2007. The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies In: Schmollers Jahrbuch (Journal of Applied Social Science Studies) 127(4), pp. 627-654.
http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.77260.de/schmoller_frick_etal_2007.pdf Accessed February 22, 2014.

Friendly, Michael and Daniel Denis. 2005. "The Early Origins and Development of the Scatterplot." *Journal of the History of the Behavioral Sciences* 41(2): 103-130.

Ganzeboom, Harry B. G. and Donald J. Treiman. 2000. Ascription and achievement after career entry. Paper presented at the 2000 meeting of the RC28.

Ganzeboom, Harry B.G.; Treiman, Donald J. (2003). "Three Internationally Standardised Measures for Comparative Research on Occupational Status." Pp. 159-193 in Jürgen H.P. Hoffmeyer-Zlotnik & Christof Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic Press.

Gartner, definition of Big Data: <http://www.gartner.com/it-glossary/big-data/> Accessed February 22, 2014.

GESIS. 2009. An illustrated user guide to the CharmCats database for classifications and conversions.
http://www.cessda.org/project/doc/D12.2c_An_illustrated_short_Usermanual_CharmCats_0.5.pdf Accessed February 22, 2014.

Granda, Peter and Emily Blasczyk. 2010. Data Harmonization. In *Cross-Cultural Survey Guidelines*.<http://ccsg.isr.umich.edu/pdf/13DataHarmonizationNov2010.pdf> Accessed February 7, 2014.

Granda, Peter, Christof Wolf and Reto Hadorn. 2010. "Harmonizing Survey Data," pp. 315-334 in *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith. New York: Wiley.

Harmonised European time use surveys (HETUS): Introduction:

<https://www.h2.scb.se/tus/tus/introduction1.html> Accessed February 14, 2014.

Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis. <http://www.ncbi.nlm.nih.gov/books/NBK132553/pdf/TOC.pdf> Accessed February 22, 2014.

Hoffmeyer-Zlotnik, Jurgen H. P. and Christof Wolf. 2003. "Comparing Demographic and Socio-Economic Variables across Nations: Synthesis and Recommendations," pp. 389-406 in *Advances in Cross-national Comparison: A European Working Book for Demographic and Socio-Economic Variables* edited by Jurgen H. P. Hoffmeyer-Zlotnik and Christof Wolf. New York: Springer Science+Business Media.

Hout, Michael and Thomas A. DiPrete. 2006. "What we have learned: RC28's contributions to knowledge about social stratification." *Research in Social Stratification and Mobility* 24 (2006): 1–20.

International Stratification and Mobility File (ISMF):

<http://www.harryganzeboom.nl/ISMF/index.htm> Accessed February 22, 2014.

Jacobs, Jerry A. 2014. *In Defense of Disciplines: Interdisciplinarity and Specialization in the Research University*. Chicago: University of Chicago Press.

Jacobs, Jerry A. and Scott Frickel. 2009. "Interdisciplinarity: A Critical Assessment." *Annual Review of Sociology* 35: 43-65.

Lee, Raymond M. 2010. "The Secret Life of Focus Groups: Robert Merton and the Diffusion of a Research Method." *The American Sociologist* 41:115-141.

Marcum, James A. 2008. "Instituting Science: Discovery or Construction of Scientific Knowledge?" *International Studies in the Philosophy of Science* 22(2): 185-210.

National Academies. 2004. *Facilitating Interdisciplinary Research*. Washington D.C.: National Academies Press.

Olenski, Józef. (2003). SSDIS. Global Standard for Harmonization of Social Statistics with special reference to transition and globalization processes. United Nations Statistics Division. ESA/STAT/AC.88/10 (April 7, 2003).

http://unstats.un.org/unsd/demographic/meetings/egm/Socialstat_0503/docs/no_10.pdf

Accessed February 22, 2014.

Platt, Jennifer. 1983. "The Development of the 'Participant Observation' Method in Sociology: Origin Myth and History." *Journal of the History of the Behavioral Sciences* 19: 379-93.

Portal on Collaboration in Research and Methodology for Official Statistics (CROS):

<http://www.cros-portal.eu/page/legal-notice> Accessed February 14, 2014.

Quandt, Markus. 2010. Data Harmonisation as a Cumulative Effort A Platform Designed to Foster the Cumulation of Knowledge. Paper presented at the Workshop on Harmonisation of Social Survey Data for Cross-National Comparison. Prague, 19th October.

http://archiv.soc.cas.cz/download/860/06_Quandt.pdf Accessed February 22, 2014.

- Rainwater, Lee and Timothy Smeeding. 1987. The Luxembourg Income Study: The Use of Telecommunications in the Social Sciences. Luxembourg Income Study Working Paper Series. Working Paper No. 12 (May). <http://www.lisdatacenter.org/wps/liswps/12.pdf> Accessed February 7, 2014.
- Schröder, Heike & Harry B.G. Ganzeboom (2013), “[Measuring and Modeling Levels of Education in European Societies](#)”. European Sociological Review
- Smeeding, Timothy and Günther Schmaus and Serge Allegrezza. 1985. An Introduction to LIS. Luxembourg Income Study Working Paper Series. Working Paper No. 1 (June). <http://www.lisdatacenter.org/wps/liswps/1.pdf> Accessed February 7, 2014.
- Smeeting, Timothy M. CV. 2013. <http://www.lafollette.wisc.edu/facultystaff/smeeding/smeeding-timothy-cv.pdf> Accessed February 7, 2014.
- The Lis Database: Guidelines. Undated. <http://www.lisdatacenter.org/wp-content/uploads/our-lis-documentation-harmonisation-guidelines.pdf> Accessed February 7, 2014.
- The standardisation routines for highest level of completed education. Undated. <http://www.lisdatacenter.org/wp-content/uploads/standardisation-of-education-levels.pdf> Accessed February 7, 2014.
- Thomas, Evan A. 1978. “Herbert Blumer’s Critique of The Polish Peasant: A Post Mortem on the Life History Approach in Sociology.” *Journal of the History of the Behavioral Sciences* 14:124-131.
- Treiman, Donald J. and Harry B. G. Ganzeboom. 2000. “The Fourth Generation of Comparative Stratification Research.” Pp. 122-150 in *The International Handbook of Sociology*, edited by Stella R. Quah and Arnaud Sales. London: Sage.
- UK’s Economic and Social Research Council: <http://www.esrc.ac.uk/> Accessed February 22, 2014.
- Wagner, Caroline S., J. David Roessner, Kamau Bobb, Julie Thompson Klein, Kevin W. Boyack, Joann Keyton, Ismael Rafols, Katy Börner. 2011. “Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature.” *Journal of Infometrics* 165: 14-26.
- Workshop “Harmonization Strategies for Behavioral, Social Science, and Genetic Research” organized by the US Department of Health and Human Services, National Institutes of Health, National Institute on Aging and Division of Behavioral and Social Research, Bethesda (MD), November 29–30, 2011. http://www.nia.nih.gov/sites/default/files/nia_bssg_harmonization_summary_version_2-5-20122.pdf Accessed February 22, 2014.