

# Making the Humanities Count

The Importance of Data

*Robert M. Solow*

*Francis Oakley*

*Phyllis Franklin*

*John D'Arms*

*Calvin C. Jones*



Making the  
Humanities  
Count: *The  
Importance  
of Data*

American Academy of Arts & Sciences

*Robert M. Solow*     *Francis Oakley*

*Phyllis Franklin*     *John D'Arms*

*Calvin C. Jones*

©2002 by the American Academy of Arts and Sciences  
All rights reserved

ISBN#: 0-87724-031-0

The views expressed in this volume are those held by each contributor and are not necessarily those of the Officers and Fellows of the American Academy of Arts and Sciences.

Please direct inquiries to:  
American Academy of Arts and Sciences  
136 Irving Street  
Cambridge, MA 02138-1996  
Telephone: (617) 576-5000  
Fax: (617) 576-5050  
E-mail: [aaas@amacad.org](mailto:aaas@amacad.org)  
Visit our Website at [www.amacad.org](http://www.amacad.org)

No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage retrieval system, without permission in writing from the publisher. The Association of American University Presses' Resolution on Permissions constitutes the only exception to this prohibition.

Cover and book design: Corey McPherson Nash

# Contents

v FOREWORD

ix ACKNOWLEDGMENTS

1 CHAPTER 1

The Value of *Humanities Indicators*  
*Robert M. Solow*

5 CHAPTER 2

Data Deprivation and *Humanities Indicators*  
*Francis Oakley*

17 CHAPTER 3

The Continuing Need for a Practical System of National  
Information and Data Collection on the Humanities  
*John D'Arms*

21 CHAPTER 4

*Humanities Indicators*: A Promising Possibility  
*Phyllis Franklin*

25 A REPORT TO THE AMERICAN ACADEMY

Evaluation of Existing Datasets for Policy Research on  
Humanities Fields  
*Calvin C. Jones*

101 CONTRIBUTORS

# Foreword

The American Academy of Arts and Sciences has long been committed to serving the humanities. During the past century it played a pivotal role in creating several important humanities institutions in this country: the American Council of Learned Societies, the Independent Research Libraries Association, and the Council of American Overseas Research Centers. In the 1960s, the Academy joined with other learned societies in the campaign to establish the National Endowment for the Humanities, and in the 1970s it led the effort to create the National Humanities Center, now one of the nation's principal centers for advanced research.

In recent decades the Academy has sponsored or conducted a number of large-scale research projects in the humanities, including an ambitious international comparative study of religious fundamentalism, *The Fundamentalism Project*, directed by Martin Marty and Scott Appleby, which resulted in a six-volume series published by the University of Chicago Press. Other publications resulting from Academy projects include *The Pursuit of Knowledge in the Early American Republic* (1976) and *The Organization of Knowledge in Modern America, 1860–1920* (1979). The Academy has also devoted special issues of its journal, *Daedalus*, to considering the status of liberal education, the changes affecting specific disciplines within the humanities, and the impact of technology on libraries.

Our current Initiative for the Humanities and Culture began in 1998, when the Academy co-hosted a meeting with the President's Committee on the Arts and the Humanities at which representatives from the major national humanities organizations expressed growing concern over the lack of public understanding and support of the humanities. As a consequence of what we learned at that meeting and frequently confirmed in the weeks and months that followed, the Academy undertook the Initiative for the Humanities and Culture.

The Initiative seeks to improve knowledge of the state of the humanities and to articulate their continued importance to national life. It consists of two distinct but related efforts. One will focus on building an infrastructure to improve the collection of data about the humanities and provide an empirical base for decision making by educators and policy makers. The other will organize a series of research projects about the evolution of the humanities disciplines, the institutional settings that support the humanities, and the influence of American cultural diversity on the development of the

humanities. This Occasional Paper explains why the first component of the Initiative—a collaborative effort to improve data collection—must ground any long-range plan to strengthen the humanities in this country.

When we consulted Academy Fellows and leaders from a range of humanities organizations, they repeatedly cited the need for a coordinated, systematic, and sustained effort to collect data about the humanities. As Academy Fellow Robert Solow writes in his essay for this volume, “The humanities community knows deplorably little about what is taught to whom and by whom, how long it takes, where graduates and post-graduates go, what they do when they get there, and how many of them there are.”

The sciences have long benefited from the availability of comprehensive, up-to-date data that describe the condition of their disciplines—as provided, for example, by the *Science and Engineering Indicators*, published every other year by the National Science Foundation. Because no comparable database exists for the humanities, there are great gaps in our knowledge of basic trends in those disciplines. In order to identify areas of expertise that lack sufficient specialists, or understand the prospects for future job growth in the humanities, or gauge the effectiveness of current academic programs, we must have accurate data about the state of the professions and the characteristics of those who teach in our classrooms or work in our libraries, historical societies, and museums. Without such data, the leaders of academic institutions and government agencies, corporate and private funders, and public humanities organizations are compromised in their efforts to recommend sound policies for the humanities and to make a cogent case for funding them.

Soon after the Academy decided to launch the Initiative, it appointed a Task Force on Data Development, chaired by Jonathan Cole, Steven Marcus, and Francis Oakley, and commissioned Calvin Jones, a statistical research consultant, to prepare a report that would evaluate existing humanities data and assess the data’s usefulness for answering the kinds of questions routinely addressed by the *Science and Engineering Indicators*. Jones’s report, *The Evaluation of Existing Databases for Policy Research on Humanities Fields*, revealed to the Task Force the strengths and weaknesses of current resources. The study makes clear that the humanities data sources we now have allow limited use for policy research because they lack uniformity in their measurement techniques, analytic methods, and reporting standards. They are additionally problematic because of small sample sizes, restriction to narrow subsets of the humanities fields, and a failure to update the materials on a regular basis or make the data available to other researchers. As a result, we cannot answer with precision a number of fundamental questions—about trends in graduate education, the long-term employment of individuals with advanced degrees (both Ph.D.’s and master’s degrees) in specific fields of the

humanities, the number of faculty members and educators K-12 in all disciplines of the humanities, and the number of individuals with degrees in the humanities who work outside their areas of specialization.

Jones's report helps to explain the uneven nature of even the best existing data resources for the humanities: "One of the major problems affecting humanities information resources stems from the fact that too many parties with different purposes and orientations are collecting too much data, rather than not enough." As he notes, some of the surveys of individual disciplines are extensive, and indeed the best of the specialized studies discussed in his report could serve as useful models for more comprehensive efforts. But without central coordination of data-collection, consistent methodology in research surveys across the range of the humanities disciplines, and sustained commitment to the surveys over time, the data derived from even the best of the existing tools possess limited value. The reality is inescapable: policy makers and professionals in the humanities are not well served by the current data collection efforts. So the Academy has resolved—in collaboration with the leading learned societies, higher education and professional associations, public humanities organizations, university research institutes, and other groups—to improve the situation.

Because data gathering is expensive and resources in the humanities limited, the Academy's activity will not duplicate existing efforts. We intend to enlist all the relevant stakeholders—including the professional associations, humanities organizations, researchers, and policy makers—in the planning and creation of the model databases, and we will continue to work closely with the many federal agencies that collect data, including the National Endowment for the Humanities, the National Science Foundation, the National Center for Education Statistics of the Department of Education, the Department of Labor, and the U.S. Census Bureau.

To supplement the report from the Academy's Task Force on Data Development, we invited four individuals familiar with the issues to explain why the *Humanities Indicators* are needed: Francis Oakley, a former president of Williams College; Robert Solow, Nobel Prize winner and professor emeritus of economics at M.I.T.; Phyllis Franklin, executive director of the Modern Language Association; and John D'Arms, president of the American Council of Learned Societies. Each brings a different disciplinary or organizational perspective to the discussion, but all agree that the time has come for an ambitious, coordinated approach. A report to the Academy by Calvin Jones follows their essays.

We view the current efforts of the Initiative as first steps toward the creation of a body of data that will grow with the years, for, as the essays and the Jones report make clear, databases become most useful when they provide consistent measures over time. Because the

Academy is an independent, national institution, it can take a long-range view of the needs of the humanities, as it did effectively in the past on the several occasions noted above. And because the Academy is an impartial convening body, it can bring together experts and representatives from a great range of scholarly associations, as it is now doing through the Initiative for the Humanities and Culture. Our common purpose is the creation of a complex set of databases, analogous to those available for engineering and the sciences, that will enable critical issues in the humanities to be addressed practically and cogently, and this report explains why the approach we are taking is both timely and essential.

As we go to press, we mourn the untimely death of our colleague, John D'Arms – a distinguished classicist, Fellow of the American Academy, and President of the American Council of Learned Societies. John D'Arms, a long-time champion of the humanities, brought his passion for scholarship to bear upon the need for better data, as his essay for this volume shows, as well as on other practical problems facing the humanities. We at the Academy join the entire humanities community in honoring his memory. We are all indebted to his leadership.

Patricia Meyer Spacks  
President

Leslie C. Berlowitz  
Executive Officer

American Academy of Arts and Sciences  
January 23, 2002

# Acknowledgments

We are grateful to our many colleagues from the humanities community, and to the social scientists and other specialists with interests in data collection, who have contributed to the conceptualization of this report and the larger Initiative. We will continue to look to their good counsel as we proceed.

Special thanks are due Academy Fellows Jonathan Cole, W. Robert Connor, John D'Arms, Denis Donoghue, Steven Marcus, Francis Oakley, Kenneth Prewitt, Stephen Raudenbush, and Robert Solow, who, along with Academy President Patricia Meyer Spacks, have lent sustained intellectual leadership to the Academy's Humanities Initiative. We are also grateful for the time and guidance contributed by Academy Fellows Norman Bradburn and Robert Post and by the many other advisors and external reviewers who have joined in the effort. Among these are Phyllis Franklin, Modern Language Association; John Hammer, National Humanities Alliance; Arnita Jones, American Historical Association; Gail Leftwich, Federation of State Humanities Councils; Felice Levine, American Sociological Association; Esther Mackintosh, Federation of State Humanities Councils; Deanna Marcum, Council on Library and Information Resources; Jeffrey Thomas, National Endowment for the Humanities; Duane Webster, Association of Research Libraries; Steven Wheatley, American Council of Learned Societies; and Kathleen Woodward, Consortium of Humanities Centers and Institutes.

We also want to thank Lynda Carlson and Ron Fecso of the National Science Foundation, Charlotte Kuh of the National Research Council, Clifford Adelman of the Department of Education, and Nils Hasselmo and John Vaughn of the Association of American Universities for sharing their extensive knowledge of current data surveys, and Stanley Katz of Princeton University, Harriet Zuckerman of the Andrew W. Mellon Foundation, and James Herbert of the National Endowment for the Humanities for their good counsel.

Corinne Schelling and Doug DeNatale of the American Academy staff played important roles in advancing this project, and the Academy's principal consultant, Calvin C. Jones, deserves special credit for his careful analysis.

The Academy thanks as well Walter B. Hewlett and the trustees of the William R. Hewlett Trust; Sara Lee Corporation, especially John Bryan, retired Chairman and CEO, and Robin S. Tryloff,

President of the Sara Lee Foundation; and the Andrew W. Mellon Foundation for their invaluable support of the Humanities Initiative in its planning stages.

Finally, we give special recognition to the Rockefeller Foundation for its support of this publication and to Gordon Conway, Raymund Paredes, and Lynn Szwaja for their early encouragement of the Initiative. The Rockefeller Foundation's assistance will also enable the Academy to implement several of the recommendations of the Task Force on Data Development and continue our work with the many organizations throughout America that share the Academy's desire to advance the cause of the humanities.

# The Value of *Humanities Indicators*

ROBERT M. SOLOW

Systematic data-collection is not a spontaneous, “natural” activity. It is costly and tedious, and has to be done for a purpose; otherwise it is unlikely to be done at all. The purpose itself is often indirect. The availability of *Humanities Indicators* will not directly lead to the writing of better history or philosophy or literary criticism, any more than the publication of *Science and Engineering Indicators* can unravel any puzzles about superconductivity or developmental biology. The data, especially in the form of consistently defined and measured time series, are a contribution to the infrastructure of the various disciplines.

They can help us to understand, in a way that anecdote cannot, what has been happening to the recruitment, training, and career chances of students of history, philosophy, or literature. Eventually, if more and better information leads to better decisions in universities, foundations, and government, over- and under-recruitment can be avoided, and students can be trained in ways that are better adapted to the functions they will actually perform and the lives they will lead. This cannot be bad for the health of the intellectual project of each discipline, and is likely to be good. On a broader scale, it is surely useful that educational resources should be better adapted to the demands that society will make on students and professionals.

So why is data-collection in the humanities still at such a primitive stage when the biennially published *Science and Engineering Indicators* is already some 1200 pages long, and still growing? I ask the question only because the answer is so obvious: it is because society at large thinks—correctly—that the science-engineering enterprise is important for the economic health and progress of the nation. So major institutions, public and private, Congress, and the press, want to know how the science-engineering enterprise is doing, how well it is gearing up to meet demands for trained people, how efficiently it is disseminating newly won knowledge about science and technology, how thoroughly it is matching demand with supply for scientists and engineers and for the products of science and engineering re-

search.

Anyone who has observed the nexus between the science-engineering enterprise and the political, economic, and other institutions of society at large knows how anxious and insecure the “basic research” community feels, with or without *Science and Engineering Indicators*. The community fears, with good reason, that support for curiosity-driven research will be neglected because politicians, foundation executives and citizens generally can see no clear connection between the activity of “pure” science and the national output of goods and services.

The apparent exception is illuminating: almost any research at the frontiers of molecular and cell biology could presumably lead, by an unforeseeable route, to useful medical discoveries. That is no doubt why Congress and successive Administrations bury the National Institutes of Health in money, but are stingier with the National Science Foundation, and often try to direct its funds away from just science into programs linked to “national needs.”

Academic economists who work at quantifying the “social return” to funds spent on industrial research and development are used to being asked by their colleagues in basic science whether one could not calculate an analogous rate of return for money spent on astrophysics and other scientific research aimed at answering questions far removed from marketable goods and services or from visible environmental amenity. The answer is probably No, because the chains of cause and effect are so numerous, long, tortuous, and diffuse, and the individual steps are so chancy. They do not lend themselves to the indirect measurements that are possible in simpler cases.

The case of the humanities is much harder, because even the final “product,” however we define it, lends itself so poorly to measurement in terms of fungible dollars. “Net contribution to GDP” is not exactly what the humanities are about. One imagines that pure science is closer to this state of affairs than anxious researchers like to admit to themselves. On the other hand, everyone *knows* that even the most ivory-tower chemistry can lead to some unexpected new material or process.

The creation of a decent data base for the humanities will not solve the problem of making the importance of humanistic education and research real to the world of legislators, governors, and citizens. That difficult job is no doubt never-ending, and not getting any easier. For anyone who does take humanistic education and research seriously, however, the Indicators project offers a major tool for evaluating and improving the institutions and routines that create our students, consumers, practitioners, and professionals, and provide them with the opportunity to study, consume, practice, and profess.

The humanities community itself, including its funders, knows

deplorably little about what is taught to whom and by whom, how long it takes, where graduates and post-graduates go, what they do when they get there, and how many of them there are. The humanities are not alone in this kind of ignorance. It would be a small miracle if there were not imbalances and inefficiencies in this process. If these could be identified, it might be possible to correct them, freeing up resources for productive use. It might even be possible to show funders where and how an infusion of new resources could have a disproportionately large effect on the capacity of the system to produce trained people and ideas.

There are bound to be opportunities for systematic self-improvement revealed by *Humanities Indicators*. Everyone needs to see how institutions and their practices create incentives, and how incentives create results. But the value of an effort to collect and publish consistent data goes far beyond the administrative imagination, important as that may be. “Know thyself” is an injunction that ought to resonate with anyone who cares about the humanities, as producer or consumer. It can only be a big step forward for those in the humanities—or any group of disciplines—to know more about and understand better the institutions that allow humanists to work and the humanities to flourish. *Humanities Indicators* would be an important step toward self-awareness.

# Data Deprivation and *Humanities Indicators*

FRANCIS OAKLEY

Why is it that those of us in the humanities seem to find it so very hard to convey to others the significance of what we do, or its importance for the national well-being, or even the status and current condition of the various fields of humanistic endeavor to which we bring so passionate a commitment? No easy question to answer, of course, nor an issue (or set of issues) susceptible in all its formidable complexity to resolution via any single or simple mode of approach. Part of the problem, I sense, is that we ourselves do not really understand what we do, why we do it, or why it might be as important as we instinctively take it to be. And part of the reason for that particular failure of understanding, or so I have come to conclude, is that even when we try to understand what we do by placing it in some larger context or seeing it from some larger perspective, we find that we lack the supportive interpretative tools provided by the systematic gathering, assembly, organization, analysis, and dissemination of the type of pertinent data long since available to those whose task it is to interpret the natural sciences not only to the larger public but also to themselves. For the humanities, perhaps surprisingly, such data are either lacking or, where they have been collected, are inconsistently assembled, hard of access, poorly disseminated, inadequately analyzed, unwittingly ignored, and routinely underutilized.

The firmness of my conviction on this score is not, I hasten to add, grounded in any professional expertise as a collector or analyst of statistical data. Far from it. In my own work as an historian of late-medieval religious and intellectual life, quantification plays little or no role at all. Instead, my preoccupation with the matter is that of a would-be user or consumer of such data. It stems from personal experience during the late 1980s and early 1990s when, as president of a liberal arts college and in an attempt to come to terms with the American undergraduate experience, I committed myself to the thick of the battle of the books then raging concerning the quality of American higher education in general and the state of the humanities

in particular.<sup>1</sup> By that time, so far as the torrent of criticism directed against the academy was concerned, the friction of constant repetition had long since come to dull the edge of surprise. Had that not been the case, most of us, I believe, would have detected something quite odd about the nature of the debate that had ensued. Much of it had been mounted in a contextual void—social, institutional, and, above all, historical. Beyond that, and very much the point that concerns me here, neither the critics nor their characteristically wounded (if occasionally truculent) respondents had shown much interest in the pertinent statistical data, even when such data were available. Instead, prescription and exhortation had been allowed to preempt description and analysis, we were being treated to a species of disheveled anecdotalism, and a free-fire zone had been cleared for eye-catching and sensationalist claims matched all too often by analytically flaccid and apoplectically sloppy responses.

Clearly, this was not good enough. If those of us in the academy in general and the humanities sector in particular were unwilling or unable to make a disciplined attempt to understand ourselves and our current plight, it was hardly surprising that, having found ourselves under attack, we were proving to be so singularly ineffective in making our case to the larger public.

In some respects the operative word was “unwilling.” Pertinent data sets actually existed capable of shedding a measure of light on some at least of the issues in hot dispute. Neither the critics, however, nor those who had undertaken to respond to them betrayed much interest in making use of such data. For the combatants the joy of battle and the clash of sweeping assertion and counter-assertion were proving, it seemed, too gratifying to encourage any craven attempt to turn for help to the relevant facts. In other respects, however, the more accurately descriptive word was “unable.” Sometimes pertinent data sets existed but had been so poorly disseminated that those whose interpretative efforts stood most to benefit from them were ignorant of their existence. In other cases, such data sets existed but were so difficult of access as severely to discourage an attempt to make use of them. And, in still others, the necessary data had never been collected, or had been collected from counterproductively small samples of respondents, or in too fragmented or chronologically limited a fashion, or had been aggregated in such a way as to render it impossible to isolate and tease out information at the needed level of specificity, or, alternatively, assembled using incompatible definitions or in response to shifting questions to such a degree that some needed modes of aggregation were precluded.

And so on. In relation to the informative listing of existing data sets pertinent to the humanities assembled in a 1998 report commissioned by the American Academy of Arts and Sciences (DeNatale

1. The outcome of that effort being my *Community of Learning* (1992).

and Fong 1998); a related report by Calvin C. Jones, which the Academy commissioned in 2000, and which forms the last section of this occasional paper, identified an even broader array of impediments standing in the way of anyone who might want to make use of those data in an attempt to understand the current state of the humanities, or to move beyond that in order to frame policy proposals looking to the future. Those reports not only describe in considerable detail the large body of data resources concerning the arts and humanities already existing, as well as the obstacles and shortcomings that so often preclude their being used effectively, they also identify the critical gaps that would continue to exist in the information at our disposal even if such impediments were somehow to vanish. That being so, it would be redundant for me to belabor the facts of the situation the reports set forth so clearly. Instead, and in order to put a few shreds of flesh on what admittedly can be pretty dry bones, it might be helpful if I were to illustrate from my own experience as a would-be interpreter of things higher educational and matters humanistic some of the obstacles to understanding one typically encounters if one seeks the broader perspective on contemporary dilemmas that only recourse to the pertinent statistical data can provide.

Two illustrations should suffice. While neither is limited wholly to the humanities, both certainly shed light on the humanistic plight. The first concerns the stated priorities and actual behavior of American faculty when it comes to balancing the research imperative with their teaching obligations. That illustration should highlight the sort of price paid when pertinent collections of data are poorly disseminated and, as a result, unwittingly ignored. The second concerns American undergraduates, what they have been choosing to study, how those choices have shifted across time, what it is we are choosing to teach them, how in differential institutional settings we seem to be more or less effective in mediating certain subjects, and why that might be so. And this latter illustration should serve to expose the veritable cat's cradle of complexly-entangled obstacles one encounters if one seeks out the pertinent data needed to comprehend the current situation of the humanities in undergraduate education, how it looks in comparison with the state of affairs in the past, and what is to be expected by way of change in the future. Those obstacles range from simple lack of data (or the need to rely on data so localized in subject, place, or time as to be of little help in gauging the big picture), via data that resist either aggregation or effective disaggregation because they have been collected piecemeal in response to differing questions or using differing and mutually inconsistent definitions, all the way to "snapshot" data collected at different points of time or in sequences too short to permit the type of longitudinal accumulation necessary if one is to appraise the degree to which the situation at any given moment is either "nor-

mal” and unexceptionable or stands out as unusual and cause for celebration or concern.

First, the priorities and behavior of faculty. One of the issues that rose to damaging prominence during the late-lamented “culture wars” and battle of the books was the so-called “publish or perish” syndrome and the threat that the university’s (allegedly characteristic) lack of commitment to teaching was seen to pose to the quality of undergraduate education. That issue had clearly “arrived” in 1995 when it was made the focus of a classic *60 Minutes* exposé on CBS, replete with fleeting footage of quasi-unintelligible instructors mumbling away in multiple classroom settings, and glancing exchanges between Lesley Stahl and sundry (but obliging) academic types willing to say the most extraordinary things about the lack of importance attached to teaching at their own particular institution—one of them, indeed, cheerfully confessing, by way of arresting wrap-up to the whole episode, that because of this he was “waiting for some powerful parent to sue the university for consumer fraud!”

Not exactly a happy comment on the current state of higher education. Not one, certainly, likely to advance its cause either with the public at large or with public officials charged with the responsibility for resource allocation. The less so in that, with an almost liturgical measure of repetition though in unusually vivid form, it involves the ventilation of a critical theme that has long since found an honored place in the potpourri of anecdote that such prominent critics of contemporary American higher education as Martin Anderson, Lynne Cheney, Page Smith and Charles Sykes have on this, (as on related matters) substituted for any real attempt, however critical, to find out and come to terms with what we can actually know about it. (See Anderson 1992; Cheney 1990; Smith 1990; and Sykes 1990.)

At their harshest, the claims such critics had been making were two-fold. First, that despite (or because of) the great emphasis that the university community places upon research, and the mounting surge of publications resulting therefrom, much of what is produced is trivial, dull, pedestrian, esoteric, or, God forbid!, some mind-numbing combination of all four. Academic writing and research, according to Anderson, constitutes “the greatest intellectual fraud of the twentieth century.” Similarly, for Sykes “much of what passes for knowledge creation . . . is merely humbug” (Sykes 1990: 103. See also Anderson 1992: 85, 106–106, 112). Second, that in our current era teaching is radically undervalued by the academy; that academics, in the frantic attempt to mount their overpriced and oversold research effort, have come increasingly to discharge their teaching responsibilities less conscientiously and less effectively than in the past; that the “academic culture” (thus Sykes) is not merely indifferent to teaching, “*it is actively hostile to it.*” (Sykes 1990:54; emphasis his).

Caught up in these charges, which are combined and recombined

in varying ways, are several distinct strands of accusation. But if one focuses specifically on just one of them, namely, the widespread complaint that the university community vastly overemphasizes research and accordingly (very much as a necessary concomitant) undervalues teaching, one confronts the fact that here the reliance on tiresomely recycled anecdotal evidence is particularly puzzling. From the late 1960s to the early 1990s, after all, we have at our disposal quite rich sets of survey data, especially those generated by the series of periodic faculty surveys that the Carnegie Commission on Higher Education mounted at intervals from 1969 onwards. The data thus generated concern (among other things) the attitude of faculty towards teaching and research, time devoted to these and related activities, actual research productivity, and so on. Although these data have been analyzed and added to by such distinguished scholars as Everett Carl Ladd, Seymour Martin Lipset, Ernest Boyer, Martin Trow and Oliver Fulton, and Howard Bowen and Jack Schuster,<sup>2</sup> they have been ignored almost entirely, not only by those critics alleging the occurrence in the academy of a flight from teaching, but also by most academics themselves, whose understanding of their own profession has been accordingly impoverished. Rarely, in my experience, do either the critics or their bruised academic respondents attempt to take into account the broad range of institutions offering higher education in America or the multiple differences in the priorities and preoccupations of the faculty who teach at them. And not least among those differences is variation in the degree of engagement in matters scholarly. “Publish or perish” may be a hallowed (and beloved) cliché among academics and educational critics alike, but one is tempted to claim that “publish *and* perish” would be no less (if also no more) accurate a description of the realities. For what statistical surveys and analyses mounted over the past thirty years have repeatedly indicated (and with only modest shifts evident between 1969 and 1989) is that the typical faculty member in the United States places a much higher priority on teaching and a lower priority on research than is often alleged and almost always conceded to be the case. Taking the profession as a whole, it turns out, a clear majority of American academics view themselves as teachers rather than as “research people.” In this respect, their actual behavior matches their attitude; they are, in fact, inactive in research but succeed somehow, nonetheless, in avoiding perishing while publishing little or nothing. Less than one fourth publish at all extensively, and, as a result, a surprisingly large proportion of the books and articles produced are the work of a smallish group (say 6 to 10 percent) of compulsive recidivists.<sup>3</sup>

2. The questionnaire and tabulated data generated by the 1969 survey are to be found in Ladd and Lipset (1975: 85–126). Results of the 1989 survey are tabulated in Boyer (1990), App. A, pp. 85–126. In addition see Trow and Fulton (1975: 39–83); Ladd (1979); and Bowen and Schuster (1986).

3. See especially Ladd (1979: 3–5, 7–11).

These data deserve far wider and more effective modes of dissemination than they have thus far received, either among academics themselves or among the broader public. For the former they could serve to undercut a damaging and constricting stereotype. For the latter, and especially those worried about an allegedly declining commitment of the academy to teaching, they could serve to reassure. In order to make it possible to monitor potential change in the attitudes and behavior to which they bear witness, the periodic surveys that generated them should be continued. But it appears that the Carnegie Commission on Higher Education has discontinued its own (particularly informative) surveys and we are threatened, accordingly, with an unhelpful return to the world of misleading (if frequently arresting) anecdote and an insistently autobiographical mode of witness which, in our discussions of the state of higher education, has served us so very poorly in the past.<sup>4</sup>

What, then, about student behavior? Over the course of the last quarter of the twentieth century, it will be recalled, during a period marked by an historic and truly massive increase in the numbers of students flooding into the undergraduate programs of our colleges and universities, our ears were assailed by a persistent drumbeat of educational commentary of the “eclipse of the liberal arts,” “crisis of the humanities,” “erosion of humanities studies,” “flight from the humanities” genre. That mordant strain of commentary evoked and lamented, among other things, what at that time appeared to be a catastrophic decline in the numbers of American students willing to commit themselves to the study of the humanities disciplines, memorializing, often with lugubrious shaking of journalistic jowl, the substitution as required reading of “trendy lightweights” for tried and true literary classics, a withdrawal from the study of Western civilization and its history in favor of more exotic (and, perhaps, more marginal) topics, as well as an accompanying substitution for more traditional interpretative modes of faddishly deconstructive approaches or those more contextual in nature and pivoting on a truly obsessive preoccupation with the categories of race, gender and class.

Certainly, and not without a degree of plangent eloquence, the more vocal critics of the academic scene made much of these alleged shortcomings, weaving them together in a powerful and regrettably influential narrative of declension. Academic overspecialization, incoherent curricula, failure “to bring the humanities to life,” the obscure, tendentiously ideological and politicized manner in which the subject matter was being mediated—such were the explanations for the “flight from the humanities” that the critics advanced with a degree of confidence that stood in almost inverse proportion to the actual dimensions of the factual base on which those explanations appeared

4. See Bowen and Sosa, (1989: 12), where they emphasize the drawbacks attendant on the fact that “[I]n thinking about education, many of us are autobiographical . . . [and] tend to make judgments about the future on the basis of our own experiences and the experiences of friends.”

to rest. “It may well be the case,” or so Roger Kimball speculated in 1990, “that the much publicized decline in humanities enrollments recently is due at least in part to students’ refusal to devote their college education to a program of study that has nothing to offer them but ideological posturing, pop culture, and hermeneutic word games” (Kimball 1990: xvii).

One of the startling things about almost all such claims is that they appear to have been based on what was purported to have been going on at little more than a dozen of the nation’s leading research universities and liberal arts colleges. No attempt was being made to take into account the extraordinary number, range, complexity, and diversity of the more than 3,500 institutions making up the American system of higher education. And another, even more startling, thing about the matter is that even if we had all attempted to take such factors into account, none of us, external critics or academic apologists alike, were really in a position to know what exactly was going on at that time. That is to say, we were certainly not in a position to know what we really needed to have absorbed in order to assess the situation calmly at a national level or to view it from any but the most impoverished and concomitantly misleading of historical perspectives. We were not even in a position to know whether or not the number of student course enrollments in the humanities really was declining. How could we be? Global, stable, reliable and comparable data simply did not (and still do not) exist. Conclusions about decline, if they were to be based on anything more than autobiographical witness or anecdotal evidence, had to be extrapolated from patterns of students majoring in humanities subjects and of degrees awarded in those fields, patterns which, in the 1970s and 1980s, were indeed marked by decline. But students do not limit (and are rarely *permitted* to limit) their course-taking to their own degree programs. Given general education requirements and the sheer increase nationally in the overall numbers of undergraduates, it may conceivably have been the case that the gross number of student enrollments in humanities courses may even have stood in that period at an all-time high.

Even if one were to exclude from consideration that arresting possibility and to concede the likelihood that the decline in majoring and numbers of degrees granted in the humanities during the 1970s and early 1980s did indeed entail a comparable decline in humanities course enrollments, it would still be difficult to know exactly how to appraise that shift without seeing it from a fairly long-term historical perspective and without teasing apart what was going on in the different institutional sectors that make up the entire higher educational universe. But as things stood at the time, the former was impossible because runs of data concerning undergraduate degrees awarded became available only as recently as 1970. And the latter was difficult

in the absence of some specially designed effort. It was only in effect, after 1989, thanks to studies undertaken first by Turner (1990), Bowen and Sosa (1989), and later by Gilbert (1995), that it finally became possible to see the “decline of the humanities” in adequate historical perspective and to analyze it paying adequate attention to shifts in the overall demography of the American student population, as well as to the differences evident in the experience of the various institutional sectors. Similarly, it was only in the 1990s that further, specially-targeted studies generated the data needed if one were to assess the validity of claims made about the alleged decline in the amount of attention being given in undergraduate education to the Western civilizational tradition and its great literary classics, as well as those made about changes for the worse in the interpretative tactic being pursued in the attempt to come to terms with literary works in general.<sup>5</sup>

And what did such specially-designed empirical enquiries reveal? Well, among other things (and by way of illustration), first, that so far as decline of majoring in humanities subjects went, different institutional sectors manifested strikingly different patterns. Kimball had had a great deal of fun at the expense of the authors of “Speaking for the Humanities” (an American Council of Learned Societies 1989 pamphlet responding to the Bennett and Cheney reports on the plight of the humanities) for their piecemeal and rather fumbling attempt to question or qualify the evidence being adduced to warrant talk about a decline in humanities enrollments.<sup>6</sup> But, Kimball to the contrary, their puzzlement was understandable. It was not, it turns out, the type of institution with which they were all affiliated—the top-ranked research university among whose faculty members what Kimball had parodied as “ideological posturing, pop culture, and hermeneutic word games” would appear to have been most deeply entrenched—that accounted, statistically speaking, for the decline in question. Instead, among the baccalaureate-granting institutions it was the comprehensive universities and colleges, and such other four-year institutions as the state colleges, that had seen the sharpest, most continuous, and most distressing declines since the early 1970s in arts and sciences majoring in general. At Cornell, for example, there had been only the most modest fluctuations between 1954 and 1986 in the percentages of arts and sciences degrees conferred, and no real change at all between 1970 and 1986. At Ball State University, on the other hand, where the arts and sciences share of degrees awarded had risen between 1954 and 1970 from 2.5 to no less than 29.9 percent, by 1986 that share had fallen back, quite catastrophically, to 13.3 percent (Turner and Bowen 1990: 518; Table 1). The message, then, was that

5. See Levine and Cureton (1992a and b); Huber and Laurence (1989); Huber (1992); Franklin, Huber and Laurence (1992); Huber (1995). For what can be gleaned from these and other (rather scattered) survey data, see Oakley (1997: 63–83).

6. Kimball (1990: 36), commenting on George Levine et al. (1989), esp. pp. 21–24, 35–37.

in relation to higher education as to other things, averages based on aggregated figures can conceal almost as much as they reveal, and that it was crucial, when analyzing these adverse trends, to distinguish carefully among the several institutional sectors.

It turned out to be crucial also, as the figures cited above already suggest, to be conscious of the degree to which the chronological point of departure settled on for purposes of statistical comparison can determine the nature, trajectory, and amount of change one perceives in the period ensuing. Because the publication of data-runs concerning numbers of degrees awarded in the arts and sciences in general and the humanities in particular began only around 1970, that was the earliest point of analytic departure available for purposes of comparison. And that turns out to have been extremely misleading. Once the effort was made (in the late 1980s and early 1990s) to recover and track the pertinent statistics back beyond that point, it became clear that the startling decline in liberal arts majoring evident in the 1970s and early 1980s represented no unprecedented decline from any long-term and stable norm. Instead, it reflected a falling-back from the all-time peak in such majoring produced by an unprecedented surge of interest lasting from the late 1950s to the early 1970s—both surge and falling-back, moreover, being in part gender-determined (Turner and Bowen 1990: 518–20). It became clear, too, that the effect of that falling-back had been to return us to something like the long-term “trajectory [of gradual decline] established over the past *century*” (Gilbert 1995: 37; emphasis mine).

As for what was being taught to those students who were still pursuing humanities subjects or the way in which those subjects were being taught, I will not dwell on the matter at any length. Instead, let me simply note how little solid information we have on such matters, so far as American higher education as a whole is concerned—at least, how little information we have of the sort that lends itself readily to reliable aggregation. But if we were tempted to take at face value the sweeping claims characteristic, during the 1980s and early 1990s, of the critics of American higher education, a cluster of targeted studies which *Change* magazine and the Modern Languages Association sponsored in the 1990s generated data that would doubtless contrive to surprise. Thus we learn, for example, that far from having declined between 1970 and 1985, the percentage of institutions requiring of all students a course not simply in Western civilization but specifically in the *history* of Western civilization had risen by a little over 5 percent, to a total of almost 50 percent (Levine and Cureton 1992b: 51, Chart 1). Similarly, we discover that far from “the trendies” dominating reading lists in introductory courses in British and American literature nationwide, such traditional or “classical” authors as Chaucer, Shakespeare, and Milton, or Hawthorne, Melville, and Whitman continue to ride high. For the better part of a

decade much prominence was given to a rather zany, throw-away remark attributed to an English department chairman in a 1988 *Chronicle of Higher Education* article to the effect that he would be willing to “bet that [Alice Walker’s] *The Color Purple* is taught in more English courses today than all of Shakespeare’s plays combined.” Once collected, however, the pertinent data revealed that, in terms of frequency of adoption in such courses, Alice Walker came in at less than 1 percent, significantly behind even Cotton Mather (2.1 percent), let alone Hawthorne (66 percent) or Chaucer (86 percent).<sup>7</sup>

All of these are indispensable facts that should long since have been at our disposal. Had we known them in the 1980s and early 1990s the vigorous debate then being conducted concerning the condition of higher education in general and the state of the humanities in particular would have been far more fruitful and would have represented something other than the good opportunity wasted that I fear it actually was. But they were facts, if already discovered, that had been very poorly disseminated, or facts only uncovered later by dint of specially mounted (and usually one-shot) efforts either at data collection or at the massaging of previously collected data not hitherto available in readily (or pertinently) usable form. And a certain poignancy attaches to that fact. Such ad-hoc efforts, after all, should really not have been needed. I hope that the two illustrations given above may have served to highlight, when it comes to understanding and making the case for the humanities, how very much may depend on the ready availability of a stable, accurate, comprehensive, timely, and continuing flow of pertinent data. Across the past half-century, our colleagues in engineering and the natural sciences have gradually succeeded in creating such a continuing flow of data in the form of the *Science and Engineering Indicators*, which the National Science Board publishes every other year. As the confusions, misunderstandings, and misrepresentations of the past have made painfully clear, it is long since time for those of us in the humanities to have at our disposal a comparable resource. All praise, then, to the American Academy of Arts and Sciences for launching, as part of its Initiative for the Humanities and Culture, the great effort to bring that much-needed resource into being.<sup>8</sup>

7. For the claim about *The Color Purple*, see Clausen (1988). This flip comment was picked up and treated as a piece of documented evidence (after all, had it not appeared in print?) by the *Wall Street Journal*, the *New Criterion*, and Dinesh D’Souza, and it was echoed glancingly or indirectly by others. For the tracking down and demolition of this particular piece of mythology, see Graff (1992: 17–21). For the pertinent statistics, see Huber (1995: 43–46, Tables 4 and 6).

8. AAAS 2000: 3–11

REFERENCES

- American Academy of Arts and Sciences. 2000. *The Initiative for the Humanities and Culture: Progress Report*.
- Anderson, Martin. 1992. *Impostors in the Temple: American Intellectuals Are Destroying Our Universities and Cheating Our Students of Their Future*. New York.
- Bowen, Howard R. and Jack H. Schuster. 1986. *American Professors: A National Resource Imperiled*. New York and Oxford.
- Bowen, William. G. and Julie Ann Sosa. 1989. *Prospects for Faculty in the Arts and Sciences: A Study of Factors Affecting Demand and Supply*. Princeton, N. J.
- Boyer, Ernest J. 1990. *Scholarship Reconsidered: Priorities for the Professoriate*. Princeton, N. J.
- Cheney, Lynne V. 1990. *Tyrannical Machines: A Report on Educational Practices Gone Wrong and Our Best Hopes For Setting Them Right*. Washington, D.C..
- Clausen, Christopher. 1988. "It Is Not Elitist to Place Major Literature at the Center of the English Curriculum." *Chronicle of Higher Education*, January 13: A52.
- DeNatale, Douglas and Elaine Fong. 1998. *Data Collection and Information Systems in the Arts and Humanities: Preliminary Survey and Recommendations*. Boston: New England Foundation for the Arts.
- Franklin, Phyllis, Bettina J. Huber, and David Laurence. 1992. "Continuity and Change in the Study of Literature." *Change* 24: 42-48 (January-February).
- Gilbert, Joan. 1995. "The Liberal Arts College: Is it Really an Endangered Species?" *Change* 27: 37-43 (September-October).
- Graff, Gerald. 1992. *Beyond the Culture Wars: How Teaching the Conflicts Can Revitalize American Education*. New York and London.
- Huber, Bettina J. 1992. "Today's Literature Classroom: Findings from the MLA's 1990 Survey of Upper-Division Courses." *ADE Bulletin* 101: 34-60.
- . 1995. "What's Being Read in Survey Courses? Findings from a 1990-91 MLA Survey of English Departments." *ADE Bulletin* 110: 40-48.
- and David Laurence. 1989. "Report on the 1984-85 Survey of the English Sample: General Education Requirements in English and the English Major." *ADE Bulletin* 87: 30-43.
- Kimball, Roger. 1990. *Tenured Radicals: How Politics Has Corrupted*

*Higher Education*. New York.

Ladd, Everett Carll, Jr. 1979. "The Work Experience of American College Professors: Some Data and an Argument." AAHE. *Current Issues in Higher Education*.

——— and Seymour Martin Lipset. 1975. *The Divided Academy: Professors and Politics*. New York.

Levine, Arthur and Jeanette Cureton. 1992a. "The Quiet Revolution: Eleven Facts about Multiculturalism and the Curriculum." *Change* 24: 25–29 (January–February).

———. 1992b. "Change Trendlines: Signs of a Changing Curriculum." *Change* 24: 49–52 (January–February).

Levine, George et al. 1989. "Speaking for the Humanities." ACLS Occasional Paper No. 7. New York.

Oakley, Francis. 1992. *Community of Learning: The American College and the Liberal Arts Tradition*. New York and Oxford.

———. 1997. "Ignorant Armies and Nighttime Clashes: Changes in the Humanities Classroom, 1970–1995," in Alvin Kernan, ed., *What's Happened to the Humanities?* Princeton, N. J.

Smith, Page. 1990. *Killing the Spirit: Higher Education in America*. New York.

Sykes, Charles J. 1990. *ProfScam: Professors and the Demise of Higher Education*. New York.

Trow, Martin and Oliver Fulton. 1975. "Research Activity in American Higher Education," in Martin Trow, ed., *Teachers and Students: Aspects of American Higher Education*. New York.

Turner, Sarah E. and William G. Bowen. 1990. "The Flight from the Arts and Sciences: Trends in Degrees Conferred." *Science* 250: 517–21.

# The Continuing Need for a Practical System of National Information and Data Collection on the Humanities

JOHN D'ARMS

In the fields of science and engineering, the National Science Foundation is the federal agency that has assumed responsibility for assessing our national capacity. Its chief instrument for this purpose is *Science and Engineering Indicators*, a biennial report that provides a thorough description and assessment of the condition of research and education in those fields in the United States. Its broad and careful analyses are extremely informative to policymakers responsible for setting national and institutional directions. In contrast, the humanities community needs, but sadly continues to lack, a similarly thorough report.

More than a decade ago, in testimony before the 101st Congress on behalf of the Association of American Universities and the National Humanities Alliance, I stressed the need to collect comprehensive, systematic data about humanities disciplines. At the time, it was evident that the data collection conducted by the National Endowment for the Humanities (NEH) was inadequate. We have now entered the twenty-first century, but no progress has been made. In fact, the situation has worsened: what was then inadequate is now nonexistent. Meanwhile, methods of research, fields of study, and educational practices in the humanities have changed drastically. My remarks from 1990 about the need for and value of data collection remain pertinent today. The substance of that portion of my testimony follows:

Responding to changes in both public and private policy requires a comprehensive system of information collection, analysis, and dissemination. In the humanities, the NEH is the single federal agency

charged to carry out such information activities. Continuous debate on the meaning and meaningfulness of statistics on college and university enrollments in the humanities has highlighted the problem of the inadequacy of our data. To address this problem, Congress utilized the occasion of reauthorization of the NEH in 1985 to direct that NEH shall “in consideration with State and local agencies, other relevant organizations, and relevant Federal agencies, develop a practical system of national information and data collection on the humanities, scholars, educational, and cultural groups, and their audiences. Such system shall include cultural and financial trends in the various humanities fields, trends in audience participation, and trends in humanities education on national, regional, and state levels.” Congress went on to specify one mode of use for disseminating the data and the analysis produced by the practical system: “Such systems shall be used . . . to prepare a report on the state of the humanities in the Nation. The state of the humanities report shall include a description of the availability of the Endowment’s programs to emerging and culturally diverse scholars, cultural and educational organizations, and communities and of the participation of such scholars, organizations, and communities in such programs.”

Congress’s formulation of the scope and nature of the federal interest in data in these areas, all of which are of great value not only to policymakers but also to scholarly, educational, and other public communities, was laudable. Unfortunately, the tasks specified by the mandate have yet to be accomplished. There remains much work to be done to develop a “practical system of national information and data collection” and to remedy the gaps in our knowledge that motivated this Congressional mandate.

In its formal response to this requirement, in December 1986, NEH stated, “The information we plan on using to conduct our assessment of the state of the humanities today is largely in place, and so we have proposed no new data collection projects as part of this system . . . ; No additional funds will be sought for implementation of the system.” To be sure, in the past NEH both collected data on the humanities and assisted others in doing so, but it did so chiefly with particular occasions and purposes in mind. It has never undertaken to establish a system with consistent standards of collection, analysis, and dissemination from year to year, nor has it taken steps to provide ready and inexpensive access to the data it collects.

NEH once provided for data collection and analysis through the Humanities Studies Program. The bulk of that funding appeared to go to two agencies: the Department of Education and the National Research Council. But most of the information gathered by the Department of Education and by the NRC was aggregate data, applicable to the humanities collectively. Consequently, it provided little insight into trends in individual humanities departments and in specific disciplines. Unfortunately, the Humanities Studies Program no longer exists. In addition, although NEH once periodically funded specialized studies of individual humanities fields by nongovernmental agencies, it no longer appears to do so on a regular basis. Yet, unless funding is available for such specialized studies, our empirical understanding of trends within the humanities will remain both sketchy and unhelpful. Three examples reinforce this point:

- We currently have longitudinal information on major fields of study planned by entering freshmen (from data collected by the Higher Education Research Institute at UCLA) and know the numbers of degrees granted in various fields, but we know virtually nothing about curricular choices between the onset and the end of the college career. Among other things, there are no easily accessible, longitudinal data on enrollments in specific humanities classes, the number of majors in specific fields, or retention and attrition rates in different fields. Such information provides a foundation for designing strategies to attract and retain majors; without it, one cannot identify whether, and when, students drop into or out of humanities disciplines.
- At the graduate level, we know the numbers of master's and doctoral degrees granted in specific humanities fields each year. But we lack evidence about the number, quality, and other characteristics of graduate students entering various humanities fields.
- The NEH has used information from its Challenge Grant program and on contributions to state humanities councils to assess private contributions to the humanities. These data have not been published and, even at their best, they provide only a very incomplete picture of private-sector giving. They cannot be said to yield the kind of systematic data on public and private funding of the humanities that is a vital underpinning for sound public policy decisions.

I concluded my 1990 Congressional testimony with three recommendations. Sadly, not only has none of these yet been realized, but from my current vantage point they seem like a wish list of impossibilities. Our recommendations on humanities data collection and analysis were as follows:

1. That the provision of the 1985 legislation calling for development of “a practical system of national information and data

collection on the humanities . . .” be retained and strengthened. While recognizing the progress NEH has achieved, we urge that the language call explicitly for the continued development of the system; increased attention to inexpensive access to NEH-compiled data by individual scholars as well as to reports and publications developed for dissemination; and for broader, more systematic consultation with relevant associations and other non-governmental organizations.

2. That NEH be instructed to reestablish, within the Humanities Studies Program, a small grants program for nongovernment organizations to encourage the collection and analysis of specialized data on the humanities disaggregated by fields of study and the secondary analysis of data relevant to understanding trends within the humanities.
3. That a standing advisory committee to the NEH on humanities statistics be established. Members of such a committee should be appointed in such a way as to ensure both broad representation of diverse humanities disciplines and institutional expertise in higher education and statistical methods. Both federal and non-federal groups concerned with the collection, analysis, and dissemination of data in the humanities should be represented on the committee. The responsibilities of the committee would include advising NEH on data needs, approaches to analysis, and dissemination.

Enactment of all of these recommendations is still desperately needed. In the intervening decade, NEH suffered a 40-percent cut in its 1996 appropriation, which increased the need for data (how do we measure the impact of such a change in funding?) while reducing the means to gather it. Are there other means to achieve this end? A study I began in 1996 of funding in the humanities demonstrated once again to me that the available data on critical subjects, collected by many different sources, were sporadic and idiosyncratic. A review of sources five years later revealed the same situation.

The initiative of the American Academy of Arts and Sciences is the most promising of recent efforts to address this continuing need. Not only does it aim to develop strategies for sustained collection and assessment of data on the humanities, but it will do so in consultation with leaders of principal humanities organizations in order to coordinate the effort and identify gaps in knowledge. Perhaps most importantly, it intends to disseminate its findings so that public- and private-sector policy decisions can be grounded in fact. ACLS is pleased to work with the Academy’s distinguished members to devise an efficient, robust, and sustainable means for gathering data that will be the basis of improved understanding and policies to advance humanistic studies in all fields of learning.

# *Humanities Indicators: A Promising Possibility*

PHYLLIS FRANKLIN

From the calls and e-mail requests I receive, I conclude that there is growing recognition among Modern Language Association (MLA) members of the importance of statistical information about the field. Data carry undeniable weight with deans, provosts, and state legislatures, and they're also useful in departmental planning. Local situations take on entirely new meaning when they're seen in a broader context, whether the issue concerns faculty salaries, enrollments, majors, Ph.D. employment, or curricular developments.

Unfortunately, statistical information about the humanities is both limited and dispersed, and so I was pleased to learn that the American Academy of Arts and Sciences is exploring how sustained, centralized data collection about the humanities might be funded and administered. I have attended two meetings the American Academy convened, and I am hopeful that this new effort will greatly improve the availability and range of information for and about the humanities.

In December 1999 Academy members Francis Oakley and Jonathan Cole convened a task force of other Academy members and representatives of the American Council of Learned Societies, American Historical Association, Council on Library and Information Resources, MLA, National Endowment for the Humanities, National Humanities Alliance, National Humanities Center, and Pew Charitable Trusts. Discussion focused on the possibility of developing *Humanities Indicators* on the model of the *Science and Engineering Indicators (SEI)*. Robert Solow, who serves on the National Science Board, talked about the administration and compilation of these indicators, which are published by the National Science Foundation every other year. A special staff oversees the publication of the *SEI*; science journalists who are paid for the work write the chapters; and a subcommittee of the National Science Board reviews the report prior to publication. Since the *SEI* staff only compiles data that other government agencies and trade associations collect, government agencies and trade associations also review the data and interpretations. The indicators are designed to show trends over time and to anticipate possible future problems, such as too few undergraduate

majors in science to meet the needs of higher education and industry, declining investments in research and development, or changes in employment opportunities for Ph.D.'s. The indicators are used in institutional planning by faculty members and college and university administrators; they are also used by the federal government, business and industry, the press, and the public, and they are credited with helping to shape public policy initiatives that fund the sciences. Solow noted that the *SEI* have been particularly effective in recent years in making a case for government support of research and development. He explained that the success of the scientific community rests on a general understanding that educating scientists and funding scientific research are essential to the national defense, the economic well-being of the nation, and the health of its people.

What are the indicators? They are quantitative and cover such matters as sectors of employment for science and engineering undergraduate and graduate degree recipients; salaries; funding for research; economic trends affecting research in the academy and industry; teacher preparation and student achievement in elementary and secondary schools; higher education enrollments and degrees granted; and numbers of research articles scientists in the United States and scientists in other countries publish. In 1998, eight chapters covered elementary and secondary education, higher education, the science and engineering work force, international research, academic research and development, research development in industry and technology, public attitudes and public understanding, and the economic and social significance of information technologies. If, like the *SEI*, the proposed *Humanities Indicators* depend entirely on data already available from government agencies, there will be substantially less information to include, but making it available regularly and providing analyses of it still would be an improvement over the current situation. Data should also be obtained from professional societies and other nongovernmental organizations, as well as state and local councils, and public humanities institutions such as library and museum groups. Furthermore, centralizing the publication of data about the humanities assumes the existence of an administrative unit and staff, who, like the staff that compiles the *SEI*, could identify and arrange for the collection of additional information. I also assume that if an administrative unit is established to compile and publish *Humanities Indicators*, the work of this unit would be supervised, as the work of the *SEI* staff is overseen, by representatives from the humanities.

As the discussion of the indicators progressed, the focus turned to the kinds of information people in the humanities might like to have. We could follow the general plan adopted for the *SEI*, which calls for quantitative data, or we could try to develop additional indicators that would say something about the quality of the field.

Jonathan Cole and Gerald Holton urged the humanities community to consider whether there should be quality indicators for the humanities. At first, I was taken with the idea, but after the meeting I read several chapters from the book that Leslie Berlowitz, the executive officer of the American Academy of Arts and Sciences, distributed at the meeting. The chapters appeared in *Toward a Metric of Science*,<sup>1</sup> a book that grew out of a conference about the development of science indicators. The invitation to the conference asked “What must one look at in order to estimate the condition of science as an intellectual activity or as a social institution?”<sup>2</sup> In response to this question, the editors of the volume expressed the wish to go beyond obvious “inputs” and “outputs,” but they also noted the impossibility of being free from—and indeed the need to develop—theoretical assumptions. While the collection of quantitative data is based on theoretical assumptions, these assumptions seem less problematic than assumptions about the value, for example, of an intellectual consensus in the field, which was one approach to a possible qualitative measure considered at the conference on science indicators. At our meeting, Jonathan Cole, a co-convenor and Provost at Columbia University, encouraged the consideration of quality indicators for the humanities, but he noted we should also learn how to improve on these materials. He pointed to the use of the number of citations a scholarly article attracts as a measure of quality that grew out of the attempt to develop qualitative measures for the sciences. The question of qualitative indicators in the humanities strikes me as intellectually interesting and possibly productive of new insights about the field, but less helpful as a measure of the condition of the humanities that institutions or the federal government could apply in judging programs and developments in the field.

Other issues were discussed at the meeting. First, the proposed project will be costly. Can a new source of revenue that does not compete with the current limited federal and private support be identified? Can funding from the private sector ensure the continuity of the enterprise? If an endowment is needed, how large would it have to be? These are difficult questions. The other concern focused on the possibility of learning what we might not want to know. Will the availability of statistical information challenge cherished beliefs? For me the answer to this question is easy. Having information is better than not having it, if our goal is to encourage vital scholarship and effective academic programs.

In addition to the goals of encouraging healthy scholarship and academic programs, the American Academy also hopes that better and more complete data can be used to make a stronger public case for supporting the humanities. But, as Gerald Holton noted during

1. Yehuda Elkana et al., eds., *Toward a Metric of Science: The Advent of Science Indicators* (New York: John Wiley, 1978).

2. Ibid. Bernard Barber, “Preface,” ix.

the meeting, data alone will not make a good public policy argument for support of the humanities. A compelling statement about why the humanities matter to individuals, communities, and the nation is also needed. Here the Academy intends to draw on the expertise of its Fellows and other scholars and scholarly organizations, as well as public humanities organizations, to create such a statement.

Since 1883, when the association was established, MLA members have collected data about the field so they could identify trends, evaluate the quality of their efforts, and plan for the future. With rare exceptions they did this work sporadically and with limited funds. We have long understood the importance of a statistical perspective, which allows one to move beyond the limited horizon of personal experience. Only recently have we begun to appreciate the public significance of quantitative information. Being counted means that you count.

A Report to the  
American Academy

Evaluation of Existing Datasets for Policy  
Research on Humanities Fields

CALVIN C. JONES

During the last two decades, concern has been mounting that teaching, learning, budgets, and a wide range of public activities in traditional humanities disciplines are caught in a downward spiral. The available evidence—mostly impressionistic in nature—has been used by leaders in the humanities to illustrate the decline and to advocate increased support for the humanities.

Assessments of the state of humanities fields and consequent appeals for greater investment—whether originating in academia or the public arena—have been hampered, however, by a lack of rigorous data analyses. The need for reliable, comprehensive, and accessible information to support more focused analyses has become increasingly apparent. If governments, foundations, and other key public and private institutions are to be convinced that the humanities need and deserve greater attention and support, we will need to improve significantly our ability to anticipate and describe important trends.

The American Academy of Arts and Sciences has developed its Initiative for the Humanities and Culture to address a critical need in our nation's intellectual and civic life and to provide the means of assessing and promoting the well-being of the humanities in the United States. The Initiative, therefore, aims to enhance understanding of the humanities and to help establish a new sense of common purpose among its practitioners and institutions. Toward that end the Academy will regularly convene humanities leaders to propose and coordinate research. More specifically, the Initiative will: (1) develop strategy for the ongoing collection and analysis of humanities data; (2) consult with leaders in humanities disciplines to identify critical gaps in knowledge and to promote research on key policy issues; and (3) improve dissemination of humanities research findings to a wide range of constituencies.

The Academy is committed to the development of an interdisciplinary data repository that will provide the basis for policy research on the condition of humanities fields and institutions and their relevance for the nation. An initial step is the design and implementation of a framework for systematic and sustained investigation and data collection in the humanities. The ultimate goal is the development of a set of *Humanities Indicators* describing the condition of both academic and public humanities and tracking trends over time.

Guided by its Task Force on Data Collection,<sup>1</sup> the Academy has launched this effort by commissioning two reports. The first, undertaken in 1998, was a preliminary survey of data on arts and humanities fields.<sup>2</sup> The second, presented here and completed in September 2000, evaluates the utility of existing data resources to serve the policy research needs of humanities fields and institutions. It builds on the earlier assessment by examining in greater depth

1. Task Force members include experts in the humanities and social sciences, and representatives of major cultural and scientific institutions. See Appendix D for a listing of this committee's members.

2. Douglas Denatale and Elaine Fong. 1998. *Data Collection and Information Systems in the Arts and Humanities: Preliminary Survey and Recommendations*. Boston, Mass.: New England Foundation for the Arts.

a collection of over a hundred major information resources to identify their strengths and limitations for policy research on the humanities.

#### REPORT CONTENTS

Section 1 summarizes key findings of the earlier Academy report and places them in the context of the current evaluation. Goals of the current evaluation are described in more detail, as are the criteria used to assess information systems.

Section 2 includes descriptions and utility assessments of over one hundred data resources listed in the *Index of Humanities Datasets*. This list was compiled in 1996 by the Modern Language Association of America for the American Council of Learned Societies, and until recently was maintained at the National Endowment for the Humanities web site. The *Index* was intended to identify datasets believed to be most relevant to the study of policy issues related to the humanities in general and to specific humanities fields and organizations (museums, libraries, etc.). For the purpose of our analysis, we have divided the *Index* resources into four groups:

- bibliographic materials
- directories and catalogs
- research datasets
- reports and publications

In addition to summarizing and updating key characteristics of each resource listed in the *Index*, we describe the potential value of each dataset for various dimensions of policy research, and detail the known limitations in their current condition.

Section 3 summarizes and analyzes findings in the qualitative review of the data resources. Three appendixes provide examples of the material contained in specific directories or surveys cited in the *Index of Humanities Datasets*.

#### THE STATE OF THE ART IN HUMANITIES POLICY RESEARCH

The main body of the report describes obstacles to conducting quantitative research from existing materials on the humanities. Such factors severely limit our ability to answer even the most basic questions about the current state of humanities fields, trends from the recent past, or conditions likely to face humanities practitioners in the future. The most serious problems with existing data resources may be classified into the following categories:

1. Fragmentation and lack of coordination (applies to both past and current data production efforts) in the quantity and quality of data currently being collected, analyzed, and disseminated.
2. Absence of data that are gathered continuously and consistently that would support research on trends affecting both current and prospec-

tive humanities practitioners, and in the numbers seeking training at various degree levels across all humanities fields. Typical causes for this problem include:

- a. limitations in the scope of data collections to narrowly targeted subsets of the complete universe (for example, studies based on members of institutional associations that cannot be generalized to the entire population of similar institutions),
  - b. population-based studies with sample sizes too small to permit separate evaluations of different humanities fields, and
  - c. disruption or discontinuation of periodic data collections necessary for time-series analyses of trends in humanities fields.
3. Limited availability of existing datasets to researchers, complicated by:
    - a. long delays between the collection and the release of national data systems,
    - b. lack of straightforward data distribution mechanisms, requiring lengthy, complex negotiations over terms and costs for researchers to gain access to certain datasets, and
    - c. proprietary restrictions on many datasets that limit use to the funders or producers of the data.
  4. Structures and frameworks for storing, managing, and accessing humanities data that impede or obviate the use of statistical research methods. This is the typical situation for the massive on-line bibliographic data systems organized for searching and listing, rather than for quantitative analysis of their contents.
  5. Persistence of inconsistencies in basic concepts and definitions employed by data producers in collecting data and/or reporting results.

The problems noted above affect each of the four functional groupings of data sets differently.

- “Bibliographic Materials” (used to analyze scholarly work and writings in the humanities disciplines) are hampered by limitations on access (problems 3b and 3c) as well as data structures and access mechanisms that create enormous startup burdens on researchers who must convert the data to a different format to conduct statistical analyses (problem 4).
- “Directories and Catalogs” (used for research on academic institutions, departments, programs, and other organizations such as libraries, presses, and museums) suffer from fragmentation and lack of coordination across disciplines (problem 1), the absence of consistent, long-term trend data (problem 2), and proprietary restrictions on access to data (problem 3c), as well as significant inconsistencies across resources in basic definitions (problem 5).
- “Research Datasets” (designed from the outset for quantitative analysis) are affected by fragmentation and lack of coordination (problem 1), absence of data at required levels of detail or disaggregation (problem 2),

all manner of limitations on data access (problems 3a-c), and inconsistencies across data systems in basic definitions (problem 5), which hinder research on the past, present, and future of the humanities labor force and job markets.

- “Reports and Publications” (compilations of statistical tables, with or without interpretive text) suffer primarily from the lack of needed types of data and definitional inconsistencies between the contents of the reports and the data needed by leaders and practitioners in humanities fields (problems 2 and 5).

We are aware that researchers studying certain specific humanities fields or practices (for example, faculty teaching literature or foreign languages, academic librarians and staff) have data resources that are well designed to answer questions that might be posed by academics, administrators, and practitioners about labor force or job market conditions in their fields. Unfortunately, researchers are unable to answer the same types of questions about the full range of humanities disciplines, or to compare current conditions with past trends across fields.

#### COMPARISON BETWEEN HUMANITIES AND SCIENCE AND ENGINEERING FIELDS

The situation for researchers on the humanities compares unfavorably with the experience of researchers attempting to answer similar questions about science and engineering fields. We can illustrate this difference by attempting to answer a few basic questions drawn from the issues listed in Section 1:

##### 1. Trends in Graduate Education

*Over the last thirty years, how many advanced (master’s and doctoral) degrees have been awarded in humanities fields?*

This question can be answered for doctorates in humanities fields using data from the *Survey of Earned Doctorates* from 1958 to the present time. However, researchers must first obtain a license from the National Science Foundation to obtain the microdata required to conduct the analyses.

Although summary reports on humanities degrees have been produced for decades, no existing reports are available that disaggregate degrees by specific humanities fields. Counts of masters and doctoral degrees awarded between the mid-1980s and mid-1990s are not reported separately for history (which is included in the social sciences), philosophy, or religion. Beginning in the mid-1990s, advanced degrees in philosophy and religion are reported separately, but degrees in history are still reported as part of the social sciences.

Individual science and engineering fields can at least obtain a reasonably disaggregated answer to this question. *The Science and Engineering Indicators 2000 (SEI 2000)*<sup>3</sup> report displays time-series graphs of the numbers of master’s

3. Data for science and engineering fields were easily found in the recent report from the National Science Board, *Science and Engineering Indicators—2000*. Arlington, Va.: National Science Foundation, 2000 (NSB-00-01).

and doctoral degrees awarded in all science and engineering fields for the thirty-year period, 1966–1996.<sup>4</sup> Separate counts over this period are available for broad scientific subfields, including

- physical and geosciences
- biological and agricultural sciences
- mathematics
- computer sciences
- engineering
- social sciences
- psychology

## 2. Higher Education Faculty in the Humanities

*How many faculty members are employed at U.S. higher education institutions in humanities fields?*

Although some associations of humanities practitioners have collected data to address this question, it cannot be answered for all humanities fields. The most systematic and current data on faculty counts are collected by the National Center for Education Statistics in the *Integrated Postsecondary Education Data System (IPEDS) Fall Staffing Survey*. This data system collects information on sex, race/ethnicity, type of institution, and position for all staff employed at postsecondary institutions. However, it does not collect any information about teaching fields of faculty.

The science and engineering fields have a slight advantage in addressing this question. The *SEI 2000* report displays data on faculty appointments of new Ph.D.'s for seven science and engineering fields. However, it does not report counts of the entire faculty work force in science and engineering. Other than this limited information, there is no source of nationally representative information on counts of faculty by field over time.

## 3. Workforce Characteristics

*How many people who hold degrees in the major humanities fields are currently employed in those fields?*

No existing publications provide answers to this question for specific humanities fields. Summary reports for the *Humanities Survey of Doctorate Recipients (SDR)* provide limited reporting on employment and unemployment for humanities Ph.D.'s, but do identify or count those employed outside their fields. No existing reports address employment patterns for holders of the master's degree in humanities fields.

Science and engineering practitioners have considerably better data resources to address this question. Tables 3.1 and 3.2 of the *SEI 2000* report display counts of persons with science and engineering degrees who were employed in science and engineering fields in 1997, disaggregated into five

4. Data points are displayed for five-year increments for 1966–1986 and one year increments for 1986–1996.

categories:

- physical and geosciences
- biological and agricultural sciences
- mathematics and computer sciences
- engineering
- social sciences and psychology

These tables also show the proportion of individuals holding science and engineering degrees in these categories at the BA, MA, and Ph.D. level who are employed in fields that are “closely related,” “somewhat related,” and “not related” to their degree fields. The proportion with occupations closely related to their degree fields increases sharply from 29 percent of those with BA’s in science and engineering fields, to 47 percent of those with MA’s and 48 percent of those with Ph.D.’s in science and engineering fields.

#### 4. Labor Market Indicators

4a. *What are the current annual average salaries earned by practitioners employed in the major humanities fields?*

The National Center for Education Statistics routinely reports on salaries of higher education faculty in humanities fields and library staff in its *National Survey of Postsecondary Faculty*. Summary data on earnings of Ph.D.’s in humanities fields were published through 1995 based on the *Humanities SDR*, the last year that survey was conducted. However, earnings were not presented separately for those working in humanities fields versus other fields. Moreover, no comparable data are routinely reported for those holding master’s degrees in humanities fields.

The *SEI 2000* report (Figure 3–11) displays median annual salaries earned by scientists and engineers in 1997 broken down into the same five categories of science and engineering fields listed under item 3 above, and also by race/ethnicity. In addition, Tables 3–10 and 3–11 of that report display details on 1997 annual salaries by field and by employment sector.

4b. *How many individuals earning new BA or MA degrees in the major humanities fields are employed in those fields?*

No systematic data are available to answer this question, either for the humanities degree earners in general, or for specific humanities fields.

The *SEI 2000* report (Table 3–5) shows the percentages of new BA and MA recipients in science and engineering fields in 1995–96 who were employed in seven different employment sectors (including two educational sectors, two for-profit sectors, and three non-profit sectors) separately for science and engineering degree earners.

4c. *What are the unemployment rates for those with degrees in humanities fields? Do humanities graduates frequently take employment outside their fields of training?*

Unemployment rates for Ph.D.'s in humanities fields may be estimated using data from the *Humanities Survey of Doctorate Recipients (SDR)* from 1977 through 1995 (biennially). With the exception of limited survey data on new Ph.D.'s in English and foreign languages, no data on unemployment or on "involuntary out-of-field employment" are available for humanities fields.

Table 3-6 of the *SEI 2000* report shows both unemployment rates and "involuntary out-of-field" employment rates for 1995 and 1997 for five separate engineering fields and twelve separate scientific fields. Table 3-9 shows similar data focusing on the most recent Ph.D. recipients in science and engineering fields.

#### 4d. *What are the prospects for future job growth in the humanities?*

No systematic data are available to address this question for the humanities in general, or for the major humanities fields in particular.

The *SEI 2000* report (Table 3-20) indicates the total numbers of jobs in nine science and engineering fields in 1998 and projected for 2008, as well as the change in expected numbers of positions over the ten-year period.

### 5. Public Awareness, Knowledge and Involvement in the Humanities

*How has public awareness of and interest in humanities issues changed over the last ten years? How does the interest of the American public in these issues compare with other countries?*

No time-series survey data are available on public awareness of or knowledge about issues of relevance to the humanities. In fact, there has been little or no systematic research to identify the issues of greatest relevance to humanities practitioners; instead, issues of public relevance are generally labeled as "political," "economic," or "social" in nature.

The *SEI 2000* report (Figure 8-1) displays a series of graphs for five surveys (1990, 1992, 1995, 1997, and 1999) showing levels of public interest and self-reported knowledge concerning eleven different scientific and technological issues, ranging from space exploration, through new scientific discoveries, to environmental pollution. Figure 8-2 uses 1999 survey data to show breakdowns of public interest and knowledge in science and technology issues by sex and educational attainment.

In addition, Table 8-1 of the report compares levels of public interest among U.S. citizens in the same eleven science and technology issues plus six other general issue areas (such as sports, politics, taxes, etc.) with levels of public interest measured in the European Union, Japan and Canada.

## CONCLUSIONS

Data resources focusing on the humanities are incapable of delivering policy-relevant information because of deficiencies in technical quality, coverage, comparability across disciplines or over time, and accessibility to researchers. Although the datasets listed in the *Index of Humanities Datasets* provide (or once provided) significant value for their producers and primary users, most

were not designed to support cultural policy research, and, in most cases, were not intended for use as research resources for quantitative analyses.

The lack of adequate data systems for policy analysis of the humanities does not result from any unwillingness on the part of humanities organizations and practitioners to gather and analyze data on their fields. On the contrary, one of the major problems affecting humanities information resources stems from the fact that too many parties with different purposes and orientations are collecting too much data, rather than not enough. The lack of articulation of effort, and in some cases the outright competition among data producers, has resulted in significant levels of duplication on the one hand, and inconsistency of measurement on the other, posing severe challenges for those who seek to combine these disparate resources into a coherent whole.

In consequence, the humanities, at present, lag substantially behind the science and engineering disciplines in developing the systems of data resources needed to monitor even the most basic trends in the number, activity levels, and productivity of, and the future prospects for, practitioners in humanities fields at all degree levels. Even in the science and engineering disciplines, despite the fact that they now possess more and better data resources, leaders remain dissatisfied with their ability to understand the state of their own fields and the roles those fields play both in academia and in the broader society. They continue to seek better data and methods to develop more accurate forecasts of supply and demand for science and engineering graduates.

In its recent report, the National Research Council's Committee on Methods of Forecasting Demand and Supply of Doctoral Scientists and Engineers recommended that the National Science Foundation support efforts to improve the quality of forecasting data and methodology by taking a leadership position in collecting and distributing improved data to researchers.<sup>5</sup> The Committee further recommended that NSF undertake a comprehensive review of data collection to support forecasting, and observed that data collected for forecasting should be openly available and distributed to researchers on a timely basis. Several specific data-gathering innovations were considered, including increases in the detail of the data compared to that collected in the *Survey of Doctorate Recipients*.

Significantly, Committee members, researchers, and policy makers agreed that existing data were not sufficiently up-to-date to support policy decisions. They also drew attention to the fact that existing data and analyses failed to serve the needs of several neglected user communities, especially those of students attempting to plan research careers and of the public and private organizations and associations that provide career guidance.

Leaders in humanities disciplines face a similar situation, but must begin their efforts from a less favorable starting position. To improve the capacity and infrastructure for policy research, humanities leaders must confront the same issues and adopt many of the same strategies now being advocated by the science and engineering fields to address similar questions. Two critically

5. National Research Council, *Forecasting Demand and Supply of Doctoral Scientists and Engineers: Report of a Workshop on Methodology*, National Academy Press, Washington, D.C. 2000.

important considerations should inform these efforts.

First, there is no one-time solution to the data resource problems described in this report. To succeed, the response must be based on a long-term perspective. Initial efforts toward the design of an infrastructure to support policy research on the humanities should be undertaken with the awareness that the development of data resources is an ongoing process that will require continuous investment of resources for the foreseeable future.

Second, as has been amply demonstrated in the science and engineering fields, sustained progress toward improving the infrastructure for policy research on the humanities must be placed within an administrative structure capable of pursuing funding for and managing the resources of an enterprise for the benefit of all stakeholders and constituencies. Without such an entity, reliance upon the self-interest of potential data users to spark organized efforts to address and surmount existing problems will leave the humanities data infrastructure in its current state indefinitely.

An initial review of available data resources in the arts and humanities was commissioned by the Academy in 1998.<sup>6</sup> Based on a review of the secondary literature, interviews with arts and humanities leaders, and an observational survey of organizations engaged in collecting data, the report summarized the initial efforts of major arts and humanities organizations to develop conceptual models, strategies, and organizational frameworks for data gathering, dissemination, and use to affect cultural policy.

While acknowledging the considerable developmental progress in recent years, DeNatale and Fong also summarized the obstacles to further advances and the extent of additional effort needed to produce a comprehensive system of cultural indicators. Although not a major focus of their report, the authors provided several pointed examples from the 1996 report (the twelfth in its series) of *Science and Engineering Indicators (SEI)* published by the National Science Board of the National Science Foundation (NSF 1996) as the best example of a currently operating model of policy relevant data, in this case focusing on science and engineering research and education in the United States, coupled with comparable data on other nations. Rather than describing a specific model to be emulated, the authors intended the examples from the *SEI* to illustrate both the feasibility and the potent value of systematic, detailed, time-series data on the condition of academic and professional fields.

DeNatale and Fong provided a realistic description of the complex, time-consuming tasks of consensus building and priority setting needed for further progress toward a system of cultural indicators, and summarized a variety of existing initiatives in government and non-profit organizations at all levels in both the arts and humanities. They observed that activity to develop new data resources needed to address national cultural policy issues “has taken place to a far greater extent among arts agencies and arts professionals than it has in the humanities.”<sup>7</sup>

Several examples indicated that consensus building and coordination efforts had already been initiated in the arts world, critical evaluations completed, and formal proposals produced and grants awarded to develop basic data structures, such as the *Universal Data Base (UDB)* of arts organizations, that would, in turn, support and spur additional systematic data gathering on arts activity in the United States. As of September 1998, no parallel examples on a similar, national scale existed in the humanities fields. The DeNatale and Fong report, followed by a series of planning meetings by the Academy task force, provided the basis for the current study.

In describing the growth of an orientation toward the value of standardized quantitative data, DeNatale and Fong identified two activities they considered critical movements toward building an infrastructure for policy-relevant data on the humanities: (1) A proposal by the Federation of State

6. DeNatale and Fong. 1998.

7. Ibid.

Humanities Councils (FSHC) for the development of a new set of standards and procedures for collecting key data elements from the 56 State Humanities Councils into a computerized database;<sup>8</sup> and (2) The existence of an “inventory” of over 100 datasets relevant to the condition of the humanities field, the *Index of Humanities Datasets (IHD)*, originally created for the American Council of Learned Societies (ACLS) by the Modern Language Association of America (MLA) with funding from the Andrew W. Mellon Foundation. Until recently the Index was available on the website of the National Endowment for the Humanities, and a list of the resources included in the index, and discussed in this report, is reproduced as *Appendix A*.

Although the information system proposed by FSHC is both an ambitious and a much-needed effort to improve the state of data on state and local humanities activities, its focus would be limited to monitoring the flows of resources through state councils and ultimately to monitoring outputs primarily, if not exclusively, in the public humanities. Even if the proposed data collection were to be realized through the FSHC initiative, however, it would not answer questions about such basic issues as the conditions and trends for the humanities labor force. To support reliable research on these issues, other types of data are required.

As it began this study, the Academy decided to take as its starting point the overview of existing resources found in the *Index of Humanities Datasets* mentioned earlier. *The Index of Humanities Datasets (IHD)* provided a logical starting point for this study for several reasons. First, it provided structured, descriptive information about a diverse collection of resources, including bibliographic data on scholarly books and articles (with some coverage of film); administrative and research data collected by humanities organizations; directories and catalogs of educational institutions, departments, and programs of study; national and special-focus surveys of institutions and individuals; and compendia of statistical tabulations. To be included in the list, a database had to include information about one or more “primary” humanities fields<sup>9</sup> and be currently available to researchers in electronic or print form. For each listed dataset, the index provided information about the sponsor or owner of the resource (with contact information), the format in which it is available, whether historical or time series data are available, and indicators of the contents of the dataset and the humanities fields covered by the data.

As DeNatale and Fong point out, the *IHD* was strictly a resource listing, with no assessment of data quality or utility, and even when listed on the

8. The FSHC initiative cited by DeNatale is a follow-on to efforts of the National Information Management Project (NIMP) initiated in the late 1980s but later stalled in implementation due to lack of financial support. The preliminary system, called Philodrudge (PHD), was produced through a collaboration between the Tennessee, Alabama, Kentucky and Missouri Humanities Councils.

9. “Primary” humanities fields were identified by NEH as history, English language and literature, foreign languages and literatures, art and art history, philosophy, religion, music and theatre, comparative literature, linguistics, American studies, and classics. Social science fields “with humanistic dimensions” include anthropology, law, political science, psychology, and sociology, and are also included in the *IHD*.

NEH website, no attempt was made to provide hypertext links to the data resources. The authors also noted that many of the resources were of little or no relevance for humanities or cultural policy.

#### PURPOSE OF THIS REPORT

In their 1998 study for the Academy, DeNatale and Fong were not charged with evaluating the quality or prospective policy value of any of the data resources in the *Index of Humanities Datasets*. As they noted, “. . . the inventory seeks to address the broadest range of research needs, and is not primarily oriented to policy needs. Nonetheless, a number of the data resources listed do have relevance to cultural policy discussions.”<sup>10</sup> Moreover, the *IHD* was highlighted because it represented a comprehensive effort to provide an overview of existing datasets and other resources. Its sponsorship by some of the leading national humanities organizations also suggested that its selection of resources represented the best thinking of the field at the time of its compilation. Consequently, the Academy requested a limited qualitative evaluation of the resources listed in the *IHD* (including bibliographic resources) for their ability to contribute to research on the critical questions of national humanities and cultural policy.

#### SPECIFIC EVALUATION OBJECTIVES

The intent of this assessment is to provide:

- A *functional classification of the types of data* included among the resources listed in the *IHD*, distinguishing the publicly available data sets designed for research from other types of information, such as proprietary or commercial data files, directories and catalogs, bibliographic compilations and indexes, and general-purpose publications, and a discussion of the implications for use of the data resources for policy research;
- A qualitative review of the utility of all data resources listed in the *IHD* to address critical human capital questions concerning the humanities, such as:
  - the number of undergraduates majoring in humanities fields each year at both two-year and four-year colleges and universities;
  - the number of graduate students enrolled in humanities programs each year;
  - the number and configuration of departments and programs in humanities fields enrolling students at all levels of postsecondary education;
  - the number of BA, MA, and Ph.D. degrees awarded annually in humanities fields;
  - the numbers of humanities centers, institutes, and other research organizations and the numbers of faculty and other staff employed or engaged;

10. DeNatale and Fong 1998: 34

- the number of faculty teaching at the secondary and postsecondary levels in humanities fields;
  - the types and numbers of libraries serving postsecondary institutions, research organizations, and the general public;
  - the numbers of humanities practitioners employed outside of academe, in what capacity they are employed, and labor force trends in the public humanities;
  - whether elementary and secondary school teachers of humanities subjects have received adequate and appropriate training to teach even the most broadly distributed subjects such as English, history, and foreign languages;
  - the annual level of scholarly activity of humanities scholars, faculty, and practitioners in terms of the numbers, forms, and content of publications in the humanities fields;
  - annual amounts of financial support from government and other public and private sources for research in humanities fields; and
  - levels of public participation in humanities-related activities, such as library membership, museum attendance, and contributions to humanities organizations.
- A description of the primary obstacles to using existing data sources to address the key questions about the state of the humanities, including such issues as:
    - inconsistency in definitions, measurement approaches, analytical methods, and reporting standards used by different producers and users of data and in published reports to identify humanities fields (and hence the students, educators and practitioners in the fields);
    - the prevalence of research and/or database designs that intend to collect data on narrow subsets of humanities fields, organizations, and practitioners (for example, association membership lists), rather than representative samples of all fields;
    - limitations of nearly all national, general purpose surveys of education and labor force participation that contain very small sample sizes for humanities students, faculty, practitioners and organizations, precluding study of separate humanities fields (and, in some cases, even preventing analysis of all humanities fields combined);
    - the absence of continuous and consistent measurement of key indicators over time, resulting from cancellations or cutbacks of data collection programs, irregular schedules for existing studies, or changes in research designs to serve ad hoc purposes;
    - restrictions in the availability to researchers of raw data that were collected for proprietary purposes;
    - problems with timeliness of large, complex measurement systems, with release of data and published reports frequently delayed for

- two or more years after data are collected;
- the need for burdensome data-preparation work to condition samples, variables, and complex indicators before many existing data systems can be used in research on the state of the humanities.

#### EVALUATION CRITERIA

Key criteria for evaluating the quality and utility of data for policy research in the humanities and culture were enumerated in a 1996 review of data available for parallel research in the arts world.<sup>11</sup> As described in that report, to be worth collecting and using, systems of data must be:

- valid and reliable (sample sizes adequate for subgroup analyses, use of rigorous data collection methods, etc.),
- comparable across organizations (consistent, appropriate standards of measurement),
- representative of the full range of individuals, organizations, activities or other output measures in all relevant fields,
- comparable (that is, regularly and consistently measured) over time,
- focused on matters that are important to managers and policy makers in all fields, and
- widely accessible and widely used for management and policy purposes.

To this list, we would add a few considerations focusing both on the characteristics of the datasets themselves, and on the behavior of data-producing organizations. To maximize utility for humanities and cultural policy research, those engaged in data-gathering activities in the humanities should:

- participate in or collaborate with efforts to develop policies and coordinate standardization of data elements and measures, data collection methods, data reporting and dissemination, and data sharing across multiple data producers;
- use advanced data design principles that permit frequent innovation in data programs while simultaneously preserving comparability of measurement across time periods;
- take advantage of emerging computing, networking, and communications technologies for data collection, reporting and dissemination that offer expanded technical capabilities and greater ease of use for data collectors, suppliers and users, while lowering the costs of surmounting existing obstacles to combining datasets produced by different sponsors;
- collect and distribute data on a continuous, regular basis in order to support the analysis of results over time and the factors that influence change in the condition of humanities organizations, practitioners and activities;

11. Kaple, Deborah, Lori Morris, Ziggy Rivkin-Fish, and Paul DiMaggio. 1996. "Data on Arts Organizations: A Review and Needs Assessment, with Design Implications." Center for Arts and Cultural Policy Studies, Princeton University. This analysis was supported primarily by Cooperative Agreement NEA DCA-96-53, with additional funding from the Nathan Cummings Foundation and the Andrew W. Mellon Foundation.

- allocate a significant share of total data production resources to constantly increasing dissemination of findings and expanding access by researchers and other users to data; and
- incorporate efforts at continuous quality improvement and continuous re-invention, driven by the requirements and preferences of customers of the data products at all levels of use.

These twelve criteria capture most of the general ideals that might form the basis of a dynamic, forward-looking system of data indicators for policy research in the humanities. They are not equally applicable to each of the datasets listed in the *IHD*, but will be applied to functionally similar groupings of data resources as appropriate.

SECTION 2. PROPERTIES OF DATA RESOURCES LISTED  
IN THE INDEX OF HUMANITIES DATASETS

FUNCTIONAL CLASSIFICATION

By “functional classification” we mean the purpose or function that a dataset was originally created to serve. As indicated in Exhibit 1, the 108 data resources<sup>12</sup> compiled by ACLS and MLA consisted of four primary types: (1) bibliographic materials, (2) directories and catalogs, (3) research or administrative datasets, and (4) publications or reports.

For some of the data resources, the functional classification is a “distinction without a difference.” For example, if a dataset classified as a directory or catalog is available to researchers in raw form—such as annual directories of college and university programs in humanities fields, or directories of museums or libraries—these data could be loaded into standard statistical analysis software packages (for example, SAS or SPSS) to permit analyses of course enrollments or majors declared for humanities fields over time, or changes over several decades in acquisition expenditures for humanities-related materials by libraries, or trends in annual counts of visitors to museums.

Unfortunately, as discussed further below, the primary obstacles to research use of datasets that were initially created as Directories or Catalogs is that they are rarely available in a form that satisfies the criteria for research uses listed above. Functional Classification, in fact, is the single best indicator of the likely utility of a dataset for policy research on humanities and culture.

Exhibit 1 shows that items in the first three categories comprise all but a few of the resources listed, and that the number of datasets in each of these three groupings is nearly equal (34 to 36). As described further below, the Publications and Reports classification includes only three items, two of which are recurring volumes of educational statistics and indicators from the National Center for Education Statistics (NCES)—the annual releases of the *Condition of Education* and the *Digest of Education Statistics*—and the third is the series of annual reports from the 56 State Humanities Councils submitted

EXHIBIT 1: COUNTS OF DATA RESOURCES BY FUNCTIONAL CLASSIFICATION

TYPE OF RESOURCE	COUNT
BIBLIOGRAPHIC MATERIALS	35
DIRECTORIES AND CATALOGS	36
RESEARCH DATASETS	34
PUBLICATIONS AND REPORTS	3
TOTAL	108

12. Appendix A actually lists 109 datasets. However, the hypertext link to the dataset entitled *NASULGC Faculty Distribution Survey by Race and Sex* brings up the same page that is linked to the subsequent entry, the *NASULGC Faculty Salary Survey*. We found no evidence either at NASULGC or at Oklahoma State University, the host site for the *NASULGC Faculty Salary Survey*, that a separate “Faculty Distribution Survey” exists. In fact, this item appears to be a single report based on data from the *Faculty Salary Survey*.

to the National Endowment for the Humanities. These resources are sufficiently unique to warrant separate treatment in this evaluation. Appendix C to this report lists the 108 datasets sorted by functional classification. The following sections discuss each of the four groupings in greater detail.

### *1. Bibliographic Materials*

Bibliographic datasets are of great potential value for policy research on the humanities and culture because they are such a rich, detailed record of the output of humanities scholars and practitioners. In terms of the scope of interest defined by each resource, nearly all the databases listed are comprehensive, sophisticated, historically inclusive, highly accurate and reliable, and internally consistent; they offer coverage of both academic and public humanities and are, in virtually all other respects, outstanding examples of data gathering, management, and distribution to users. U.S. bibliographic resources are among the finest in the world and are considered key resources by libraries and researchers the world over.

The Bibliographic Materials group includes only those resources that consist of data records about publications (books, articles, reviews, monographs, dissertations, etc.), and in a few cases about collections or archives of published material, or about films. It is nevertheless a diverse collection that includes such resources as the *Arts and Humanities Citation Index*; *Current Contents*; the Educational Resources Information Center (ERIC); *Historical Abstracts*; *Newsletters in Print*; and *Women's Studies Abstracts*. Also included in this group are systems that include document distribution services (electronic or print) in addition to bibliographic information retrieval.

The bibliographic data systems listed in the *IHD* vary enormously in size and comprehensiveness, ranging from *Magill's Cinema Surveys*, with approximately 500 data records, and the *Music Article Guide*, with 350 articles, at the small end, to the Library of Congress' *MARC (Machine Readable Cataloging)* system with over 4.5 million publication records, *UnCover*, with over 5 million records, the Online Computer Library Center's *OCLC Online Union Catalog*, with approximately 33 million records, and the *Research Libraries Information Network (RLIN)* system, maintaining over 78 million data records.

In theory, bibliographic databases offer the raw material for detailed analyses of both the quantitative and qualitative aspects of published work in humanities fields in the United States, trending back over many decades, and in some cases centuries, of work. In theory, they offer the possibility of linking to other data systems, such as educational institution records, expenditure data, subscription and readership data, and so on, in order to build a system of indicators that connects a wide range of inputs into the labor force and organizational structure of humanities endeavors to at least one important measure (published work) of outputs from humanities professions. In theory, these massive data systems for tracking the production of intellectual capital could serve as key anchors in the organization of humanities policy

research data.

Unfortunately, what seems reasonable in theory must overcome daunting obstacles to become feasible in practice. With a few exceptions, the primary use of these databases is to locate bibliographic data, often including abstracts, and in some cases to view the materials in electronic form or order delivery of printed copies. The typical use of these resources is to generate responses to user queries (search commands) by retrieving one or more data records that meet the user's search criteria. In general, these data systems were not designed for (and consequently are rarely used for) any type of quantitative or statistical analysis of their contents. Most of the listed systems employ sophisticated, computerized text-search engines to carry out their searches. A significant number fail to code their data records with content keywords or other classifiers of the material (even those provided in the original publication) that might permit reliable identification of the humanities field(s) represented by the work. Several resources that do use keyword classifiers do not distinguish works in the arts and humanities, and lump both together in a single category.

The bibliographic datasets listed in the *IHD* are quite diverse in scope and content. As shown in Exhibit 2, the 35 resources are about evenly divided into two groups. We found that 16 of the datasets are large, general purpose bibliographic systems that cover the full range of publications and also include items related to humanities fields (for example, *Current Contents*, *Dissertation Abstracts*, and *Nexis*). The first section of Exhibit 3 lists these general bibliographic systems. Although the descriptive material in the *IHD* sometimes indicates the approximate proportion of data records that cover humanities fields, for the most part this ratio is unknown.

EXHIBIT 2: COUNTS OF BIBLIOGRAPHIC DATA RESOURCES BY TYPE

BIBLIOGRAPHIC MATERIALS	COUNT
GENERAL RESOURCES THAT INCLUDE ITEMS IN HUMANITIES FIELDS	16
RESOURCES THAT INCLUDE ONLY ITEMS IN HUMANITIES FIELDS (16 OF 19 FOCUS ON A SINGLE FIELD)	19
TOTAL	35

The remaining 19 resources focus entirely on humanities fields (*MLA Directory of Periodicals*, *Bibliography of the History of Art*, *RILM Abstracts of Music Literature*, etc.—see the second section of Exhibit 3 below). Even a cursory examination of the list of humanities bibliographic resources reveals that 16 of the 19 humanities-focused bibliographies focus on a single humanities field (e.g., art history, ethics, religion, philosophy). The remaining three are the *Arts and Humanities Citation Index* (which includes items in the arts, but is not a general purpose bibliographic system), and the two MLA resources that focus on English and the full range of foreign languages and literatures.

EXHIBIT 3: BIBLIOGRAPHIC MATERIAL INCLUDING ARCHIVES (35)

*General Resources that Include Humanities Materials (16)*

AAUP On-Line Catalog  
Archives USA  
Current Contents  
Directory of Archives and Manuscript Repositories  
Dissertation Abstracts International  
Educational Resources Information Center (ERIC)  
Expanded Academic Index  
Faxon Finder  
LCMARC Database  
Newsletters in Print  
Nexis  
OCLC Online Union Catalog  
RLIN  
Social Sciences Citation Index  
Ulrich's International Periodicals Directory  
UnCover

*Resources Explicitly Focusing on One or More Humanities Fields (19)*

MLA Directory of Periodicals  
MLA International Bibliography  
Arts and Humanities Citation Index  
ATLA Religion Database  
America - History and Life  
ARTbibliographies Modern  
Avery Index to Architectural Periodicals  
Bibliography of the History of Art  
Ethics Index  
George Eastman House Interactive Catalog  
Historical Abstracts  
Linguistics and Language Behavior Abstracts  
Magill's Cinema Surveys  
Music Article Guide  
New Testament Abstracts  
Old Testament Abstracts  
Philosopher's Index  
RILM Abstracts of Music Literature  
Women's Studies Abstracts

Evaluating the completeness of the compilation of bibliographic resources listed in the *IHD* is beyond the scope of this report.

The data resources in both of these groups vary considerably in size.

Because of the presence of such massive resources as *LC MARC*, *OCLC Online*, *RLIN*, and *UnCover*, the general-purpose bibliographies average nearly 8 million data elements (with an unknown fraction being applicable to the humanities), while the humanities-specific resources average about 110,000.

All but three of the resources in the Bibliographic Materials group (the *Directory of Archives and Manuscript Repositories*, the *Music Article Guide*, and *New Testament Abstracts*) are available to researchers in the form of a database that may be obtained on tape or diskettes (12 of the datasets), CD-ROM media (23 datasets), or online (29 datasets). Only six of the 35 datasets are available exclusively online.<sup>13</sup>

Despite the wide availability of these resources to researchers, the formats of bibliographic databases vary considerably, and a substantial number are not stored in a format that can be loaded into standard statistical analysis programs. In fact, many bibliographic databases are sold for use only with proprietary data management or query software, which is provided in a purchased or leased package, typically on CD-ROM media. For example, the Library of Congress sells CD-ROM versions of the *LC MARC* database, which is embedded in the proprietary Folio Views database management software. The functionality provided by Folio Views serves well the requirements of library professionals, especially cataloging staff; however, it is not designed to generate the types of output needed for policy research, such as counts by year of various types of published items in specific humanities fields or subfields, counts of volumes published by types of presses, and so on. By running repeated retrieval queries, researchers could keep manual counts of the numbers of "hits" for each query, and through that process build up the cell counts to populate a simple cross-tabulation that might be relevant to policy issues. However, using that approach would entail a massive loss in research productivity compared to using software expressly designed to create contingency tables or to calculate mean (average) counts of attributes across analytical subgroups.

The technology exists to convert bibliographic databases (or subsets of the data that would be most relevant for policy research) from proprietary formats into standard raw formats (typically, simple ASCII data records). However, converting the data in this way will require negotiating with the owners to obtain license rights for this type of use.

The six online bibliographic resources referred to above present somewhat greater obstacles for research use. By definition, online databases are accessed and managed with complex, proprietary software systems and cannot be converted to formats suitable for use with statistical analysis software. Moreover, costs of use, whether borne by libraries or other institutions or individual users, can be considerable, since charges are typically levied for connect-time to the resource and for each query performed on the database.

13. These are the *AAUP Online Catalog*; *Nexis*; the *OCLC Online Catalog*; *RLIN*; *UnCover*, and *Magill's Cinema Surveys*. All are available via the Internet using standard World Wide Web browsers as well as through more primitive telnet connections.

Given the enormous inefficiency of query and retrieval routines for calculating statistics, the use of these resources for policy research on the humanities and culture is, not surprisingly, rare to nonexistent.

Thus, while the collection of bibliographic datasets is a vast, historically rich source of information about a substantial fraction of the key outputs of the humanities enterprise, the fact that they were designed for an entirely different purpose imposes severe limitations on their value for the types of statistical policy analysis that is common in, for example, the science and engineering, health care, and education fields.

## 2. *Directories and Catalogs*

The Directories and Catalogs category consists of items that, from the perspective of the present evaluation, are frequently closer in purpose and typical use to the Bibliographic Materials category than to the Research Dataset category. That is, while the majority of these resources are created by collecting data directly from institutions, organizations, and individuals, the original purpose for creating the directories was to permit users to obtain and examine information one record at a time, or to compare information for small numbers of entities by visually reviewing all of the information collected for each. The original conception did not typically anticipate using the data for statistical or other policy research on academic or professional fields.

The 36 resources classified as Directories or Catalogs may be further subdivided into four groups based on the type of entities described in the dataset as shown in Exhibit 4.

EXHIBIT 4: COUNTS OF DIRECTORY AND CATALOG DATA RESOURCES BY TYPE

DIRECTORIES AND CATALOGS	COUNT
ACADEMIC INSTITUTIONS (DEPARTMENTS, PROGRAMS, PRESSES, ETC.)	20
FUNDING SOURCES	10
HUMANITIES ORGANIZATIONS (PRIMARILY NON-ACADEMIC)	4
INDIVIDUAL HUMANITIES PRACTITIONERS (PHILOSOPHERS)	2
TOTAL	36

Exhibit 5 lists the 36 datasets in the Directories and Catalogs category, sorted into the four groups listed in Exhibit 4.

EXHIBIT 5: DIRECTORIES AND CATALOGS (36)

### *Academic Departments, Programs, Presses (20)*

Association of American University Presses Directory  
College Blue Book

GRE/CGS Directory of Graduate Programs, Volume D  
Minority On-Line Information Service  
Peterson's College Database  
Peterson's Gradline  
APA Guide to Graduate Programs in the Classics  
Departmental Administrators in the Modern Languages  
Directory of Departments and Programs of Religious Studies in North  
America  
Directory of Graduate Programs in American Studies  
Directory of History Departments and Organizations  
Directory of Programs in Linguistics in the U.S. and Canada  
Doctoral Programs in Theatre Studies, Performance Studies, and Dance  
Folklife Sourcebook – A Directory of Folklife Resources in the United  
States  
Graduate Programs in Art History and the Visual Arts  
Guide to Ethnomusicology Programs in the U.S. and Canada  
Guide to Graduate Degree Programs in Architectural History  
Guide to Graduate Programs in Philosophy  
Guide to the Field of Folklore  
Guide to U.S. Graduate Programs in History of Science

*Directories of Funding (10)*

Annual Register of Grant Support  
Awards Almanac  
Corporate 500 – Corporate Philanthropy  
Directory of Grants in the Humanities  
Foundation Directory  
Foundation Grants Index  
Giving USA  
Grants on Disc  
Nonprofit Almanac – Dimensions of the Independent Sector  
Prospector's Choice

*Directories of Nonacademic Organizations (4)*

American Art Directory  
American Library Directory  
Directory of Historical Organizations in the United States and Canada  
Official Museum Directory

*Directories of Individuals (2)*

Directory of American Philosophers 1998–1999  
The Philosophers Phone and Email Directory 1998–1999

## *Directories of Academic Institutions*

Twenty of the 36 directories focus on academic institutions, including colleges and universities, academic departments, programs of study, or academically affiliated organizations such as university presses. Examples of resources in this group range from the relatively large *Peterson's Gradline*, containing profiles of more than 30,000 graduate and professional education programs offered by about 1,500 U.S. and Canadian postsecondary institutions, and the MLA's *Departmental Administrators in the Modern Languages*, which lists all 5,750 known English and foreign language departments in U.S. and Canadian schools, to such small datasets as the American Society for Theatre's *Directory of Doctoral Programs in Theatre Studies, Performance Studies, and Dance*, containing information on all 44 known departments in North America, and the Society of Architectural Historians' *Guide to Graduate Degree Programs in Architectural History*, with data on 39 academic departments.

In their review of available humanities data, DeNatale and Fong combined the resources that we have categorized here as Directories and Catalogs together with those we classified as Research Datasets into a descriptive table of available datasets.<sup>14</sup> At first glance, combining directories and research datasets (especially for directories of academic institutions) appears reasonable because of the substantial overlap in the specific data elements they contain (e.g., enrollees by department, faculty size, degrees awarded, amounts of financial aid provided, and so on). While directories and research datasets may thus appear to support the same sorts of analysis of the educational system and the “input” side of humanities and cultural policy, collapsing these types of resources together blurs an important distinction about the relative availability of the two categories of data for research uses. For example, of the 20 directories of Academic Institutions, only eight of the resources are made generally and fully available to researchers by their producers. These are [1] the *College Blue Book*, [2] the *Minority Online Information Service*, [3] *Peterson's College Database*, [4] *Peterson's Gradline*, [5] the *Directory of Graduate Programs in American Studies*, [6] the *Folklife Sourcebook*, [7] the *Guide to Graduate Programs in Architectural History*, and [8] the *Guide to US Programs in the History of Science*. Seven of the eight are available only online, and the eighth (the *College Blue Book*) is available only on CD-ROM.

Of the remaining twelve that are not generally and completely available to researchers, four resources can be made available in response to special requests from researchers for access to the data: the *APA Guide to Programs in the Classics*; *Departmental Administrators of the Modern Languages*; the *Directory of History Departments and Organizations*; and the *Guide to Graduate Programs in Philosophy*. The remaining eight “data” resources exist only in printed form. To use these resources for policy analysis would require reducing the publications or other printed materials to machine-readable form, which requires significant time and expense on the part of researchers and may also require negotiations with the data owners concerning license to

14. DeNatale and Fong 1998: Appendix A

convert the information in printed directories for this use.

Limitations on data availability also preclude rigorous evaluations of the quality of data in directories and catalogs. Most of these directories are compiled using censuses of the membership of professional organizations or societies—questionnaires completed by all units in a defined population.

However, only a few actually report response rates to the data collection-effort, such as the AAUP, the American Philological Association, the MLA, the Council of Societies for the Study of Religion, the American Historical Association, the Linguistic Society of America, and the History of Science Society, all of which report 90 to 95 percent participation in their data-gathering efforts.

Some larger efforts, such as ETS's GRE/CGS *Directory of Graduate Programs* (Volume D) which attempts to gather information about some 13,000 graduate programs, fare less well, with response rates typically of 83 percent, but falling to only 65 percent in 1995 (the last year reported). *IHD* listings for several academic directories do not display information that addresses population coverage and response bias issues. The dataset descriptions simply assert (without providing evidence) that these directories include all known elements in the population, and therefore bias “is not an issue.”

Another obstacle to policy analysis afflicts datasets in the Directories and Catalogs group, especially those focusing on academic institutions. Because the data systems were originally designed to support printing in annual publications, the data-maintenance routines typically involve over-writing data values for the last reporting period with new values for the current reporting period during each update—annually or on some alternative schedule. Consequently, the active datasets include only the current year values, and generally do not retain historical values for such data elements as faculty size, student enrollments, degrees awarded, financial aid conferred, or even programs and courses offered. To obtain historical data for time-series analysis for these datasets, in nearly all cases, researchers would either have to attempt to obtain prior years' data files (non-active, outdated files), or would have to abstract values from printed reports of the directory contents for prior years. Either of these approaches would require a sustained and costly effort to plan and execute the construction of a longitudinal or time-series data system that does not currently exist for any of these academic institution directories.

### *Directories of Funding Sources*

The *IHD* also contains a total of ten directories of grantmakers and other potential funding sources for work in humanities fields. Despite the small number, the list nevertheless includes diverse materials. At the largest scale is the highly general *Nonprofit Almanac*, containing some 171,000 data records on “the independent sector in American society,” only a small percentage of which are grant-making foundations. Of moderate size are both the *Foundation Directory* and *Prospector's Choice*, each of which includes 10,000 to 12,000 philanthropic foundations, most of which fund work and projects in

other than humanities fields.

The list also includes several directories with greater focus on humanities related fields, including the *Annual Register of Grant Support* (3,000 records), the *Awards Almanac* (1,500 records), and the *Directory of Grants in the Humanities* (3,500 records), all of which document grants in the full range of humanities disciplines. The remainder of the directories in this category list foundations and other funding organizations that support smaller subsets of humanities fields.

Of the ten directories of funding organizations, two datasets, the *Annual Register of Grant Support* and the *Awards Almanac*, are not fully available to researchers and a third, *Giving USA*, requires a special request from researchers to gain access. We note that *Giving USA* is the only survey of funding sources for which a response rate has been included in the *IHD* (50%), a rate low enough to create serious reservations about selection bias in the dataset. Otherwise these three resources are available only in print form.

The remaining seven directories are available to researchers through a variety of media, with *Corporate 500*, the *Foundation Directory*, and the *Foundation Grants Index* available only online, *Grants on Disc* only available on CD-ROM, the *Nonprofit Almanac* available only on tape, *Prospector's Choice* available on either tape/diskette or CD-ROM, and only the *Directory of Grants in the Humanities* available on tape, CD-ROM as well as online. Based on a cursory review of the web sites established by producers of these resources, it appears that very few of these information systems are available in a format that can be easily converted for use in statistical analysis software. Consequently, the use of these resources for research on trends in grant funding for humanities fields and subfields, if possible at all, would require a substantial investment of time and effort to convert them from systems designed for query and retrieval to systems that support statistical summarization and time-series plotting. The degree of effort normally associated with these conversion processes, and the possible legal barriers to doing so, is a plausible explanation for the lack of use of these datasets in humanities policy research.

Similar to the case for directories of academic institutions, the use of directories of funding organizations for time-series analyses is especially problematic. The typical data-maintenance design for information directories is to modify data records by replacing previously reported information with new data each time the system is updated (most update annually but some producers use a longer or irregular cycle). And as for the academic directories, in most cases the only way to obtain historical data is to build entirely new data records for grant awards and expenditures that add rather than replace new data for each annual update cycle. For many of the listed directories of funding organizations, historical time-series data on grant awards in humanities fields can only be obtained from sequences of printed reports, again adding substantially to the start-up costs of policy analysis of investments in humanities fields and activities.

### *Directories of Nonacademic Organizations*

As shown in Exhibit 5, four of the directories describe groups of organizations that are either entirely or primarily nonacademic. These include the *American Art Directory*, with about 7,000 listings of museums, libraries, schools, and corporations with art holdings, as well as affiliated museums, galleries, and libraries; the *American Library Directory*, with some 36,000 records on public, government, academic, and specialized libraries; the *Directory of Historical Organizations in the U.S. and Canada*, listing about 13,000 societies, museums, archives, and special collections; and the *Official Museum Directory*, which lists over 7,300 museums and cultural institutions in the United States.

Two of the four data resources in this group are fully available to researchers, *The American Art Directory* (tape and online) and the *American Library Directory* (tape, CD-ROM, and online). The *Directory of Historical Organizations* and the *Official Museum Directory* are available in print form only. The level of difficulty in using these datasets for research purposes must be considered moderate to high, since considerable conversion effort must be expended to create machine-readable datasets.

A significant quality and utility issue with this group of directories is the uncertainty about coverage of the populations they purport to describe. Of the four, only the *American Library Directory* is described as including “all available units” in its database; the other three are described as using no specific procedures to address or account for possible coverage bias.

Gleaning historical data from these databases will be problematic. Although the *American Library Directory* and the *Official Museum Directory* are updated annually, the *American Art Directory* is updated biennially, and the *Directory of Historical Organizations* is updated infrequently (the last update was for the 1990 release, and the next subsequent update is now in preparation). In addition, like all the other directories covered above, data-maintenance practices involve updates that replace old data with new, rather than adding information to preserve the historical sequence. Consequently, monitoring the characteristics of the listed organizations over time will require abstraction of historical data from printed volumes.

### *Directories of Individuals*

The final two datasets in the *IHD* under our Directories category are two listings of information about philosophers and their academic departments maintained by the Philosophy Documentation Center at Bowling Green University. The *Directory of American Philosophers* (now in the 20th edition for 2000–2001) is primarily an alphabetical listing of some 12,500 philosophers in the United States and Canada, with addresses, telephone numbers, email addresses, and personal web sites. In addition to individual listings, this directory also contains data records for nearly 1,900 university and college

departments and programs, assistantships, and fellowships, for 134 centers and institutes, 186 philosophical societies, 239 publishers, and 292 journals that publish scholarly work in philosophy. The *Philosophers' Phone and Email Directory*, 2000–2001 is offered as a lower-cost extract of information from the *Directory of American Philosophers*, and wholly duplicates the information on individual philosophers contained in that database, supplemented by additional entries culled from philosophical societies, web sites, and forms collected at philosophy conventions.

Both of these volumes are updated every two years. *IHD* descriptions claim that the directories are 75 percent complete. However, the Philosophy Documentation Center provides no evidence supporting this estimate. Neither of these resources is available as a dataset to researchers, but is published in print form only.

### 3. *Research Datasets*

The *Index of Humanities Datasets* identified 34 resources that are specifically designed as research databases to support administrative and policy research. Except where noted, these datasets are provided to users in raw data format suitable for loading into standard statistical analysis software, such as SAS and SPSS, and are accompanied by data dictionaries and documentation of collection methodology.

Typically, the quality of these datasets is enhanced through careful editing and “cleaning” of the data to resolve reporting errors or inconsistencies by using logical rules and/or by recontacting respondents to correct errors. Missing data rates for individual survey items are typically low; in some cases all missing data are eliminated by the process of “imputation” of values (estimating values for one data record based on empirical relationships in other records where the data are not missing). In addition to raw survey responses, some of the datasets—especially those produced by NCES—also include “derived” or “composite” variables—summary items based on raw survey items. Creation of composite or summary variables by the data producers saves researchers the time and effort of doing the task themselves, and produces useful summary measures that are consistent across research uses. Most of these data systems are based on probability samples selected from defined populations, although in several cases the population consists of the membership of an association, typically a subset of a larger population; a few are complete censuses of membership organizations.

Most frequently, data are collected using standardized questionnaires, typically self-administered paper questionnaires delivered and returned by mail. With the exceptions noted, most of the surveys listed here typically achieve reasonably high response rates, reducing the risk of selection bias in the database. Selection bias that remains in any of these 34 datasets usually results from population coverage bias—for example, a study that includes probability samples of students or faculty from a small, non-representative

group of postsecondary institutions rather than the entire population of institutions. In some instances we identify below, population coverage bias seems to be so severe that the datasets cannot be used alone to generate accurate population estimates. However, even datasets that cannot produce generalizable findings may be useful for studies of relationships among key policy variables, such as the impact of faculty salary levels or trends in grant funding on university department enrollments, degrees awarded, or publications produced in humanities fields.

As shown in Exhibit 6, the 34 research datasets fall naturally into four broad subgroups based on the populations of interest, including Academic Institutions (5 datasets); Libraries and Museums (7 datasets); Postsecondary Faculty (5 datasets); and Postsecondary Students (17 datasets). We have further divided the relatively large Student dataset category into four subgroups based on research design features, separating those based on samples selected from populations of postsecondary students (8 datasets), those that include postsecondary students as a subset of a sample drawn from a larger, more general population (4 datasets), those based primarily on collections of administrative records of postsecondary students such as school transcripts or standardized test scores (3 datasets), and those based on samples of Ph.D. recipients (2 datasets).

Many of the research datasets are based on complex, multi-stage sampling and data gathering designs. For example, the AACU *Curriculum Database* was created by selecting a representative sample of colleges and universities, then collecting transcript records and other administrative data on all 1991 graduates from the institutions sampled. Because the primary units of analysis are college graduates, we have classified this dataset under the category “Samples of Administrative Records of Postsecondary Students.” It should be kept in mind that a dataset of this kind may also be suitable for institution-level analysis, that is, by calculating summary statistics across all the graduates in each of the 81 participating institutions, postsecondary institutions could be made the units of analysis. We review the characteristics of each of these seven groupings below, focusing our attention on the primary units of analysis for which the dataset was initially created.

Exhibit 7 provides a complete list of the 34 Research Datasets sorted into the seven categories listed in Exhibit 6.

EXHIBIT 6: COUNTS OF RESEARCH DATA RESOURCES BY TYPE  
 RESEARCH DATASETS IN THE INDEX OF HUMANITIES DATASETS

RESEARCH DATASETS	COUNT
ACADEMIC INSTITUTIONS (SCHOOLS, DEPARTMENTS AND PROGRAMS)	5
LIBRARIES AND MUSEUMS	7
SAMPLES OF POSTSECONDARY FACULTY	5
SAMPLES OF POSTSECONDARY STUDENTS	7
SAMPLES OF ADMINISTRATIVE RECORDS OF POSTSECONDARY STUDENTS	4
SAMPLES THAT INCLUDE POSTSECONDARY STUDENTS	4
SAMPLES OF PH.D. RECIPIENTS	2
TOTAL	34

EXHIBIT 7: SORTED BY TYPE

*Academic Institutions (5)*

National Survey of Graduate Assistantships  
 Research-Doctorate Programs in the United States  
 IPEDS Completions Survey  
 Annual Survey of Colleges  
 Community College Curriculum Studies

*Libraries and Museums (7)*

Public Library Statistics  
 ACRL University Library Statistics  
 ARL Annual Salary Survey  
 ARL Statistics  
 Academic Library Survey  
 Museum Financial Information Survey  
 National Museum Survey of 1989

*Samples of Postsecondary Faculty (5)*

NASULGC Faculty Salary Survey  
 National Faculty Salary Survey by Discipline and Rank (CUPA)  
 National Study of Postsecondary Faculty  
 National Survey of Faculty (Carnegie)  
 American College Teacher

*Samples of Postsecondary Students (7)*

AAU/AGS Project for Research on Doctoral Education  
 Baccalaureate and Beyond Longitudinal Study  
 Beginning Postsecondary Student Study  
 CGS/GRE Survey of Graduate Enrollment

IIE Reports on Foreign Students  
National Postsecondary Student Aid Study  
American Freshman - National Norms

*Samples of Administrative Records of Postsecondary Students (4)*

AACU Curriculum Database  
Advanced Placement Summary Reports  
Characteristics of GRE Test-Takers  
College Bound Seniors: Profiles of SAT and Achievement Test Takers

*Samples that Include Postsecondary Students (4)*

National Longitudinal Study of the High School Class of 1972  
High School and Beyond  
National Education Longitudinal Study of 1988  
Survey of Income and Program Participation

*Samples or Censuses of Ph.D. Recipients (2)*

Humanities Doctorates in the United States (subset of the Survey of  
Doctorate Recipients dataset)  
Survey of Earned Doctorates (census of Ph.D. recipients)

## *Academic Institutions*

Five datasets contain data records on Academic Institutions. The *National Survey of Graduate Assistantships* is conducted and maintained by the Office of Research and Graduate Studies at the University of Nebraska-Lincoln; the datasets for *Research-Doctorate Programs in the United States* are maintained by the National Research Council; the *Integrated Postsecondary Education Data System (IPEDS) Completions Survey* is sponsored and managed by the National Center for Education Statistics (NCES); the *Annual Survey of Colleges* is conducted by the College Board; and *Community College Curriculum Studies* is maintained by the Center for the Study of Community Colleges, UCLA.

The *National Survey of Graduate Assistantships* has been conducted every two years from 1988–89 through 2000–01 and has collected data from all academic departments in over 100 universities on the numbers and characteristics of teaching assistantships available to graduate students. Information on graduate and postdoctoral fellowships was added in the 1990–91 wave. The participating universities are members of the Council of Graduate Schools (CGS). Over 3,800 academic departments participated in recent surveys. The response rate in the first year was low (54 percent), but in subsequent years has been in the 70- to 75-percent range. Four years of data (1991–92, 1993–94, 1995–96, and 1997–98) are currently available to researchers on diskettes (for ASCII files) or by email in the form of an Excel spreadsheet. The 1999–2000 dataset will be available in September 2000.<sup>15</sup>

The utility of this dataset for policy research is limited primarily by the relatively narrow focus of the data collected and is somewhat compromised by the non-probability sampling method. However, these data are still useful for comparing specified types of graduate-student support across humanities fields, and between humanities and non-humanities fields throughout the 1990s. If institution identifiers are included in the dataset, by matching these data records to other institutional datasets it may be possible to estimate relationships across fields of study between departmental student enrollments, progress, and degree awards on the one hand, and graduate student support via teaching assistantships, on the other.

*Research-Doctorate Programs in the United States* collects survey data from 3,634 academic departments in 274 postsecondary institutions that satisfied the National Research Council's eligibility criteria for inclusion—mainly that they must have awarded a minimum of 500 degrees in 50 fields during the period from 1986 to 1990. The survey was last conducted in 1993 as a follow-up to an initial survey completed in 1982. A wide array of policy-relevant data on departmental characteristics (highest degree awarded, areas of specialization, academic rank, etc.), aggregate faculty and student counts and characteristics (enrollments, degrees awarded, time to degree, etc.), and selected institutional characteristics (institutional type and level, type of funding/control, institutional expenditures, etc.) were collected in each of the surveys.

15. This updated information was provided by staff of the Office of Research and Graduate Studies of the University of Nebraska-Lincoln.

Virtually all humanities fields are covered separately in this dataset. The complete data files are available on tape or CD-ROM media to all researchers.

Although these surveys benefit from intelligent design and high data quality, like the *National Survey of Graduate Assistantships*, their policy uses are limited by their age (collected in 1993, most of the information collected summarized 1991–92 department characteristics). If a new survey were conducted using a replication of the research design employed by the two earlier surveys, the comparisons of 1982, 1992, and 2000 data on research-doctorate programs would be of exceptional value.

The *IPEDS Completions Survey* is conducted annually by NCES. The research design is a census of over 9,200 postsecondary institutions in the United States. The *Completions Survey* collects data about all levels of degrees awarded by postsecondary institutions and the characteristics of those degrees (such as level and field), the departments conferring them, and the individual awardees (sex, race/ethnicity, age). Data from the *Completions Survey* may be merged easily with other IPEDS components, such as the *Institutional Characteristics* survey, in order to expand the possibilities for policy analysis. Complete IPEDS data files are made available to researchers on both tape and CD-ROM media. Data on all fields of study, including humanities fields, are coded in great detail using the Classification of Instructional Programs taxonomy. Few other datasets contain as much detail on degree characteristics for humanities fields. However, by itself, IPEDS data files are suitable only for the analysis of entire postsecondary institutions and groups of institutions, as well as analyzing trends over time in national and regional data. Disaggregating IPEDS data to the academic department level, if feasible at all, would be an extremely labor-intensive step.

The primary limitation in the use of *IPEDS Completions* data for policy research is the relatively minor issue of the time between data collection and release of data to the public. NCES quality control and review procedures (internal and external) are painstaking and consequently time-consuming. Until this year, all IPEDS data were collected using paper forms distributed to participating institutions and returned for processing by mail. Data processing and cleaning requirements to minimize measurement error were therefore extensive. All participating institutions (as well as state coordinators of the IPEDS program) were given an opportunity to review and correct information in their data records after cleaning but prior to public release. The result was that the first published reports and the first data file releases typically occurred several years after each round of data collection. At the beginning of the academic year 2000/01, the most recent release of the *IPEDS Completions Survey* available to the public included data covering degrees awarded during the 1996–97 academic year. NCES is now in the process of implementing a completely revamped system for collecting all IPEDS data. The conversion of IPEDS to a Web-based collection and data management system is expected to reduce dramatically the response burden on participating institutions and also to reduce (to less than one year) the time interval from completion of data collection to the release of public data

files. This change will further improve the attractiveness of IPEDS data for policy research in the humanities and culture.

The College Board's *Annual Survey of Colleges* collects standardized information on a wide variety of institutional characteristics from over 3,000 colleges and universities. Annual surveys have been conducted continuously since 1983, with very high response rates. In addition to standard institutional data (level, type of control, degrees awarded, tuition and fees, admissions procedures, financial aid, etc.), the survey also collects data about academic departments (fields offered, counts of faculty, etc.) and students (enrollment counts, majors, time-to-degree, demographics, etc.).

Limited data are collected by fields of study, especially by specific humanities fields. Most of the data about students are collected at the institution level, or for broad field categories; humanities fields are combined into a single category in a manner determined by the College Board. Data from the *Annual Survey of Colleges* is used in the preparation of the College Board's publication, *Index of Majors and Graduate Degrees*, which provides information about some 600 fields of study at 3,000 institutions. However, this aspect of the dataset serves a function akin to the directories reviewed above, more than it supports policy research uses. The dataset is available on CD-ROM as a resource to purchasers of the annual *College Handbooks* published by the College Board. The raw dataset may also be made available to researchers in response to special requests. However, compared to other data resources, the *Annual Survey of Colleges* offers little unique information for policy research on humanities fields.

*Community College Curriculum Studies* is not a single dataset but is actually a diverse collection of studies carried out on an ad hoc basis in accordance with the research agenda of the Center for the Study of Community Colleges (CSCC) at UCLA. Although data have been collected for various purposes from 1975 through the present, they are not made available to the public, or to researchers via special requests. The topics covered are limited to institutional and departmental data on student enrollments, course offerings, and degree requirements. Fields of study covered in the surveys include virtually all humanities fields.

Unfortunately, little evidence exists to permit evaluation of either the quantity or the quality of the data (selection bias, coverage bias, response rates, data completeness and consistency, etc.). Several reports produced by the CSCC are directly relevant to key policy issues in the humanities, such as "Curricular Trends in Community Colleges: Implications for Transfer."<sup>16</sup> However, unless arrangements can be negotiated for these datasets to be accessed by other researchers, additional work using these resources will be restricted to issues selected by the CSCC at UCLA.

16. ERIC ED 354050

## *Libraries and Museums*

### LIBRARY DATASETS

The Libraries and Museums category includes seven datasets (see Exhibit 7), of which four pertain to university libraries, one describes public libraries, and two are focused on museums.

*Public Library Statistics* is an annual data collection program sponsored by NCES that collects 53 basic information items from the universe of about 9,000 public libraries on such topics as staffing, service locations, revenues and expenditures, sizes of holdings and collections, measures of services, and descriptions of the populations served. The first survey was conducted in 1988 and annual surveys have continued through the present. Response rates are extremely high (98–100 percent), eliminating selection bias from the dataset. *Public Library Statistics* is a unique, well-organized, valuable data resource for studying the characteristics and activities of public libraries across the nation. Data are measured consistently and continuously, supporting detailed analyses of trends in library conditions and operations. The entire series of datasets is available to the public on tape media. Data files and documentation for recent years may also be downloaded from the NCES web site (<http://nces.ed.gov/surveys/>) at no cost.

As for many other NCES datasets, the primary limitation for use in humanities policy research is the delay in public release of the data from each new round of collection. For example, the most current final release version of the *Public Library Survey* data file available today is for Fiscal Year 1996. However, in an effort to provide faster access to its customers, NCES has also released a preliminary version of the survey data for Fiscal Year 1998 on its web site. The preliminary release versions are updated on a flow basis during the data editing and cleaning process until the final version is declared ready.

*ACRL University Library Statistics*, managed by the Association of College and Research Libraries (ACRL), is a system of biennial surveys begun in 1979 and continued through the present time. The survey compiles operating data from over 105 participating libraries. Institutions surveyed are in Carnegie Classifications Research I and II and Doctoral Granting I and II, except those that are members of the Association of Research Libraries. Data reported include collection information, such as volumes, monographs, serials, and microforms; staffing levels; expenditures; and interlibrary loan activities. Institutional data include degrees offered, enrollment size, and faculty size. The most recent available data are for 1996–97. The survey datasets are available to researchers on tape/diskette media. ACRL has two additional data products of potential relevance, a database of *Academic Library Trends and Statistics*, based on a data compilation from 1,064 academic libraries carried out by the Survey Research Center at the University of Virginia, which covers institutions in all Carnegie Classifications. The availability of these data (other than in a printed report for 1998) is unknown at this time. The second is a database that combines ACRL and IPEDS library data into a single database. ACRL's commercial product is a CD-ROM that includes a software system (*Scholarstat Libraries*) designed to help librarians compare their libraries

to real or composite “peer institutions.” Whether ACRL will make this combined data system available in raw (ASCII) format is also unknown at this time.

ACRL products (with the possible exception of the database that includes IPEDS data) are limited in their utility for policy analysis primarily by their use of non-probability selection methods for institutions to include in the databases. The data can be analyzed in some detail for the population of interest, but statistical estimates from *ACRL University Library Statistics* cannot be generalized to all academic libraries. For the defined population, however, the dataset will support point estimates, trends over time, and assessments of relationships among key policy variables such as investments, expenditures, and service outcomes.

The Association of Research Libraries (ARL) *Annual Salary Survey* has since 1993 gathered salary and other core data (sex, race/ethnicity, position, years of employment, etc.) on approximately 12,000 professional staff in ARL’s 121 member libraries, with participation rates of 100 percent each year. Although the data provide the basis for detailed annual reports (which present point estimates as well as trends over time), the data are considered proprietary and are not released to researchers outside ARL. To expand the use of ARL Salary Survey data for policy research would require negotiating access to raw data with ARL, making arrangements with ARL to obtain specific types of analyses, or attempting to extract statistical data from printed reports (which may also require negotiation concerning fair use of the information).

The potential utility of ARL’s second major dataset, *ARL Statistics*, is considerably higher, since data from these annual surveys of ARL’s 121 member institutions are made available to researchers both via tape/diskette and on line. However, because the *ARL Statistics* dataset contains basic measures of library characteristics, such as holdings, acquisitions expenditures, staffing, reference transactions, interlibrary loans, and use of electronic media, as well as basic institutional characteristics, such as institutional enrollments, doctorate degrees awarded, etc., recent editions of this dataset are largely duplicative of other, more comprehensive data programs, such as the IPEDS *Academic Library Survey* (see below). However, the ARL dataset has the distinction that it has been collecting core data since 1906, and historical data are available from 1908 to the present, both on tape media and online (from a host facility at the University of Virginia). This feature, by itself, is enough to make the ARL Statistics dataset important for humanities policy research.

The *IPEDS Academic Library Survey* is conducted by the National Center for Education Statistics on a biennial basis. Part of the IPEDS data system since 1988, *Academic Library Surveys* have been conducted since 1966 through the present. Data are collected from postsecondary institutions on over 3,400 academic libraries and provide an overview of their status nationally and by state. The *Academic Library Survey* captures information for a combination of library and parent institution characteristics, with primary emphasis on library holdings, functions, and activity levels across a wide range of services. Response rates have ranged from 87 percent to 93 percent over the last few survey waves. Data are made available to the public via tape media and via

download from the NCES Website.

Data quality is very high in the *Academic Library Survey* system; missing data are virtually eliminated by recontacting participants or imputing values. However, data availability delays for this dataset appear to be even worse than for most other NCES products. Researchers may now download the final *Academic Library Survey* dataset for 1994 and the preliminary dataset for 1996. The shift from mail-out/mail-back collection methods to Web-based data collection and online editing is expected to reduce the time to availability very substantially. It is not yet clear whether this survey will convert from paper to online data collection in the 2000 or 2002 iteration.

#### MUSEUM DATASETS

The *National Museum Survey of 1989* is described as a “comprehensive baseline survey” based on a stratified sample of more than 3,000 museums and conducted by the American Association of Museums (AAM) in the stated year. However, documentation in the *IHD* indicates that this dataset contains 1,077 records and that the response rate was 70.4 percent (implying an initial sample size of 1,530 museums). We assume, therefore, that a sample frame containing over 3,000 museums in 1989 served as the universe for this survey. The survey covered a wide range of topics including such institutional characteristics as year of founding, organizational structure, location(s), sources of support, measures of size (staff, space, collections, visitors, etc.), connected facilities, and the like. Also covered were museum activities, such as research, teaching, and programs, as well as financial measures such as revenues and expenditures of various types. The data are available on tape/diskette media to researchers from AAM. However, they are somewhat out-of-date, and no comparison data are available for assessing trends over time.

The *Museum Financial Information Survey* is conducted annually by the AAM with random samples of just over 800 museums selected from a universe that currently includes over 3,600 institutions. The dataset includes many items similar to the 1989 “baseline” survey, but also includes many more financial measures, ranging from admission fees to endowments, funding, overall operating income (broken down by source), investments, operating expenses, surpluses/deficits, capital expenditures, and so on. Unfortunately, data from the *Museum Financial Information Surveys* are proprietary to the AAM and are not currently released to the public in any form, including printed reports. At present, AAM is the only organization with access to and capable of undertaking policy analyses of these data.

#### *Samples of Postsecondary Faculty*

The *Index of Humanities Datasets* lists six items relevant to postsecondary faculty. However, one of the items, the *NASULGC Faculty Distribution Survey by Race and Sex*, is not a dataset, but is the printed report from the *NASULGC Faculty Salary Survey*. Consequently, we examine only the latter

dataset here.<sup>17</sup>

The NASULGC *Faculty Salary Survey* collects annual salary information about faculty members in three academic ranks on a biennial basis from member institutions of the National Association of State Universities and Land-Grant Colleges that award doctoral degrees in at least five fields. *Faculty Salary Surveys* have been conducted from 1983 to the present, and a database containing “odd-year” data for the entire time span is available to researchers.

There are several apparent limitations to this dataset for humanities policy research. First, the universe of institutions employing the faculty about whom salary data are collected is defined as the 100 members of the NASULGC. While data for these institutions may be a useful indicator, they cannot be used to generalize to salary trends for other institutional sectors. Second, although the teaching/research fields covered by the survey include all humanities fields, the specific fields for which salary data are collected vary from year to year, so that the continuity of data for a given field will vary considerably, making it difficult to compare humanities fields and to calculate an overall average across humanities fields for a particular year. Finally, there are some indications of quality problems with this dataset. *IHD* documentation indicates that there are 49 institutions (providing data on some 50,000 faculty) included in the datafile, and that the several surveys completed during the 1990s have had response rates averaging 50 percent.<sup>18</sup> This implies a total population of about 98 institutions; NASULGC currently has 208 member institutions. However, it is possible that only about a hundred of these are large enough to award doctoral degrees in at least five fields. In any event, prior to using these data for policy research in the humanities, considerable further effort should be made to identify the nonresponse patterns to the surveys and to what degree nonresponse bias will compromise the accuracy of results.

The *National Faculty Salary Survey by Discipline and Rank* is conducted annually by the College and University Personnel Association. It collects summary data on faculty salaries broken down by academic rank and field, and includes virtually all humanities fields. The Salary Survey has been conducted since 1982 from a sample of four-year public and private colleges and universities. Data for 914 institutions, representing over 177,000 faculty members, are included in the current data file.

The primary potential limitations for policy research of this data system are known coverage bias (the frame excludes most land-grant universities) and the fact that the dataset is not made available to researchers outside CUPA (although printed reports are released each year). To be useful for policy analysis, annual salary data would have to be linked to institutional char-

17. This accounts for the fact that the *IHD* lists a total of 109 datasets, but we have examined only 108 in this report.

18. Although it appears that the 50 percent completion rates are for institutions, it is not clear what these results imply for the coverage and completion rates for faculty employed at the participating institutions. On the one hand, the 50 percent of the institutions that participate may employ a larger share of all faculty employed by member institutions; on the other, it is equally possible that the participating institutions employ a below average number of faculty.

acteristics; however, there are no evident provisions for this in existing documentation. In addition, no response rate information is provided in the descriptive material on this dataset, an issue that requires further scrutiny.

NCES has conducted the *National Study of Postsecondary Faculty* (NSOPF) three times (1988, 1993, and 1999). However, the sample and data collection designs for the 1988 and 1993 waves differed substantially. The 1988 round included three levels of sampling—postsecondary institutions (424), department chairs (about 2,400), and faculty (about 8,400). The 1993 survey included only two levels of sampling—postsecondary institutions (over 800), and faculty within institutions (nearly 26,000). Moreover, while the 1988 survey sampled faculty classified as “instructional faculty” by sampled institutions, the 1993 survey sampled all faculty, without regard to primary classification of duties. NCES also tailored the design of the 1993 NSOPF survey to accommodate sample supplements funded by the National Science Foundation and the National Endowment for the Humanities. NEH supplements were also included in the 1988 and 1999 surveys.

Despite the methodological differences noted above, the NSOPF surveys are among the best available data on postsecondary faculty available to researchers. Data are collected about faculty demographics, educational history, academic rank, tenure, field, teaching loads, institutional responsibilities, salaries and benefits, attitudes and values, job satisfaction, and several other topics. Faculty data can be coupled with data collected via the institutional questionnaire for more detailed breakdowns. Compared to other postsecondary faculty surveys listed in the *IHD*, they are less subject to coverage bias, and survey response rates are significantly higher for the NCES surveys (about 90 percent for the institutional questionnaire and 75 to 82 percent for the faculty questionnaire), reducing possible error from nonresponse bias.

There are three limitations to the NSOPF datasets worth noting. First, because of differences in the approach to sampling faculty noted above, the 1988 and 1993 datasets are not strictly comparable. However, the precise differences in statistical estimates that might be due to the methodological changes between the two studies have not been formally documented. Informal checks carried out by one of NCES’s contractors for an analysis of 1988–1993 trends in faculty data uniformly confirmed that the 1993 results were closely comparable to those from other contemporaneous surveys that used sampling methods resembling those of the 1988 survey. Compared to other data problems identified in this review, this issue seems relatively minor. A second issue is the substantial difference in sample sizes between the 1988 and 1993 studies. Estimates from the 1988 data will have larger sampling variances than those for 1993, particularly for the analysis of faculty in specific fields. For example, while the 1993 study collected data from nearly 4,000 faculty in humanities fields, the 1988 survey included only about 1,500 faculty members across all humanities fields. This relatively small sample size would limit the number of subgroups (for example, tenure status by sex for each humanities field) that could be analyzed separately with acceptable accuracy. Finally, as mentioned for other NCES datasets, the intensive review and

quality assurance procedures impose significant time delays between data collection and data release. For example, the datasets and users manuals for NSOPF 1993 were not released until December 1997. Delays of this magnitude reduce the value of the otherwise high-quality data resources for policy research in the humanities and culture.

The Carnegie Foundation has sponsored four *National Surveys of Faculty*, with surveys completed in 1969, 1975, 1984, and 1989. The stratified samples for each survey were modest in size (about 5,500 cases in the most recent data file) but attempted to cover a variety of disciplines. Topics covered in the questionnaire are similar but not identical to those used in the NCES surveys, and include demographics, teaching experience, working conditions and characteristics of their institution and department, research/teaching loads, salaries and benefits, attitudes and values, and related topics. Datasets are available to the researchers for the last three surveys conducted (1975, 1984, and 1989).

Three primary limitations affect the value of these data for policy research. First, the most recent data are over ten years old. With the advent of the NCES NSOPF system, it is unlikely that any more surveys will be conducted by Carnegie. Second, the sample sizes are too small to support separate analyses of humanities fields. In fact, the documentation indicates that the dataset contains only a broad indicator that the faculty member teaches in a humanities field, but does not specify which field. Third, response rates for the last three surveys have been between 50 percent and 55 percent; the risk of significant nonresponse bias is therefore much higher than for the NCES surveys.

The final faculty data resource is the *American College Teacher* dataset produced by the Higher Education Research Institute at UCLA's Graduate School of Education. Surveys are conducted under this program on an irregular basis "every few years." The data are collected from faculty at a self-selected group of public and private 2- and 4-year institutions that participate in the Cooperative Institutional Research Program (currently about 400 institutions), not from probability samples. Almost 34,000 faculty members appear in the most current data file. The data producers have generated weights for data records to match them to known population totals (NCES data). Surveys have been completed in 1989–90, 1992–93, 1995–96, and 1998–99. Topics covered include most of the same types of topics addressed in the other faculty surveys reviewed here: demographics, institutional and department characteristics, experience, tenure, academic rank and field, teaching/research responsibilities, perceptions, attitudes, values, etc.

Three significant limitations affect the value for policy research in the humanities of the *CIRP Faculty Surveys*. First, by design, the data from CIRP surveys are made available only to researchers at collaborating institutions. Second, despite the self-selected nature of the institutional participants, response rates for the faculty surveys have been relatively low, ranging from a high of 61 percent in 1992–93 to a low of 42 percent in 1995–96, raising the prospect of significant nonresponse bias. Third and most important, the fac-

ulty surveys are based on non-probability samples of institutions. Despite the strategy for weighting the data to approximate the distributions on key variables in the entire faculty population, and despite the fact that hundreds of percentage tables are included in published reports, even the data producers tend to avoid quantitative descriptive or analytical statements in their summaries of findings, preferring qualitative language, such as, “Regardless of tenure status, faculty of all ranks have become increasingly supportive of tenure in recent years. Not surprisingly, the strongest advocates for the tenure system are the most highly ranked tenured faculty.”<sup>19</sup>

### *Samples of Postsecondary Students*

High-quality, continuous measurement of the numbers and diversity of postsecondary students enrolled in humanities fields at the undergraduate and graduate levels is essential primarily to support research both on the demand for academic teaching staff to develop and deliver programs of study and also on the potential future supply of humanities graduates and practitioners. We describe here eight datasets that were designed specifically around samples of postsecondary students—some covering the undergraduate population only, some focusing only on graduate students, and the remainder including both levels.

The eight data systems vary considerably in design features—including one-time, cross-sectional surveys, continuing time-series surveys, and longitudinal surveys of various types of student cohorts; the quantity and richness of the types of data they collect; sample sizes; and both continuity and recentness of measurement. Although several are available to researchers, others remain proprietary to the sponsoring organization.

Although all eight resources include students enrolled or majoring in humanities fields, none of the eight was specifically designed for research on humanities students. As a result, for several of the datasets, the subsamples of humanities students are too small to be of practical value for policy studies. Moreover, several of the datasets described here do not attempt to cover the entire population of postsecondary students in the United States, and consequently cannot yield generalizable statistics on student characteristics, behaviors, attitudes, educational outcomes, or career activities.

The *AAU/AGS Project for Research on Doctoral Education* focuses on graduate education in selected fields of study (biochemistry, economics, English, mathematics, and mechanical engineering from 1989 through 1997, and chemical engineering, history, physics, political science, and psychology from 1992 through 1997) at a self-selected group of institutions that are members of the Association of Graduate Schools (AGS) in the Association of American Universities (AAU).<sup>20</sup> All students enrolled in graduate programs in the specified field/year combinations were included in the study at each

19. From the Summary section of “An Overview of the 1998–99 Faculty Norms,” posted on the CIRP web site at <http://www.gseis.ucla.edu/heri/cirp.htm>.

20. Although the project is open to all AGS members, only a limited number of the 62 member institutions contribute “annual project dues” and staff time to the conduct of an annual data collection. The number of participating institutions has varied from 41 in 1991 to 32 in 1994 and only 21 institutions in 1997.

participating institution. Data were assembled from computerized student-level administrative records maintained by the participating institutions—students were not surveyed directly.

The AAU/AGS Project for Research on Doctoral Education covers several topics of relevance to local and national policy concerning the supply and demand of Ph.D.'s in the selected fields, such as results of admissions processes, academic talent of applicants that are admitted/rejected, amounts and kinds of student financial assistance offered, and factors affecting program persistence and completion. However, while the topics covered by the data collection vary over the years, the survey fails to address the full range of issues each year, in either data collection or analytical reports. The project decided to stop data collection on new student cohorts as of 1997—with the exception that longitudinal data will continue to be collected on cohorts already in the system. Data records exist in the system for approximately 20,000 individual graduate students and are available to researchers on tape/diskette or CD-ROM.

The primary limitations of the AAU/AGS dataset for humanities policy research are methodological in nature. In particular, coverage bias is inherent in the focus only on AAU/AGS members and the incorporation of data on the basis of self-selection by member institutions, and an unknown but probably substantial nonresponse bias due to declining participation by member institutions over time (only 21 of 62 institutions participated in the 1997 round). In addition, the sketchy documentation available at the project's web site (<http://www.aau.edu/agsindex.html>) does not reveal whether or which topics are covered consistently in time-series measurement, or the percentage of the institutions or students that have longitudinal data available. Further, little information is available about data-collection practices used by participating institutions. The project web site includes a single technical report from 1994 prepared by staff and students of the University of California at San Diego about the procedures used to compile and transmit data by that institution; there is neither a direct statement nor an implication that the same procedures were used by any other participating institution—raising questions about variability in data quality. Possibly as a consequence of these methodological concerns, the publication stream based on AAU/AGS data is limited to very brief annual reports on the ten disciplines covered, and neither AAU/AGS project staff nor independent researchers have made frequent or intensive use of these data.

Sponsored by NCES, the initial *Baccalaureate and Beyond* (B&B) survey is a longitudinal study of a sample of college graduates in 1992–93. It replaced the Recent College Graduates cross-sectional studies. The baseline B&B student cohort was identified in the 1993 round of the *National Postsecondary Student Aid Study* (NPSAS), a nationally representative sample of all postsecondary students described in more detail below. This cohort will be resurveyed biennially for 12 years: the first follow-up occurred in 1994, the second follow-up in 1997, and the third is scheduled for 2000.

The B&B sample design retains nearly all of the college graduates in the 1992–93 academic year identified in the NPSAS:93 cross-sectional survey ( $n=10,080$ ) and is generalizable to the entire population of 1993 college graduates. The B&B dataset includes a great deal of baseline data collected through self-administered questionnaires and records abstraction, focusing on respondents' postsecondary institution, student financial assistance, academic field and performance, educational and occupational goals, and detailed demographics. In addition, detailed information was collected via telephone surveys from parents of most of the B&B cohort members about financing postsecondary schooling. Course-taking patterns and other academic information were extracted from college transcripts. Follow-up surveys will collect longitudinal data by telephone surveys on labor force participation, graduate and professional school attendance and completion, career development, family formation, and related issues. Baseline and follow-up survey completion rates have been high and the files have very little missing data on survey measures. Documentation on the project design and execution is extensive and user-friendly. The files are usually readily available on tape and CD-ROM to the public for secondary analysis.<sup>21</sup> Public release data files take approximately two years for preparation and review before they are distributed outside NCES.

The B&B design is well suited to tracking longitudinal outcomes for college graduates, including comparisons across students majoring in different fields. However, the baseline B&B sample of just over 10,000 individual graduates contains only just over 1,000 cases who earned degrees in any humanities field. Of this total, about 230 received degrees in Letters, English and American, just under 190 received undergraduate degrees in History, and about 125 received degrees in "Liberal Studies." Only about 35 to 50 sample members have degrees in Philosophy or Religion. Moreover, as of the second follow-up (1997) only about 25 percent of the sample (just over 2,500 individuals) had enrolled in any form of postgraduate education. Of this group, fewer than 500 had enrolled in degree programs in any humanities field. Consequently, despite the excellent design and execution of the B&B study, this large national dataset has relatively little data for policy research on the humanities and culture. The 2000 B&B cohort will almost certainly have the same distributional properties. Even if the research community were able to persuade NCES to modify its sampling methods for NPSAS and B&B to yield larger numbers of cases with degrees in humanities fields, the first opportunity to do so will probably be with the class of 1996 or 1997.

Much like B&B, the *Beginning Postsecondary Student Study* (BPS) is a longitudinal data program sponsored by NCES and is based on a subsample of first-time, full-time students identified in the *National Postsecondary Student Aid Study*. B&B and BPS derive their baseline samples from alternating rounds of NPSAS: the first B&B cohort was selected from NPSAS:93 and the first BPS cohort from the earlier 1990 NPSAS wave. Follow-up surveys

21. A check of the NCES web site revealed, however, that the CD-ROM for public use, the Data Analysis System for the B&B baseline, and first and second follow-up data, was "out of stock." A limited Data Analysis System is available on-line from the Web site, but this system is capable of only simple tabular analyses, not sophisticated data manipulations and analyses.

were completed with the BPS:90 cohort in both 1992 and 1994. A new BPS sample cohort was generated from the NPSAS 1996–97 survey and is referred to as BPS:96. One survey follow-up has been completed in 1998 with the BPS:96 cohort and a second is to be completed in 2000.

Similar to B&B, the BPS dataset includes rich data on a wide range of postsecondary experiences including topics associated with admissions, financing, and academic pursuits and outcomes, as well as measures of students' goals and plans, demographics, employment while attending school, and so on. Financial aid data and academic records are also collected for BPS cohort members, and telephone surveys are completed with the parents of a large subsample of BPS cases. BPS:90 included about 8,000 sample members, and BPS:96 included about 11,000 full-time first-year students in the sample. As with other NCES projects, response rates for the baseline and follow-up data collections are high and the quality of the individual data elements exceeds that of most other publicly available data systems. Documentation of survey procedures and methodology is detailed and complete. Data are made available to researchers in tape and CD-ROM format approximately two years after each round of data collection is completed.<sup>22</sup>

Despite the high quality of this system, the utility of BPS data for policy research in the humanities is limited by the same problem affecting B&B, namely small subsample sizes for students in humanities fields. For example, BPS:90 contained about 8,000 sample members, but only slightly over 700 cases were identified as enrolled in humanities fields (including “liberal studies” majors within the humanities). For BPS:96, the situation is hardly better—with about 900 out of 11,000 cases indicating humanities majors. The sample sizes for individual humanities fields are thus far too small for any detailed analyses. Moreover, even if the follow-up studies are continued with BPS cohorts beyond the time customarily associated with earning bachelor's degrees, the numbers of BPS-sampled cases entering graduate programs in the humanities will be extremely small.

Thus, while both B&B and BPS samples are powerful tools for studying broad factors affecting postsecondary attendance, persistence, and completion, and in particular the contribution of student financial assistance to those processes, neither of these data systems uses a sample design that is adequate for research on specific humanities fields, and they are barely adequate for research on the humanities as a group of fields.

The *CGS/GRE Survey of Graduate Enrollment* is a collaboration between the 680-member Council of Graduate Schools and the Graduate Record Examinations Board that has collected annual data on the characteristics of students who applied to and were enrolled in graduate departments in member institutions between 1986 and the present. In addition, the fall survey also collects information about degrees conferred in the previous twelve months. Data are collected from institutions on total numbers of applications (counting separately those accepted and those not accepted), enrollments, and degrees awarded.

22. As for the B&B dataset, the NCES website indicates that the most recently released BPS dataset containing data from the BPS:96 baseline and BPS:96–98 follow-up is also “out of stock.”

Application, enrollment, and degree data are collected for 51 separate fields of study. However, in annual reports on enrollment and degree trends, disciplines are collapsed into nine broad fields. Institutional characteristics are also maintained in the dataset. Well over one million data records are maintained in a consistent format covering the life of the program. Response rates of member institutions have been consistently in the 93- to 94-percent range over the years. Because reporting of enrollment and degree data is done only by broad field, use of this data resource for policy research in the humanities could not rely on reported results alone. Although the *Index of Humanities Datasets* suggests that the CGS/GRE datasets are available to researchers in their entirety, this is not currently the case, and while CGS staff have frequently provided special tabular breakdowns to researchers, there are no time-series data products routinely packaged for sale or other form of distribution. However, CGS has stated that the Council is willing to consider requests for access to the dataset by researchers for non-proprietary policy purposes.

The primary limitations of the CGS/GRE Graduate Enrollment dataset for policy research are the potential for population coverage bias (the survey participants are limited to member institutions) and the lack of open availability of the dataset. Possibilities exist for overcoming the availability limitation, and further research is necessary to determine the potential undercoverage in the graduate student population (across fields of study) resulting from the restriction of the project to CGS member institutions.

The *Institute of International Education* (IIE) Reports on Foreign Students consists of annual surveys and accompanying reports on the numbers of students studying abroad each year, both U.S. students enrolled overseas and foreign students enrolled in accredited U.S. postsecondary institutions. The data collection and reporting program was initiated in 1949. The 1997–98 report includes responses from nearly 2,600 postsecondary institutions and reports aggregate data on over 480,000 foreign students studying in the United States.

The survey of foreign student enrollments succeeds in gathering aggregate counts of foreign student enrollments from a high percentage (typically over 90 percent) of the institutions it surveys. However, the attempt to collect individual student characteristics, such as demographics and field of study, is considerably less successful, with response rates typically in the 50-percent range. The lower response rates for demographic and field-of-study data compounds the basic difficulty of using these data for humanities policy research: the survey only collects field of study in the broadest terms. The survey documentation in the *IHD* records indicates that data are not collected on enrollments in History or its subfields, nor is History included in the “Humanities” category for which data are presented in annual reports.

Raw data files constructed by IIE are generally not available to researchers. However, data in tabular form (spreadsheets or text files) for the last five surveys prior to the current year (in this case ranging from 1993–94

through 1997–98) may be downloaded from the IIE web site (<http://www.opendoorsweb.org/datadownload/download9798.htm>). The spreadsheet tables clearly show the numbers of cases for which individual student counts are available (about 248,000 in 1997–98), compared to the number for which aggregate institutional counts are available (about 481,000 in 1997–98). Like the printed reports, these spreadsheet tables present breakdowns for only 12 explicit fields of study (of which one is a partial collection of humanities fields), consequently the available data are of very limited use for rigorous policy analysis on the humanities and culture.

The *National Postsecondary Student Aid Study* system is a major data program sponsored by the National Center for Education Statistics designed to track the distribution and impact of student financial assistance on postsecondary school choice, enrollment, persistence, and completion. It is conducted every three to four years with very large multi-stage samples of postsecondary institutions and students of all types and at all levels. Student samples are drawn in two stages, the first being nationally representative samples of postsecondary institutions in all sectors, including proprietary, public and private two- and four-year colleges and universities, and graduate and professional schools; and the second being probability samples of students attending the selected institutions.

NPSAS has been conducted in 1987, 1990, 1993, and 1996; a fifth round was conducted in 2000. Approximately 900 to 1,100 postsecondary institutions and 50,000 to 60,000 students are sampled for each NPSAS survey.<sup>23</sup> Each round of NPSAS has collected data from students' administrative and financial aid records as well as conducting interviews with sampled students. Interviews are also conducted with parents of a substantial subsample of students. Finally, institutional data are also collected on tuitions, fees, and all other costs of enrollment and attendance.

The primary focus of the NPSAS studies is on how students and families finance postsecondary schooling. Therefore, the bulk of the data collected concern costs of postsecondary school attendance and how those costs are met. In addition, the survey collects detailed demographic information, educational activities and goals, field of study, institutional characteristics, test scores, academic performance, extra-curricular activities, and attitudes and values.

As noted above, the large NPSAS student samples provide the basis for selecting two subsamples of students for longitudinal follow-up. The *Beginning Postsecondary Student* study (BPS) is designed around the subsample of NPSAS students who are first-time, full-time students in any type of postsecondary institution (including welding schools and other vocational training institutions, as well as law and medical schools). To date, NPSAS surveys have selected between 8,000 and 11,000 students who have been selected (or retained) for the BPS samples. The *Baccalaureate and Beyond*

23. In the most recent round of NPSAS (1996), administrative and financial aid records data were collected from institutions for about 50,000 students and interviews were completed with about 30,000. Parent data were collected for only a small subsample of about 3,500 students.

(B&B) study is the longitudinal continuation of the research on the subsample of college graduates from accredited 4-year institutions in the NPSAS sample. The 1993 NPSAS round yielded about 11,000 current graduates for the B&B sample. NPSAS:2000 is expected to include a similar number. All data collected in the NPSAS studies are available to the public, and most can be obtained via the NCES web site (<http://nces.ed.gov/surveys/>). The data systems are very large and complex, but are of generally high quality and are well documented. Response rates have hovered in the 90 percent range for all rounds of NPSAS to date.

Although the overall student sample sizes for NPSAS are very large, the numbers of cases of special relevance for policy research on the humanities is somewhat limited. First, a substantial proportion of each NPSAS sample is selected from less-than-two-year schools in the proprietary sector. These schools offer little or nothing by way of humanities curriculum or concentration. Second, the portion of the sample attending two- and four-year colleges and universities contains many students who have not yet determined a field of study or concentration. The resulting proportion of humanities students in the NPSAS sample, then, is similar to that for BPS and B&B—or approximately 10 percent.

For the 1996 round of NPSAS, of the 50,000 students for whom administrative records were collected, there were about 5,200 students in the entire sample who were majoring in humanities subjects. For the subset of about 30,000 students from whom interview data were collected, about 3,100 were students majoring in humanities disciplines. Assuming the subset of the sample who were interviewed about educational goals and career plans is of greatest interest, the numbers of students majoring in specific humanities fields were generally small, ranging from approximately 30 in Philosophy, to 240 in History, 420 in English and American Letters, but as many as 1,920 in the rather vague “Liberal Studies” category. Thus, even the large NPSAS samples do not yield sufficient numbers of students in specific humanities subjects to carry out detailed analyses for more than a handful of fields. Moreover, public data files for each round of NPSAS are typically not available until more than two years after the data were collected, also limiting their utility.

The final resource listed in the *IHD* based on a sample of postsecondary students is the *American Freshman: National Norms*, sponsored by the Cooperative Institutional Research Program (CIRP) and implemented by the Higher Education Research Institute at UCLA, with the support of the American Council on Education (ACE). The freshman survey program was originally designed as a probability sample of institutions and students, but evolved into a self-selected participation process within a few years after initiation. It has collected data annually since 1966 and shares many characteristics with the American Teacher program described above.

The CIRP web site (<http://www.gseis.ucla.edu/heri/cirp.htm>) indicates that in addition to the American Teacher and American Freshman programs, HERI has recently designed and implemented two other student surveys.

The College Student Survey (started in 1993) is similar to the freshman survey, but is designed to provide data on upper level college students, and the *Entering Student Survey* (added in 1999) is a parallel to the freshman survey for students entering two-year colleges (but is not limited to first-time entrants).

All CIRP surveys operate on the same principle: postsecondary institutions elect to participate in the program by notifying the CIRP office and paying a fee (typically a fixed fee of \$400 to \$450 plus \$1 per processed questionnaire). Institutions are provided survey materials and conduct the student surveys using local staff (proctored survey administration sessions are advised, to promote the best response rates). Approximately 700 institutions participated in the most recent data collections. However, the number of student participants varies from about 350,000 for the freshman and entering student surveys, to about 150,000 for the college student survey. Moreover, institutions participating in the freshman survey are not necessarily the same from year to year, although the overlap between adjacent years has ranged from about 85 to 90 percent.

Completed forms are returned to CIRP offices for processing and analysis, which includes an attempt to add statistical weights to records from institutions providing “the most representative data.”<sup>24</sup> The weighting process is, however, not based on probability sampling but is used as a means of expanding the data from the self-selected groups of institutions to estimates of the total population from the U.S. Department of Education.<sup>25</sup> Although CIRP publications have included estimates of sampling errors for their reported percentages, the sampling variances can only be viewed as extremely crude approximations of what the statistical precision levels might have been if probability sampling had been used in the survey design. There is no scientific basis for using standard error estimates with percentages that are based on self-selected or other nonrandom sample designs.

The data collected in the freshman and other three student survey programs are a mix of standard items asked of all participating students and “locally determined” items added by each participating institution. The standardized data tend to concentrate on student biographies and demographics, attitudes, values, aspirations, and experiences, and are not always closely tied to academic issues (for example, the most recent report on the freshman survey explored differences in perceived “stress levels” among different types of students and how they coped with it). A significant proportion of the data elements capture data about high-school experiences. However, even the

24. The weighting process is designed to permit the reporting of “national norms,” or benchmark statistics against which participating institutions can compare their own response data. CIRP analysts selectively eliminate some institutions from the process of developing weights if their student-level response rates fall below a threshold related to institution type. Typically, only about 70 percent of the institutions and students participating in the annual survey have been used in calculating the CIRP norms.

25. The IPEDS Fall Enrollment data have typically provided the population counts to which CIRP data are matched, with weighting cells created from a combination of institution type and size, and sex of student. Weight values frequently vary enormously across cells, from very small values (between 1 and 2) in larger private colleges and universities to very large values (between 70 and 80) for large public two-year colleges.

standardized portion of the questionnaire is not consistent from year to year, but is revised annually to meet the interests and data requirements of the participating institutions. The survey item on probable major field of study is asked in insufficient detail to examine all major fields within the humanities (all foreign languages and literatures are a single category).

The principal limitations of the CIRP data for high quality policy research on the humanities are design features. First, the nonrandom nature of the selection process for participating institutions and the subjective nature of the weighting process undercuts the credibility of the data. Coverage bias and nonresponse bias in the raw data may be very large. The attempt to eliminate the bias with non-statistical weighting schemes introduces the risk that the weighting alone could be responsible for year-to-year trends in the findings.

While not optimal for developing estimates of population parameters of interest to humanities policy research, the large numbers of student observations means that the CIRP datasets will have more students with probable majors in the humanities than most of the other datasets described here. These data may therefore be of considerable value for studying relationships among choice of field and a wide variety of other student characteristics and school experiences. Second, the data are, in general, not made available to the public, but only to researchers at participating institutions. The HERI research staff will produce custom tabulations on a fee basis; however, this arrangement places severe limits on the extent of data reorganization or manipulation that may be included in the analyses. Finally, as illustrated by the example above focusing on “academic stress,” the topics covered by the CIRP surveys, while important to postsecondary education providers, are frequently irrelevant to the issues of greatest concern for humanities policy. In addition, although the question about choice of major field of study is asked in reasonable detail, the data available to analyze relationships between choice of field and other student traits change frequently across the years of the program. This will present some level of obstacle for the study of such relationships over time.

### *Samples of Administrative Records of Postsecondary Students*

Four of the data resources listed in the *IHD* are based primarily on samples of administrative records associated with individuals who were, or who sought to become, postsecondary students at the undergraduate or graduate level. That is, these resources contain data that were originally created for purposes other than research, including official academic transcript records and data collected with standardized tests scores used for ranking applicants for admission to educational programs. The primary characteristic they share is that all four are based on unique subpopulations of postsecondary students—such as individuals who elected to take a particular standardized test, or college graduates in a particular year and from a sample of institutions. As we indicate below, these four defined populations are of varying relevance for addressing the key issues of humanities policy.

The *AACU Curriculum Database* was created in the early 1990s for the Association of American Colleges and Universities by the Institute for Research on Higher Education (IRHE) at the University of Pennsylvania. The dataset consists of information coded from college transcripts for a total of 42,007 individuals who graduated in 1991 from a stratified probability sample of 100 colleges and universities drawn from a universe of 1,360 eligible institutions that granted bachelor's degrees in the arts and sciences.<sup>26</sup> A total of 81 institutions participated in the project, submitting transcript and demographic data for over 42,000 graduates. Available evidence indicates that this dataset was created using a scientifically valid design, that institutional participation rates were high (although the within-school transcript-level completion rates are unknown), and that the course-taking data are detailed and potentially useful. However, to date, this dataset has provided the basis for only a single analysis and report.<sup>27</sup> A likely reason is that for several years the dataset was not generally available to the public; requests for access to the data were evaluated by the IRHE "on a case-by-case basis."

In May 1995, NSF, a primary sponsor of the study, indicated that the data were being made available on the World Wide Web. However, a search of the World Wide Web and the NSF web site (at both <http://www.nsf.gov/sbe/srs/stats.htm> and <http://caspar.nsf.gov/>) in July 2000 revealed no evidence that the dataset is actually available online, or whether control over access to the data has been transferred from IRHE to NSF or another organization.

Assuming the data are more widely available to policy researchers, an additional primary limitation is the age of the dataset. The analysis of transcript data for even a large sample of 1991 college graduates will have little to offer concerning curriculum activities of undergraduates in the "Internet Age" either for the sciences or the humanities.

The remaining three datasets are created and maintained by the Educational Testing Service and consist primarily of standardized test scores accompanied by a small amount of demographic and biographical data. *College Bound Seniors: Profiles of SAT and Achievement Test Takers* includes test data and information from the Student Descriptive Questionnaire (SDQ) for over one million high school (and some college) students who took the SAT or Achievement tests between 1973 and the present time (an average of about 45,000 per year). The questionnaire covers primarily demographic and educational topics, including high school performance and experiences (academ-

26. Data published by the National Center for Education Statistics indicate that there were over 2,100 4-year colleges and universities in existence during the late 1980s and early 1990s. Information was not readily available to explain why the sample frame for the AACU Curriculum Database was limited to only 1,360 institutions (less than two-thirds of the total). Without further information on the sample design, we cannot be certain that all institutions granting bachelors degrees in 1991, and all 1991 college graduates, had a chance of selection into this sample. The possibility remains that omissions could represent possible bias in the data for graduates from selected fields.

27. See Institute for Research on Higher Education, University of Pennsylvania. 1994. "Final Report: Curriculum Assessment Service Database: Estimates of Student Curricular Activity From a National Survey of Colleges and Universities." Philadelphia. The report has been quoted in NSF's *Science and Engineering Indicators: 1996* report as well as a NSF DataBrief in 1995.

ic and extra-curricular), and educational plans and aspirations. Some 94 percent of test takers also complete the SDQ. SAT scores exist for the math and verbal components; Achievement test scores are available for all fields in which the test is offered, which includes most humanities fields. Available documentation does not indicate how many observations exist for Achievement tests in humanities fields.

These data are a potentially rich source of historical information about the characteristics of students who take particular Achievement tests as well as the SAT. Data for students who took only the SAT may be of interest for policy research specifically on humanities issues, since students were asked to select their probable college major field of study, and specific humanities fields were coded in some detail. Ultimate degree-attainment goals were also asked of all SAT takers. These data are available to researchers on tape/diskette for a fee. Over one million data records have been compiled on AP test takers over the 45-year period, suggesting an average of almost 27,000 test-takers per year (about 60 percent of the number of SAT/Achievement test takers). In theory, the availability of appropriate identifying information might permit linking of AP score data with SAT and Achievement test data. However, this capability is not described in the dataset descriptions. The Advanced Placement Summary data are also available to researchers on tape/diskette from ETS for a fee.

The *Advanced Placement Summary Reports* program is also compiled by ETS. This dataset consists primarily of AP scores for all subjects tested (which includes humanities disciplines) and limited demographic data on the test-takers, who are primarily high school students. Historical data are available from 1955 to the present.

The final resource based on test-takers is the *Characteristics of GRE Test-Takers*, also maintained by ETS. Like the others, this dataset is a combination of test scores (the general test plus tests in 14 separate fields), limited demographic information, and a few items on educational goals (degree goals and plans for advanced study). The number of observations now exceeds 500,000. However, the information relevant to policy research in the humanities is limited by two significant factors. First, the test takers' advanced study and degree plans are only coded into very broad fields, of which "Arts and Humanities" is one. Thus, the data support the study of characteristics of students in separate humanities disciplines only if they took GRE subject tests in specific fields. More important, GRE data are not generally made available to researchers. However, ETS and the Graduate Record Examinations Board will consider special requests from researchers for access to the data.

#### *Samples of Larger Populations that Include Postsecondary Students*

Four data resources fall into this category: the *National Longitudinal Study of the High School Class of 1972* (NLS-72); *High School and Beyond: A National Longitudinal Study for the 1980s* (HS&B); the *National Education Longitudinal Study of 1988* (NELS:88); and the *Survey of Income and Program Participation*

(SIPP).

The first three datasets are core studies of NCES' national longitudinal studies program and are all based on multi-stage samples of student cohorts<sup>28</sup> beginning in the eighth grade (NELS:88), tenth grade (HS&B) or twelfth grade (HS&B and NLS-72).<sup>29</sup> On average, the sampled cohorts in the NCES studies were followed with approximately biennial surveys for about ten to twelve years after the baseline data collection. All three studies are complex data systems that include survey and cognitive test data from students as well as administrators and teachers, parents, and data from high school and postsecondary school transcripts and financial aid records. They represent some of the richest resources in the nation for studying the process of middle- and high-school education, passage from high school to the labor force, entry and progress through postsecondary schooling, marriage and family formation, and a wide variety of other social, political, and economic activities.

All three studies began with relatively large samples of high-school or middle-school students (roughly 25,000 students per cohort). However, for reasons associated with funding limitations, all of the education longitudinal study cohorts were reduced significantly for the follow-up surveys, typically to between 8,000 and 15,000 individuals per cohort. While these may seem like large numbers, the samples include the entire range of middle-school or high-school students, and therefore have significantly smaller samples of postsecondary-school attendees, and typically very small samples that continue their education and earn advanced degrees. For example, the third follow-up of the NELS:88 cohort was conducted in 1994 with a sample of almost 16,000 of the original 25,000 sample members. In 1994, "normal progression" college students in the cohort would have been completing their second year of college. Of the 12,000 individuals who participated in the 1994 round, although 44 percent (5,280) were taking academic courses at a two- or four-year college, only 30.5 percent (3,660) indicated that they expected to earn a bachelor's degree, 23.1 percent (2,772) expected to earn a master's degree, and only 6.6 percent (792) planned to earn a doctorate.

The follow-up surveys in 1992 of the HS&B younger cohort (high-school sophomores in 1980) is even more telling. By the time of the Fourth Follow-up survey, members of this cohort were more than ten years past their high-school graduation year (1982). Of the approximately 12,600 respondents to the 1992 survey, just over 9,000 had engaged in some form of postsecondary education. However, many of these had attended less than two-year programs seeking skill certification rather than academic degrees, and a substantial additional percentage attended only two-year colleges. In terms of degree attainment, by 1992 only 0.2 percent (30 individuals) had earned doctorate

28. Stratified random samples of public and private high schools (HS&B and NLS-72) or middle schools (for NELS:88) were selected in the first stage. Samples of students within schools were selected in the second stage.

29. While NELS:88 was initiated with a single cohort of eighth graders in 1988, and NLS-72 selected twelfth graders in 1972, HS&B was initiated with two cohorts—tenth graders and twelfth graders in the same high schools in 1980.

degrees, another 0.9 percent (133) had earned professional degrees, another 2.6 percent (385) had earned master's degrees, and 19.4 percent (2,886) had bachelors degrees as their highest educational attainment. In total, 23.1 percent of the HS&B younger cohort (3,425 individuals) had earned at least a BA degree by 1992. However, only 2.1 percent of the entire cohort (311 individuals) had BA degrees in any humanities field, including only 44 with BA degrees in history, and only 104 individuals with BA degrees in English or American letters. The data distributions are comparable for the HS&B older cohort (1980 seniors) and for the NLS-72 cohort, and will almost certainly be comparable for the NELS:88 cohort at the time of the 2002 follow-up (assuming a follow-up survey is conducted by NCES in that year).

Exemplary as the NCES longitudinal studies are for supporting research on the factors affecting educational access and choice through high school and all levels of postsecondary education, the sample designs used in these surveys simply do not support detailed research on educational activity in low-incidence fields, such as the humanities. Transcript data collected for all postsecondary attendees may be of some value for examining the course-taking behavior, especially for those who attended collegiate institutions. Although transcript data about course-taking in the humanities will be relatively sparse, it may be of value to compare patterns for college and advanced students in the 1970s, 1980s, and 1990s.

It should be kept in mind, however, that because the NCES longitudinal study cohorts are samples of students in a specific year of middle or secondary school, the cohorts are not representative of all students enrolled in postsecondary schools in a particular year (or group of years). These projects were designed to study the dynamics affecting education and labor-force participation of grade-level cohorts. College and university "grade levels" are far more diverse with respect to age than middle- or high-school grades. By the time members of any NCES longitudinal cohort reach a certain level of postsecondary education (for example., college seniors), they constitute only a sizeable percentage of all students enrolled at the same level—with the remainder being primarily students who started college at an older age than they did. For example, researchers could not treat the HS&B older cohort in 1984 (when normal progression sample members would be college seniors) as representative of all college seniors in 1984. This subsample of the cohort would only be representative of that portion of college seniors in 1984 who had been high-school seniors four years earlier.

The *Survey of Income and Program Participation* (SIPP) is a large national survey program sponsored by the U.S. Census Bureau and conducted since 1984 through the present. SIPP is designed as a continuous series of national panels based on probability samples of households in the United States. Samples have varied in size from about 14,000 up to nearly 37,000 households. Each sampled cohort is followed for a period of time (ranging from two and one-half to four years with data collected on a continuous basis using a four-month recall period (with about a fourth of the sample being

interviewed each month). New panels are introduced in February each year.<sup>30</sup> Unlike the NCES cohorts, which are samples of school-grade cohorts, each SIPP sample represents all noninstitutionalized individuals age 15 or older living in households.

The primary purpose of the SIPP is to collect information about economic activity of individuals and families, including amounts of income from all sources, labor force participation, and eligibility for and participation in various income support programs. Other core data collected to support analyses include demographics, education and training, and family characteristics. In addition, “topical modules” are included in each wave to support research on other policy issues, such as personal histories, child care, disability, school enrollment, and tax burden. Data are collected primarily by in-person household interviewing, supplemented as necessary by decentralized telephone interviewing. The 1996 panel introduced several new design features, including questionnaire revisions, a four-year participation span for sampled households (through March 2000), and the use of computer-assisted interviewing with laptop computers in the field. The Census Bureau releases SIPP data periodically in cross-sectional, longitudinal, and topical module datasets. Datasets are available currently for all waves of the 1984 through 1993 cohorts, plus the first wave of the 1996 panel.<sup>31</sup>

The primary limitations of using SIPP data for policy analyses in the humanities and culture begin with the daunting size and complexity of the data systems—including the relational data structures used through the 1990 panel, and the more conventional, rectangular datasets produced since 1990. No fewer than twelve data file-users’ manuals are published to guide researchers in the use of the multiple file formats.

The most common data structure currently in use for economic policy analyses is the so-called “person-month” file, a format initially recommended by analysts in the Congressional Budget Office and also independently adopted by several policy research organizations. However, although these files permit more detailed analyses of income flows, the original files contained a bewildering array of weekly, monthly, quarterly, and annual financial data, and included (by design) a substantial duplication of data across fields, making them even more complex than the rectangular person-level files. To respond to this complexity, a redesign of the SIPP person-month datasets was undertaken by the Census Bureau in 1990 to simplify the structure of each data record (one month’s data for each person in the file). Much of the duplication of data elements has been eliminated, as well as data quality flags, leaving a single set of individual and household demographic items, an extensive series of measures of labor-force participation, past jobs held, and separate income data from up to 35 possible sources (plus data imputation flags for these items), followed by data elements on asset income. The current file structure, while much cleaner and simpler than prior to 1990, will still require significant effort in data preparation for use in humanities policy

30 During the SIPP redesign effort, it was decided that no new panels would be introduced in 1994 or 1995. To continue the flow of data, the 1992 panel was extended to ten waves (re-interviews) and the 1993 panel was extended to nine waves.

31. Some longitudinal datasets are still in production for the cohorts introduced in 1990 and later.

research.

The second major issue involves the timing of data release. At present, the most recent available data are for the 1993 panel (which runs through 1996), and the first wave of interview data from the 1996 panel. SIPP data therefore cannot be used to track up-to-the-minute trends in labor force attributes or earnings profiles of the various professions. At best, the data can be applied to understanding the impact of policies implemented at some time in the past (such as the early 1990s)—and even then, only if relevant measures are available through the “topical modules,” as described further below.

The final, and most important, issue concerns the identification of sampled cases in the files that may be considered relevant for research on humanities policy issues. Examples of use might be to track employment and earning data for practitioners in humanities fields—including the academic and public humanities, or to examine financial support for postsecondary education at various levels across fields of study. Unfortunately, the SIPP system does not collect data at an adequate level of detail to meet even general requirements for research on practitioners in humanities fields. The most recent analysis of economic returns to advanced degree attainment in various fields of study (U.S. Census Bureau 1990) was based on analyses of data from the “topical module” fielded during the second wave of interviewing the 1987 panel (June–September 1987). As part of this report,<sup>32</sup> analysts compared differences in monthly income between individuals earning bachelors’ degrees and advanced degrees for major fields of study.

However, in the comparison, only 20 fields were defined in the response categories of the survey question. Individuals with degrees in most humanities fields are contained in a single category “Liberal Arts/ Humanities,” and those with degrees in history are classified with all other social science degrees. Even using these broad categories, the numbers of cases with advanced degrees is too small to assess the earnings differences of interest. The “Liberal Arts/Humanities” category contained only 26 individuals with Ph.D. or professional degrees, 439 with MA degrees, and 1,907 with BA degrees.<sup>33</sup> With such small subsample sizes, even the 90 percent confidence intervals used by Bureau analysts in this report were much too large to conclude that the small monthly income difference between humanities practitioners with advanced degrees and those with only bachelors degrees was statistically reliable. Given the distributions illustrated, even with the significantly larger 1996 SIPP cohort sample, the power of analyses of even all humanities advanced degree holders combined will not be much improved. There is no realistic prospect of using SIPP data to examine practitioners in separate humanities fields.

### *Samples of Ph.D. Recipients*

32. U.S. Census Bureau. Household Economic Studies. 1990. What’s it worth? Educational Background and Economic Status: Spring 1987. Series P-70, No. 21. Washington, DC.

33. The corresponding numbers in the Social Sciences category were 93 Ph.D.’s, 479 MAs, and 1,585 BA degrees.

The *Index of Humanities Datasets* includes two resources that focus entirely on recipients of doctoral degrees, the *Survey of Earned Doctorates (SED)*,<sup>34</sup> and *Humanities Doctorates in the United States*.

Sponsored primarily by the National Science Foundation with additional support from the National Institutes of Health, the National Endowment for the Humanities, the U.S. Department of Education, and the U.S. Department of Agriculture, the SED has been conducted annually since 1958. For most of this period, the National Research Council (NRC) of the National Academy of Sciences managed the collection and analysis of SED data. In 1997, the SED contract was moved to the National Opinion Research Center (NORC), which began collecting SED data as of the 1998 round.

The SED collects basic information from all individuals who received research doctorates from U.S. universities and colleges each academic year. The survey seeks modest amounts of data for relatively few topics, including demographic characteristics, field of degree, educational history, types of financial aid received, educational debt, time to degree, and plans for further postgraduate education and career choices. Response rates have historically been very high (about 94–95 percent), virtually eliminating bias in the data. Because the SED is a census of all doctoral recipients, sampling variability is also eliminated. Significantly, respondents to the SED are asked to classify their field of dissertation research using a list of 281 detailed categories.

The documentation in the *IHD* predates the shift of responsibility for the SED from NRC to NORC and is incorrect in several other respects. For example, it suggests that SED data are not available to researchers, other than through several publications of annual and time trend data. However, NSF maintains a Web-based facility called WebCASPAR<sup>35</sup> where aggregate data tables from the SED and many other surveys (primarily institutional surveys) are available for review and customization for specific users' needs. Because of the need to protect confidentiality agreements, SED datasets are not generally made available to the public. However, NSF will provide access to microdata for qualified researchers under a special licensing agreement. This level of data access is essential to producing detailed statistics on Ph.D. recipients in separate humanities fields.

34. As described below, the *Survey of Earned Doctorates (SED)* has been conducted annually since 1958 as a complete census of new Ph.D. recipients at US educational institutions. Procedures used in the SED permit the creation of a complete list of Ph.D. holders referred to as the Doctoral Records File (DRF), which is maintained by NSF. The DRF serves as the universe or sampling frame for the *Survey of Doctorate Recipients (SDR)* conducted biennially since 1973. *Humanities Doctorates in the United States* is the subset of survey data collected in the SDR on Ph.D. recipients in humanities fields. The listing in the *Index of Humanities Datasets* includes only reference to the *Humanities Doctorates in the United States* because the remainder of the SDR data apply to only non-humanities fields. Our discussion of the *Humanities Doctorates in the U.S.* resource attempts to clarify the relationship between this resource and the SDR.

35. WebCASPAR may be visited at <http://caspar.nsf.gov/cgi-bin/WebIC.exe?template=nsf/srs/webcasp/start.wi>. Tables for the SED on WebCASPAR display the summary number of degree recipients in humanities fields. Detailed data are reported each year since 1973 in summaries entitled "Doctorate Recipients from United States Universities: Summary Report."

The second dataset listed in the *IHD* focusing explicitly on this population is referred to as *Humanities Doctorates in the United States*, more accurately called the *Humanities Survey of Doctorate Recipients (Humanities SDR)*<sup>36</sup> because the design and methods parallel those used in the *Science and Engineering Survey of Doctorate Recipients*, a system of biennial surveys begun in 1977 conducted with individuals who received degrees in science and engineering. As for the SED, the NRC conducted the S&E SDR and *Humanities Doctorates* surveys through 1995. NEH was unable to sponsor the *Humanities SDR* after its funding was reduced in the mid-1990s. The last round of the *Humanities SDR* was conducted in 1995.

The *Humanities SDR* collected longitudinal data from a rolling sample of individuals who received doctoral degrees during the prior 50 years, were under the age of 76, and lived in the United States in the year of the survey. At the time of each survey, the longitudinal sample was adjusted by dropping cases who were older than 75 and by sampling the newest doctoral recipients who had earned degrees since the last prior survey from the Doctorate Records File. The survey was conducted by mail with telephone follow-up, and collected basic information, such as demographics, employment characteristics, earnings, publications, field mobility and retention, and related topics. The *Humanities SDR* typically included about 9,000 respondents. Biennial data exist from 1977 to 1995; reports were published for each year the survey was conducted.

Although every effort has been made to enhance data quality, nonresponse bias is occasionally a problem, with response rates ranging from a high of 87 percent (1993) to as low as 54 percent (1989). More important, like the SED, the *Humanities SDR* data files are not generally available to the public. However, NSF has a policy of permitting access to qualified researchers under controlled conditions.

#### 4. *Publications and Reports*

The final category of datasets included in the *IHD*, Publications and Reports, consists of only three entries. Two of these, the *Digest of Education Statistics* and the *Condition of Education*, are annual reports from the National Center for Education Statistics containing a broad variety of statistical indicators of the state of the U.S. educational system at all levels. The third is the series of *Annual Reports to NEH* by the 56 State Humanities Councils.

The *Digest of Education Statistics* was initiated in 1962 and has been produced annually except for the years 1977–78, 1983–84, and 1985–86, when combined volumes were released. The *Digest* contains data on all levels of education, including a major section on postsecondary institutions, faculty, students, and graduates. The *Condition of Education* has been produced since 1975 and contains information on no more than 60 indicators of the state of education in the United States based on data in approximately 400 tables. Both reports draw on numerous studies, most of which are sponsored and managed by the National Center for Education Statistics. A substantial pro-

36. Sample members for the 1995 *Humanities SDR* were sent a questionnaire entitled “Survey of Humanities Doctorates.” Sample members for the 1995 Science and Engineering SDR were sent a questionnaire entitled *Survey of Doctorate Recipients*.

portion of the indicators are presented as trends in percentages or average values over time, with historical spans ranging from a few years to several decades, depending upon the topic. Both of these reports are available both in printed volumes and as electronic documents. In addition, tables used in the documents are available in spreadsheet format. Because of the publication cycle and the internal review requirements at NCES, both of these reports are based on data that are at least two years old and frequently older.

Although these annual reports contain a wealth of information about pre-school, elementary and secondary education, postsecondary schooling, and graduate and first professional education, both reports contain modest amounts of data of direct relevance to humanities policy. However, because of the design of the tables, it is difficult to use the data as reported for research on humanities policy. First, many of the tables in both volumes that report on postsecondary education by field of study provide data on the humanities only as a single broad category, with no breakdown by humanities fields. This decision is further complicated by the decision rules used by NCES to construct its broad reporting categories. For instance, history is included in the social sciences, rather than the humanities.

In many tables that use a standard set of 31 “field of study” categories, several humanities fields are reported separately (for example, area and ethnic studies, foreign languages, letters, philosophy and religion), but others (history, anthropology, classics, linguistics) are subsumed within other categories and cannot be separated out or recombined with other humanities fields. The result is that these tables are virtually useless for developing detailed profiles of educational conditions and trends that would be of value to humanities leaders and practitioners. Moreover, even the broad classifications schemes used for “field of study” are not consistent across tables in the *Digest*. For example, in Table 207 reporting “Enrollment in institutions of higher education, by major field of study, age, and level of student,” the listing of fields shows subcategories of social sciences that permits history to be counted separately. Unfortunately, in this table, there is no category for Foreign Languages.

We found only two tables in the *Digest of Education Statistics* that contain sufficient detail in the breakdown of fields of study to extract data for nearly all humanities or humanities-related fields. These were typically tables 240 and 241 of each volume, reporting “Bachelor’s, master’s, and doctor’s degrees conferred by institutions of higher education, by sex of student and field of study” for the two most recent academic years for which data are available.

In the *Condition of Education*, only a small number of indicators are typically reported by field of study. All of the tables in the recent volumes we examined collapsed all humanities fields into a single category that included the broad category “liberal/general studies,” but excluded anthropology, archaeology and history, which were included in the social sciences. (By way of contrast, it was not clear if arts fields were included at all, or where if so.) To develop its indicators, the *Condition of Education* also includes many tables and figures produced by other agencies that use their own, possibly different

definition of the fields to be included under the heading “humanities.” However, no explanation is provided in the *Condition of Education* of how the taxonomies differ across tables or what impact it might have on the trends reported. Consequently, we conclude that relatively little data of value for policy research on humanities and culture can be gleaned from these volumes. Extracting those data that are of value will require substantial effort to determine category definitions in order to ensure that the data elements obtained from these sources have a consistent meaning and operationalization in order to compare them with data from other sources.

The remaining item in the Publications and Reports category consists of the *Annual Reports of the 56 State Humanities Councils* on various types of financial data, staffing, operations, and activities. These resources are of potential value because they are among the few that provide the potential to analyze and report flows for the public (nonacademic) humanities. At this time, the reports are public documents that exist only in printed form. However, a significant difficulty of using these reports, one that has been the source of some frustration for NEH staff, has been the lack of consistency in the quantity, quality, format, inclusiveness, and timeliness of the data reported across the 56 Councils. As described in the introductory section of this report, in recent years considerable development work has been done under the leadership of the Federation of State Humanities Councils to address these issues, and to launch a program to collect core data from the state councils in a standardized way using web-based data-capture technology.

*The State of the Art in Humanities Policy Research*

The main body of the report has described obstacles to conducting quantitative research on the humanities. Such factors severely limit our ability to answer even the most basic questions about the current state of humanities fields, trends from the recent past, or conditions likely to face humanities practitioners in the future. The most serious problems with existing data resources may be classified into the following categories:

1. Fragmentation and lack of coordination (applies to both past and current data production efforts) in the quantity and quality of data currently being collected, analyzed, and disseminated.
2. Absence of data gathered continuously and consistently that would support research on trends affecting both current and prospective humanities practitioners, and in the numbers seeking training at various degree levels across all humanities fields. Typical causes of this problem include:
  - a. limitations in the scope of data collections to narrowly targeted subsets of the complete universe (for example, studies based on members of institutional associations that cannot be generalized to the entire population of similar institutions);
  - b. population-based studies with sample sizes too small to permit separate evaluations of different humanities fields; and
  - c. disruption or discontinuation of the periodic data collections necessary for time-series analyses of trends in humanities fields.
3. Limited availability of existing datasets to researchers, complicated by:
  - a. long delays between the collection and the release of national data systems,
  - b. lack of straightforward data distribution mechanisms, requiring lengthy, complex negotiations over terms and costs for researchers to gain access to certain datasets, and
  - c. proprietary restrictions on many datasets that limit use to the funders or producers of the data.
4. Structures and frameworks for storing, managing, and accessing humanities data that impede or obviate the use of statistical research methods. This is the typical situation for the massive on-line bibliographic data systems organized for searching and listing, rather than for quantitative analysis of their contents.
5. Persistence of inconsistencies in basic concepts and definitions employed by data producers in collecting data and/or reporting results.

The problems noted above affect each of the four functional groupings of data sets differently.

- “Bibliographic Materials” (used to analyze scholarly work and writings

in the humanities disciplines) are hampered by limitations on access (problems 3b and 3c) as well as data structures and access mechanisms that create enormous startup burdens on researchers who must convert the data to a different format to conduct statistical analyses (problem 4).

- “Directories and Catalogs” (used for research on academic institutions, departments, programs, and other organizations such as libraries, presses, and museums) suffer from fragmentation and lack of coordination across disciplines (problem 1), the absence of consistent, long-term trend data (problem 2), and proprietary restrictions on access to data (problem 3c), as well as significant inconsistencies across resources in basic definitions (problem 5).
- “Research Datasets” (designed from the outset for quantitative analysis) are affected by fragmentation and lack of coordination (problem 1), absence of data at required levels of detail or disaggregation (problem 2), all manner of limitations on data access (problems 3a-c), and inconsistencies across data systems in basic definitions (problem 5), which hinder research on the past, present, and future of the humanities labor force and job markets.
- “Reports and Publications” (compilations of statistical tables, with or without interpretive text) suffer primarily from the lack of needed types of data and definitional inconsistencies between the contents of the reports and the data needed by leaders and practitioners in humanities fields (problems 2 and 5).

We are aware that researchers studying certain specific humanities fields or practices (for example, faculty teaching literature or foreign languages, academic librarians and staff) have data resources that are well designed to answer questions that might be posed by academics, administrators, and practitioners about labor force or job market conditions in their fields. Unfortunately, researchers are unable to answer the same types of questions about the full range of humanities disciplines, or to compare current conditions with past trends across fields.

### *Comparison between Humanities and Science and Engineering Fields*

The situation for researchers on the humanities compares unfavorably with the experience of researchers attempting to answer similar questions about science and engineering fields. We can illustrate this difference by attempting to answer a few basic questions drawn from the issues listed in Section 1:

#### *1. Trends in Graduate Education*

*Over the last thirty years, how many advanced (master’s and doctoral) degrees have been awarded in humanities fields?*

This question can be answered for doctorates in humanities fields using data from the *Survey of Earned Doctorates* from 1958 to the present time.

However, researchers must first obtain a license from the National Science Foundation to obtain the microdata required to conduct the analyses.

Although summary reports on humanities degrees have been produced for decades, no existing reports are available that disaggregate degrees by specific humanities fields. Counts of masters' and doctoral degrees awarded between the mid-1980s and mid-1990s are not reported separately for history (which is included in the social sciences), philosophy, or religion. Beginning in the mid-1990s, advanced degrees in philosophy and religion are reported separately, but degrees in history are still reported as part of the social sciences.

Individual science and engineering fields can at least obtain a reasonably disaggregated answer to this question. *The Science and Engineering Indicators 2000 (SEI 2000)* report displays time-series graphs of the numbers of master's and doctoral degrees awarded in all science and engineering fields for the thirty-year period, 1966–1996.<sup>37</sup> Separate counts over this period are available for broad scientific subfields, including

- physical and geosciences
- biological and agricultural sciences
- mathematics
- computer sciences
- engineering
- social sciences
- psychology

## 2. *Higher Education Faculty in the Humanities*

*How many faculty members are employed at U.S. higher education institutions in humanities fields?*

Although some associations of humanities practitioners have collected data to address this question, it cannot be answered for all humanities fields. The most systematic and current data on faculty counts are collected by the National Center for Education Statistics in the *Integrated Postsecondary Education Data System (IPEDS) Fall Staffing Survey*. This data system collects information on sex, race/ethnicity, type of institution, and position for all staff employed at postsecondary institutions. However, it does not collect any information about teaching fields of faculty.

The science and engineering fields have a slight advantage in addressing this question. The *SEI 2000* report displays data on faculty appointments of new Ph.D.'s for seven science and engineering fields. However, it does not report counts of the entire faculty work force in science and engineering. Other than this limited information, there is no source of nationally representative information on counts of faculty by field over time.

37. Data points are displayed for five-year increments for 1966–1986 and one-year increments for 1986–1996.

### 3. Workforce Characteristics

*How many people who hold degrees in the major humanities fields are currently employed in those fields?*

No existing publications provide answers to this question for specific humanities fields. Summary reports for the *Humanities Survey of Doctorate Recipients* (SDR) provide limited reporting on employment and unemployment for humanities Ph.D.'s, but do identify or count those employed outside their fields. No existing reports address employment patterns for holders of the master's degree in humanities fields.

Science and engineering practitioners have considerably better data resources to address this question. Tables 3.1 and 3.2 of the *SEI 2000* report display counts of persons with science and engineering degrees who were employed in science and engineering fields in 1997, disaggregated into five categories:

- physical and geosciences
- biological and agricultural sciences
- mathematics and computer sciences
- engineering
- social sciences and psychology

These tables also show the proportion of individuals holding science and engineering degrees in these categories at the BA, MA, and Ph.D. level who are employed in fields that are “closely related,” “somewhat related,” and “not related” to their degree fields. The proportion with occupations closely related to their degree fields increases sharply from 29 percent of those with BAs in science and engineering fields, to 47 percent of those with MAs and 48 percent of those with Ph.D.'s in science and engineering fields.

### 4. Labor Market Indicators

4a. *What are the current annual average salaries earned by practitioners employed in the major humanities fields?*

The National Center for Education Statistics routinely reports on salaries of higher education faculty in humanities fields and library staff in its *National Survey of Postsecondary Faculty*. Summary data on earnings of Ph.D.'s in humanities fields were published through 1995 based on the *Humanities SDR*, the last year that survey was conducted. However, earnings were not presented separately for those working in humanities fields versus other fields. Moreover, no comparable data are routinely reported for those holding master's degrees in humanities fields.

The *SEI 2000* report (Figure 3-11) displays median annual salaries earned by scientists and engineers in 1997 broken down into the same five categories of science and engineering fields listed under item 3 above, and also by race/ethnicity. In addition, Tables 3-10 and 3-11 of that report display details on 1997 annual salaries by field and by employment sector.

4b. *How many individuals earning new BA or MA degrees in the major humanities fields are employed in those fields?*

No systematic data are available to answer this question, either for the humanities degree earners in general, or for specific humanities fields.

The *SEI 2000* report (Table 3–5) shows the percentages of new BA and MA recipients in science and engineering fields in 1995–96 who were employed in seven different employment sectors (including two educational sectors, two for-profit sectors, and three non-profit sectors) separately for science and engineering degree earners.

4c. *What are the unemployment rates for those with degrees in humanities fields? Do humanities graduates frequently take employment outside their fields of training?*

Unemployment rates for Ph.D.'s in humanities fields may be estimated using data from the *Humanities Survey of Doctorate Recipients* (SDR) from 1977 through 1995 (biennially). With the exception of limited survey data on new Ph.D.'s in English and foreign languages, no data on unemployment or on “involuntary out-of-field employment” are available for humanities fields.

Table 3–6 of the *SEI 2000* report shows both unemployment rates and “involuntary out-of-field” employment rates for 1995 and 1997 for five separate engineering fields and twelve separate scientific fields. Table 3–9 shows similar data focusing on the most recent Ph.D. recipients in science and engineering fields.

4d. *What are the prospects for future job growth in the humanities?*

No systematic data are available to address this question for the humanities in general, or for the major humanities fields in particular.

The *SEI 2000* report (Table 3–20) indicates the total numbers of jobs in nine science and engineering fields in 1998 and projected for 2008, as well as the change in expected numbers of positions over the ten-year period.

##### *5. Public Awareness, Knowledge and Involvement in the Humanities*

*How has public awareness of and interest in humanities issues changed over the last ten years? How does the interest of the American public in these issues compare with other countries?*

No time-series survey data are available on public awareness of or knowledge about issues of relevance to the humanities. In fact, there has been little or no systematic research to identify the issues of greatest relevance to humanities practitioners; instead, issues of public relevance are generally labeled as “political,” “economic,” or “social” in nature.

The *SEI 2000* report (Figure 8–1) displays a series of graphs for five surveys (1990, 1992, 1995, 1997, and 1999) showing levels of public interest and self-reported knowledge concerning eleven different scientific and technological issues, ranging from space exploration, through new scientific discoveries, to environmental pollution. Figure 8–2 uses 1999 survey data to show

breakdowns of public interest and knowledge in science and technology issues by sex and educational attainment.

In addition, Table 8-1 of the report compares levels of public interest among U.S. citizens in the same eleven science and technology issues plus six other general issue areas (such as sports, politics, taxes, etc.) with levels of public interest measured in the European Union, Japan, and Canada.

## CONCLUSIONS

Data resources focusing on the humanities are incapable of delivering policy-relevant information because of deficiencies in technical quality, coverage, comparability across disciplines or over time, and accessibility to researchers. Although the datasets listed in the *Index of Humanities Datasets* provide (or once provided) significant value for their producers and primary users, most were not designed to support cultural policy research, and, in most cases, were not designed for use as research resources for quantitative analyses.

The lack of adequate data systems for policy analysis of the humanities does not result from any unwillingness on the part of humanities organizations and practitioners to gather and analyze data on their fields. On the contrary, one of the major problems affecting humanities information resources stems from the fact that too many parties with different purposes and orientations are collecting too much data, rather than not enough. The lack of articulation of effort, and in some cases the outright competition among data producers, has resulted in significant levels of duplication on the one hand, and inconsistency of measurement on the other, posing severe challenges for those who seek to combine these disparate resources into a coherent whole.

In consequence, the humanities are, at present, lagging substantially behind the science and engineering disciplines in developing the systems of data resources needed to monitor even the most basic trends in the number, activity levels, and productivity of, and the future prospects for, practitioners in humanities fields at all degree levels. Even in the science and engineering disciplines, despite the fact that they now possess more and better data, resources leaders remain unsatisfied with their ability to understand the state of their own fields and the roles those fields play both in academia and in the broader economy. They continue to seek better data and methods to develop more accurate forecasts of supply and demand for science and engineering graduates.

In a recent report, the National Research Council's Committee on Methods of Forecasting Demand and Supply of Doctoral Scientists and Engineers recommended that the National Science Foundation support efforts to improve the quality of forecasting data and methodology by taking a leadership position in collecting and distributing improved data to researchers (NRC 2000). The Committee further recommended that NSF undertake a comprehensive review of data collection to support forecasting, and observed that data collected for forecasting should be openly available and distributed to researchers on a timely basis. Several specific data-gathering innovations were considered, including significant increases in the detail

of the data compared to that collected in the *Survey of Doctorate Recipients*. Significantly, Committee members, researchers, and policy makers agreed that existing data were not sufficiently up-to-date to support policy decisions. They also drew attention to the fact that existing data and analyses failed to serve the needs of several neglected user communities, especially those of students attempting to plan research careers and of the public and private organizations and associations that provide career guidance.

Leaders in humanities disciplines face an identical situation, but must begin their efforts from a less favorable starting position. To improve the capacity and infrastructure for policy research, humanities leaders must confront the same issues and adopt many of the same strategies now being advocated by the science and engineering fields to address similar questions. Two critically important considerations should inform these efforts.

First, there is no one-time solution to the data resource problems described in this report. To succeed, the response must be based on a long-term perspective. Initial efforts toward the design of an infrastructure to support policy research on the humanities should be undertaken with the awareness that the development of data resources is an ongoing process that will require continuous investment of resources for the foreseeable future.

Second, as has been amply demonstrated in the science and engineering fields, sustained progress toward improving the infrastructure for policy research on the humanities must be placed within an administrative structure capable of pursuing funding for and managing the resources of a developmental enterprise for the benefit of all stakeholders and constituencies. Without such an entity, reliance upon the self-interest of potential data users to spark organized efforts to address and surmount existing problems will leave the humanities data infrastructure in its current state indefinitely.

## Appendix A.

Replica of the *Index of Humanities Datasets* Page  
formerly at the Web Site of the National Endowment for the Humanities

### *Index of Humanities Dataset Descriptions*

AACU Curriculum Database  
AAU/AGS Project for Research on Doctoral Education  
AAUP On-Line Catalog  
ACRL University Library Statistics  
APA Guide to Graduate Programs in the Classics  
ARL Annual Salary Survey  
ARL Statistics  
ARTbibliographies Modern  
ATLA Religion Database  
Academic Library Survey  
Advanced Placement Summary Reports  
America - History and Life  
American Art Directory  
American College Teacher  
American Freshman - National Norms  
American Library Directory  
Annual Register of Grant Support  
Annual Reports of the State Humanities Councils  
Annual Survey of Colleges  
Archives USA  
Arts and Humanities Citation Index  
Association of American University Presses Directory  
Avery Index to Architectural Periodicals  
Awards Almanac  
Baccalaureate and Beyond Longitudinal Study  
Beginning Postsecondary Student Study  
Bibliography of the History of Art  
CGS/GRE Survey of Graduate Enrollment  
Characteristics of GRE Test-Takers  
College Blue Book  
College Bound Seniors: Profiles of SAT and Achievement Test Takers  
Community College Curriculum Studies  
Condition of Education  
Corporate 500 - Corporate Philanthropy  
Current Contents  
Departmental Administrators in the Modern Languages  
Digest of Education Statistics  
Directory of American Philosophers 1998–1999  
Directory of Archives and Manuscript Repositories

Directory of Departments and Programs of Religious Studies in North America  
 Directory of Graduate Programs in American Studies  
 Directory of Grants in the Humanities  
 Directory of Historical Organizations in the United States and Canada  
 Directory of History Departments and Organizations  
 Directory of Programs in Linguistics in the U.S. and Canada  
 Dissertation Abstracts International  
 Doctoral Programs in Theatre Studies, Performance Studies, and Dance  
 Educational Resources Information Center (ERIC)  
 Ethics Index  
 Expanded Academic Index  
 Faxon Finder  
 Folklife Sourcebook - A Directory of Folklife Resources in the United States  
 Foundation Directory  
 Foundation Grants Index  
 GRE/CGS Directory of Graduate Programs, Volume D  
 George Eastman House Interactive Catalog  
 Giving USA  
 Graduate Programs in Art History and the Visual Arts  
 Grants on Disc  
 Guide to Ethnomusicology Programs in the U.S. and Canada  
 Guide to Graduate Degree Programs in Architectural History  
 Guide to Graduate Programs in Philosophy  
 Guide to U.S. Graduate Programs in History of Science  
 Guide to the Field of Folklore  
 High School and Beyond  
 Historical Abstracts  
 Humanities Doctorates in the United States (parallels Science & Engineering Survey of Doctorate Recipients)  
 IIE Reports on Foreign Students  
 IPEDS Completions Survey  
 LCMARC Database  
 Linguistics and Language Behavior Abstracts  
 MLA Directory of Periodicals  
 MLA International Bibliography  
 Magill's Cinema Surveys  
 Minority On-Line Information Service  
 Museum Financial Information Survey  
 Music Article Guide  
 NASULGC Faculty Distribution Survey by Race and Sex  
 NASULGC Faculty Salary Survey  
 National Education Longitudinal Study of 1988  
 National Faculty Salary Survey by Discipline and Rank (CUPA)  
 National Longitudinal Study of the High School Class of 1972  
 National Museum Survey of 1989

National Postsecondary Student Aid Study  
National Study of Postsecondary Faculty  
National Survey of Faculty (Carnegie)  
National Survey of Graduate Assistantships  
New Testament Abstracts  
Newsletters in Print  
Nexis  
Nonprofit Almanac - Dimensions of the Independent Sector  
OCLC Online Union Catalog  
Official Museum Directory  
Old Testament Abstracts  
Peterson's College Database  
Peterson's Gradline  
Philosopher's Index  
Prospector's Choice  
Public Library Statistics  
RILM Abstracts of Music Literature  
RLIN  
Research-Doctorate Programs in the United States  
Social Sciences Citation Index  
Survey of Earned Doctorates  
Survey of Income and Program Participation  
The Philosophers Phone and Email Directory 1998-1999  
Ulrich's International Periodicals Directory  
UnCover  
Women's Studies Abstracts

## Appendix B.

### Dataset Description for the AACU Curriculum Database from *Index of Humanities Dataset Descriptions*

*Name of Dataset:* AACU Curriculum Database (#05536)

*Name and Address of Organization Maintaining Dataset:* University of Pennsylvania; Office of Institutional & Policy Planning; 4200 Pine Street, 5A; Philadelphia, PA 19104; Phone: 215-898-5897; Fax: 215-898-9876;

*Description:* The Association of American Colleges and Universities, in conjunction with the University of Pennsylvania's Institute for Research on Higher Education, has assembled a computerized database containing transcript and other data on the 1991 graduating classes at a nationally representative sample of 81 colleges and universities. Requests for access to the database are evaluated on a case-by-case basis.

*Humanities Disciplines Covered:*

*Specific Humanities Fields:* Yes

*Humanities Subfields:* No

*Broad Disciplines:* Yes

*Humanities Name:*

*Fields Covered:* Visual/Studio/Fine Arts; Foreign Languages and Literature (includes classical languages); History; Performing Arts; Anthropology; Political Science, Policy Studies, & International Relations; Sociology; Psychology; Other Social Sciences; Education; Professional Fields; Life Sciences; Natural Sciences.

*Humanities Fields in Broad Discipline Category Include:* Art (and Architectural) History and Criticism; English & American Language and Literature; Comparative Literature; Linguistics; Philosophy; Religion; Other Humanities.

*Topics Covered:*

Sex; Racial and ethnic identity; Age/year of birth; Year in which began college/graduate study; Courses taken; Years of study and/or number of credits completed in specific areas; Major or probable major; Grades earned in specific courses; Grade point average; Placement test scores and/or other test scores; Source of funding/control; Institutional type and level (e.g., Carnegie type or two- & four-year); Enrollment (i.e., number of students); Academic calendar, type used; Number of degrees awarded by institution/department; Departments' degree requirements.

*Entire data set available:* No

*Access to information in data set is provided through:*

*Specially prepared computer runs or tables:* No

*Special requests by researchers for access:* Yes

*Published reports:* Yes

See Below for Details About Available Formats, Size of Data Set, Period Covered, and Sampling Procedure.

*Data Set Elements*

*Number in data set:* 42,007

*Type:* Recent college graduates

*Description:* N = all 1991 graduates at 81 institutions

*Procedures used to correct for bias:* Weighting

*Type of data set:* Discrete

*Intervals in which data collected:* Every few years

*Are comparison data available?:* No

*Format(s) in which data set is available*

*Special Requests by Researchers for Access:*

*Is a formal written application required?:* Yes

*Must request be reviewed by oversight body?:* Yes

*Address (if differs from address at beginning of form):*

*Published Reports:*

*Are data collection procedures delineated?:* Yes

*Are the nature of the data used in individual tables and their limitations explained?:* Always

*Frequency of Use For:*

*Broad Disciplines:* Usually

*Humanities Fields:* Sometimes

*Humanities Subfields:* Not relevant

*Address for ordering publications (if different from address at beginning of form):*

*Titles of Major Reports:*

*Curriculum Assessment Service Database: Estimates of Student Curricular Activity from a National Survey of Colleges and Universities*

*Record last updated:* 10/10/96 18:17:56

## Appendix C.

### Resources in the *Index of Humanities Datasets* Sorted by Functional Grouping

#### Bibliographic Material including Archives (35)

##### *General Resources that Include Humanities Materials (16)*

AAUP On-Line Catalog  
Archives USA  
Current Contents  
Directory of Archives and Manuscript Repositories  
Dissertation Abstracts International  
Educational Resources Information Center (ERIC)  
Expanded Academic Index  
Faxon Finder  
LCMARC Database  
Newsletters in Print  
Nexis  
OCLC Online Union Catalog  
RLIN  
Social Sciences Citation Index  
Ulrich's International Periodicals Directory  
UnCover

##### *Resources Explicitly Focusing on One or More Humanities Fields (19)*

MLA Directory of Periodicals  
MLA International Bibliography  
ATLA Religion Database  
America - History and Life  
ARTbibliographies Modern  
Arts and Humanities Citation Index  
Avery Index to Architectural Periodicals  
Bibliography of the History of Art  
Ethics Index  
George Eastman House Interactive Catalog  
Historical Abstracts  
Linguistics and Language Behavior Abstracts  
Magill's Cinema Surveys  
Music Article Guide  
New Testament Abstracts  
Old Testament Abstracts  
Philosopher's Index  
RILM Abstracts of Music Literature  
Women Studies Abstracts

*Directories and Catalogs (36)*

*Academic Departments, Programs, Presses (20)*

Association of American University Presses Directory  
College Blue Book  
GRE/CGS Directory of Graduate Programs, Volume D  
Minority On-Line Information Service  
Peterson's College Database  
Peterson's Gradline  
APA Guide to Graduate Programs in the Classics  
Departmental Administrators in the Modern Languages  
Directory of Departments and Programs of Religious Studies in North America  
Directory of Graduate Programs in American Studies  
Directory of History Departments and Organizations  
Directory of Programs in Linguistics in the U.S. and Canada  
Doctoral Programs in Theatre Studies, Performance Studies, and Dance  
Folklife Sourcebook - A Directory of Folklife Resources in the United States  
Graduate Programs in Art History and the Visual Arts  
Guide to Ethnomusicology Programs in the U.S. and Canada  
Guide to Graduate Degree Programs in Architectural History  
Guide to Graduate Programs in Philosophy  
Guide to the Field of Folklore  
Guide to U.S. Graduate Programs in History of Science

*Directories of Funding (10)*

Annual Register of Grant Support  
Awards Almanac  
Corporate 500 - Corporate Philanthropy  
Directory of Grants in the Humanities  
Foundation Directory  
Foundation Grants Index  
Giving USA  
Grants on Disc  
Nonprofit Almanac - Dimensions of the Independent Sector  
Prospector's Choice

*Directories of Nonacademic Organizations (4)*

American Art Directory  
American Library Directory  
Directory of Historical Organizations in the United States and Canada  
Official Museum Directory

*Directories of Individuals (2)*

Directory of American Philosophers 1998–1999

The Philosophers Phone and Email Directory 1998–1999

*Research Datasets (34)*

AACU Curriculum Database

AAU/AGS Project for Research on Doctoral Education

Academic Library Survey

ACRL University Library Statistics

Advanced Placement Summary Reports

American College Teacher

American Freshman - National Norms

Annual Survey of Colleges

ARL Annual Salary Survey

ARL Statistics

Baccalaureate and Beyond Longitudinal Study

Beginning Postsecondary Student Study

CGS/GRE Survey of Graduate Enrollment

Characteristics of GRE Test-Takers

College Bound Seniors: Profiles of SAT and Achievement Test Takers

Community College Curriculum Studies

High School and Beyond

Humanities Doctorates in the United States (parallels Science and

Engineering Survey of Doctorate Recipients)

IIE Reports on Foreign Students

IPEDS Completions Survey

Museum Financial Information Survey

NASULGC Faculty Distribution Survey by Race and Sex

NASULGC Faculty Salary Survey

National Education Longitudinal Study of 1988

National Faculty Salary Survey by Discipline and Rank (CUPA)

National Longitudinal Study of the High School Class of 1972

National Museum Survey of 1989

National Postsecondary Student Aid Study

National Study of Postsecondary Faculty

National Survey of Faculty (Carnegie)

National Survey of Graduate Assistantships

Public Library Statistics

Research-Doctorate Programs in the United States

Survey of Earned Doctorates

Survey of Income and Program Participation

*Publications or Compilations (3)*

Annual Reports of the State Humanities Councils

Condition of Education

Digest of Education Statistics

Appendix D.

Task Force on Data Development  
[*Humanities Indicators*]

Jonathan Cole	Columbia University
Robert Connor	National Humanities Center
John D'Arms	American Council of Learned Scientists
Phyllis Franklin	Modern Language Association
John Hammer	National Humanities Alliance
Arnita Jones	American Historical Association
Steven Marcus	Columbia University
Francis Oakley	Williams College
Robert Post	University of California, Berkeley
Kenneth Prewitt	New School University
Stephen Raudenbush	University of Michigan
Robert Solow	Massachusetts Institute of Technology
Jeffrey Thomas	National Endowment for the Humanities
Steven Wheatley	American Council of Learned Societies

*Ex officio*

Leslie Berlowitz	American Academy of Arts and Sciences
Malcolm Richardson	American Academy of Arts and Sciences/ National Endowment for the Humanities

# Contributors

Robert M. Solow is Professor Emeritus of Economics at the Massachusetts Institute of Technology and the 1987 Nobel Prize recipient for his work in economics. Professor Solow is a Fellow of the Academy and a member of the Academy's Task Force on Data Development.

Francis Oakley is President Emeritus of Williams College and a Fellow of the American Academy of Arts & Sciences. He leads the Academy's Task Force on Data Development and serves on the steering committee for the Academy's Initiative for the Humanities and Culture.

John D'Arms was President of the American Council of Learned Societies and a Fellow of the Academy. He served on the advisory committee for the Academy's data development project.

Phyllis Franklin is Executive Director of the Modern Language Association and serves on the advisory committee for the Academy's data development project.

Calvin C. Jones is President of Statistical and Evaluation Research, a consulting firm based in Fairfax, Virginia. He serves as consultant to the Academy on database development and is the author of the report to the Academy evaluating current data collection efforts in the humanities.

