

International Symposium 2009 REPORT

Data Archives and International Collaboration
for Research, Public Policy, and Business:

As a witness of the Launch of Educational Data Archive of
Educational and Social Survey Research Center
Hyogo University of Teacher Education

Date: November 22nd, 2009 Sunday 10:00-16:00
Venue: International Conference Room 301,
Kobe International Conference Center
6-9-1 Minatojima-nakamachi, Chuo-ku, Kobe
650-0046, Japan

	International Symposium2009 PROGRAM P. 132
	OBJECTIVE P. 133
	OPENING MESSAGE Eiichi Kajita, President, HUTE P. 135
	Chapter 1 –I Current status and future prospects of leading data archives in Japan, the United States, and Europe, and explanation of the ESSRC, HUTE Educational Data Archive	
	● PETER GRANDA	
	Challenges in Archiving Education Data: The ICPSR Experience P. 140
	● EKKEHARD MOCHMANN	
	From National Data infrastructures to International Data Infrastructure Networks P. 151
	● HIROSHI ISHIDA	
	SSJDA (Social Science Japan Data Archive) and Activities at the Center for Social Research and Data Archives P. 165
	● YASUO WATANABE	
	Launching of Educational Data Archive “JEDI” and the Networking of Academic Data Archives P. 173
	Chapter 1 –II Assessment of and comments on the Educational Data Archive (Conference minutes) P. 186
	Chapter 2 Presentation and Discussion	
	● MASUO KOYASU	
	The significance of data preservation and dissemination in the area of longitudinal cohort studies P. 192
	■ Discussion I: Data Archives and Academic Research (Conference minutes) P. 194
	● TAKANORI MIYAKE	
	Using Data to Formulate Education Policies – With reference to nationwide surveys of academic ability and learning context (Conference minutes) P. 206
	■ Discussion II: Data Archives and Public Policy on Education (Conference minutes) P. 210
	● KENICHI ARAI	
	Data Archives and Educational Industry P. 219
	■ Discussion III: Data Archives and the Educational Industry (Conference minutes) P. 223
	OVERVIEW Yasuo Watanabe, Professor, HUTE P. 229
	CLOSING COMMENTS Eiichi Kajita, President, HUTE P. 230
	SYMPOSIAST CVs P. 234

International Symposium 2009

Symposiast:

PETER GRANDA	Archivist and Assistant Archival Director, Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan
EKKEHARD MOCHMANN	Former Director International Data Service, GESIS-Leibniz Institute for the Social Sciences
HIROSHI ISHIDA	Professor, Institute of Social Science, The University of Tokyo Vice president, Japanese Association for Mathematical Sociology
MASUO KOYASU	Professor, Kyoto University President, Japan Society of Developmental Psychology
TAKANORI MIYAKE	Senior Specialist, Office for Assessment of Academic Ability, Elementary and Secondary Education Bureau, MEXT-Japan
KENICHI ARAI	Corporate Senior Vice President, Benesse Corporation, Japan Director, Benesse Educational Research and Development Center
YASUO WATANABE	Professor, Hyogo University of Teacher Education Director, Educational and Social Survey Research Center

Program:

10:00 am - 10:10 am	Opening message (Eiichi Kajita, President, HUTE)
10:10 am - 11:45 am	Current status and future prospects of leading data archives in Japan, the United States, and Europe, and explanation of the ESSRC, HUTE Educational Data Archive (20-min. presentation for each archive)
11:45 am - 12:00 noon	Assessment of and comments on the Educational Data Archive
12:00 noon - 1:00 pm	Lunch
1:00 pm - 1:50 pm	Discussion I: Data Archives and Academic Research (40 min. after 10-min. presentation by Masuo Koyasu, Kyoto Univ.)
1:50 pm - 2:00 pm	Intermission
2:00 pm - 2:50 pm	Discussion II: Data Archives and Public Policy on Education (40 min. after 10-min. presentation by Takanori Miyake, MEXT-Japan)
2:50 pm - 3:00 pm	Intermission
3:00 pm - 3:50 pm	Discussion III: Data Archives and the Educational Industry (40 min. after 10-min. presentation by Kenichi Arai, Benesse Corp.)
3:50 pm - 3:55 pm	Overview (Yasuo Watanabe, Professor, HUTE)
3:55 pm - 4:00 pm	Closing comments (Eiichi Kajita, President, HUTE)
(4:30 pm - 6:30 pm)	Reception [admission: J¥2,000]



Educational and Social Survey Research Center (ESSRC),
Hyogo University of Teacher Education (HUTE)
Outline of International Symposium 2009

Theme: Data Archives and International Collaboration for Research, Public Policy, and Business:
As a Witness of the Launch of Educational Data Archive of ESSRC, HUTE

Objective:

The ESSRC was established in April 2005 as a research institution affiliated with HUTE based on a Special Education Research Grant from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT-Japan), with a five-year term that finishes at the end of March 2009. During this period, ESSRC has focused on developing an educational data archive, as well as on gathering empirical data and making it available, and has worked to implement surveys related to education and society, hold international symposia and training sessions on survey methods, train young researchers, and develop an educational survey curriculum.

As a way to wrap-up the five-year activities of ESSRC and to pass down its ideas and legacy to its successors, we plan to hold the International Symposium 2009 on Sunday, November 22, 2009 in the autumnal setting of Kobe. The theme will be “Data Archives and International Collaboration for Research, Public Policy, and Business: As a Witness of the Launch of Educational Data Archive of ESSRC, HUTE.” We will focus particularly on the development and operation of educational data archive systems and invite researchers representing advanced data archives from the United States, Europe, and Japan as well as representatives of academia, government, and the educational industry in Japan. The symposium will provide the opportunity to exchange information and opinions on promoting the utilization of data archives and collaboration among related institutions in Japan and abroad.

Our aim is to gather and save the valuable survey data concerning education that would be scattered, lost, or obsolete if the ESSRC’s educational data archive development project were to be abandoned in its current state, to make these data available on the Internet and elsewhere, and to encourage their secondary use for empirical research. The data archive system developed by the ESSRC enables four levels of disclosure and access methods, ranging from free access and download of stored data over the Internet to restricted use under surveillance with no duplication whatsoever permitted, and including conventional usage methods intermediate between these levels, to meet the diverse needs of both those who use and those who deposit the data. The use of stored data is free of charge and limited to the objectives for research, education, or public policy, and data may not be used for commercial purposes. The educational data archive system constitutes an indispensable foundation if educational discussion is to move forward from its existing basis in individual experience and conjecture or speculation to empirical debate based on objective evidence and robust analysis.

In Europe and the United States, an archive system that links the archives of a number of universities and research institutions over a network has accumulated results and contributed in a wide-ranging manner to the development of academic research, policy proposals, and statistical and mathematical education. In Japan, however, few data archives have as yet been established and their history is still slight, in addition to the fact that they all function independently. The ESSRC provides the data archive system it has developed to a wide range of Japanese universities and research institutions free of charge, with the objective of contributing to advancing academic data archives in Japan and to their networking.

This fall's International Symposium 2009 brings together the diverse range of panelists listed below to speak to an audience including students, faculty, researchers, administrators, and private business operators, where they will offer a range of reports and engage in discussion about data archives. During the morning, representatives of the advanced European data archive GESIS, the Inter-university Consortium for Political and Social Research (ICPSR) in the United States, and the University of Tokyo's Social Science Japan Data Archive (SSJDA) will report on the current status and future prospects of their own archives. The ESSRC will also report on the functioning of the educational data archive and make a statement, and will sum up its five years of activities in light of the comments made by the speakers. In the afternoon, representatives of the Japan Society of Developmental Psychology, MEXT-Japan, and Benesse Corporation will raise questions and make statements concerning data archive utilization from their respective standpoints representing academia, the government, and industry, after which a discussion will take place involving all the panelists. The reception following the symposium will offer the chance for participants to mingle with panelists and exchange information and opinions. This event will reaffirm the role of data archives, in the sense that they enable academic research, teaching practice, and policy promotion in Japan to be based on empirical proof, and we anticipate that it will provide a highly significant opportunity as a call for collaboration to develop data archives and their networking further in future.

Languages: Simultaneous interpretation between Japanese and English

Organizer: Educational and Social Survey Research Center, Hyogo University of Teacher Education

Supporters: MEXT-Japan, Boards of Education of six prefectures (Osaka, Kyoto, Hyogo, Nara, Shiga, Wakayama) and two cities (Osaka and Kobe) in the Kinki region, Japan Association of Universities of Education (JAUE), Association of ICPSR Users in Japan, Benesse Corporation, NEC Corporation

Opening message

Eiichi Kajita

President, Hyogo University of Teacher Education

Good morning. Thank you for joining us for this 2009 international symposium in spite of today being a holiday. The Hyogo University of Teacher Education (HUTE) organizes an international symposium annually at this time of the year, but it has special significance this year.

This is because our activities at the Educational and Social Survey Research Center (ESSRC), which was founded five years ago with a fixed five-year mission, as a research organization affiliated with HUTE, will culminate at the end of this fiscal year upon the termination of a special grant from the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Even though the ongoing activities will continue, they will no longer be covered by the special grant from MEXT. So, today's international symposium will serve as a final review of our activities so far at ESSRC. Besides conducting various research and survey projects, ESSRC has had another important mission of creating a data archive. ESSRC has already developed the framework of the data archive, and today we will describe its features. We welcome your critiques and comments.

From a very broad viewpoint, our attempt to create this educational data archive has been part of our effort to revolutionize Japan's educational policies and practices, teacher training and educational research activities in Japan. Unfortunately, past discussions on education have rarely been based on evidence.

When people discuss education, they tend to base their opinions on impressions. This "impressionism" is a very common pitfall. People may say, for example, that they saw stars in the eyes of children or that all students in the classroom looked "very alive." When they describe such impressions, they believe that they have said something very important. However, do stars in the eyes of children prove that they have gained some real understanding? This is not always the case. If children in the classroom look very alive, we should find out what makes them look so alive. It may simply be because they were in a mischievous mood, but such aliveness will not help the children to grow. In spite of this truth, expressions such as "very alive" and "stars in the eyes of children" have been used too often when discussing education. This is dangerous because phrases such as "stars in the eyes of children" have dominated not only discussions among teachers at schools but also discussions on educational policies. This is the background of the *yutori* education policy ("relaxed education" or "education without pressure" with shorter school hours and a smaller curriculum) that the Ministry of Education promoted in the 1990s. In our opinion, this is the result of "impressionism" in education. It is utterly inappropriate to discuss education on the basis of impressions.

Another pitfall is what we may call “emotionalism.” Since educationalists are often highly idealistic, they tend to put much emotion into what they say. They may say, for example, that we should care about the growth of children with whole-hearted commitment and true love. When expressing such opinions, they believe they are saying something very important about education, but such assertions do not really help. Even though it is true that love and whole-hearted commitment are necessary, these may not always support the academic achievement and human growth of children. To understand this, let’s think about our own childhoods. When teachers and parents were too keen to train us, didn’t we sometimes resist? It is important to properly determine the depth of our engagement with others, in this case children, based on love and whole-hearted commitment. But such an engagement is meaningless unless it truly supports the academic achievement and human growth of children. Imposing love on children is like harassment. We often hear people say that teachers should seek “close intimacy” with children. This might be good in winter but too sweaty in summer.

In my opinion, education should not be discussed in such terms. Discussions on education should be more reasonable, logical and coherent. Education should produce concrete results. But results may not come immediately; some results may come only after one year or even after 30 years. With a broad perspective on multiple timeframes, we must assess the results achieved in each specific period. For example, we must evaluate the results from each school year, from the six years in elementary school, from the nine years of compulsory education up to graduation from a junior high school, or from the total period of education up to graduation from high school or university. Only after working hard to determine the results of education in each specific timeframe can we successfully plan future actions and educational policies. Only by proceeding in this manner can we be successful.

Japan has not successfully followed this approach. Therefore, our educational data archive is designed to help evaluate the results of education in each specific timeframe and to make the data available as widely as possible as a common asset of all those who are interested. The data will assist discussions on educational policies and the governmental administration of education, and help teachers think about educational activities at their schools and about their efforts in the classroom. In short, our educational data archive is intended to ensure a correct awareness and sound judgment on education on the basis of firm evidence and to improve the planning of future actions.

Even though the creation of this educational data archive is an important achievement in itself, ESSRC has been pursuing the larger aim of revolutionizing attitudes toward education in Japan. Prof. Katsuno, now President of Gifu Pharmaceutical University, served as the Director of ESSRC with a deep understanding of the mission that I have described until last year after the Center was founded. Prof. Watanabe succeeded him as the next Director of ESSRC and brought the project to its final stage.

Today, we will report on our efforts to create this data archive at HUTE, and we would greatly appreciate your frank opinions. Our data archive still has some technical issues to be solved,

and we must consider how to continue our work and solve these issues in the next step. But we must also think of how best to use the data archive as a catalyst for revolutionizing the attitudes of the entire educational establishment in Japan. With all these topics in mind, we would like to discuss and hear your views on our future activities at ESSRC. I look forward to your active participation in today's discussions.



Chapter 1-I

Current status and future prospects of leading data archives in Japan, the United States, and Europe

Current status and future prospects of leading data archives in Japan, the United States, and Europe

Challenges in Archiving Education Data: The ICPSR Experience

Peter Granda

Archivist and Assistant Archival Director
Inter-university Consortium for Political and Social
Research (ICPSR) University of Michigan

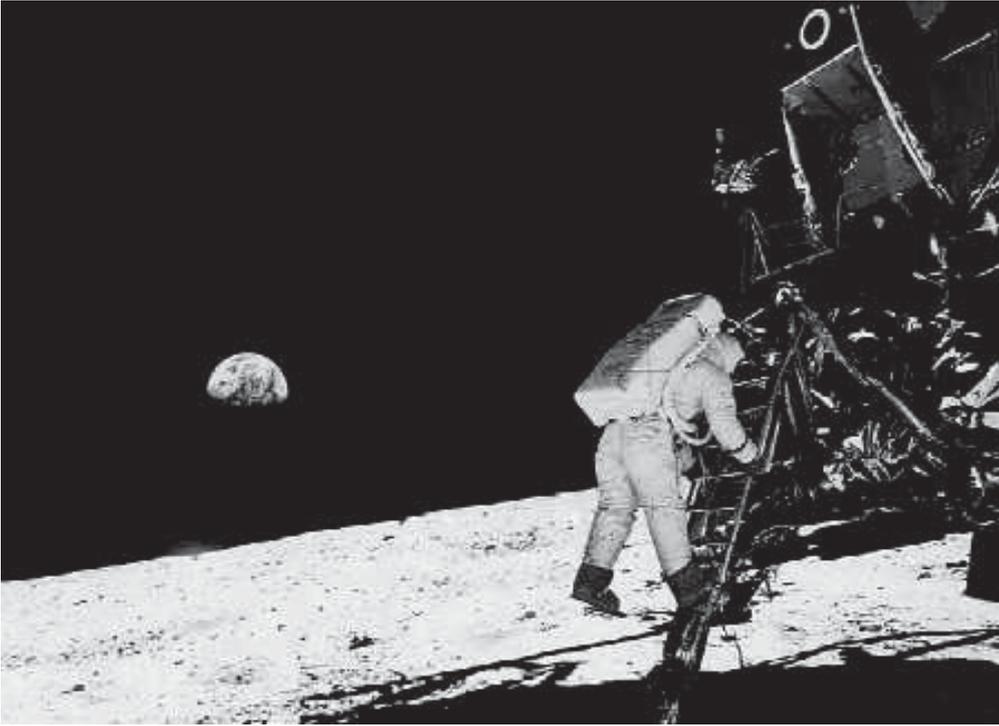
You might wonder why I am showing a slide of an event that happened 40 years ago that at first glance looks to have little to do with the theme and purpose of this symposium. But our theme does focus on “international collaboration” and no collaboration of any sort is possible without both preservation and sharing.

Perhaps this second photo will help. It shows the three Apollo 11 astronauts looking at film that they took during their mission to the moon in 1969. The story of what happened to the film is one that all archives can understand and appreciate. It turns out that the original film was inadvertently erased and reused to record data from subsequent space missions. Perhaps, the original footage was not properly labeled or catalogued. Perhaps someone was just careless but a valuable historical artifact would have been permanently lost except for the fact that other institutions, such as news organizations, had made copies of the original film. A restoration project was then able to recover the original footage and even use modern technologies to improve the quality of the images.

This is yet one more example, even though more are unnecessary, dramatizing how ephemeral the historical record might be if insufficient thought is given to the seemingly mundane questions of preservation and data sharing.

[2nd photo: Apollo 11 astronauts Neil Armstrong, Michael Collins and Buzz Aldrin examine film taken of their mission.
Credit: NASA]





ICPSR AND EDUCATION DATA

When education data is preserved and shared with others, it can make a very big difference in not only how researchers do their work but how government policies are enacted and implemented. How have these issues affected the acquisition, preservation, and dissemination of education data for the archival community? Let me present the experience of one social science data archive, the Inter-university Consortium for Political and Social Research (ICPSR).

Established in 1962, ICPSR is the world's largest archive of digital social science data with a mission to:

- Collect and preserve research data for future generations
- Disseminate data economically for further analysis and instruction by researchers, scholars and instructors
- Serve as a leading quantitative analysis training resource

I can think of no better example to illustrate the themes of data preservation, sharing, and international collaboration than an education survey done in the United States just a couple of years before the Apollo astronauts set out on their journey. This survey currently is archived at ICPSR.

A CASE STUDY: JAMES COLEMAN

The Equality of Educational Opportunity Study (EEOS) was funded by the U.S. Department of Health, Education and Welfare in response to provisions of the U.S. Civil Rights Act of 1964. The study was conducted by James Coleman, an eminent sociologist of education, who became so closely associated with this project that it became known as the "Coleman Study". Completed in 1966 this study sought to assess the availability of equal educational opportunities to children of different races, colors, religions, and national origins. The EEOS consists of test scores and questionnaire responses obtained from first-, third-, sixth-, ninth-, and twelfth-grade students, and questionnaire responses from teachers and the administrative heads of schools (called principals in the U.S.). These data were obtained from a national sample of schools in the United States. Data on students include age, gender, race and ethnic identity, socioeconomic background, attitudes toward learning, education and career goals, and racial attitudes. Scores on teacher-administered standardized academic tests are also included. These scores reflect performance on tests assessing ability and achievement in verbal skills, nonverbal associations, reading comprehension, and mathematics. Data on teachers and principals include academic discipline, assessment of verbal facility, salary, education and teaching experience, and attitudes toward race.

The collection comprises 7 datasets which contain almost 640,000 records. Its impact was immediate and long-lasting. This immense set of observations about the major actors in the task of educating students in public schools created some very controversial findings for both politicians and educators:

- When comparing students who had similar economic and family backgrounds, the quality of the school itself contributed little to student achievement
- The family background of the student was the most important factor in determining how well a student performed in school
- Minority (mainly non-White) children benefitted from attending high schools with white students
- The performance of the white students did not change if they attended schools that were racially mixed

Such a report, which questioned the value that schools and teachers had on students, was heavily criticized by academics and education practitioners alike.

But the more important point is that the original data was available for others to use and test these results. Many did and came to the same result as had James Coleman. The sharing of data combined with archival preservation permits others to test research hypotheses and review published results. In other words the concept of “replication” is encouraged (see, Barbara Schneider, “Building a Scientific Community: The Need for Replication”, Teachers College Record, Vol. 106, No. 7, July 2004, pp. 1471-1483). This is an extremely important principle in all of social science research and is only made possible when data are easily available as they are from the archival community. This continued availability encourages scholars to pursue “evidence-based education” that is the aim and purpose of the Educational and Social Survey Research Center (ESSRC) at Hyogo University for Teacher Education and all other important education archives throughout the world. These archives permit research investigations that result in policy decisions that may affect how national educational systems operate.

It also allows other researchers to take retrospective looks of this same dataset and draw new and different conclusions (see, James M. Towers, “Twenty-five Years After the Coleman Report: What Should We Have Learned?”, Contemporary Education, Vol. 63, No. 2, Winter 1992, pp. 93-95).

More importantly, these data are now a permanent part of the international educational research infrastructure. They are still accessible more than 40 years after they were produced and may invoke new kinds of research uses not dreamed of by the original investigators.

I would hazard a guess that almost all countries can point to a similar dataset or datasets that have served key roles in the history of their own educational systems.

This dataset is also emblematic of the long connection of between ICPSR and the American educational research infrastructure over the years.

ICPSR AND U.S. EDUCATIONAL STATISTICAL DATA

The Department of Education in the United States reports that it can trace its origins to 1867 with its original purpose the collection of information on schools and teaching that would help all of the individual states to

establish effective school systems. Over the last half decade, a specific branch within the Department of Education, the National Center for Education Statistics (NCES), has collected vast amounts of data on the infrastructure, policies, and effectiveness of the teaching and learning experience and to suggest possible improvements. Added to this we must include the vast amount of education-related data collected by academic researchers, commercial institutions, and those who devote their full energies to the practical work of teaching and learning.

Despite its importance to the well-being of all societies, educational research and the preservation of its underlying data sources have provided formidable challenges to the archival community. First, there is the great variety of these data sources: years ago they focused on statistical and survey data but increasingly there is a growing community of scholars who generate qualitative data in such forms as videos of classroom interactions and audio interviews. Secondly, there is variety of population groups which are the objects of researcher investigations: students, teachers, parents, communities, governments, all of whom have important roles in making decisions that directly affect individuals at all stages of their lives. Thirdly, there are very important confidentiality issues surrounding much educational research data. This concern does not affect the national and regional statistical summaries of the educational establishment itself, e.g., the number of schools in a particular area, the student population, the type of degrees awarded each year, the financial resources of the institution, and the composition of its staff. But whenever surveys are conducted of sample populations or, in more extreme cases, when audio and visual records of specific individuals are collected, the guarantees made to preserve the identity of each respondent, while allowing some form of researcher access, can be problematic. Finally, and perhaps because all of these reasons, the educational research community has not always demonstrated a culture of data sharing that is equal to that of their colleagues in such fields as political science and sociology.

ICPSR began to collect education data in the late 1970's and early 1980's. Not surprisingly, most of it centered on the United States with a smattering of data from other countries. The entire collection on the general theme of education encompasses about 400 distinct data collections which comprise several thousand data and documentation files. What was collected was an amalgam of government datasets and studies produced from individual investigator research grants. There was no systematic effort to collect certain types of data since in those days archiving education data was in its early stages. ICPSR was happy to acquire anything on this subject that it could obtain.

Beginning in the 1990's all kinds of data began to appear as CD-ROM products and on the World Wide Web for researchers to download. In the United States, the National Center for Education Statistics (NCES) took the lead in making their statistical and survey data available to the research community. A visit to their Web site now provides access to a wealth of information about the educational character of the entire country. Users can look for information on colleges that they might want to attend through a sophisticated browsing tool, compare the financial resources of different public school systems, and download a variety of datasets and reports. But, in 1996, NCES also had significant amounts of data stored on archival tapes which were largely inaccessible to researchers.

It was at this point that ICPSR joined with NCES in an effort to retrieve this information and put it in a form that would be useful for research purposes.

NCES initially provided funding to ICPSR to ingest hundreds of data collections on removable media (tapes, CD-ROMs) and several thousand pages of NCES paper documentation. The archiving process included the transfer of all digital materials to disk, the conversion of all paper materials into electronic form, and the processing of data and documentation files along with descriptive metadata for release to the research community.

During the course of this project which ran for almost a decade, ICPSR processed 330 data collections, which comprise 1,112 data files and 1,895 documentation (e.g., codebooks, user guides, SAS and SPSS setups) files. All were available free of charge for downloading to the research community.

As part of its arrangement with NCES, ICPSR processed these data collections to conform to the most recent set of NCES statistical standards (NCES Publication 2003-601; Section 7: Dissemination of Data). ICPSR also added value by creating additional formats and products such as the creation of “full-suite” collections that not only provide ASCII data but also SAS, SPSS, and Stata setup files and SPSS Portable, SAS Transport, and Stata System files so that researchers can simply download, click, and open files directly in their favorite statistical software package.

RECENT EDUCATIONAL ARCHIVE INITIATIVES AT ICPSR

More recently ICPSR has expanded its involvement into other areas of education research moving beyond the world of aggregate and survey data. More than five years ago ICPSR obtained funding from the Child Care Bureau, Office of Family Assistance, and the Office of Planning, Research and Evaluation, Administration for Children and Families, in the U.S. Department of Health and Human Services to create a new data archive entitled Child Care and Early Education Research Connections (CCEERC; www.researchconnections.org) whose goal is to promote high quality research and the use of that research in policymaking. The emphasis on reaching the policymaking community reaches purposely to a different audience for archival products – one that is particularly important in the field of education where research and policymaking is often closely intertwined.

This new archive is operated by ICPSR in collaboration with the National Center for Children in Poverty at the Mailman School of Public Health, Columbia University. Through the project’s Web site, researchers can quickly search the full text of thousands of resources relevant to the field of child care and early education. Interactive tools allow users to refine their searches, download full text documents, build customized tables on state policies, compare state demographics, and analyze research data online.

This comprehensive collection includes scholarly research, policy briefs, government reports, data, and instruments from a wide range of disciplines and sources, including multiple federal agencies. *Research Connections* compiles and distributes bibliographies, develops issue briefs, and synthesizes research on key topics. *Research Connections* also offers public access to child care and early education research data, some of which have never before been publicly available. Researchers can download analysis-ready data directly to their desktop or analyze selected data online free of charge.

Data training workshops and an archive for research projects serve the entire field. The integrated database links written documents to the actual instruments and data files on which their findings are based, provides links from the data files to the publications based on them, and links resources on similar topics, by the same author, or funded by the same agency. This database is up to date and continually growing. Currently, the database provides access to over 11,000 resources, including 132 datasets.

Other major subjects of research in the collection include, but are not limited to:

- Parents and families using child care and early education services
- The early childhood and school-age child care work force
- Child care and early education settings, including center care, Head Start, public prekindergarten, family child care, family-friend-neighbor care (i.e., kith and kin care), in-home care
- National, state, and local child care and early learning policies

The collection centers on the United States but also includes research on early care and education in other countries.

Research Connections accepts all research and related documents disseminated in the field, without judging the quality of their design, methods, findings, and general content. It is, therefore, essential that visitors evaluate the quality of research they find on the site.

Even more recently, ICPSR added yet a third special education-related archive to its holdings: The PreK-3rd Data Resource Center. This is an online resource center designed to expand the knowledge base and provide tools for the access and handling of Prekindergarten through Third Grade longitudinal data in the United States. The goal of this project is to inform the Foundation for Child Development's PreK-3rd initiative and build the research field by facilitating the analysis of rich, complex, longitudinal datasets that contain a wide range of variables on the child, family, school, and neighborhood. The Web site disseminates datasets and user guides developed to provide researchers with detailed guidance in creating their own extract datasets.

The PreK-3rd Data Resource Center provides researchers with a single point of access to longitudinal datasets for conducting research to inform PreK-3rd approaches to education.

Datasets include:

- Early Childhood Longitudinal Study, Kindergarten Class of 1989-1999
- National Head Start/Public School Early Childhood Transition Demonstration study, 1991- 1999
- National Longitudinal Survey of Youth, 1979: Child Surveys
- Panel Study of Income Dynamics, Child Development Supplement
- Study of Early Child Care and Youth Development, Phases I-IV

Each Resource Guide provides an overview of a specific dataset, with detailed instructions on how to manage and combine data files to facilitate state-of-the-art research.

PreK–3rd research focuses on children’s learning experiences across the years PreK through at least Third Grade. Research questions that examine relations between the accumulation of these experiences and children’s educational outcomes and well-being are particularly well-suited to inform how policymakers and educators can organize the first six years of education from PreK through Grade 3.

All of these archival initiatives indicate the breadth and depth of resources available for educational research. What ICPSR is planning to do is to reorganize these three separate projects into a broader and centralized single archive which will focus on all types of education data.

CONFIDENTIALITY ISSUES

All data collections, whether they are produced by statistical agencies, academic research centers, or individual investigators, by definition are concerned with confidentiality issues. One recent study summarized the “trade-off” dilemma for statistical agencies (protecting confidentiality vs. maximizing the research value of the data) as follows:

“Each is charged with collecting high quality data to inform national policy and enable statistical research. This goal necessitates dissemination of both summary data and microdata. Each is also charged with protecting the confidentiality of survey respondents – not only because of legal and ethical mandates, but because public trust and perceptions of that trust are important contributors to data quality and response rates.” (Pat Doyle, Julia I. Lane, Jules J.M. Theeuwes, and Laura V. Zayatz, Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies. Amsterdam: Elsevier Science, 2001).

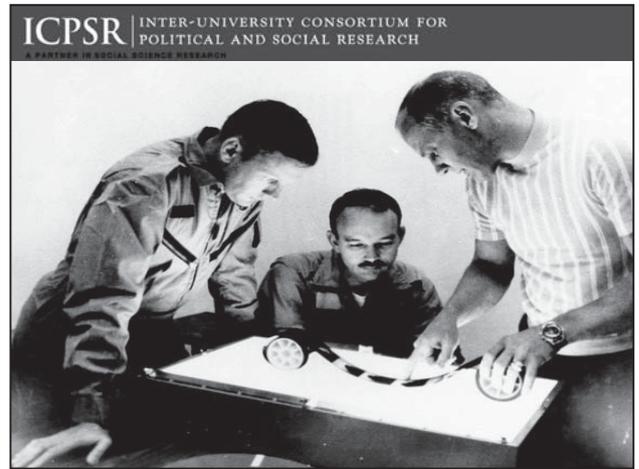
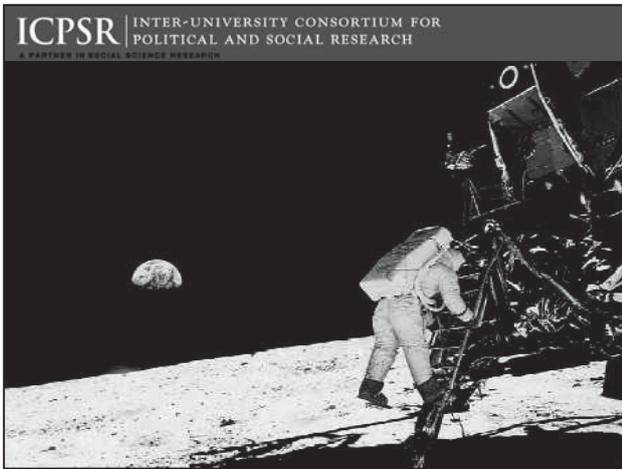
This dilemma affects all researchers who collect data and is especially acute with education data since its potential public value is so great while, at the same time, the identity of its institutional and individual respondents are at risk if too much information about them is made accessible.

One option increasingly in use nowadays is to provide both public and restricted versions of the data files to researchers. Restricted data contain confidential information that could lead to disclosure of the identity of respondents. ICPSR protects these data by maintaining them on a secure non-networked server. Individuals who apply for and complete a Restricted Data Use Agreement may obtain access to these data for legitimate research purposes. The application and agreement spell out the conditions of use pertaining to respondent confidentiality, as well as measures required for the physical safekeeping of the restricted datasets when in the researcher's possession. After both ICPSR and a responsible official at the researcher's institution sign the agreement, a copy of the data is supplied directly to the researcher. The agreement requires the applicant to destroy the data after a set period of time and to provide ICPSR with proof that the data have been destroyed. As the amount of information about individuals and institutions increases all the time and is often made readily available on the Internet, we may

expect to see more of these types of non-public access to research data in the future.

An ongoing challenge throughout the ICPSR experience centers on the geographic scope of the data that we have been able to collect. As I have described, the vast majority of it reports on the character of educational institutions in the United States with much less coverage from other parts of the world.

Nowadays, of course, researchers can access data from archives all over the world. It is not necessary for data to reside in just one location or in a few locations. The most important point to make is that researchers everywhere should have opportunities to access data from everywhere. For many good reasons this may not always be possible, but I believe that those of us who work in the archival community and for whom service to researchers is our main mission must continually strive to improve access to data wherever they may reside. Over the last few years, it is clear that the Educational and Social Survey Research Center has now become a very important source of data on the Japanese education system. The establishment of this archive has made a key contribution not only for Japanese scholars but for their colleagues all over the world who are engaged in international research.



James Samuel Coleman



May 12, 1926 - March 25, 1995

James S. Coleman served as the 63rd President of the American Sociological Association. His Presidential Address, "The Rational Reconstruction of Society", was delivered at the Association's Annual Meeting in Pittsburgh in 1992. A profile of James Coleman was published in the September 1991 issue of Footnotes.

EQUALITY OF EDUCATIONAL OPPORTUNITY

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
 John W. Colwell, Secretary
 OFFICE OF EDUCATION, Harold Swiney, Director

International Archive of Education Data

International Archive of Education Data

ABOUT IAED
 DATA & SURVEYS
 DATA ANALYSIS SYSTEM
 RELATED SITES
 HELP
 WHAT'S NEW
 CONTACT US
 SEARCH
 SITE MAP

IAED was established through funding from the National Center for Education Statistics (NCES) and is currently supported by the Inter-University Consortium for Political and Social Research (ICPSR).

ICPSR

Child Care & Early Education: Research Connections

Child Care & Early Education
RESEARCH CONNECTIONS

Search the Collection

Research Connections

Infusing a comprehensive, up-to-date, and easy-to-use collection of more than 11,000 resources from the many disciplines related to child care and early education.

NEW RESEARCH

• The longitudinal national work activities in day care research in the context of the Early Head Start program on research that "did not wait"

• How do teachers' education and training affect program quality?

• And children gain from New Jersey's National Professional Program?

ICPSR

PreK-3rd

PreK-3rd Data Resource Center
 The First Six Years of Learning and Beyond

ABOUT THE PROJECT
 DATABASES & RESOURCE GUIDES
 PUBLICATIONS & RESOURCES
 CONTACT US

About the Project

The PreK-3rd Data Resource Center is an online resource center designed to expand the knowledge base and provide tools for the access and handling of Prekindergarten through Third Grade longitudinal data. The goal of this project is to allow the Foundation for Child Development's PreK-3rd collection and build the research field by facilitating the analysis of rich, complex, longitudinal datasets that contain a wide range of variables on the child, family, school, and neighborhood. The Web site disseminates datasets and user guides developed to provide researchers with detailed guidance in creating their own extract datasets.

ICPSR

Human Subject Protection and Disclosure Risk Analysis

Human Subject Protection and Disclosure Risk Analysis

November, December 16, 2007

Mission
ICPSR
HUTE

Contact Us
ICPSR
HUTE

Projects
Project 1: Informed Consent and Perceptions of Risk and Harm in Survey Participation
DISC00000

Related Sites
Project 2: Evaluation of Disclosure Risk and Statistical Methods for Disclosure Limitation
DISC00000

Conferences
Project 3: Statistical Disclosure Control Best Practices and Tools for the Social Sciences
DISC00000

Web Site Tools
Browse the HSP Database Database

HSP Projects

What is HSP?

Funding Opportunities

Acknowledgements

ICPSR
HYOGO UNIVERSITY OF TEACHER EDUCATION
EDUCATIONAL AND SOCIAL SURVEY RESEARCH CENTER

Current status and future prospects of leading data archives in Japan, the United States, and Europe

From National Data infrastructures to International Data Infrastructure Networks

Ekkehard Mochmann

Former Director International Data Service,
GESIS-Leibniz Institute for the Social Sciences

The unexpected does happen- fortunately more frequently at the beginning of innovative developments and it is likely to happen less frequently if an infrastructure is set up to change the situation to the better. This could be a very brief summary of the history and impact of data infrastructures for the social sciences and obviously, for other disciplines as well. It refers to the anecdotes about the river Rhine swamping away from the basements of the American embassy at Bonn the punched cards of the first surveys taken in post war Germany or the loss of the first European Community wide survey, which was the starting point to the ongoing Eurobarometer survey series since 1972. Even though this might be a tempting opening to expand on the importance of base-line studies and longitudinal studies, let us rather focus- as invited by the organizers- on the development and functions of a data infrastructure, taking the example of the German Data Archive for the Social Sciences, the former Central Archive for Empirical Social Research (ZA) (<http://www.gesis.org/en/institute/gesis-scientificsections/data-archive-for-the-social-sciences/>).

1. Data infrastructure development and integral functions

ZA was founded in 1960 as an interdisciplinary institute by the Department for Economics and Social Sciences of the University of Cologne. The impetus to create this archive was triggered by the observation that decision makers and industry managers were in need of evidence based scientific analyses to overcome widespread ideologies. Primary objectives were archiving of representative sample surveys from the German speaking area and making these data available to the scientific community. Right from the beginning the philosophy was to develop the data archive in close cooperation and to engage in international comparative research. This could best be supported by a research team connected to the data archive as part of the University and by active participation as partner in international survey projects.

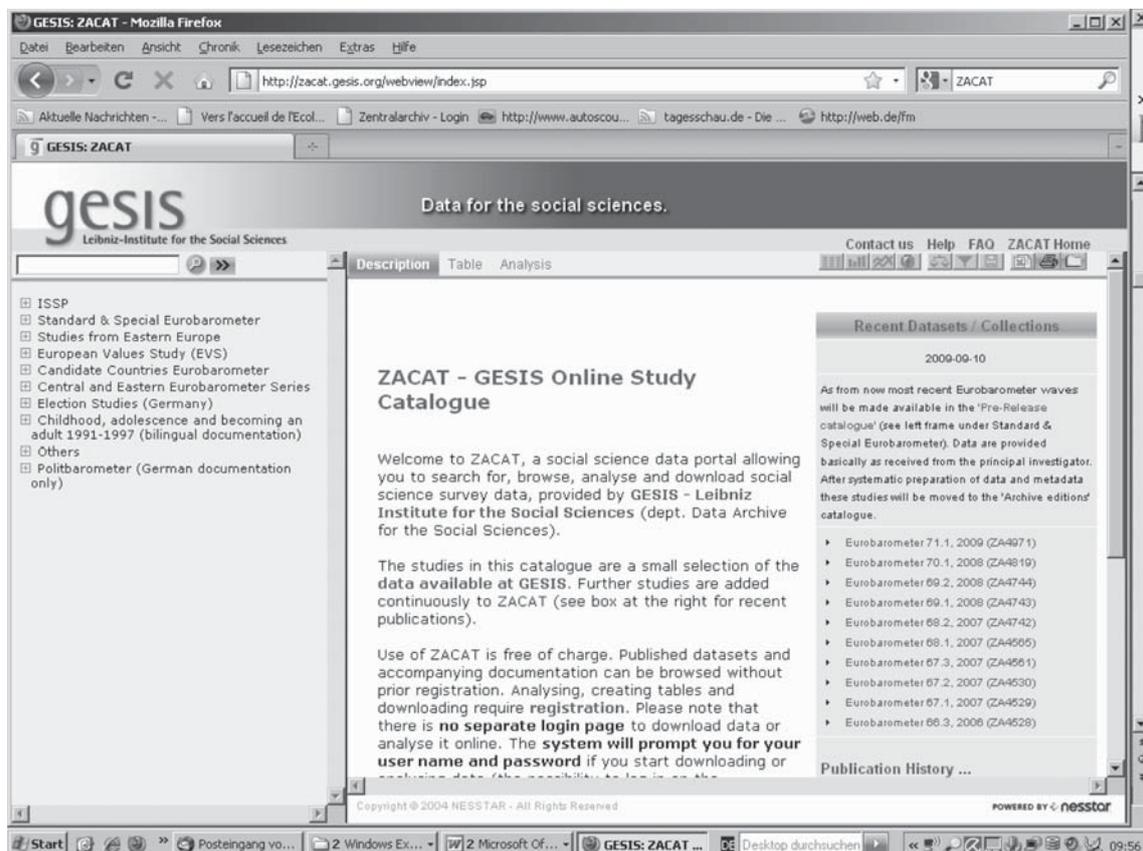
In order to base data acquisition on systematic procedures ZA started a research project documentation of ongoing and completed social research projects in Germany, Austria and Switzerland. All quantitative projects based on representative surveys were identified as basis for follow up data acquisition negotiations. This happened at times, when the Social Sciences considered itself “data poor” and, as repeatedly observed in different countries, those who were lucky data owners were not inclined to share their data treasures. Those providing information on their researcher did profit from access to the compiled results of this documentation of research projects.

In addition the uses of secondary analysis and data literacy, i.e. the expertise to master statistical data analysis were a rare competence. To improve this situation the archive published articles on the new techniques of secondary

analysis in highly visible methods books (Scheuch in König, Klingemann/ Mochmann in Koolwijk) and started the spring seminars, i.e. training seminars in advanced techniques of quantitative data analysis (<http://www.gesis.org/forschung-lehre/veranstaltungen/fruehjahrseminar/>) in 1972.

For this purpose the first statistical package for the social sciences available at that time (SPSS) was implemented and used with specially prepared data sets to teach the new analysis techniques. Participants were allowed to take the Statistical Package and the data sets back home and set up their own local university teaching courses leaning on the materials provided in the spring seminar. Parallel to this a series of re-analysis seminars was run in the normal university curriculum, using classical studies of eminent scholars, such as Lazarsfeld's "The Academic Mind" or Lipset's, Trow's and Coleman's "Union Democracy". The snowball effect of training seminars and advanced teaching networks helped to gradually foster a culture of data sharing and evidence based research.

With the steadily growing holdings and ever more comprehensive information, efficient tools for data access were required. In the late 1960ies ZA developed the first data retrieval system for study descriptions and the individual variables included in a survey, along with question and answer texts as well as the frequencies in absolute numbers and percentages. The development of the Z.A.R.-System was made possible by a generous research grant from Volkswagen foundation. This system was later coupled with the statistical Analysis system SAS and was offered as first interactive data retrieval and analysis system accessible world wide via the EARN/BITNET, which was a forerunner to the Internet. Today the most frequently used studies are accessible via ZACAT, a social science data portal allowing to search for, browse, analyse and download social science survey data, provided by the GESIS data archive (<http://zacat.gesis.org/webview/index.jsp>).



ZACAT holdings itself are linked to the CESSDA portal, which offers on line access to the data of participating European data archives (<http://www.CESSDA.org/accessing/catalogue/>).

This historical sketch summarizes the kind of activities, the archive engaged in:

- data acquisition, data management and archiving as basis for qualified data services;
- development of methods and techniques to provide efficient access tools.
- high level research and methodological developments demonstrating the potentials of secondary analysis
- training in advanced methods of data analysis and data management

The ZA soon became a central node in international research networks and helped to set up similar data infrastructure institutes in other parts of the world. A new challenge to data sharing arose when the concepts of “privacy” and “data protection” became popular. The ZA participated in international projects and conference to emphasize its way forward arguing for the “balance between the need to access data and the protection of the interests of the individual”, which was taken up by data protection legislators in the legal terms of “informational self determination” and “The research privilege to access data” under specified circumstances. A major breakthrough was the “Concept of factual anonymity” based on a project of systematic analyses of disclosure risks using a combination of micro census and sample survey data in different scenarios.

Today, the data archive is a department of the German Social Science Infrastructure Services (GESIS, <http://www.gesis.org/>). GESIS also included as new departments the former Centre for Survey Research and Methodology, in Mannheim (ZUMA) and the former Social Science Information Centre in Bonn (IZ), which is responsible for social science literature and research project documentation. GESIS is funded by the German government at state and federal levels to establish a social science infrastructure for the scientific community. It has more than 200 staff members and a budget of roughly 15 million euro per year. GESIS itself is the Leibniz Institute for the Social Science of the Gottfried Wilhelm Leibniz Association, a scientific community of some 80-institutes working in close cooperation with universities (<http://www.wgl.de/>). At the international level, the data archive works closely with the Council of European Social Science Data Archives (CESSDA, http://www.CESSDA.org/related/other_org/) and the International Federation of Data Organizations for the Social Sciences (IFDO, <http://www.ifdo.org/>). It is an active partner in the establishment of new research infrastructures in Eastern Europe (EDAN, <http://www.gesis.org/en/research/networks/data-infrastructure/edan/>). Recognizing the growing importance of metadata standards for data documentation GESIS strongly supports the Data Documentation Initiative (DDI, <http://www.ddialliance.org/>).

In 1996 ZA has created the European Data Laboratory (ZA-EUROLAB) as access facility to more than 5000 data sets and data collections in its holdings from all fields of empirical social research. Since its establishment more than 300 researchers from over thirty countries all over the world have worked at the EUROLAB for more than 4.000 days altogether with support from the Large Scale Facility and Access to research infrastructures program of the European Commission.. For almost 50 years, the data archive supported national and international empirical social research by making data available to the scientific community for further analysis. It offers data collections

in all fields of the Social Sciences as well as including data from history, economy and psychology. Topical fields covered in the more than 5000 data sets of its holdings are for example: family and gender, electoral research, democracy, travel and leisure, work life, mass communication, education, values and European integration.

The ZA holds a unique collection of comparative and longitudinal data. Principal investigators and research networks have chosen the ZA as their resource centre for creating integrated data bases. Currently, ZA serves in this function for the Eurobarometers in co-operation with Inter-University Consortium for Political and Social Research (ICPSR), Ann Arbor (<http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>), the International Social Survey Programme (ISSP) with ASEP/JDS (<http://www.jdsurvey.net/jds/jdsurvey.jsp>), Madrid, and the European Values Survey (EVS) in co-operation with Tilburg University. The major election studies to national parliaments in Europe (ICORE) were compiled by ZA with assistance of other European data services (CESSDA). Furthermore, the ZA became the holding archive of the original documents from the “Comparative Party Manifestos Project”. In co-operation with the principal investigators, the data archive co-ordinates and creates European data bases, which could otherwise not be made available to the scientific community. Further international comparative data sets including several European societies in the holdings are the surveys of the Comparative Studies of Electoral Systems (CSES), surveys from the United States Information Agency (USIA), the Central and Eastern Eurobarometer (CEEB), Candidate Countries Eurobarometer (CCEB), the World Values Survey (WVS), the Political Action survey, and the surveys from the Founding Elections in Eastern Europe. Important studies such as the German General Social Survey (ALLBUS), a national reference study, is documented and made available to the scientific community world wide. In this context particularly noteworthy is, that the Japanese General Social Survey (JGSS) made its series available to the GESIS data archive and the Social Science Japanese Data Archive (SSJDA) at the University of Tokyo cooperates in the IFDO context with the German Data archive.

In addition to this emphasis on fostering comparative social research by providing well documented data, ZA played a significant role in opening the European discourse on infrastructure needs for European comparative research with particular emphasis on integration of the European data base, data harmonization and data access, as well as on data protection issues in international perspective, topics which became central working packages in subsequent large scale projects, funded by the European Commission.

2. Contributions to Educational Research

Over past decades the educational sector came up with a remarkable number of longitudinal and replicated cross section surveys, which served as an important evidence base on educational careers and impact factors on educational processes. Alarming results of educational assessment studies, such as PISA intensified the interest in reliable data to correct long established stereotypes and to provide evidence based orientation for future reforms of the educational system.

To improve access to widely scattered educational data the project “Documentation of Longitudinal Studies into the Education System” was commissioned by the Federal Ministry of Education and Research and completed by

the University of Erfurt / Chair for Empirical Education Research under the direction of Prof. Weishaupt. One first product was based on studies already archived and documented according to the standards of the data archive. Panels and replicative studies were identified and documented on CD-ROM. This contains searchable study descriptions and further information on altogether 40 longitudinal and quasi-longitudinal studies into the education system. The studies, which contain relevant information on education trajectories and factors which influence education processes, were organized into the following subject areas:

- Childhood/youth and family
- School research
- Higher education research
- Qualification and occupation
- Continuing education/adult studies
- Omnibus surveys

This documentation is available from the GESIS data archive free of charge ([http://www.gesis.org/en/services/data/survey-data/subject-based-study-pools/longitudinalstudies-into-education-system/?sword_list\[0\]=weishaupt](http://www.gesis.org/en/services/data/survey-data/subject-based-study-pools/longitudinalstudies-into-education-system/?sword_list[0]=weishaupt))

Other studies could be identified in a variety of institutes. While some were well documented on web sites, others were hardly visible and could only be identified when quoted in some publications. Also the culture of data sharing was not well developed over all. While some researchers were highly cooperative, others were reluctant or not willing at all to make “their data” accessible. The inclusion of those studies into the documentation are a starting point to open access to further relevant studies. This project also shows, that longitudinal studies had many blank spots in particular re personnel in the educational sector. As a consequence it was hard to find any data on careers in the educational sector, self perception or attitudes of teachers to their profession.

A new challenge was posed by the data heaps of large international studies. To improve data access in this area a Research Data Centre at the Institute for Educational Progress was established (<http://www.iqb.hu-berlin.de/arbberreiche/fdz>). This institute has the mandate to collect and document the data sets produced in large national and international assessment studies such as PIRLS or PISA, making them available to the scientific community for reanalyses and secondary analyses. GESIS supported the starting phase of this new research data centre.

Currently the RDC contains the data of these studies:

- ELEMENT Survey of reading and mathematics comprehension developments in grades 4 to 6 in Berlin
- IGLU/PIRLS 2001 Progress in International Reading Literacy Study 2001
- MARKUS Cumulative Assessment Study on Mathematics in Rhineland-Palatinate: Competencies, Classroom Teaching Features, School Context
- PISA 2000/PISA 2003 Programme for International Student Assessment

and a pilot study of French language competence.

Members of the scientific community are given access to the required data upon request. More detailed information on data sets available at the IQB RDC as well as on data application procedures and other relevant topics is accessible via the web site.

Early in 2008 another research data centre was established to accommodate a wide range of surveys on education, employment and further training as well as on the transitions between these phases. The Research Data Centre of the Federal Institute for Vocational Education and Training (BIBB-FDZ) offers several services for researchers:

- Standard access to firm- and individual-level data on skill attainment and its use in the (vocational) education and labour market system
- Documentation of these data sets, i.e. a description of their central characteristics, main issues and variables, data collection, anonymisation, weighting and recoding etc.
- Advisory service on data choice, access and handling of the data, as well as on research potential, the scope and validity of the data.
- Supply of several data tools, e.g. standard measures and classifications in the fields of education, occupations, industries and regions (if possible also cross-national ones), formally anonymous data for remote data access, or references to publications with the data.
- Assistance and support for visiting researchers at the safe centre at BIBB-FDZ in Bonn

As this centre is still in the start up phase not all information is available in English yet (for further details see: <http://www.bibb.de/en/50113.htm>).

The research data centres were primarily designed to allow researchers on site access to sensitive data in controlled environments. Increasingly they also provide scientific use files of anonymous data to the scientific clientele. GESIS recently established as new units three research data centres: The research Data Centres on “ALLBUS-the German General Social Survey”, “International Survey Programmes” and “Electoral Research Data”, thus contributing to the network of research data centres operating according to the standards of the German Council for Social and Economic Data (<http://www.ratswd.de/eng/index.html>).

3. Open access to publicly financed data

Creating a culture of data sharing has been a laborious but –in the end- rewarding effort in many societies, which can afford a functioning data infrastructure. The OECD Initiative “Re-Inventing the Social Sciences” led to a ministerial decision stipulating that publicly funded research data should be openly available to the maximum extent possible. This principle was endorsed by the declaration of OECD research ministers in January 2004 (OECD 2007) as well as by several declarations that explicitly include original scientific research results, raw data and

metadata, for example, the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (http://www.mpg.de/pdf/openaccess/BerlinDeclaration_en.pdf). They all agree that availability should be restricted only by legitimate considerations of national security; protection of confidentiality and privacy; intellectual property rights and time-limited exclusive use by principal investigators. Publicly funded research data are a public good, produced in the public interest. As such, they should remain in the public realm. Expanding the adoption of this principle to national and international stages will enable researchers, empower citizens and convey tremendous scientific, economic and social benefits. The widespread national, international and cross-disciplinary sharing of research data is no longer a technological impossibility (Arzberger et al 2004, p.36).

Recent developments in cyber-infrastructure and e-science point to great potential for rapid progress in widening and accessing the evidence base (Atkins 2003, Berman, Brady 2005, Mochmann in RatSWD 2009). This will not succeed by technical fiat. The scientific community has to co-ordinate its increasingly trans-disciplinary efforts in agreement with funding agencies and governments, with an eye on those parts of the world that are still blank spots in the global data landscape.

Literature

Atkins, D.E. et al. (2003): Revolutionizing Science and Engineering through Cyberinfrastructure. Report of the National Science Foundation. Blue-Ribbon Advisory Panel on Cyberinfrastructure. January 2003.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, Max-Planck-Gesellschaft, Berlin 2003.

Berman, F. and Brady, H. (2005): Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. Hoffmeyer-Zlotnik, J. H. P.; Wolf, Chr. (eds.): Advances in Cross-National Comparison. New York et al.: Kluwer/Plenum 2003

Hyman, H. H. 1972. Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities. New York

Jagodzinski, W. 2004 Methodological problems of value research. In: Vinken, H.; Soeters, J.; Ester, P. (eds.): Comparing Cultures. Dimensions of Culture in Comparative Perspective. Leiden, Boston: Brill 2004, pp. 97-121.

Jensen, U., Mochmann, E. 2003 MetaDater: Towards standards and tools for the description of comparative surveys. In: ZA-Information, 2003, Nr. 52, S. 191-198.

Jowell, R. 1998. How Comparative Is Comparative Research? American Behavioral Scientist, Vol. 42, No. 2, 168-177 . Sage Publications

Jowell, R., Roberts C., Fitzgerald, R., Eva, G. 2007: Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey. London, Sage

Kaase, M. 1999. Quality Criteria for Survey Research. Berlin

Klingemann, H.-D.; Fuchs, D. (eds.) 1995. Citizens and the State. Beliefs in Government, Vol. 1. New York: Oxford University Press

- Klingemann, H.-D.; Hofferbert, R. I.; Budge, I. 1994. *Parties, Policies, and Democracy*. San Francisco: Boulder; Oxford: Westview Press
- Klingemann, H.D., Mochmann E. (1975): „Sekundäranalyse“ in van Koolwijk, J and Wieken-Mayser, M.: *Techniken der empirischen Sozialforschung*, Oldenbourg Verlag, München.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (ed.) 2001. *Wege zu einer besseren informationellen Infrastruktur*. Baden-Baden: Nomos
- Merritt, R. L. 1968. *USIA Surveys*. In: R. L. Merritt, D. J. Puchala (eds.): *Western European Perspectives on International Affairs. Public Opinion Studies and Evaluations*. New York: Frederick A. Praeger, pp. 3-31
- Merritt, R. L.; Rokkan, S. 1966. *Comparing Nations. The Use of Quantitative Data in Cross-National Research*. New Haven, London: Yale University Press
- Mochmann, E. 2009: *Improving the Evidence Base for International Comparative Research*, in *International Social Science Journal* 193 (forthcoming)
- Mochmann, E. 2009: *e- Science Infrastructure for the Social Sciences*, Working Paper 115, German Council for Social and economic Data (RatSWD), Berlin
- Mochmann, E. 2005 *From Metadata Standards to the Data GRID for Comparative Social Research*. 1st. International Conference on e-Social Science - 22-24 June 2005, Manchester, UK. <http://www.metadater.org/presentations.htm>
- Mochmann, E. 2002: *International Social Science Data Service Scope and Accessibility*. Report for the International Social Science Council (ISSC). Cologne
- Mochmann, E. 1999. *ZA-Eurolab: A Virtual Observatory for Comparative Social Research*. Report at the ESF Exploratory Workshop “The Virtual Observatory: Access to, and Use of, Information and Data Astronomical, Earth and Environmental Observations. European Science Foundation, Strasbourg, September 20-21
- Mochmann, E. 1999. *European Infrastructure Needs for Comparative Socio-Economic Research*, EU Socio-Economic Research Conference, Brussels, April 28-30
- Mochmann, E. 1998. *European Cooperation in Social Science Data Dissemination*. In: R. Walker, M. F. Taylor (eds.): *Information Dissemination and Access in Russia and Eastern Europe: Problems and Solutions in East and West*. Amsterdam: IOS Press, pp. 33-43
- Mohler, P. Ph.; Uher, R. 2003 *Documenting Comparative Surveys for Secondary Analysis*. In: Harkness, J. A.; van de Vijver, F. J. R.; Mohler, P. Ph.: *Cross-Cultural Survey Methods*. Hoboken: Wiley, pp. 311-327
- Moschner, M. 2007. *Tabular History of Comparative Survey Research*, http://www.gesis.org/en/data_service/eurobarometer/handbook
- NESSIE 2004, Roundtable 4. *Access to Quality Comparative Data for European Comparative Socio-Economic Research* http://www.nessie-essex.co.uk/roundtable_4Agenda.asp
- Oyen, Else 1999. *Transnational and cross-national social science research*. World social science report, UNESCO p.187-191
- Reif, K.-H.; Inglehart, R. (eds.) 1991. *Eurobarometer: The Dynamics of European Public Opinion. Essays in Honour of Jacques-René Rabier*. London: Macmillan

- Rockwell R. C., 1999. Data and statistics: empirical bases for the social sciences. World Social Science Report. UNESCO
- Rokkan, S.; Urwin, D.; Aarebrot, F. H.; Malaba, P.; Sande, T. 1987. Centre-Periphery Structures in Europe. An ISSC Workbook in Comparative Analysis. Frankfurt, New York: Campus Verlag
- Rokkan, S. 1964. Introduction. The Use of Sample Surveys in Comparative Research. In: International Social Science Journal, Vol. XVI, No. 1, pp. 7-18
- Rokkan, S. 1964. Archives for Secondary Analysis of Sample Survey Data: An Early Inquiry into the Prospects for Western Europe. In: International Social Science Journal, Vol. XVI, No. 1, pp. 49-62
- Rokkan, S. 1962. The Development of Cross-National Comparative Research: A Review of Current Problems and Possibilities. In: Social Sciences Information, Vol. 1, No. 3, pp. 21-38
- Schaber, G. 1994. Developing Comparative Databases. Prepared for the 2nd CIESIN User Workshop. Atlanta, Georgia, October 11-13. Luxembourg: CEPS/INSTEAD
- Scheuch, E. K. 2003 History and Visions for the development of data services for the social sciences” (International Social Science Journal Vol. 177)
- Scheuch, E. K. 1967. Entwicklungsrichtungen bei der Analyse sozial-wissenschaftlicher Daten. In: R. König (ed): Handbuch der empirischen Sozialforschung, Bd. 1, Stuttgart, 2. Auflage, S. 136-196
- Scheuch, E. K.; Brüning, I. 1964. The Structure and Mode of Operation of the Zentralarchiv. In: International Social Science Journal 16, pp. 86-96
- Scheuch, E. K.; Stone, P. J. 1964. The General Inquirer Approach to an International Retrieval System for Survey Archives. In: American Behavioral Scientist, Vol. 8, No. 10, pp. 23-28
- Schupp, J.; Wagner, G. 1995. The German Socio-Economic Panel: A Database for Longitudinal International Comparisons. In: Innovation, pp. 95-108
- Silberman, R. 1999. Rapport Silberman: Les Sciences Sociales et leurs Données. Paris : Ministère de l'Éducation Nationale, de la Recherche et de la Technologie
- Solga., H., Wagner, G. G. 2007 A Modern Statistical Infrastructure for Excellent Research and Policy Advice Report on the German Council for Social and Economic Data during its First Period in Office(2004 – 2006), http://www.ratswd.de/download/workingpapers2007/02_07.pdf
- Steering Committee and Methodology Committee of the ESS 1998. Blueprint for a European Social Survey (ESS). An ESF Scientific Programme. European Science Foundation, Strasbourg,
- Tanenbaum, E.; Mochmann, E. 1994. Towards a European Database for Comparative Research. In: International Social Science Journal 142
- The Royal Statistical Society & the UK Data Archive (eds.) 2002. Preserving & Sharing Statistical Material. Colchester: University of Essex
- Thomassen, J. (ed.) 2005 Titel The European Voter. A Comparative Study of Modern Democracies. Oxford, New York: Oxford University Press
- Treadwell, W. 2002. DDI, The Data Documentation Initiative: An Introduction to the Standard and its Role in

Social Science Data Access. ALCTS Papers on Library Technical Services and Collections, No 10. Chicago: American Library Association

van Deth, J. W.; Scarbrough, E. (eds.) 1995. The Impact of Values. Beliefs in Government, Vol. 4. New York: Oxford University Press

van Langenhove, L. 1999. Re-Inventing the Social Sciences. Opening Speech at the first OECD International Social Sciences Workshop "The Social Sciences in a Digital Earth". Ottawa, October 6-8

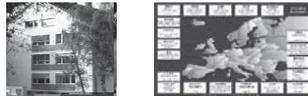


Educational and Social Survey Research Center (ESSRC)
 Hyogo University of Teacher Education (HUTE)

2009 International Symposium:
 Data Archives and International Collaboration for Research,
 Public Policy, and Business: As a Witness of the Launch of
 Educational Data Archive of ESSRC, HUTE

**From National Data Infrastructures to
 International Data Infrastructure Networks**

Ekkehard Mochmann
 e.mochmann@web.de



Kobe International Conference Centre, Nov. 22, 2009

Topics

The Cologne Data Archive in GESIS Context

European and International Data Service
 Networks

Co-operation with Research Networks (ComNet-
 Scenario)

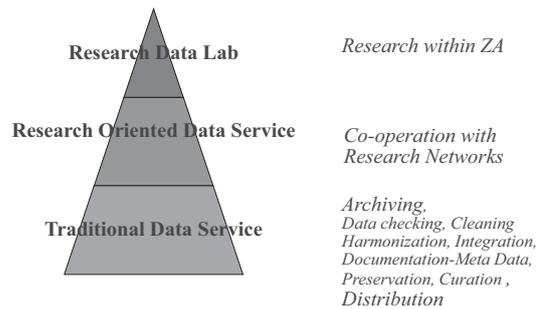
Research - and Service Data Centres

Open Access and Data Science

**1. The Cologne Data Archive in
 GESIS Context**



Focus on: Data Archiving and Data Service



GESIS Context

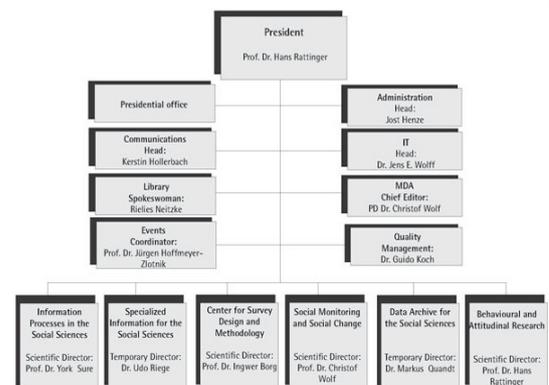
www.gesis.org

GESIS – ZA, Cologne Data Service and Data Analysis
 EUROLAB, Historical Social Research

GESIS – IZ, Bonn Literature and Research Project
 Documentation, IT- RTD

GESIS – ZUMA, Survey Research Methods, Long
 Mannheim Term Observation, MicroDataLab
 ALLBUS, ISSP, ESS

GESIS Service Department Berlin – East EUROPE



International Seminars and Summer Schools

- ZA Spring Seminar (since 1972)
- ZA/ZHSF Autumn Seminar
- ZUMA Workshops
- Data confrontation seminars
- Teaching in International Summer Schools
- Guest Professorships

ZA-EUROLAB

Access to research Infrastructures (ARI) objectives:

- Support all researchers of EU member states to access infrastructure services, which:
 - offer world class service, essential for high level research
 - are unique or rare in Europe and are not easy to replicate
 - which can support external users scientifically, technically and logistically

2. European and International Data Service Networks



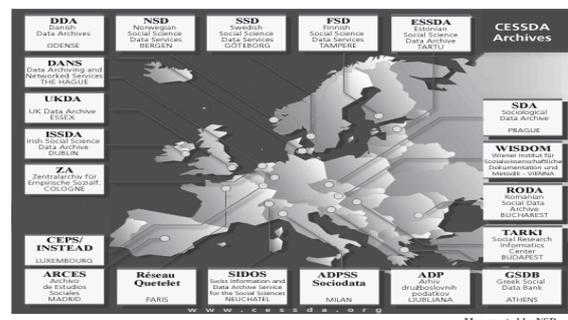
International Federation of Data Organisations for the Social Sciences (IFDO) www.ifdo.org



IFDO Profile

- IFDO organises the international data transfer between member archives
- IFDO supports new archives world wide
- Cooperates with CESSDA and IASSIST

European Social Science Data Archives (CESSDA) www.cessda.org



CESSDA network

- CESSDA has provided networked infrastructural services for the social sciences for the past 30 years through the acquisition, support and supply of data for the European social science and humanities research
- Over this period it has grown both in geographical and substantive terms. The network now extends to 21 countries across Europe.
- CESSDA works with data infrastructures in emerging and new member countries (e.g EDAN co-ordinated by GESIS).

EDAN: East European Data Archive Network

Participating Archives:

- Ungarn: TARKI (1985)
- Russland (UdSSR): DB (1985)
- Estland: ESSDA (1994)
- Litauen: LSZDA (1996)
- Slowenien: ADP (1997)
- Tschechien: SDA (1998)
- Russland: RSDA (2002)
- Rumänien: RODA 2001
- Bulgarien: SSSA (2003)

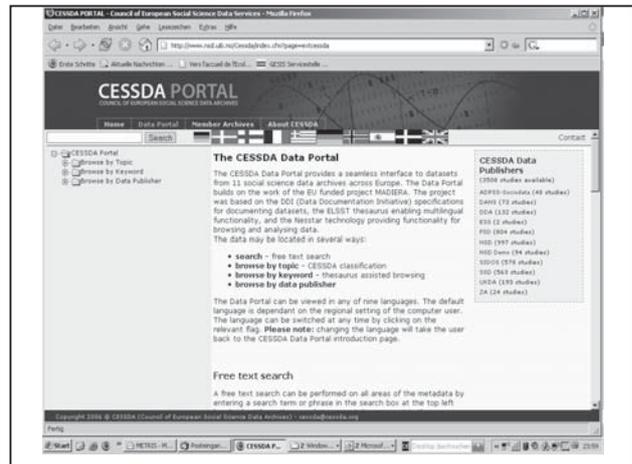


Other Participants: Initiatives in Slovakia, Poland, Belarus, Ukrain, Latvia, Serbia + Ireland, Portugal

Services provided

Collectively the constituent CESSDA member organisations

- serve some 30,000+ researchers within the ERA each year,
- provide access to and deliver some 50,000+ data collections per annum and
- acquire a further 1,000 data collections each year.
- In addition, the CESSDA organisations provide access gate-ways to important SSH data materials such as the European Social Survey (NSD), the Eurobarometers (GESIS-ZA with ICPSR), the International Social Survey Programme and the European Values Surveys (GESIS-ZA with ASEP/JDS).



CESSDA Expert Seminars

- The development of CESSDA data bases,
- new data management methods and technologies are discussed in CESSDA expert seminars.
- CESSDA staff is closely connected by internal discussion lists, regular expert meetings and cooperation in transnational RTD projects.

3. Research Co-operations (COMnet-Scenario)

- Eurobarometer (EB)
- European Election Studies (EES)
- European Values Study (EVS)
- Comparative Study of Electoral Systems (CSES)- International Committee for Research into Elections and Representative Democracy (ICORE)
- International Social Survey Programme (ISSP)
- World Values Survey (WVS)
- Comparative Party Manifestoes (CMP)

4. RDC/ SDC Functions

- **Research Data Centers** make individual data accessible for scientific research by and large through the creation of factually anonymized data sets (Scientific Use Files) released to research institutions. In exceptional cases, where research concerns particularly sensitive data, or where it is not possible to adequately anonymize data without the loss of information, data access is possible through the creation of workplaces for guest researchers at specific Research Data Centers, or through the development of a system for controlled remote data access.
- **Data Service Centers** are designed to facilitate empirical research in the social and economic sciences. They are charged with implementing a broad range of services, including improved data documentation that meets scientific requirements, the construction of a metadata portal, and the provision of trained support for data users.

RDCs

Research Data Centres

- Research Data Center of the Federal Statistical Office (FDZ-Bund)
- Research Data Center of the Statistical Offices of the Länder (FDZ-Länder)
- Research Data Center of the Federal Employment Agency at the Institute for Labor Market and Occupational Research (FDZ-IAB)
- Research Data Center of the German Pension Insurance (FDZ-RV)
- Research Data Center of the Federal Institute for Vocational Education and Training (BIBB-FDZ)
- Research Data Center of the Institut for Educational Progress (IQB)
- Research Data Center of the Socio-Economic Panel Study (FDZ-SOEP)
- The Research Data Centre ALLBUS at GESIS
- The Research Data Center "International Survey Programmes" at GESIS
- Research Data Center "Voting Behavior database" at GESIS
- SHARE Research Data Center

SDCs

Data Service Centres

- GESIS' Service Centre for Microdata at GESIS-ZUMA / MISSY
- International Data Service Center of the Institute for the Study of Labor (IZA)

Open Access and Data Science

OECD:

Open Access to publicly financed data

Berlin Declaration:

Open Access to all research results,
i.e. including data

-> „Data Science“ and e- infrastructure

Research Data Centers

Research Data Center of the Federal Statistical Office (FDZ-Bund)
Research Data Center of the Statistical Offices of the Länder (FDZ-Länder)
Research Data Center of the Federal Employment Agency at the Institute for Labor Market and Occupational Research (FDZ-IAB)
Research Data Center of the German Pension Insurance (FDZ-RV)
Research Data Center of the Federal Institute for Vocational Education and Training (BIBB-FDZ)
Research Data Center of the Institut for Educational Progress (IQB)
Research Data Center of the Socio-Economic Panel Study (FDZ-SOEP)
The Research Data Centre ALLBUS at GESIS
The Research Data Center "International Survey Programmes" at GESIS
Research Data Center "Voting Behavior database" at GESIS
SHARE Research Data Center

Data Service Centers

GESIS' Service Centre for Microdata at GESIS-ZUMA /
International Data Service Center of the Institute for the Study of Labor (IZA)

Current status and future prospects of leading data archives in Japan, the United States, and Europe

SSJDA (Social Science Japan Data Archive) and Activities at the Center for Social Research and Data Archives

Hiroshi Ishida

Professor, Institute of Social Science, The University of Tokyo
Vice president, Japanese Association for Mathematical Sociology

Roles and Significance of Data Archives

Data archiving organizations gather, compile, save and make available data from social surveys. Data archives are like libraries for data, but whereas there are book libraries throughout Japan, data archives have attracted little public attention. Even though many social surveys, public opinion surveys and attitudes surveys are conducted every year, few people in Japan have recognized the need to make the survey data easily accessible until very recently.

Like libraries for books, data archives have important functions (Sato and Ishida and Ikeda, 2000; Sato and Sato, 2006). First, data archives prevent the loss of data by safely storing valuable data. Not only for individual researchers and research institutions that conduct surveys but also for others, data archives provide secure data storage and minimize the risks of accidental data loss. Second, data archiving organizations reduce the work for individual researchers and research institutions that conduct surveys by compiling and making the data available to others on their behalf. Many individual researchers and research institutions recognize the importance of making data available to others, but often find the task too burdensome because it would involve advance preparation of materials such as general descriptions of data sets and code books to enable the survey data to be used by others. This would require extra time and money in addition to that spent on the survey and analysis. Data archiving organizations carry out this task on behalf of them.

Third, data archives enable others to perform secondary analyses. Researchers conducting secondary analyses may use a very different perspective from that of the original researchers who conducted the primary analyses. Secondary analyses, therefore, often lead to new results and findings. Even more importantly, making survey data available to others assures the reproducibility of analysis. In social sciences, unlike in natural sciences, there is no established practice of ensuring that primary analyses of survey data are verified by others through secondary analyses of the same survey data, and such verification by others is rare. Assured reproducibility will greatly improve the standard of social science research as empirical studies.

Fourth, data archives prevent unnecessary repetition or duplication of similar surveys. This not only saves time and cost for the researchers but also reduces the burden on the target population, and helps identify the kinds of surveys that are really required.

Fifth, data archives will help improve the quality of social surveys conducted in the future, since survey researchers need not start from scratch but can consult materials and analyze data from earlier surveys, and so can conduct preliminary studies on the significance of questions and the validity of working hypotheses. This helps them better identify the hypotheses that could be proven by surveys. Moreover, they can analyze the feasibility

of comparisons with earlier surveys or analysis concerning trends over time. Furthermore, researchers who are familiar with data archives are likely to consider the eventual disclosure of survey data when planning and implementing surveys. Therefore, they may aim for higher quality surveys and ease of understanding by others, and prepare code books that describe data cleaning and coding procedures.

Thus, data archives contribute to society in many ways including safe storage of data, promotion of secondary analyses and higher quality of social surveys. This report outlines the activities concerning the SSJDA as an example of data archives in Japan, and also describes future challenges.

Outline of SSJDA (Social Science Japan Data Archive)

The Institute of Social Science, University of Tokyo, has been running the Social Science Japan Data Archive (SSJDA) since April 1998, making available social survey data to researchers. The data archive has been managed by the Information Center for Social Science Research on Japan, which was established in 1996 as an affiliated institution of the Institute of Social Science (in April 2000, the Information Center for Social Science Research on Japan was reorganized into the Center for Social Research and Data Archives). Data that had been kept at the Institute of Social Science, as well as questionnaire data (micro data) from social surveys conducted by private research institutions, researchers at universities and others, were collected and compiled using a unique methodology, and made available for secondary use by academic researchers.

“SSJ” stands for Social Science Japan. This acronym, which had been used to represent major activities at the Institute of Social Science, University of Tokyo, was chosen to be a part of the name of our data archive. One of the missions of the Institute of Social Science, University of Tokyo, is to serve as an interface with the global community of social scientists studying Japan. Accordingly, the Institute of Social Science seeks to promote networks of overseas researchers who study Japan and actively supports the activities of these scholars. The Institute of Social Science has not only been inviting such overseas researchers conducting studies on Japan to work with us as guest professors and researchers, but has also been conducting various “SSJ” programs to promote social scientific studies on Japan by overseas researchers. Examples of activities include:

- Publishing SSJ Newsletters in English
- Running the SSJ Forum on the Internet, which provides a listing of discussions concerning modern Japan
- Editing Social Science Japan Journal (SSJJ) for publication by Oxford University Press, which is a fully-referred journal dedicated to social scientific studies on Japan.

Therefore, running the SSJDA represents one of our efforts to promote social scientific studies on Japan.

Activities concerning SSJDA

Figure 1 illustrates the flow of activities concerning SSJDA. By referring to this figure, I would like to briefly describe the functions of SSJDA (Sato 2006; Sato and Sato, 2006)

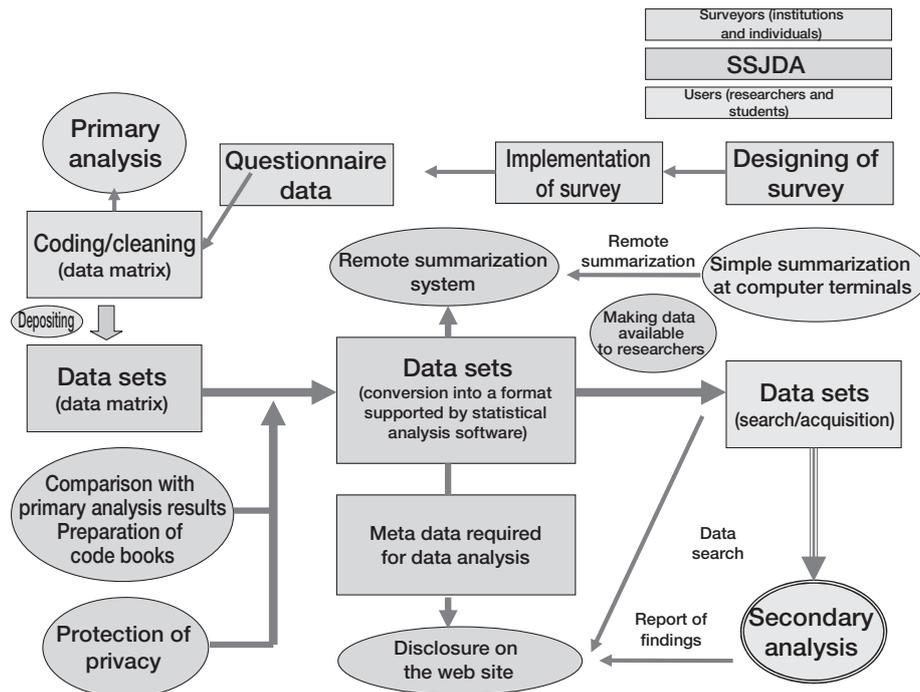


Figure 1 Functions of SSJDA

The first step is the depositing of survey data. We accept micro data sets which are collected, cleaned, coded, and analyzed by the researchers who originally conducted the survey. We advertise our services to research institutions to encourage them to deposit valuable survey data to the SSJDA. In the second step, we compare the deposited data sets with published research reports and the results of the primary analysis in order to check if there are no discrepancies between output from the deposited data sets and the simple tabulations results reported in the published materials. We also check data sets against code books to ensure no discrepancies in variable names, categories and missing values. If any discrepancy is found in this verification process, we contact the data depositor to identify any need for data correction. To protect the privacy of respondents, we ensure that personal information cannot be identified (by grouping the detailed district codes into a more aggregated category, for example). Next, we convert data into a format supported by statistical analysis software, then the converted data can be processed immediately by statistical analysis software. At present, we attach variable names and labels supported by the SPSS statistical analysis software package.

The third step is the preparation of meta data for data sets. Meta data contain basic information on data sets such as general descriptions, variable lists and simple frequency outputs. We prepare meta data and make them accessible on our web site, so that the users can use the search function to look for particular data sets. They are able to browse through meta data, which provide general information on data sets, to find data sets that could be useful to their studies. Only after that do they submit a request to use selected data sets. Some meta data are available in English. The fourth step is to make data available to researchers, which can be achieved in two ways at present: researchers may download data directly from the SSJDA web site using the SSJDA-Direct system (Sato, 2009), or fill in a request form downloaded from the web site, mail it to us, then receive the requested data sets on CD-ROM through the post. In the past, all data were sent by post on CD-ROM, but to improve the efficiency of responding to an increased number of requests, we launched the direct download system in April 2009. However, the number

of data sets downloadable using the SSJDA-Direct system is still limited. As an alternative way of accessing the data set that does not involve downloading micro data, we provide services of remote analysis system, with which researchers can analyze micro data in the SSJDA remotely from terminals at their universities (Sato, 2008). The remote analysis system allows simple analyses of questionnaire data from a remote terminal, and is often used in undergraduate courses on social surveys and quantitative analyses.

Finally, users submit data utilization reports and research findings to the SSJDA after completing their secondary analyses using data from the SSJDA. The SSJDA management team publishes the list of research findings on the web site along with statistics on data utilization requests and publications.

Major Collections of Survey Data Available on SSJDA and Major Depositors

The first collection of data that became available on the SSJDA was the data that had been kept by the Institute of Social Science, University of Tokyo. From various collections of surveys conducted in the 1950s, which had been kept in the underground storage room of the Institute, we picked up materials concerning a survey of junior high school seniors who graduated in 1953 in Kanagawa Prefecture, and produced electronic data from papers, and made them available on the SSJDA. By making public our own data set first, we wished to motivate others to share their survey data and encourage secondary use of the data.

Later, we received tremendous support from many private and governmental research institutions that understood the significance of data archives. Major depositors and survey projects that provided data to the SSJDA include the following:

- Questionnaire survey on the work and life of working citizens (Research Institute for Advancement of Living Standards)
- Survey on life insurance (Japan Institute of Life Insurance)
- Survey on the life of university students (National Federation of University Cooperative Associations)
- Survey on small and medium sized enterprises in Japan (General Research Institution of the Japan Finance Corporation)
- Survey on the employment and labor condition of the youth (Japan Institute for Labor Policy and Training)
- Working person survey (Works Institute sponsored by Recruit)
- Monographs on elementary school students, junior high school students and high school students (Benesse Corporation)
- Survey on pachinko players (Entertainment Business Institute)

We also receive data from academic surveys conducted by governmental institutions and research groups:

- Attitudes survey on various working styles (Quality-of-Life Policy Bureau of the Cabinet Office)
- Japanese General Social Surveys (JGSS Research Center, Osaka University of Commerce)
- National Family Research of Japan (NFRJ Committee, Japan Society of Family Sociology)
- 1995 Social Stratification and Social Mobility (SSM) Survey (1995 SSM Research Study Group)

Data Sets Available on SSJDA and Number of Data Utilization Requests

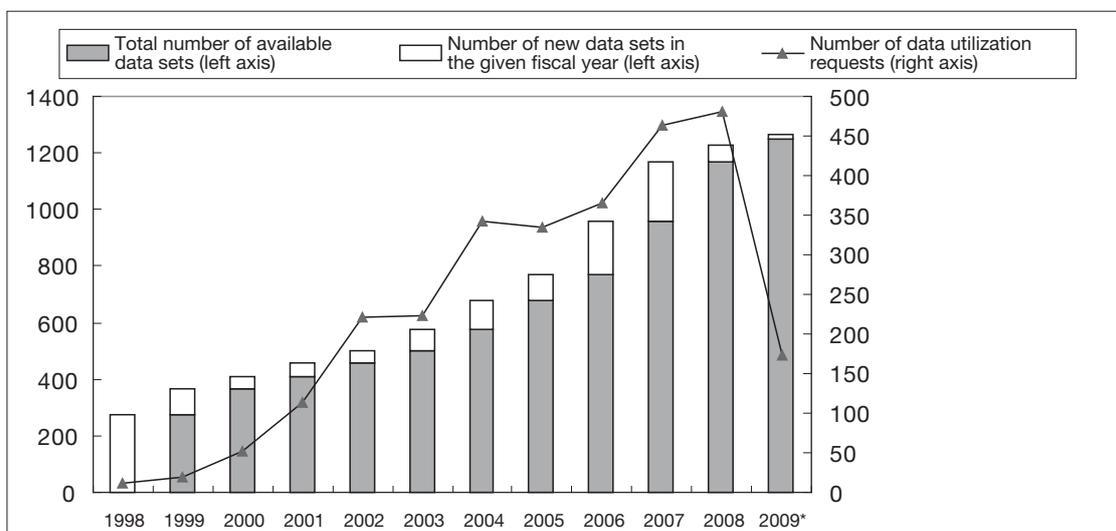


Figure 2 Total number of available data sets, number of new data sets made available, and number of data utilization requests

(Note: The figures for fiscal 2009 are totals in the period from April to August.)

Figure 2 shows the number of data sets newly made available in each fiscal year, the total number of available data sets and the number of data utilization requests. In the first year, fiscal 1998, 276 data sets were made available on the SSJDA. In each subsequent fiscal year, about 50 new data sets were released, reaching 499 in total by fiscal 2002. Starting from fiscal 2004, with the exception of fiscal 2008, about 100 new data sets were made available in each fiscal year, with the total reaching 1247 by the end of August 2009.

The number of data utilization requests was 11 in fiscal 1998 and 20 in fiscal 1999, but exceeded 50 in fiscal 2000, 100 in fiscal 2001 and 200 in fiscal 2002. In each subsequent fiscal year, the number of data utilization requests remained above 200, and reached 481 in fiscal 2008. In fiscal year 2009, we received 173 data requests between April and August. The total number of requests since the outset now exceeds 2800; the number grew rapidly from fiscal 2001 not only because the SSJDA became more widely known to the public, but also because the JGSS data, which can be used for multiple purposes, became available on the SSJDA and because of secondary analysis study group meetings which will be mentioned later.

In addition to the number of data utilization requests, we also keep track of the number of data sets made available upon request, even though this is not shown in the figure. In fiscal 2008, for example, we made 2034 data sets available in response to 481 requests, in other words, four data sets per request on average. The breakdown of data sets we provided upon request in fiscal year 2008 shows that 52% were for educational purposes and 48% for research. Not all data sets in the archive are available for educational purposes, but they do include comprehensive survey data such as JGSS data, and so the proportion of usage for education is relatively high.

The number of academic papers and books written using data sets from the SSJDA was 113 in fiscal 2008, while the total from the outset to August 2009 is 529. These figures cover only those cases reported by archive users or confirmed by the SSJDA management team, so the actual number could be much higher.

Center for Social Research and Data Archives

Activities concerning the SSJDA are closely connected with a wide variety of activities conducted at the Center for Social Research and Data Archives of the Institute of Social Science, University of Tokyo. Following a reorganization in April 2009, the Center staff presently consist of 14 academic staff (six professors, six associate professors and two research fellows), some of whom have additional posts, one technical expert (data archivist), one administrative worker and several part-time staff. The technical expert specializes in work related to the data archive itself, and so is equivalent to a data archivist in the West. However, the profession of data archivist, as a category of specialists like librarians, is not well established yet in Japan, so there is an urgent need to train experts in this field. In the meantime, professors and part-time staff at the Center try to substitute for data archivists. The Center consists of the following four divisions:

- (1) research base division which handles the administration of the SSJDA;
- (2) social research division which produces unique survey data from panel surveys conducted by the Institute of Social Science
- (3) quantitative analysis division which handles the promotion of secondary analyses and education on quantitative studies; and
- (4) international research division which promotes international cooperation on social research and data archives.

The Center is involved in three main activities concerning the SSJDA, as shown in Figure 3. The Center’s main activity is running the SSJDA as already described. As additional related activities, we conduct new social surveys to make the high-quality data available to public (preparation of primary data) and also encourage secondary analyses and education on quantitative analyses (encouragement of secondary analyses).

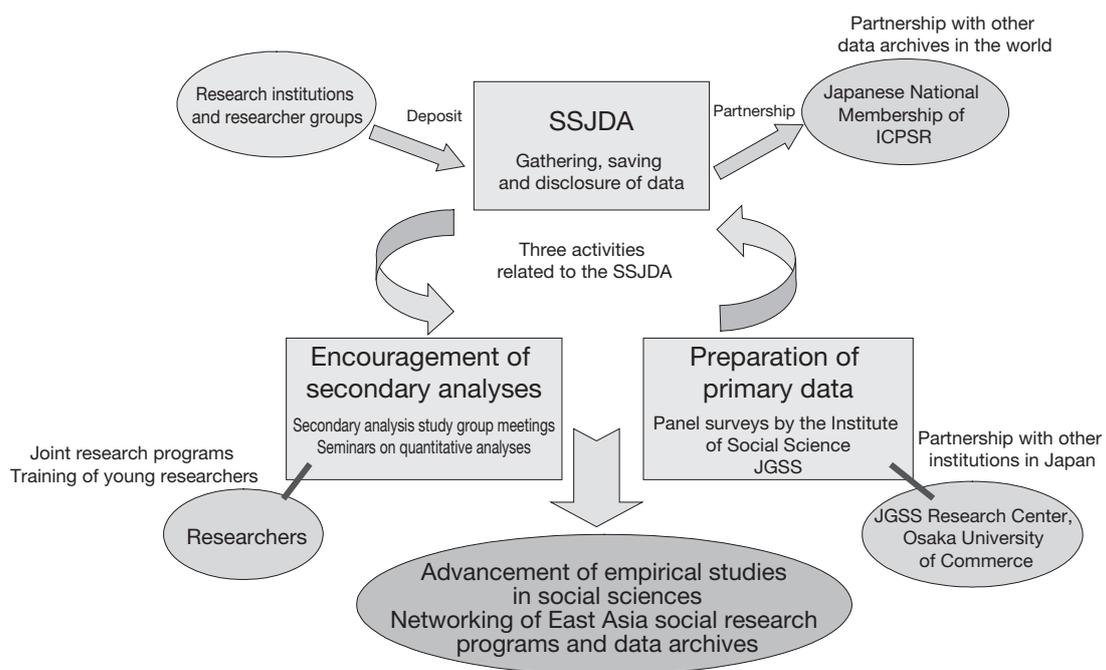


Figure 3 Activities at the Center for Social Research and Data Archives

The Center helps to prepare primary data by conducting JGSS jointly with the JGSS Research Center at the Osaka University of Commerce, and by conducting panel surveys as unique projects of the Institute of Social Science. Both are national surveys designed to make the survey data broadly available for academic purposes. Through careful cleaning and coding, we prepare high-quality primary social survey data. For basic items in questionnaires such as industry and occupation, respondents are asked to supply information descriptively. We then translate these descriptions into codes using industrial and occupational code systems. Keeping a data archive going requires a constant supply of high-quality data sets sufficient for advanced quantitative analyses. It is essential that such primary data be prepared by a research institution with sufficient capacity for conducting surveys. This condition ensures that data archives will be sustainable.

We started encouraging secondary analyses to help researchers who found it difficult to make full use of valuable data sets obtained from data archives. Even though we began to make data available from the SSJDA in fiscal year 1998, we only received 11 data requests in the first fiscal year, as mentioned before. This was because Japanese researchers were not aware of the concept of secondary analysis. Most researchers had little understanding of the methodology of secondary analysis, or quantitative analysis required for secondary analysis. Therefore, the Center helped to compile a basic textbook on secondary analysis (Sato, Ishida and Ikeda, 2000), and in fiscal year 2000, began to organize annual meetings of a secondary analysis study group. Researchers from institutions that deposited data to the SSJDA attended these meetings and gave lectures on the characteristics of data sets and how they can be used. Participants were invited to write papers on findings from secondary analyses of these data sets, which can take a whole year. At the end of each fiscal year, meetings are held to report the findings from secondary analyses. Study group members report their findings and external researchers comment on them, and the papers are published as part of the SSJDA's research paper series. The annual meeting of fiscal 2008 was held on the theme "Empirical Studies Based on the World Value Surveys." As a guest, we invited Seiko Yamazaki from Dentsu Soken. The meeting was attended by 24 researchers from 11 institutions.

In one of our other programs to encourage secondary analyses, we have been holding public seminars on quantitative analysis since fiscal year 2006. These seminars are held annually, for one to two weeks continuously in early March, and include lectures on various statistical techniques for quantitative analysis and on secondary analysis procedures. Some lectures involve practical laboratory training. The seminar in fiscal year 2008 consisted of five courses: "Introduction to Secondary Analysis," "Basics of Advanced Secondary Analysis," "Regression Analysis : Sharpen Your Sword," "Longitudinal Analysis of Micro Data" and "Basics of Quantitative Analysis with STATA." In total, 121 researchers from 80 institutions participated in the seminar.

From fiscal year 2005 to 2008, we presented SSJDA Excellent Paper Awards for distinguished research papers using data sets from the SSJDA. These awards, including both excellence awards and encouragement awards, have been given to 14 papers.

Thus, while concentrating on activities concerning the SSJDA, we focus on the systematic cycle of activities: namely, the preparation of data, the storage and distribution of data, training on using the data, and research activities using the data. In this way, we intend to contribute not only to the advancement of empirical studies in social sciences, but also to the networking of social research and data archives particularly in East Asia.

Further Challenges

Finally, I would like to describe the challenges that lie ahead for the SSJDA and the Center for Social Research and Data Archives.

First, we must maintain and expand the SSJDA. In response to rapid growth in the number of new data sets and number of users, we need to expand our manpower and services accordingly. We must urgently train experts such as data archivists and increase the efficiency of data distribution by making better use of SSJDA-Direct (“enhancement of data archive services”). We also need to convert the existing data sets to be compatible with the Data Documentation Initiative (DDI) format, which is a world standard format for databases. Conversion to the DDI format will greatly improve compatibility and linkage with other data archives in the world (“Global standardization of SSJDA”).

Second, regarding preparing the primary data, we must continue the ongoing panel survey projects. Since panel surveys keep track of the same individuals, they must be long-term to reveal the long-term evolution of careers and lifestyles of each individual. Distinct from cross-sectional surveys that are snapshots of the situation at a particular time, these projects become more valuable the longer they continue. Our challenge is to obtain institutional support to ensure the continuity of such projects (“Preparation of social survey data”).

Third, we must promote joint research programs that use data deposited at data archives, with open calls for such programs. We have already been supporting the training and skill-development, particularly of young researchers, by organizing secondary data analysis seminar meetings and quantitative analysis seminars. We will continue to encourage secondary data analyses and help upgrade the research (“Promotion of secondary analysis and human resource development”).

Fourth, we must build and strengthen partnerships with other institutions worldwide involved in data archives. It is particularly important to strengthen partnerships with data archives and research institutions in East Asia for sharing data and mutual cooperation (“International partnership”).

Even though many challenges remain, we will continue our activities with a long-term vision, supported by the researchers and research institutions who conduct surveys and use the archive data.

REFERENCES:

Tomohiko Sato and Hiroki Sato (2006), Roles of Data Archives and the Present Status of SSJDA - For Assuring the Reproducibility of Empirical Studies, *Nippon Rodo Kenkyu Zasshi*, Vol. 551, pp. 42–54

Hiroki Sato, Hiroshi Ishida and Kenichi Ikeda (ed.) (2000), *Secondary Analysis in Social Research*, University of Tokyo Press

Hiroki Sato (2006), Present Status and Challenges of SSJDA, *ESTRELA* Vol. 147, pp. 2–5

Keiichi Sato (2008), Remote Analysis System for SSJDA, *ESTRELA* Vol. 175, pp. 2–9

Hiroki Sato (2009), Development of Data Download System for SSJDA, *ESTRELA* Vol. 185, pp. 20–23

Explanation of the ESSRC, HUTE Educational Data Archive

Launching of Educational Data Archive “JEDI” and the Networking of Academic Data Archives

Yasuo Watanabe

Professor, Hyogo University of Teacher Education
Director, Educational and Social Survey Research Center

Introduction: Objective and Basic Policies

The Educational and Social Survey Research Center (ESSRC) was established in April 2005 as a research institution affiliated with the Hyogo University of Teacher Education (HUTE) with the intent of “*setting up and running a worldwide data archiving organization to play a pivotal role in the field of research on education in the age of globalization and information technology.*” The setting up of educational data archives and the collection, safe-keeping and disclosure of survey data are important missions at the heart of all ESSRC’s activities. ESSRC has established two basic policies concerning the setting up and management of educational data archives: first, on behalf of data depositors, we aim to handle as much of the saving and disclosing of valuable data as possible. The task of managing data properly to ensure it is always available for use and the task of disclosing data in an easily accessible manner are too burdensome to be undertaken by individual researchers, who may be unable to do so for many reasons. By offering to undertake these tasks on behalf of all researchers, ESSRC intends to decrease the overall social cost of surveys and researches concerning education. The second policy is to make data as freely accessible as possible. Since our aim is to share valuable data with all members of society, it is important not to place unreasonable restrictions on the free access to data, but to institute only the minimum restrictions required for management, while making arrangements that encourage data owners to deposit data.

“JEDI system” for the Educational Data-archive

Putting these policies in practice, ESSRC developed an online catalog system for use with new educational data archives, and started a web-based service for distributing social survey data concerning education. This service, which we started in July this year, is called the Japan Educational Data-archive Initiative (JEDI). The JEDI online catalog system (“JEDI system” hereafter) is based on the DSpace open source software which was jointly developed by the Massachusetts Institute of Technology and Hewlett-Packard Development Company, L.P. in the United States, with necessary changes to support the policies of ESSRC. DSpace is the global standard software used by institutional data repositories for collecting, managing, safe-keeping and disclosing various digitized resources. DSpace is already used by many national and private universities, libraries and research institutions in Japan as well. Since DSpace (including the engine, components and database) is freely available as open source software, it has greatly helped reduce our development cost. Moreover, as the de-facto world standard, the DSpace software is user-friendly, and supports unique control and management practices at each institutional data

repository. With many users around the world, DSpace archiving systems are expected to remain in use for a long time.

ESSRC developed the JEDI system for the storage and disclosure of all available survey data concerning education (including both raw data and related documents). However, we aimed to provide a universal architecture that allows the system to be used also for archiving data in academic fields other than education. The system is hardware-independent, so if a new dedicated server is unaffordable, the software can be installed on an existing server provided there is sufficient space.

Free Distribution of the JEDI Systems and the Networking of Archives

HUTE is ready to distribute the JEDI system to universities, colleges and research institutions in Japan free of charge, as universal software for academic data archiving, complete with related documents such as agreements, protocols and procedure descriptions. By making the JEDI system freely available, we aim to assist the management of valuable survey and experiment data presently kept at universities, colleges and research institutions in Japan, to prevent loss due to various causes, and to encourage the disclosure of these data on the Internet in Japan and worldwide, thus contributing to the progress of research and education. We expect many educational and research institutions will take up our offer and use the JEDI system. This will help us to retrieve valuable data that have been kept in isolation or left neglected throughout Japan. It is important to network as many data archives as possible to enable search and download across different archives. This will enable one-stop browsing of survey data via a personal computer in the laboratory or at home without having to visit many sites. By creating an independent network of data archives for each academic field including education, we will soon be able to produce in Japan a comprehensive academic data archiving system comparable with advanced networking systems for academic data, such as the one run by ICPSR in the United States and the one run by GESIS in Europe, which were described in the first half of the morning session. Japanese researchers have benefited greatly from these sophisticated overseas data archives. We should quickly prepare such data archives also in Japan to satisfy the needs of researchers in Japan and in the world. When building such a network of academic data archives, use of the same software at as many institutional data repositories as possible should be encouraged. This will not only minimize the development cost for each archive, but also will greatly facilitate the control, management and use of the entire network. To maximize the benefit of the activities conducted at ESSRC, HUTE encourages the use of this software and proposes the networking of academic data archives within each academic field such as education, not only to the participants of this symposium but also to others working at universities, colleges and research institutions in Japan.

Features of the JEDI System: Four Types of Access Controls

Next, I would like to describe specific features of the JEDI system. The most prominent feature of the JEDI system as a data archiving system is the availability of four types of access control arrangements. With the JEDI system, data deposited at the archive for disclosure are categorized into four access control levels. Level-1 data are “open

data" that anyone can freely browse and download. Metadata, composed of the index and catalog descriptions of deposited data, are usually handled as open data. Most metadata can be searched by Google. However, with some undisclosed data, the metadata may also be kept undisclosed. At present, the JEDI service provides a great amount of raw data as Level-1 data. Level-2 data require user authentication before browsing and downloading, and so are accessible only to registered users. Each student and faculty member at HUTE can access such data by using his or her mail address at the HUTE mail server as the user ID and entering the same password used to access the HUTE mail server. For someone who is not a student or faculty member at HUTE, entering an e-mail address on the web page normally results in receiving an e-mail, which is sent automatically for identification and allows the receiver to proceed to a user registration page, and then enter his or her name and other information. However, the JEDI system may restrict access using a blacklist, which may contain individual e-mail addresses as well as definitions of e-mail address groups. Following user registration, each user can freely browse and download data in this category, for which user authentication is mandatory. Level-3 data are accessible upon request. Before access to such data is allowed, the user must submit a specific request for approval by the data manager, using the online request form at the JEDI site. The user is requested to describe the intended purpose of use; ESSRC examines the request, and if it is reasonable, delivers the requested data files by e-mail, postal mail or other appropriate means. The distribution of Level-1, Level-2 and Level-3 data involves data duplication in some form, and users thus obtain duplicates of deposited data. Conventionally, in almost all such cases of allowing access to deposited data, the only way to prevent unrestricted redistribution was to require the user to sign an agreement prohibiting the redistribution of data to a third party. However, such arrangements for preventing redistribution may not be effective. The JEDI system does not allow any duplication of Level-4 data, which are made available for browsing only in a supervised environment. The request procedure for Level-3 data (upon request) and Level-4 data is similar. With Level-4 data, however, the permitted user must come to the university that keeps the data, where he or she is allowed to analyze the data under the supervision of management staff using a terminal that does not allow the duplication of data. Later on, the user receives hard copies of analysis results only. Yet even with such data that are made available only in a supervised environment, metadata are disclosed on the web. Except for that, DVD, MO disks and other media containing Level-4 data are stored in fire-proof safes that are specially designed to protect such electronic media, with absolute isolation from the data archiving system server. In addition to these four categories of data with different degrees of access control, there are undisclosed data protected from any external access (such data can be browsed only by the management staff). Therefore, if we include this arrangement for undisclosed data, the JEDI archival data fall into five categories according to the five levels of access control. In terms of accessibility on the web (or the availability of data on the web server), Level-1 and Level-2 data fall into the category of online data, while Level-3, Level-4 and undisclosed data fall into the category of undisclosed or offline data. In view of ESSRC's policy of allowing as much unrestricted access to the educational data archive as possible, it is desirable that most of the archival data are made available as Level-1 or Level-2 data.

I have described the five types of access control arrangements, including the four for disclosed data and the one for undisclosed data, from the standpoint of data users. Next, I would like to describe the five types of access control arrangements from the standpoint of data depositors, who are on the other side of the archive. Data depositors decide to deposit data to an archive under a wide variety of circumstances. Since the standpoints of

academic societies, the government and industries will be discussed in the afternoon session, I would like to focus on the standpoint of individual researchers who own data. Let us suppose that a researcher, who may later deposit his or her data to an archive, conducts a survey. He or she will conduct this survey according to a deliberate plan, focusing on a certain subject that matches his or her academic interest. The researcher will obtain various data from his or her survey, and analyze them in a preferred theoretical framework. If fortunate, the researcher will arrive at academically valuable findings or interesting conclusions, and then will present or publish the results. The whole project from planning a survey to publishing findings requires much time, labor and cost. However, typically, some parts of the data remain unanalyzed or are left without conclusions. This is why data should be made available for secondary use by others. However, researchers are often keen to complete as much of the analysis by themselves, even if very time-consuming, and so are naturally reluctant to disclose the data too early. The five types of access control arrangements supported by the JEDI system allow careful reconciliation of such conflicts of interests between the original producers and secondary users of survey data.

Data depositors are always free to choose any of the four access control levels for disclosure, and are even allowed to keep their data undisclosed. Some researchers may prefer to save their data to the archive as soon as possible to be released from the task of having to manage and protect the data from loss and obsolescence. Still, they may wish to keep their data undisclosed until they complete further analysis and publish their findings. In such cases, researchers may choose to specify a date until when their data should remain undisclosed. The access control level (including the choice of nondisclosure) does not have to be the same for all blocks of data provided by the depositor; different blocks of data can be assigned different access control levels. The whole body of data provided by a depositor usually contains not only raw data files and basic summary data files, but also a wide variety of supplementary materials such as original questionnaire sheets, reports, parameter code lists, survey implementation manuals, data editing or processing records, reference materials concerning target individuals, and an index of such materials. Since the personal information of target individuals must be kept strictly confidential, some parameters in data files must be protected from leakage. Not only is it necessary to protect these parameters as such, but it is also necessary to prevent the guessing of personal information by combining data from a certain parameter with other pieces of information. Such misuse must be prevented by carefully handling such data. The access control level (including the choice of nondisclosure) can be assigned differently to different data files, parameters and supplementary documents. To enable exact and efficient handling of such complex assignment of access control levels, a form titled "Index and Terms of Use of Deposited Files" in Excel format is provided as an attachment to the "Data Deposit Agreement" form. Based on the access control requirements specified using this form, ESSRC works upon the deposited data as required, such as protecting the confidentiality of personal information, before making the data available at the specified access control level.

Deposited data may need to be reworked in various ways depending on how they are going to be used. However, such reworking, even if done to protect personal information, decreases the value of the data. Therefore, ESSRC recommends data depositors to hand over the raw data as they are; the raw data are kept at ESSRC as master data in an offline environment. Even Level-1 and Level-2 data are reworked before being made available on the server. We are committed to maintaining and improving such arrangements for strict control and management of

information, as well as the robustness of system security, in order to retain the trust of data depositors.

Other Features of the JEDI System

In addition to the four types of access control arrangements, the JEDI system has several other features. The portal window presently displayed by the JEDI system is very simple. Like a Google screen, it only contains an input field for searches. Clicking on Browse on the menu bar at the top left of the screen displays a menu of frequently-used tag descriptors and sort keys. The main display area at the center of the screen displays general messages from ESSRC and a menu with links to pages that display policies concerning the control and management of the archive. A frame on the right provides links to JEDI-related news articles, starting with the latest one, as well as links to external databases.

Instead of a hierarchical index structure in the style of library catalogs, the JEDI system uses tags for cataloging data. The difference between the two is like the difference of e-mail sorting methods between Microsoft Outlook and Google Mail. Tags are sort keys (or catalog headings) that help identify the categories to which the data belong. Tags are parts of metadata, and are written using world-standard descriptors such as ERIC, MeSH and APA. In addition, the JEDI system supports a unique metadata description scheme that ESSRC developed specifically for empirical education studies. The JEDI search engine looks for matching entries not only in metadata but also in data files and document files that are available online. The search operation extends even to scanned materials in PDF format provided their contents have been read out using OCR and attached to original documents as text files. Tag contents are handled as Level-1 data, and therefore are accessible to ordinary search engines on the web such as Google and Yahoo. This facilitates search and access from all over the world.

Following the completion of development activities planned for this fiscal year, the JEDI's online system is going to be equipped with R, a free software environment for statistical analysis. R provides a software environment for data analysis using the S language, which was developed by the Bell Telephone Laboratory. R is freely distributed as open source software, and can run on Linux, Unix, MacOS and Windows. Many experts around the world are involved in the development of R, adding new methodologies and algorithms from time to time. R is highly functional, capable of fast computing, and supports many types of graphical presentation. Once R becomes available on the JEDI's online system, users will be able to perform various statistical analyses on the web without having to download data from the archive. We are going to make the JEDI system support the R-compatible data format. Therefore, as the system is developed and services are enhanced, users will become able to download data in the R-compatible data format, analyze the downloaded data immediately in the laboratory or at home, and use the analysis results.

We also plan to prepare help screens for the JEDI system and begin to support major languages other than English. And similar to shopping carts at Amazon, we plan to add a feature that allows each user to display his or her browsing history with a list of browsed data sets, and a space to temporarily store multiple target files before batch downloading.

At the end of our system development activities this fiscal year, we are going to ask reliable experts to perform system vulnerability testing to ensure there are no security holes in the JEDI system. In order to maintain and improve system security, this vulnerability testing will be periodically repeated by professionals after major system modification and development work, or to respond to changes in the security environment.

Three Categories of Collected Data and Their Secondary Use

Data collected by ESSRC fall into three categories. The first category is data collected by researchers in the pursuit of specific research objectives. A single collection of such data could be bulky in certain cases, but usually is nothing more than a relatively small collection of survey data concerning a specific field of interest, which is often limited in geographical coverage, timeframe and scope. Data in this first category are collected by individual researchers who conduct surveys with limited funds from the university's basic research budget, or supported by external financial aid such as a scientific research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology, and the results of such surveys are reported at academic conferences and published in academic journals. The second category is data systematically collected by administrative agencies, typically of the national government, governmental research institutions, or by private corporations and affiliated research institutions. With data in this category, such as national census data, a single collection of data is often enormous and extensive in scope, containing the results of surveys continued over a long period. Data in this category may be called "flagship survey data." The third category, unique to ESSRC which specializes in surveys concerning education, is data collected by teachers for practical use in the classroom and for verifying the effectiveness of lessons. A single collection of such empirical data from teachers is limited in scope, and usually pertains to students in the teacher's classroom, in a specific school year, or in a whole school. Nevertheless, such collections of data are significant because they provide detailed information on individual students that is often inaccessible to researchers at universities. Comparing the characteristics of these three categories of data, data in the first category excel in the robustness of survey design and analytical framework, but are handicapped by the limited scale and scope of surveys. Data in the second category, which is flagship survey data, are enormous and can be used for multiple purposes because they come from major continuous survey programs. Data in the third category are handicapped by the limited scale and scope of surveys, but are valuable because they often contain detailed information on individual students. In conclusion, the data of all three categories are valuable in their own ways.

It is widely acknowledged that data archives help protect valuable data from the risk of loss, neglect and obsolescence, and also encourage secondary use. I mentioned in the beginning that the two basic policies of ESSRC are to undertake the task of data storage and management on behalf of data depositors and to make the data as freely available as possible. Both of these policies are meant to encourage the secondary use of data. However, it is important to ask: "Is such secondary use really possible?" or "What is the significance of such secondary use?" Since researchers often express skepticism concerning the secondary use of data, I would like to take this opportunity to describe the importance of secondary use and the role that archives are expected to play.

With regard to the second of the three categories of data, namely, flagship survey data, I don't think that anyone

doubts the significance of their secondary use. The GSS data managed and disclosed by ICPSR of the United States, come from the General Social Surveys, which are like national surveys in the United States. The GSS data have been put to secondary use in many different ways, often to the reporting of important research findings. As something similar, Japan General Social Surveys (JGSS) are conducted with a strong initiative from the Osaka University of Commerce. The JGSS data are made available at ICPSR and SSJDA.

However, a secondary user may find that even flagship survey data from complete enumeration surveys do not provide information on a few of the parameters of interest. The secondary user may then choose to conduct additional surveys on such missing parameters. In early August this year, newspapers reported that a group of researchers led by Professor Hiroaki Mimizuka of Ochanomizu University discovered a positive correlation between the academic achievements of children and the income of their parents as a result of analysis that the group conducted after an additional survey on the income of the parents of about 5,800 students in the sixth grade of elementary schools, as a subgroup of students who had been covered by the national academic achievement survey of the last fiscal year. The correlation between the socio-economic status of parents and the academic achievements of their children is accepted overseas as a basic hypothesis, and has been supported by many empirical studies. However, this was the first time that this hypothesis was verified in Japan by a major study with the involvement of the Ministry of Education, Culture, Sports, Science and Technology. Merging data from different sources, like in this example, opens up possibilities for new types of analysis not possible without such merging. To facilitate such merging of data, it is important to pool many collections of data at a single repository, where each single collection of data can be from a survey that is limited in scale and scope. ESSRC may make some chosen data sets available in merged format, or may provide a data merge service to prepare merged data sets according to requests from users. Even if each of the merged data sets contains confidential information such as personal information, and even if the given piece of information is required by the merging operation as a key parameter, we can make the merged data set available without including such confidential information. This is an advantage that may encourage the secondary use of empirical data from teachers by general researchers. In conclusion, data archives provide functions and services that greatly enhance the possibility and significance of secondary use by removing many obstacles that researchers will encounter if they try to collect data themselves.

More explanations will be required to resolve doubts about the secondary use of data from researchers and empirical data from teachers. Regarding data from researchers, the disclosure of such data is very important as the issue of scientific method because it assures the reproducibility of primary analyses. In other words, this allows other researchers to verify the results. Moreover, the same data can be analyzed from different angles, using different theoretical frameworks. For example, as reported by Professor Ishida today, the Institute of Social Science, University of Tokyo, organizes a study group for secondary analysis, the outputs from which are reported at conferences of sociological societies and published as contributed articles in the academic journals. Equally important is the secondary use of empirical data from teachers, which may be used as training materials that provide insights on survey techniques and analytical methods. For researchers and teachers who are planning to design and implement a new survey, materials from earlier surveys, such as questionnaire sheets, descriptions of analytical methods and reports on survey results are valuable reference materials when designing surveys.

However, data in these two categories, namely, data from researchers and data from teachers, are limited in the scale and scope of surveys, and if not combined, will not attract many secondary users.

Since new surveys are often designed after earlier surveys, similar surveys are repeated in different localities at different timings. The term “meta analysis” is used in connection with methodologies and statistical analysis approaches that aim to reach more reliable conclusions by combining data from two or more past surveys. While the analysis of data from a single survey alone may not lead to sufficiently reliable conclusions due to insufficient samples, meta analysis allows higher significance and lower standard error. Even though meta analysis can be useful, there are some problems unique to this approach. One such problem is known as “publication bias” or “neglect of unpublished data.” The problem is that the whole collection of data sets addressed by meta analysis may be composed too exclusively of published data, that is, data that have passed the screening process with the acknowledgment of statistical significance. As a more fundamental problem, the comparative analysis of data from different surveys that address different groups according to the approach of “multi-population analysis” is discouraged because raw data from earlier surveys, which served as models of new surveys, mostly remain unavailable to general researchers except in rare cases. Therefore, in spite of the availability of new survey data that could be compared with earlier survey data, the new survey data can be compared only with embedded statistical values and indices such as mean values that are available in published literature. When the JEDI and other archives begin to collect and make available large quantities of all types of data, including not only data from advanced research projects such as those published in journals, but also data from surveys that are too minor to be able to assert statistical significance on their own, this may solve the above-mentioned problems with meta analysis such as publication bias and the difficulty of multi-population analysis.

Another important mission of data archives is to reduce the risk of data loss due to disasters and accidents. Particularly in Japan which is frequently hit by natural calamities such as earthquakes, it is important that duplicates of master data on electronic media are stored in different locations. It is also important to make arrangements to ensure that, whenever data on the system’s server is lost or rendered irretrievable due to an accident or error, a backup is provided by a server at another institution. As announced at the beginning of this presentation, HUTE proposes free distribution of the JEDI system and the networking of archives. This is an important proposal for reducing the risk of data loss. It goes without saying that flagship survey data must be protected because they were collected by surveys at enormous cost. But data from smaller surveys must also be protected because, with any type of data, a single accident may result in permanent loss. There is thus an urgent need to network academic data archives.

Conclusion: the Future of the Educational Data Archive

Thanks to the completion of the development of a new online system, the educational data archive at ESSRC was launched in July this year. Ours still cannot be compared with the advanced data archives of the United States, Europe and Japan, as described in the first half of the morning session. We must admit that our efforts to collect more data for storage in our educational data archive JEDI have lagged behind so far. Lacking in achievements,

we can only speak about our dreams. Therefore, please allow me to finish my presentation by speaking about our hopes for the future.

With your kind understanding and support, ESSRC will continue to expand the collection of the three categories of data that I mentioned earlier. It is particularly important to encourage the depositing of flagship survey data owned by administrative agencies and education industries. Such data may serve as the nucleus that will attract data from researchers and empirical data from teachers and also as the source of strong motivation for secondary use. We understand the need to keep some data undisclosed for the time being, so owners are invited to use the educational data archive rather like a bank’s safety deposit box. By that, I mean they can deposit their data into the educational data archive and be freed from the task of data management and risk of loss. Representatives of academic societies are encouraged to make the depositing and disclosure of survey data as a condition for publishing academic papers in their journals. The disclosure of survey data endorses the reproducibility of analyses reported by academic papers. This practice, therefore, will build global trust and respect toward academic research in Japan. Furthermore, researchers are invited to deposit their survey data to make public the results of surveys conducted using public funds, aided by a scientific research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology, for example, and to enable the sharing of valuable research findings and data. Finally, teachers, undergraduate students and graduate students are invited to use data and materials in the educational data archive to stimulate objective, realistic and substantial discussions on the subject of education based on the results of empirical studies and also to facilitate problem-solving in the classroom. Moreover, they are encouraged to conduct their own surveys and deposit the data and analysis results into the archive for use by children and younger researchers.

In this presentation, I have reported our activities at ESSRC concerning the educational data archive. With the cooperation and support from the participants of this symposium and others involved in researches concerning education, teaching activities, administration and industrial activities, we will continue to work for the advancement of education in Japan and the development of international communities that share the same interest. Now, I would like to invite the panelists to give us honest feedback about the educational data archive project of ESSRC.

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 1



Data Archives and International Collaboration for Research, Public policy, and Business:
 As a witness of the Launch of Educational Data Archive of ESSRC, HUTE

ESSRC
JEDI

Launching of **JEDI** for evidence-based education Educational Data Archive and the Networking of Academic Data Archives

Held on November 22nd, 2009, Sun. at Int'l Conference Room 301, Kobe Int'l Conference Center
 Sponsored by Educational and social Survey Research Center, Hyogo University of Teacher Education

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 2



Objective & Basic Policies

ESSRC's Objective

Setting up and running a worldwide data archiving organization to play a pivotal role in the field of research on education in the age of globalization and information technology

JEDI's Basic Policies

- ① To store and disclose the valuable data on behalf of data depositors.
- ② To make data as freely accessible as possible.

↓

Maximizing the benefit for both depositor and user

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 3



JEDI System for the Educational Data Archive Japan Educational Data-archive Initiative

- Based on DSpace developed by MIT and HP.
- Freely available as open-source software.
- Introduced by many universities, libraries, and research institutions also in Japan.
- Global standard for institutional repositories.
- Easy to use; standardized control/management.
- Expected to remain in use for a long time.

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 4



Free Distribution of the JEDI System and the Networking of Archives

- HUTE is ready to distribute the **JEDI** system free of charge, complete with related documentation such as agreements, protocols and procedure descriptions.
- HUTE proposes the establishment of the Japan Academic Data-archive Network (tentative name).

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 5



Features of JEDI: Four Types Access Controls

Publicly Disclosed	Level1	Open Data	Online	Can be copied
	Level2	Data requiring Verification		
	Level3	Data requiring Approval	Offline	Cannot be copied
	Level4	Data under Surveillance		
Undisclosed	up to a specified date for an indefinite period			

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 6



Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 7

Three Categories of Collected data

① Researchers' Data	② Flagship Data	③ Teachers' Data
Collected by individual researchers; Published in academic journals	Systematically collected by institutional efforts	Collected by teachers for practical purposes
Limited in scale and scope of survey	Large scale and longitudinal study	Limited scale and scope of survey
Robustness of theoretical and analytical frameworks	Ideal for secondary use	Contains detailed information on individual pupil

Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 8

Possibility of Secondary Use

① Researchers' Data	② Flagship Data	③ Teachers' Data
Support follow-up surveys	Support supplemental research	Training materials for survey techs. and statistics
Meta-analysis Multi-population analysis	Provide data-merging services	Reference to a new survey design

- Hyogo University of Teacher Education | Educational and Social Survey Research Center | International Symposium 2009 9
- ### Request for Data Depositing
- 【 Flagship data 】 To administrative agencies and education industries
 - As a core to attract other data; strongly motivate secondary use.
 - As a bank's safety deposit for data that must remain undisclosed.
 - 【 Researchers' data 】 To individual researchers and academic societies
 - To return the benefit from survey aided by public grant such as the scientific research grant from the MEXT, Japan to society.
 - As a condition for publishing papers in academic journals.
 - 【 Teachers' data 】 To teachers, graduates, and undergraduates
 - To stimulate the “evidence-based” arguments on education.
 - To help children, teachers and researchers in the future.



Chapter 1-II

Assessment of and comments
on the Educational Data Archive

Assessment of and comments on the Educational Data Archive (Conference minutes)

(Granda) Prof. Watanabe-san's paper covers all the important aspects of the data archiving, preservation, and dissemination processes. I think it is particularly significant that he stressed the importance of metadata for data discovery among members of the research community. I was very impressed with the discussion in the paper which described the four types of access controls of the JEDI system.

Many archives began their existence with the clear objective of distributing open or Level One data. This, of course, was a generally accepted principle years ago when data sources were relatively rare and privacy was not nearly such a sensitive topic as it is today. Today, great care must be taken in disseminating data because of the existence of large amounts of publicly available administrative and commercial data. Because of enhanced computing power, these types of data have a greater potential to endanger respondent confidentiality if they can be merged with microdata.

ESSRC has developed a system to classify data into five different types that allows access to researchers on a scale from open on one side to a secure enclave on the other side where the user must analyze the data under very restrictive conditions. Now these varieties of data availability exist in many archives only now after a long development period, but the ESSRC has been able to create such a sophisticated system as they built their basic dissemination system from the very beginning. This allows them to address the needs of all types of depositors and to assure them that their data will be handled appropriately.

I am also very pleased to see that the ESSRC is planning to use the R software in their proposed online analysis system. The use of open source software products is the direction that many members of the international data archive community are moving as they try to integrate their services more closely with one another.

Prof. Watanabe-san also describes three categories of data; those collected by researchers for a specific purpose, those collected by administrative or government agencies which he called "flagship survey data" and those collected by teachers in the classroom. While he rightly emphasizes the importance of the flagship survey data because of its completeness, it is also important to note that many educational researchers are now very interested in the value of both the quantitative and the qualitative data collected in the classroom and are doing much more research with these data than ever before. The use of such data raises very probing questions about confidentiality, particularly where video data is concerned.

I only would disagree with Prof. Watanabe-san on one point that he did not mention specifically in his presentation because of time. He says that one 'ESSRC' still cannot be compared with the advanced data archives of the United States, Europe, and Japan, as described in the first half of the morning session. On the contrary, I believe that ESSRC has made a wonderful beginning as an educational data archive and can indeed be compared very favorably with other archives elsewhere in the world.

(Watanabe) Thank you very much for your generous comments and your kind evaluation of our work. Next, I would like to hear from Dr. Mochmann.

(Mochmann) As frequently it is quite comfortable to follow after ICPSR has spoken, as they address the major points and then you can pick up what is left. So I can be happy to state that I agree very much with what Peter just said.

I want to emphasize four points. The one is that it is certainly a clever idea to combine standard software and to build it on DSpace and combining it with R. I know that 'R' is being picked up in Europe as well, it is part of the teaching in the famous Essex Summer School, but it is not yet a standard which is being used by researchers, so this means that you may face an educational task in that field for social researchers and offer training in this analysis technology, if you want your system being used.

The second point in that context is that software solutions, even if they are platform independent, will change quite rapidly. They will be around for three to five years and then a new generation will come. Therefore, it is particularly important that you put emphasis on the standards of the metadata and the data formats themselves. As Prof. Watanabe pointed out, the methodological challenge is not only to retrieve data, but also to combine data of different origins and accumulate data, etc. So you have a need on the data management side which is metadata interoperability, but also data interoperability and I think that is at the core of the archival work. Even though this is not very much visible to the outside that is where lots of the investments are going. So, the question is to what extent are those metadata standards being used in the JEDI system compatible to the other archival standards or can they be transformed?

I share the perspective of classifying the different types of data. We are used to talking about reference data which is normally the data from big continuous data collection programs, which are widely published in the literature and are acknowledged. We are talking about research data which are in the research process, but not yet so well known and not yet a standard and we are talking about data which is not yet processed to archival standards. In all those categories, of course, you will have to distinguish between data that can be accessed without problems of freedom of information but there also will be some types of sensitive data at least on some variables which must be taken particularly care of.

Finally, this is one of my specialties, as I am also working in the field of Computer-Aided Content Analysis, the big question about hierarchical classifications and tagging according to Google style. I am not so well familiar with your constituency which is the teachers making use of the data. As sociological researchers, we like to access data via the language of sociological theory looking for data to test particular assumptions and hypotheses, but the theory is not what you find in the target language of the questionnaires and the measurement instruments.

So, how do you assess that the theory language is compatible to the object language of your measurement and is that done as an automatic extraction of meaning process or is there some intellectual process involved? Thank you very much.

(Watanabe) Thank you for your comments, Dr. Mochmann. As an early pioneer in the field of data archives, you have a deep knowledge of the future of data archives. Your advice on our system reflects your vision of how data archives may evolve in the future. Next, I would like to hear from Prof. Ishida from the University of Tokyo.

(Ishida) Even though Dr. Mochmann said that it was easy to follow after ICPSR had spoken, it is not easy for me to add something after listening to comments from the representatives from ICPSR and GESIS, which are two of the world's famous data archiving organizations. Still, I would like to comment on three points.

First, I would like to say that data archives are like libraries for data. There are libraries for books everywhere in Japan. Each university has one. Some university libraries are unique, such as having a great collection of books on a particular subject. Ideally, data archives should be like that. It may be too much to expect each university to have a data archive, but I would like several universities to have a data archive with certain uniqueness.

In that respect, it is very encouraging to hear about the opening of an educational data archive at HUTE. We have been running the Social Science Japan Data Archive (SSJDA) for about 10 years. Even though we were not entirely without support, we always expected others to start up a data archive like we did. HUTE decided to set up an educational data archive, and I think we should celebrate today's memorable event.

I would like to comment on two other points based on our experience at SSJDA. First, as mentioned by Dr. Mochmann, much of the work done by a data archive is invisible to the receivers of data services. In other words, a data archive is a hidden infrastructure. There is much work to be done in the process between accepting deposited data and then providing the data to the public. Moreover, making data properly available on the web using various technologies is rather like building invisible mechanisms or an invisible infrastructure, and it requires a huge investment. Now that HUTE has set up a data archive after a tremendous effort, we hope they will continue to maintain and improve the infrastructure.

Second, I would like to say something about our early experience: we only received 11 requests for data in 1998, the first year of SSJDA. Earlier, when I was a post-graduate student at a university overseas, it was a real struggle to obtain the survey data from Japan that I needed for my doctoral thesis. I returned to Japan with a wish to help create an archive of social science data, and that is how I came to participate in the startup of SSJDA. I expected to receive many requests to use data when we opened our data archive, but we received so few requests, which was very disappointing. It turned out that people were reluctant to use our data archive because they did not know how to use it or how they could conduct secondary analyses.

The HUTE's data archive specializes in collecting educational data. To be able to make the best use of such data, people will probably have to learn particular ways of using or analyzing the data. It is not enough simply to make the data available to people. I encourage you to give people advice and training on how to analyze the data. Then more people will begin to use your data archive and you

will begin to hear about excellent studies conducted with the help of your data archive. I make these two comments in the hope that you may attain this ideal.

(Watanabe) Thank you very much for your suggestions, Prof. Ishida. You said it is not enough that we simply make data available at our data archive and that we should teach people how they can use the data. I remember that Dr. Mochmann also emphasized this point. Thank you very much.



Chapter 2

Presentation and Discussion

Discussion I: Data Archives and Academic Research

The significance of data preservation and dissemination in the area of longitudinal cohort studies

Masuo Koyasu

Professor, Kyoto University
President, Japan Society of Developmental Psychology

Profile of the topic presenter:

Masuo Koyasu was born in Kyoto, in 1950. He graduated from the Faculty of Education of Kyoto University (majoring in educational psychology) in 1973 and studied at Kyoto University's Graduate School of Education until leaving the school partway through the doctorate course in 1977. After serving as assistant professor and associate professor at Aichi University of Education, he has been working at Faculty of Education, Kyoto University, as associate professor since 1988 and as professor since 1997. Masuo Koyasu has a PhD in psychology. He currently specializes in "cognitive psychology in education", conducting researches on the development of cognitive ability in young childhood and childhood. He is known for his recent publications and papers on the "theory of mind."

1. Intent of the Topic Presentation

- Emphasis on the following message concerning the purpose of the project: *"In pursuing the educational data archive project, our aim is to gather and save the valuable survey data concerning education that would otherwise be scattered, lost, or obsolete, to make these data available on the Internet and elsewhere in properly compiled or edited formats, and to encourage their secondary use for empirical research."*
- Summary of request: The topic presenter wishes to speak for 10 minutes on topics related to survey data concerning education, data archives, academic researches and academic societies.
- Suggestion from the standpoint of developmental psychology: emphasis on issues that concern longitudinal cohort studies.

2. Archiving of Data from Longitudinal Cohort Studies

- The term "longitudinal cohort study" refers to a research study that involves long-term tracking of the target population. Such a project, for example, may involve the implementation of a certain educational, training or treatment program as a part of the project, followed by monthly or annual surveys on how the benefits from the program may change after the program ends. More broadly, the same term may refer to study projects that involve the long-term tracking of developmental changes manifested by the target population with or without the implementation of a certain program as a part of the project. If the target population shares the same birth period or experience of a particular social event, the researches fall into the category of cohort studies.

- Examples of cohort studies in the world (1) - Canada:

In Canada, two longitudinal study projects are in progress. The National Longitudinal Study of Children and Youth, which started in 1994, targets 22,831 children (born 1983-1994, divided into 10 cohort groups), and involves a survey once every two years. The Québec Longitudinal Study of Child Development, which started in 1998, targets about 2,000 children.

- Examples of cohort studies in the world (2) - Great Britain:

In Great Britain, the UK Millennium Cohort Study was started in 2000 by the Longitudinal Study Center of the Institute of Education, University of London. The project targets 18,818 children in 398 localities of Great Britain, and involves the tracking of these children at nine months from birth and at the ages of three, five and seven.

- Examples of cohort studies in the world (3) - Australia:

- Data collection activities for the Longitudinal Study of Australian Children started in 2003. The project targets about 5,000 infants of up to 1 year old and about 5,000 young children at the age of four or five, and involves a survey of these children once every two years.

- Examples of cohort studies in the world (4) - “Study of Broadcasting Programs Good for Children” by the NHK Broadcasting Culture Research Institute

The NHK (Japan Broadcasting Corporation) Broadcasting Culture Research Institute started in 2002 a 12-year project for studying broadcasting programs good for children. The target population was selected as follows: out of about 6,000 infants born in Kawasaki City, Kanagawa Prefecture, in the six months from February to July, 2002, about 1,600 infants were chosen at random, out of which the infants in about 1,368 households that agreed with the project were finally selected as targets. The project will run for 12 years, and involves annual questionnaire surveys (by mail), usually in January. The topic presenter is involved in the project.

3. Need for the Disclosure of Data Archives

The importance of disclosing data archives, containing data from completed study projects, is as follows:

- Data verifiability: Data should be made available for reanalysis by researchers other than the main researchers.
- Data accessibility: The main researchers alone will not be able to study the enormous amount of data in full detail.
- Historical importance of data: The archives contain data that would have remained unavailable if they had not been collected at the time they were.
- Making research results available to research sponsors: Researches are conducted using public funds (e.g. taxes). Research results, therefore, should be disclosed to the



Chapter 2 Presentation and Discussion

Discussion I: Data Archives and Academic Research (Conference minutes)

(Watanabe) Thank you for your presentation, Prof. Koyasu. As something related to the importance of panel data discussed in the morning presentation by Prof. Ishida from the University of Tokyo, Prof. Koyasu emphasized the importance of longitudinal data from long-term surveys. He also spoke on the theme of how data archives may be used for stimulating the activities of academic societies. Does anyone wish to speak on related topics or have any questions for Prof. Koyasu?

(Mochmann) Just a short observation. I fully agree that longitudinal data are of high importance and we are also putting emphasis on long-term observational data. In our holdings of our archive, we do have some 40 longitudinal studies from research on educational topics in Germany.

(Watanabe) Dr. Mochmann spoke about the availability of longitudinal data in Germany. Prof. Koyasu, do you have any questions about data archives in Germany?

(Koyasu) Not particularly about the situation in Germany, I believe that Dr. Mochmann acknowledged the general importance of making longitudinal data available at data archives.

(Watanabe) Isn't there any data archive in Japan that offers longitudinal data? Prof. Ishida, you mentioned that you have already been collecting panel data that will be kept at SSJDA (University of Tokyo). What about longitudinal data?

(Ishida) My knowledge is not very comprehensive, but as far as I know, there is very little data from panel surveys in Japan. The panel surveys that we conduct described earlier are available at the SSJ data archive. There are other panel surveys. For example, the Institute for Research on Household Economics releases its long-term panel data among Japanese women, and Keio University makes available Keio Household Panel Survey and Japan Household Panel Survey. However, I agree with Prof. Koyasu that the accumulation of panel data is limited and that the importance of panel data is not well understood. Overseas, not only is there more data available from panel surveys but also the data are beginning to be analyzed more extensively. Recently in some countries, at least in the field of sociology in which I specialize, most researches are based on data from longitudinal studies. This is because if you really wish to identify changes in people's behaviors and their ways of thinking, you naturally have to keep on tracking the same individuals for a long time. For example, even though a single survey may indicate that married individuals are happier than unmarried individuals, this does not necessarily suggest that marriage, as an event, contributes to the difference in happiness observed between married and unmarried individuals.

Only by continually tracking the same individuals to compare their levels of happiness before and after marriage can you determine how marriage contributed to the level of their happiness. Such a longitudinal approach enables more accurate tracking of changes in society. Therefore, I find longitudinal data invaluable for social science studies. We expect more research organizations to conduct such longitudinal studies and make the survey data available to others. This will lead to rapid progress in researches.

(Koyasu) I would like to add a few words concerning the importance of longitudinal studies. I believe that the development of the concept of cohort led to awareness of the importance of longitudinal studies. Let us assume that a single survey revealed differences among age groups (10-19, 20-29, 30-39, 40-49, etc.). Is the difference observed between the 20-29 age group and the 60-69 age group attributable to aging or to differences in their experiences in society such as differences in the education they received in childhood? This cannot be determined by analyzing survey data. Thus, the concept of cohort began to challenge the conventional classification by age group.

I would also like to mention something about the national surveys on academic achievement that are conducted today in a format developed under the guidance of President Kajita. These surveys target students in the sixth year of elementary school and the third year of junior high school. If the difference in scores between these two times can be analyzed with reference to each individual student, the data from these surveys point to real changes that must have taken place in the person during the given period.

When analyzing data from panel surveys of recent years or data from multiple-year surveys, researchers often use the structural equation modeling approach. This approach is found to be the best because it helps clarify cause-and-effect relationships. By factoring in the elapse of time, it is possible to narrow down the list of potential causes because subsequent events cannot be a cause while preceding events can be a cause. Thus, the inclusion of a time variable is a great advantage. To allow researchers to use a structural equation model, it is important that a time variable is added. In addition, it is important that the reliability of analysis is maintained by observing changes of the same individuals over time. In that sense, longitudinal data are very valuable.

Data from the national surveys on academic achievement could be even more valuable if they allowed analysis of changes between the sixth year of elementary school and the third year of junior high school with reference to each individual student, but unfortunately the survey data are lacking in this respect.

(Granda) It is about something that Prof. Koyasu said at the very beginning. He mentioned that he was not an expert in data archiving, but it is so important for the producers of the data to be at least a little familiar with data archiving and to be in touch with the data archivists when they are collecting the data from the very beginning. This will only lead to have better quality data and better documentation for the end-users if the researchers at the very beginning of the data collection process are aware that there will be secondary use of their data later on. So, I think it is very valuable that he and other data producers are here today.

(Koyasu) Thank you, Dr. Granda, for your valuable comments.

(Watanabe) Dr. Granda manages an educational data archive at ICPSR. Does ICPSR offer longitudinal data from long-term observation of the same individuals? If you make such data available, can you give us some examples of how they are used?

(Granda) We have longitudinal data, collected both by our public institutions and by our national government. There is the Early Childhood Longitudinal Study in the United States which is done by the US Department of Education which has gone on for many years and follows groups of students from when they are very young to when they are very old, and we also have at the Institute for Social Research at the University where I work, another survey called 'Monitoring The Future' in which they have studied also going back to the 1970's students and follow them through until they are adults.

All of these data sets have received very heavy use by researchers. They are available publicly, some through ICPSR, but the producers of Monitoring the Future also maintain their own very rich web site and user support system and have educational researchers on staff who can give lots of advice to users about the data. Because as you know much better than I do, this longitudinal data is much more difficult and complex to use than ordinary survey data and requires more training and experience, so the responsibility to document it correctly is very high I think.

(Watanabe) Thank you very much. Next, I would like to hear from Mr. Arai, Director of the Benesse Educational Research and Development Center. Since our field of vision is often confined to academic interests, I would like to ask him if Benesse Corporation has longitudinal data concerning education and if Benesse Corporation conducts longitudinal studies. Mr. Arai, do you have such longitudinal data at Benesse Corporation? If you do, can you describe them and the challenges you have experienced in conducting longitudinal studies?

(Arai) We have been experimenting a little with cohort studies that track the same targets, which, in our case, are infants. One problem that we face is that the number of infants under observation gradually diminishes over time. Concerning the long-term preservation of data, I would first like to mention that our multiple-year survey programs usually involve conducting a major survey every five years. The longest survey program has gathered data since around 1980. When we began these survey programs, we did not plan for advanced data archiving of the kind discussed by Dr. Granda. Therefore, some of our data sets from the early surveys are not preserved in a good condition. Therefore, as a future challenge, we must think about ways for long-term storage of long-term survey data. We must find how to keep the data for 50 years, for example. Your discussions reminded me of the services provided by data archives, and the institutional arrangements made to preserve deposited data.

(Watanabe) Mr. Arai, Director of the Benesse Educational Research and Development Center, pointed out the difficulties of keeping data for a long time to allow long-term survey data to be used. This is a new question raised by a surveyor, to be solved by archivists. Dr. Mochmann is an expert and highly experienced in data preservation. Dr. Mochmann, can you please tell us your view on the key to long-term preservation of data? We would appreciate your suggestions. This topic could be related to the importance of metadata that you discussed in the morning session. What do you think?

(Mochmann) I think there are several aspects to that question. Let me demonstrate it with respect for instance to panel data. Panel data is the case where you have one sample and you are continuously monitoring the same sample over time that is, you know, people also will die sometime and that will mean if the oldest person in the household, who normally will be the head of household, has died, then in the next measurement of this particular family somebody else will be head of the household. This means that you have to have an open eye on what is happening in the sample as such and you have to document it according to the changes in the natural panel. So, this is tracking the changes over time on both the measurement and the metadata level.

The other thing is also being compatible with the documentation over time and including information about events in time which may have an influence on the topics being measured in the particular survey. This is something which is frequently forgotten, but if you want to interpret the data at a later point in time, you want to be aware of the events which did have an impact on the measurement and the attitude which is the topic of the particular survey. I think this has major concerns apart from the physical preservation which is basically governed by the same laws and rules as is in preservation of cross-sectional data.

(Watanabe) Thank you very much. Does anyone from the audience have a question to ask? President Katsuno, please go ahead.

(Katsuno) You have been discussing about long-term survey data that typically come from cohort studies. Data collected by such studies could be large in the beginning but become smaller over time. Moreover, the number of samples is much smaller when compared with data from cross-sectional studies. Therefore, making such long-term survey data available at data archives is very valuable. However, since such data come from surveys that keep on tracking the same individuals, there always will be concerns about privacy. I believe that efforts to address such concerns have been made in the United States and Europe. In Japan, Prof. Ishida has been making the same effort. What measures have been taken to protect the privacy of respondents?

(Watanabe) About this question from President Katsuno, I would like to add that he holds data from a health-related long-term survey program, which has continually tracked children in a certain locality for about 25 years. Dr. Granda, Dr. Mochmann and Prof. Ishida, could you reply to this question?

(Granda) Only in the sense that I think the United States is fortunate because of the size of its population and the size of its area. Most of the longitudinal surveys that are done are based on national samples, so even as you correctly describe what happens in a longitudinal study, still because the sample is a national sample and efforts are made to minimize respondent identifying characteristics, whether they be direct or indirect characteristics, I think you are still able to minimize the risk of identifying any individual respondent. On the public use files, you just have variables that are not going to endanger respondent confidentiality.

(Watanabe) Dr. Mochmann, do you have something to say about security, confidentiality, or any related issue?

(Mochmann) I think Peter mentioned the problems of cohort studies. It is even more a problem when you are going to panel studies because here you are dealing with the same people and that implies that you have to keep track of the addresses in order to get in touch with the same people again. And that model is only working if you are working with informed consent of the people answering your questionnaire. There are no organizational models which are good to take you around to solving that problem. But there is one problem intrinsic to that kind of procedure which, at least according to German Data Protection Law, could create problems: if you are cooperating as a researcher with a field organization and those two parties form the group of people who are authorized to tackle this data under the Data Protection Law because they are the principle investigators. Now sometimes, and this was an unforeseen case, it happens that the field organization has to close down and if only the field organization was entitled as principle investigator vis-a-vis the interviewees, this would mean the end of the process, and therefore, it is important to team up with the principle investigator and to give the principle investigator the same rights as to the field research institute so that it can survive by the continuity of the principle investigator.

(Watanabe) Thank you very much. I would like to hear also from Prof. Ishida because he has been collecting data from panel surveys. Prof. Ishida, would you please comment on the difficulties you have experienced, particularly on the issue of privacy?

(Ishida) I think it is better to separate issues involving the disclosure of panel data from issues involving the implementation of panel surveys. With regard to disclosure, the procedure for disclosing panel data does not basically differ from that for disclosing cross-sectional data. In either case, we disclose data in a way that no one can identify any individual respondent. We must respect and protect the privacy of respondents. Of course, to continue the surveys, we need to keep the names and addresses of respondents in order to maintain contact with the same individuals. In the case of a cross-sectional survey, we need to identify the target population only when implementing the survey and then we can discard any information related to the privacy of respondents. This information, however, has to be retained for a long time in a panel survey, raising concerns about the

protection of privacy. This is a major difference between cross-sectional surveys and panel surveys.

Let me tell you what we have been doing in our panel surveys. When conducting the first survey in the program, we tell the candidate respondents that it is a continuous survey program and that they should participate only if they agree to be covered by a series of surveys. Only if they agree are they selected to be respondents. Some people may accept coverage by a single survey but not by repeated surveys, but such people are not chosen as respondents. This procedure has the disadvantage of lower response rate, but it is an advantage to target only those who have agreed to respond to repeated surveys over a long period.

Let me give you another example of what we do. The selection of respondents is consigned to a research company, which does the work by referring to the Basic Resident Register, for example. We ensure that the names and addresses of the respondents are never revealed to us (client researchers) by the research company. Only the research company keeps that information; it remains hidden from us. In the survey data, respondents are represented by ID numbers and nothing else.

At the end of the questionnaire sheet we prepare for our survey, there is a box in which the respondent can write any comment. Among various comments, there are sometimes complaints about the impoliteness of the surveyor. Some respondents report being angered by a certain question. In such cases, we write a letter to the respondent. We hand over the letter to the research company, giving them the ID number of the respondent and asking them to send the letter to the respondent because we received a complaint from the person. This is how we protect the personal information of individual respondents. The research company keeps the personal information without revealing it to the researchers.

(Watanabe) Thank you very much. President Katsuno of Gifu Pharmaceutical University, I hope you are satisfied with the explanation. I have just learned for the first time that researchers do not hold personal information on individual respondents. I was rather surprised.

(Koyasu) What I am going to say may overlap with what others have said. I would like to tell you what we did at the NHK Broadcasting Culture Research Institute when conducting surveys. This may be useful for our discussions about some challenges that are particular to longitudinal studies and about the handling of longitudinal data.

In the survey we started in Kawasaki City, we referred to the whole of the Basic Resident Register to identify children who were born in the period between January and June of a certain year. We sent a letter to the parents of all those children, asking them to participate in our 12-year survey program. Only the children of the parents who accepted the conditions were selected. This procedure itself may have produced a strong bias because it is rather unusual to be a child in a family that would accept such conditions. Nevertheless, a considerably large number of families agreed to take part in the survey.

In this way, we obtained from each family informed consent for 12 years. In addition, we confirm the willingness of each family before conducting each annual survey. We conduct a survey in January

each year. In the period between surveys, we try to maintain contact with each family by sending a birthday card and sending greetings cards at particular times of the year. Through such efforts, we try to maintain the population, which has fallen from 1300 to 900 by now. Seven years have already passed since this survey program was started, so I guess we have been rather successful in maintaining the population. There are many families who have moved from Kawasaki City, but most of them remain in the survey program. So, it is important for us to provide respondents with a detailed explanation of the survey program and obtain their informed consent.

As for the contents of the survey, we have an ethics committee outside the researcher community, which checks the contents of each annual survey for any controversial or problematic implications. With regard to the processing and analysis of data, it is important to establish a firewall between the data handling section and the data analysis section (the researcher community). We firmly maintain such a firewall. By using the data handed over to researchers, it is impossible to identify any individual respondent.

Finally, I would like to add that, at present, our policy is to disclose nothing more than statistics from these survey data. We do not have concerns about the unique comments made by individual respondents because they are outside the scope of analysis. If unique comments such as those made in a free-description box were to be disclosed, they could facilitate the identification of individual respondents. Therefore, our basic rule is not to disclose any data that could be used to identify individual respondents.

(Watanabe) I understand that there are many difficult challenges concerning the protection of confidentiality. Am I right to understand that data in archives do not contain the names of individual respondents? You said that the names of individual respondents are kept at the research company without being revealed to others.

(Koyasu) Yes, this serves as the firewall.

(Watanabe) I understand. But when a data archiving organization provides a data merge service as we discussed this morning, doesn't the lack of names of individual respondents cause problems?

(Koyasu) Each individual respondent is given a unique ID number. Using this as a match key, you can pick up data on the same individual from different data sets from different years.

(Watanabe) I see. Only the research company that picked up the respondents can match unique ID numbers with the names of respondents. Is there any question about this from the standpoint of a researcher who may use the data for a presentation at a meeting of an academic society?

(Ishida) Since we have been talking about panel surveys, I would like to speak more about the difference between tracking surveys and single surveys. I would also like to explain some of the

precautions we must take before disclosing survey data.

Since we should be tracking the same individuals, certain parameters, such as the gender and the year of birth of each individual respondent, should remain the same. As long as the identity of the respondents is maintained, they cannot change. However, in actual surveys, there are strange cases such as someone's gender changing from male to female between the surveys in the first year and second year. There are also similar cases of inconsistencies in age. It is easy to make an error when giving one's age, but not about gender.

What is happening? In our survey, a respondent who was identified to be a male in the first year's survey was identified as a female in the second year's survey. In fact, the respondent himself filled in the questionnaire in the first year's survey but his wife filled in the second year's survey. We reached this conclusion because, in the first year's survey, we asked about the respondent's marriage status and about the occupation of the spouse. The occupation of the respondent reported in the second year's survey matched the spouse's occupation reported in the first year's survey. This suggested that the wife of the respondent had filled in the questionnaire with information about herself. You may be interested to know what we did with the data provided. We switched the entries between "occupation of the respondent" and "occupation of the spouse." We had to discard all information concerning the opinions of the respondent but kept information about the household.

There is another type of inconsistency that we may encounter. In the questionnaire we prepare for a survey, we may repeat some questions that we asked in the past. The same respondent should respond similarly to the same question, but a questionnaire returned ostensibly from the same respondent may contain a different response. How should we handle such inconsistencies in data? This is a very difficult issue.

Before disclosing survey data, we must ensure, for example, that the gender of each respondent is consistent among different data sets. To remove any exceptions to this rule, we must clean up the survey data. We have to examine the survey data even in the case of a single survey, but we only do it once. In the case of a survey program that keeps track of the same respondents again and again, there is a higher chance of cases in which the identify of the respondents becomes doubtful. We must clean up the survey data to remove such cases. This is a unique difficulty particular to the analysis and disclosure of data from panel surveys, and our experience has shown us the greatness of this difficulty. I am sure that organizers of panel surveys overseas have also faced the same problem.

(Watanabe) When we look at a data set in a data archive, it appears as an assortment of numeric strings. They may look simple, but in fact it requires great effort to prepare these numeric strings. As one of the measures for stimulating the activities of academic societies proposed by Prof. Koyasu in the second half of the program today, the depositing of survey data in a data archive should be a condition for accepting the publication of academic papers in journals. However, many academic societies find this difficult. What is the cause of this difficulty? Has this been practiced in other countries like the United States and Germany? If the director of an academic society in Japan tries to follow this procedure, what difficulties is he or she likely to face?

There is something that I would like to ask Mr. Miyake from MEXT even though he may find it difficult to answer because he is not directly in charge of the issue. Many researchers conduct their studies with the help of a scientific research grant. Would there be a problem in requiring researchers to agree to deposit their data in a data archive as a condition for accepting a grant?

(Miyake) In my previous post, I was engaged in allocating not scientific research grants but the budget for supporting the research and development of science and technology. We were aware of this question of how to ensure that the results of government-aided activities are given back to the government or to people. The submission of a report or the like is a condition for receiving budget. However, we still have not established any particular rule on how researchers should repay by making available the data they collect in their research. Therefore, I would like to know whether, in any other country, the authority in charge of supervising data archives has introduced a rule on how researchers aided by a government grant should make their data available to others. I hope someone here can answer this question.

(Watanabe) This is an interesting question. Dr. Granda, is there any rule in the United States that requires researchers who receive a government grant to deposit their data in a data archive?

(Granda) The data that is funded by the public and is collected by our government, that is usually made available at no charge like our census data. Since the data collection process is usually very costly, most academic researchers get money from our funding agencies like our National Science Foundation and our National Institute of Health. Part of what they must promise when they receive funds from such organizations is that they will make the data publicly available too.

We must remember that whether it is the government making data publicly available, academic researchers making their own data publicly available, or archives making their data publicly available, it is only data that has maintained the confidentiality of the respondents that would be made publicly available. So, the archives and the data producers must make different decisions to create a public data file depending upon how the data is collected in the first place. If you have a national sample for example, you may not have to do as much processing of the data set as you would for a data set that is collected in Kawasaki City, where everyone knows that the participants in that survey come from one city instead of from the whole of Japan.

So, each data set presents different challenges to make a public use data set. And I think that is the responsibility of those who collect the data and those who disseminate the data as well.

(Watanabe) Thank you very much. Dr. Mochmann, what is the situation in Germany? Is there any rule that requires researchers who receive a government grant to deposit their data in a data archive?

(Mochmann) I think here we are entering the field of creating a culture of data sharing and this has been a long-term process in Germany. For the first decades after the foundation of the central

archive in Cologne which was in 1960, it was us going around and trying to convince people about the advantages of sharing their data and getting data as a gift data and what you can do with data and giving new life to old data by setting up a comparative research at a particular point in time.

We then started some systematic efforts to talk to the funding authorities that they should implement the requirement with granting research money that the data shall be archived. The German National Science Foundation did comply to that request when a case became prominent in biogenetic research, their research assistant found out that the Prof. had faked the data. From that time onwards the National Science Foundation thought this is a nightmare if the research cannot be replicated. Therefore, they made it binding that all data being used in research by their grants shall be archived. They did not say it must be archived in this or that place, but they said it must be properly archived at least for 10 years to allow for replication and they made a recommendation that it should be the central archive as a preferred place, but likewise it could be with the institute itself that made proper procurements.

Now, this is for research data and we have recommendation by the German Association of Social Research Institutes to donate the data to the central archive. Peter rightly pointed to the fact what is happening with the data which is collected under the authorship of public authorities and this was an open situation and a difficult situation. We made a move to ask the Statistical Bureau to make data publicly available and we run big projects in terms of how to assess anonymity and this led to the concept of factual anonymity which means data sometimes in theoretical identifiable, but in practice it is an incredible effort and you would need about the same type of resources to hire somebody to investigate who is the individual behind it. So I think that this is a pretty typical German solution of acknowledging the concept of factual anonymity.

On a global level, I was lucky to participate in a special panel set up by the OECD which came up with a ministerial agreement of the research ministers of the OECD countries in 2004 regulating access to publicly financed data and the research ministers, following our recommendations, clearly agreed that publicly financed data should be made available. However, there are problems in implementation because some of the countries noticed what this means for the national budget if you make publicly financed data available. Fortunately, there are lots of publicly financed data, but they agreed that it is only meaningful to make this data available if the metadata, that is the documentation, is appropriate and does allow for later interpretation, and that was not self understood in the past. Frequently you had data, but no appropriate metadata. Creating that needs additional resources and that item has to be included in the budgets of the future.

(Watanabe) Thank you very much. Dr. Mochmann told us about systematic efforts in Germany. In his presentation, Dr. Mochmann emphasized the importance of creating a culture of data sharing. This is a difficult challenge, but Dr. Mochmann and his colleagues have been making various attempts to address it. The providers of grants can set a rule that grant recipients must deposit their data in a data archive. But we must ask ourselves if our data archives are sufficiently well established to be trusted by the providers of grants. Do our data archives have the systems and organizational

arrangements that can be trusted? We will face this question more frequently in the future.

Now I would like to ask a question to Mr. Arai, Director of the Benesse Educational Research and Development Center. Mr. Arai, your company, Benesse Corporation, spends a lot of money on conducting surveys or commissioning surveys to outside organizations. Would there be any problem or concern for your company to deposit the data from these surveys in a data archive?

(Arai) In principle, perhaps with the exception of some data, data confidentiality should not be a problem when we deposit our data in a data archive because it is our basic policy to make all our survey data available to the public. Data from all of our recent surveys are well prepared for depositing in a data archive because that was the intention from the outset. It might be more difficult to deposit data from earlier surveys because they haven't been prepared for archiving. However, I don't expect there would be any serious problem.

(Koyasu) I talked of incentives rather than rules. We should start by simply suggesting that people deposit their data in a data archive. We should not try to impose a rule from the very beginning.

(Mochmann) Two more comments, I fully agree on the observation of the developmental psychology to base it on motivation rather than on forcing people. We found that one big motivation for in particular prominent researchers was to give visibility to the kind of research they have done and to gain visibility by publicizing data. But I also would like to put this discussion into the context of the Open Access Movement, which nowadays is a worldwide movement, which emanated from the intention to make all published articles available not only via commercial publications but also on the open market and you know about the golden and the green way to open access.

There have been several declarations on that count and one very prominent one is the so-called Berlin Declaration which implies that not only published articles should be made available openly, but all research results- and research results explicitly include research data! This is stipulating that ways are being found to put the data on servers, and what is being done at the moment in Germany is that several prominent institutes are setting up servers which can be used to publish not only the articles, but also the data of the staff members.

(Watanabe) Thank you very much for your comments. Mr. Miyake, your question concerning the scientific research grant provoked these responses. Did they give you some useful information?

(Miyake) Yes, I can clearly see that many people are involved in the discussion about various ways in which the results of government-aided researches should be made available to the public. In the case of researches conducted with the help of a scientific research grant and other types of grants, the research data are disclosed after the final results are obtained. In the case of researches conducted with the help of a scientific research grant, research data are saved to a database that is accessible to the public. As for the question of how much we should control such data, we are still discussing it

from various standpoints. While listening to the responses to my question, I realized that we need to think more about this question in the future.

(Watanabe) Thank you very much.



Discussion II: Data Archives and Public Policy on Education

Using Data to Formulate Education Policies -With reference to nationwide surveys of academic ability and learning context

Takanori Miyake

Senior Specialist, Office for Assessment of Academic Ability,
Elementary and Secondary Education Bureau, MEXT-Japan

(Conference minutes)

Good morning/afternoon, ladies and gentlemen.

My name is Miyake, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT). At MEXT, I am in charge of the national surveys on academic achievement and learning.

For these national surveys on academic achievement and learning, we have been receiving tremendous support from HUTE, and we are particularly grateful to President Kajita who has been assisting us since the very beginning of the project. Therefore, I feel a little embarrassed to be speaking about these surveys in front of him, but I must do so having been kindly asked by the organizers.

I am going to speak on the theme of data utilization in the process of educational policy development, with a particular focus on the national surveys on academic achievement and learning.

These academic achievement surveys are not only for helping the national government to develop policies. They have other purposes such as to help schools improve education based on national-scale educational data and to promote improvement by identifying challenges with academic achievement data and the learning of each student covered by the surveys. Since this symposium is about data archives, I would like to focus today particularly on the contribution of these surveys to the making of educational policies.

The national surveys on academic achievement and learning have been conducted annually since three years ago. Behind the decision to begin such surveys, there was a need to better understand the facts and issues concerning school education. In addition, worldwide academic achievement surveys, such as the OECD Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), revealed that the academic achievement and willingness to learn of Japanese students have been declining. So, it was necessary to establish a system for guaranteeing the quality of compulsory education.

With this background, the government started (or restarted) these nationwide academic achievement surveys in fiscal 2007. The surveys have been conducted annually for three years up to fiscal 2009. In these three fiscal years up to 2009, the scope of these surveys has covered all students in the sixth year of elementary school and the third year of junior high school. Thus, each annual survey covered more than two million students. The surveys looked at the subjects of national language and primary arithmetic or mathematics, and included questionnaire surveys on students. Since we had been alarmed by the drop in willingness to learn, we conducted such questionnaire surveys in addition to straightforward examinations.

The questions we prepared included those focusing on knowledge, that is, questions that required calculations or the spelling of Kanji characters, as well as questions to assess the capability of application, or the ability to apply

knowledge and skills in real life situations.

Given our interest in data archives, an important question is how we can then use the data from such academic achievement surveys. The results from surveys must always be fed back in one way or another. Survey data, therefore, naturally needs to be made available to the public. However, during sessions at the national assembly and at the council of experts, it has been argued that special care must be taken to prevent surveys from contributing to stratification and excessive competition as a result of ranking the scores, for example. Therefore, we decided to control how surveys are conducted by writing a procedural document, which contains precautions on the handling of survey data.

The lower part of the slide shows specific rules concerning the disclosure of survey data. If survey data are going to be disclosed, it is important to clearly establish who is going to disclose the data and when. It is also important to take appropriate measures to prevent the survey data from contributing to excessive competition and stratification, as stated in the upper part of the slide. The procedural document establishes rules for such things, and all participating organizations must agree to comply with these rules. It is important to make these rules clear when we ask for participation in the surveys. Among various surveys conducted by the national government, our academic achievement surveys are conducted under such an arrangement.

I agree with Dr. Koyasu about the importance of preparing a system for protecting personal information. For the academic achievement surveys, we made careful arrangements to ensure the protection of personal information. One example is our rule that MEXT, or any subcontractor organization, must never use a format that includes the names of individual students. The rule was formulated a little differently in fiscal 2007, but this is how the rule was set for the academic achievement survey in fiscal 2009. As in the other surveys discussed earlier, we use ID numbers instead of names. ID numbers are assigned to student names at schools, and during the subsequent data processing, MEXT and subcontractors never handle student names; they only handle ID numbers and test data. We have thus created a firewall like in the cases of other surveys discussed earlier. Through a variety of careful arrangements including this, we ensure the protection of personal information.

I would like to say something about the government's position, but this is not a subject I am directly in charge of. So, what I am going to speak about now is based on my personal understanding. Well, major government surveys are conducted under the Statistics Law. The Statistics Law was amended recently and came into effect in April 2009. This amendment greatly changed the institutional framework for using statistical data from government surveys. In the past, statistical data from government surveys were viewed as data to be used by administrators for policy-making. Now, such statistics are expected to provide fundamental information for society. This is a great change in perspective. Before the amendment, statistical data from government surveys existed to be used by the government, even though the published parts of such statistical data could also be used by the public. This principle was changed by the amendment. For example, the amended Statistics Law calls for the establishment of a system that allows statistical data to be flexibly used to help improve academic research and higher education. So, there is a general trend in Japan toward greater use of statistical data from government surveys.

The academic achievement surveys are not controlled under the Statistics Law. Nevertheless, the purpose of these surveys is not limited to helping the government verify the effectiveness of educational programs and plan for improvements. These surveys are also meant to help the boards of education and schools improve their educational programs, and to help teachers improve their teaching activities. Therefore, we need to ensure the

data from academic achievement surveys can be used effectively. Apart from considerations about the Statistics Law, surveyors need to support the use of survey data for broad purposes. This is true in all surveys including, of course, our academic achievement surveys.

The challenge is how to solve the conflict that we have been discussing between the need for careful handling of survey data and the need for allowing the survey data to be used. This trade-off dilemma, as mentioned earlier by Mr. Granda, is of course faced also by the organizers of academic achievement surveys.

Concerning this challenge, I would like to exchange good ideas with you about the collection and utilization of survey data. I also would like to discuss the question of how to gain acceptance on collecting data on something as personal as the academic achievements of individual students. How and with what reasons can we gain such acceptance? Even though we have many ideas about systems for protecting data, we must also ensure that people, who are not experts, understand how data are protected. The presence of a data protection system will not guarantee the immediate donation of data or support for surveys. What must we do to gain such support? What is the best way to build support? Another question is how to foster understanding on the value of the researches that can be conducted using survey data among depositors of data and also among the people covered by surveys. These are the topics that I propose for discussion.

Now I would like to speak from another perspective. Earlier I said that Japan “restarted” the national surveys on academic achievement. In fact, such surveys were restarted in fiscal 2009 after a long blank since 1966. In the intervening time, the government conducted several nationwide surveys on curriculums, but they somewhat differed from academic achievement surveys in purpose. Therefore, as I have said, the national surveys on academic achievement were restarted after a blank of almost half a century. In that long interval, there is almost no data, which surely made it difficult to adopt an evidence-based approach to planning educational policies. Given this administrative situation, it was difficult to build understanding on the importance of evidence-based policy-making, on the need to speed up the development of competency, and on the need to collect data on academic achievement.

Since it should be possible to gain more support from academic researchers on educational policies, for example, I would like to show an example of how research findings may contribute to the making of educational policies.

Let me explain how the Basic Plan for the Promotion of Education is drawn up, which is an overall five-year plan concerning the national educational programs of Japan. In this process, actual discussions began in February 2007, and continued for nine or ten months up to the finalization of the plan in April 2008. The process included hearings from related organizations and the collection of public opinions. We collected opinions from academics, among others. So, I am interested in discussing how researchers can quickly summarize findings from their data analysis, how to ensure that such findings are used effectively for making government policies, and what arrangements we can make to support this process.

The effective use of data archives not only encourages research based on the analysis of deposited data, but also efforts to ensure that research results are then used by the government to make educational policies, for example. As mentioned by President Kajita at the beginning, people are likely to fall into the trap of “impressionism” or “emotionalism” particularly when they discuss education. In making educational policies and all other policies, how can we encourage the use of an evidence-based approach? For example, as mentioned in the lower part of the slide, how can we ensure that findings from educational researches are used when drawing up educational

policies? While different entities may use such research findings for different purposes, how can the government and administrators use the findings effectively? Should there be a mechanism for administrative organs to draw conclusions from data? Or, is it better to have administrators use inputs from academics as they develop policies? What mechanisms do other countries have for ensuring that findings from policy researches are fed back to the policy-making process? Should such mechanisms be set up in Japan? These are the topics that I propose for discussion and the exchange of information. Thank you for listening.



Chapter 2 Presentation and Discussion

Discussion II: Data Archives and Public Policy on Education (Conference minutes)

(Watanabe) Thank you for your presentation, Mr. Miyake, on the national surveys on academic achievement and their problems. Do any of the panelists wish to give feedback or make a comment?

(Granda) One of the points made was convincing the respondents about the value of doing research. I know this is a difficult question at times, but I think sometimes future respondents respond very favorably when they know that the data has produced research and that this research is really valuable to the society as a whole and is going to possibly result in changes that would be beneficial to many, many people. If they believe that their responses as an individual to the survey add to that objective and that can be demonstrated to them, I think then they are more likely to participate.

(Mochmann) Two observations. I think in your presentation you also briefly touched upon qualitative materials and that is an important aspect, so far archiving predominantly has been focusing on quantitative materials. Firstly, because they were more frequently used, and secondly, it is somewhat easier to archive well-structured data as contrast to qualitative materials which tend to come as text in different forms with different interpretations. But we observed a strengthening of the qualitative paradigm and the convergence with the quantitative paradigm. So, I agree we must have an eye on archiving qualitative materials as well.

Then you posed a question which is very difficult to answer. Actually, the question was how can the archives be used to inform policies? We currently are conducting a big research project which is called “Social Science and Humanities Futures,” which is funded by the European Commission and the concern here is what can social sciences and humanities contribute to inform policies. There is one point which I want to make in the first. It is always good to accumulate data and to accumulate wisdom and knowledge and to provide it to the society and to the policies, but we must be aware that the archives cannot have better data than are given from the research process and the principle investigators. So, I think the principle investigators must be aware of the needs of the politicians and the policy makers.

And second thing is my observation that frequently the politicians are interested in clear-cut answers, which are not very complicated and not depending on ‘If Then’ statements, whereas researchers tend to be quite differentiated in their argumentation and they do not like to say this will happen ‘B’ if ‘A’ is given. Their statements strictly are more complicated.

Now, the expertise in the ministerial bureaucracies is not so well staffed anymore that they would know how to deal with complicated statements and this is what we call in our project the translation problem of translating empirical knowledge based on clear evidence into knowledge which could be easily consumed by bureaucracies and by politicians. And I cannot resist the temptation to quote the famous movie “Lost in Translation.” I hope that the knowledge which will be available in the

archives will not be lost in translation into policies. Therefore, I welcome very much the openness of evidence-based policies.

(Ishida) It was a very interesting presentation, and I think it is rather unusual for an officer from MEXT to make such honest remarks. There were several remarks that I found particularly interesting.

For example, he mentioned they are finding it difficult to collect information due to growing concerns about the confidentiality of personal information. If even officers at MEXT are facing such difficulty, it is not surprising that we too have suffered the same difficulty in our surveys. Unlike MEXT or the Ministry of Health, Labour and Welfare, we do not have any authority behind our surveys. We can collect information thanks only to the sincerity of visiting surveyors, who almost beg for cooperation, and the goodwill of respondents. So we probably face greater difficulties in trying to collect information in our surveys.

What can we do to deal with this difficulty? As mentioned earlier by Dr. Granda, it is important to fully inform people about the usefulness and significance of our surveys. In the case of a panel survey program, the same respondents are contacted for repeated surveys. We should tell them how the data from earlier surveys have been used. After each survey, as soon as we have cleaned up the data, we prepare a quick report on the survey and distribute it to the respondents. We prepare a four-page leaflet to provide them with the summary of findings from the survey. We sent a letter with the leaflet asking respondents to inform us of any change in their mailing address. In this way, we ensure that the findings from each survey are reported back to the respondents.

We also prepare press releases. If our surveys are covered by the media, such as newspapers, we report it to the respondents so that they know the survey data are well utilized. We try to provide people with as much information as possible to tell them about the usefulness of surveys not only for academic studies but also in a broader context. This is the only thing we can do at the moment, but it is important to make such efforts.

It is difficult to foretell how survey data are going to be used in policy making even though we may learn about it later. So, I don't think that anyone expects their data will be used in policy making when making them available at a data archive. In many cases, we must first make our data available to the public, and only later on, we may find out how the data may have helped policy making. Our data archive contains data from some governmental surveys, which have come to us in various ways. For example, a certain governmental department gave us their data because they wanted to use various data in our data archive when drawing up a white paper. In another case, a researcher at our institute was involved in a survey conducted by a governmental department, and this led to them depositing the survey data in our data archive.

How will such data deposited with us be used? At present, it is still not clear and much depends on chance. It is natural that depositors expect their data to be used in policy making and we also share this expectation. This is an ideal. However, it is difficult to foresee the impacts of survey data in a variety of scenarios. I am aware that I am making a vague statement. Maybe I should say that

depositors should not expect too much.

Still, there are advantages in making data available to the public. It may take a long time before these advantages become visible, and we even may not be able to find any visible example of these advantages. Still, I ask data owners to trust the long-term advantages as they try to make their data available to the public.

(Koyasu) The topic under discussion is related to more than just how data archives contribute to policy making. That is, we are discussing how research activities and research data may contribute to policy making. I mentioned earlier that longitudinal studies over a long term have been conducted as major projects overseas, particularly in developed countries, while such studies are still at a primitive stage in Japan. Then I also mentioned that a major motive behind such longitudinal studies over a long term is a commitment to clarifying how poverty may affect academic achievement or the growth of children based on the clear assumption that there is a relationship between the two. Therefore, overseas researchers try to produce various data concerning this relationship to encourage administrators to develop policies based on a recognition of this relationship, even though their survey data may not have a direct impact on policies.

If taxpayers believe that survey data will contribute to policy making, there will be a better understanding about the value of various study projects. When I talked with Prof. Ishida during the break, we agreed that it had become very difficult, particularly at schools, to collect information concerning the social economic status (SES) of parents. In fact, it is now impossible to collect information about the family's income, the parents' academic background or whether the family owns a house or not, for example. In international or joint research projects, overseas researchers ask us why no SES data is available from Japan, and it is difficult to give them a satisfactory answer. SES data are not available from Japan because people do not believe that such data contribute to their welfare or to the improvement of child education. Why do people lack this trust? The government, researchers and many other factors are surely responsible. In any case, we must do something about it as a challenge unique to Japan.

I would like Dr. Granda to correct me if I am wrong, but I heard the United States started something called the Head Start Program in 1965, which is more than 40 years ago. What does "head start" mean? It refers to the ideal of having all children aligned on the start line as they begin primary education at primary schools. In horse racing, the word "head start" means that the heads of all the horses are aligned on the start line at the beginning of a race. In reality, however, children do not start from the same start line when starting primary education at primary schools. Children receive different training during preschool education particularly in respect of their readiness for education or their preparedness for learning. Some children are well prepared to study at school while others are not. This develops into a major difference in their capacity to receive the benefits of education at primary schools. In recognition of this reality, the Head Start Program offers various help to preschool children so that they can make a head start upon entering primary school. It is well known that the educational TV program "Sesame Street" came out of the Head Start Program.

There have been analyses on the effectiveness of the Head Start Program, and I know some reports say the program is not very cost-effective. However, whether the Republicans or Democrats have been in power, the United States government has continuously increased the budget for the Head Start Program. So, it is important that we promote such activities. We cannot be sure whether such activities will greatly help eliminate poverty or reduce poverty's impacts on the academic achievements of children in the short term. Nevertheless, the US government has demonstrated its commitment to solving this problem by continuing to increase the budget for the Head Start Program. An example like this may help people believe that surveys and their findings have some impact on policies. Then people will be more willing to give information about themselves to surveyors in the hope of contributing to making better policies.

While it is difficult to decide the first action we should take, a group led by Prof. Hiroaki Mimizuka of Ochanomizu University has set a good example by studying the results of the national surveys on academic achievement. The results of their analysis of the relationship between parents' socio-economic status and the academic achievement of their children were reported by newspapers. Their findings may affect various policy making to address the problems identified, such as free education at high schools. It is very important that the public is convinced that surveys are not simply for collecting data but that administrators have been using or will begin to use survey data when making policies.

(Watanabe) Dr. Granda, do you wish to say something about Prof. Koyasu's view on how survey data have contributed to policy making in the United States?

(Granda) Just to agree with Prof. Koyasu because it has been a program that has been going for a long period of time and it has created some benefits, but one of the other points I was going to make though is there are other kinds of education data that I think have a very direct policy making effect even though they may not be particularly related to government policies. Our Ministry of Education collects data that just reports on the aggregate nature of the schools. So they will report, for example, on the graduation rates or the finances of the schools, giving very aggregate information. And this information is then made available so that parents who are very interested in the quality of the schools that their children attend will look at these aggregate data and for them it is a policy decision; they will decide, "I can send my child to this school because it is rated very highly on some of these measures." So, education data is used everyday by parents and so it is very important to them and they make policy decisions based on this information.

(Watanabe) Thank you very much. I think that the business community, particularly the education business community, has a strong interest in using data from the national surveys on academic achievement. For today's symposium which commemorates the opening of our data archive, we were keen to have the participation of Mr. Arai, Director of the Benesse Educational Research and Development Center. This is because, in a symposium like this organized by a university, we rarely

have a chance to listen to opinions from industry. I think that the education business community's role in public education has become much greater recently. So, I would like Mr. Arai, Director of the Benesse Educational Research and Development Center, to say something about the use of the national surveys on academic achievement data in the creation of educational policies or business policies, or about the use of such data for business purposes.

(Arai) First, I would like to say that the last series of national surveys on academic achievement was very important. This was a complete survey program implemented after a long interval that included the administration of achievement tests to all students, even though we are not sure if such a complete survey will ever be repeated again. The surveys involved not only the administration of achievement tests but also the collection of various profile data. I have high expectations that the analysis of test data in combination with the profile data may produce valuable findings. I don't know the original purpose of collecting such data or the policy behind the survey design, but I am glad to know that considerable quantities of basic data have been collected.

At present, people only pay attention to things such as the name of the prefecture that came top or second. However, the data can be analyzed in terms of various attributes. It is also possible to analyze changes over the three years covered by the surveys, and I am very interested in such analyses. On the other hand, as mentioned earlier, there are issues concerning the confidentiality of personal information. While we also understand the importance of collecting and analyzing data that pertain to the profiles of individuals, our ability to collect such data in our own surveys is very limited.

I think that educational policies have two major purposes: increasing the value of education for society and decreasing disparities in education. In view of these two purposes, which were also mentioned earlier by Prof. Koyasu, it is necessary to collect various information concerning the profiles of individuals. Even though the collection of personal information is highly controversial, it is essential to link it with reliable personal information to yield results that will be useful for making various policies. To what extent is this possible? This is a difficult question.

The education business community is very interested in findings from further analyses of the test data. I think it is important that researchers go beyond mundane interest in the ranking of scores and expand their attention to qualitative aspects. We are concerned about what may happen if future surveys no longer cover all students. The government administered achievement tests to all students for three years. If the survey program could be continued a little longer, it could provide a longer-term perspective. About the future of these surveys, we don't yet know what the government is going to decide under the new administration. In any case, we are interested in the findings from analyzing the survey data.

(Watanabe) Thank you very much. Now, I would like to invite President Kajita of our university, to speak. I am sure that many of you have been waiting to hear from him.

(Kajita) Hearing your discussions, I would like to make sure you understand the purpose and intent

of these national surveys on academic achievement. I would like to tell you that the government has been sending us three different messages.

Firstly, by deciding to implement such surveys and by choosing a certain format for them, the government has sent us a certain message. Throughout the 1990s, the Ministry of Education (presently MEXT) instructed the boards of education to refrain from such actions as conducting academic achievement surveys. As briefly mentioned earlier by Mr. Miyake, it was revealed in 2000, as a result of PISA tests and the earlier UNESCO-sponsored IEA tests, for example, that the academic ability of Japanese students had dropped considerably in the 1990s. In 2000, under the leadership of Prime Minister Obuchi and Prime Minister Mori, the government formed the National Commission on Educational Reform to undertake major educational reforms. Over the next one year, the members of the National Commission held 15 sessions at the prime minister's official residence.

In 2001, based on the results of these discussions, Minister Machimura of MEXT, which was formed by the integration of the Ministry of Education and the Science and Technology Agency, asked the ministry to draw up the 21st Century Educational Reform Plan (commonly known as the Rainbow Plan). This demonstrated MEXT's total commitment to guiding school education from 2001 in a direction that would strengthen the academic ability of students. As a part of such policy, MEXT decided to organize these surveys that involved administering achievement tests to all students in the sixth year of elementary school and in the third year of junior high school. Thus, they gave the message that the academic ability of students should be taken seriously.

These tests included not only general questions that addressed the knowledge, understanding and skills of students but also descriptive questions, which were like questions in PISA tests that are meant to assess the comprehension of students. Such questions, which might have been criticized as being too difficult or unique if they had been used a few years ago, were added because they wanted to assess the students' ability to solve problems by active thinking. The solving of such questions requires flexible, rather than straight-forward, application of learning. What does this mean for teachers? It is a message that, at elementary schools and junior high schools, teachers should help students to develop the ability to solve such questions. So, they are telling us that tests and academic ability should be taken seriously, and that it is not enough for students to develop knowledge, understanding and skills that can be assessed simply by objective tests. In addition, they should develop their ability to apply their learning not only in a straight-forward manner but also in more sophisticated ways; such abilities may be very difficult to measure but can be assessed to a certain extent by including descriptive questions in achievement tests. This is the first message.

As the second message, they expect the prefectural governments, municipal governments and schools to analyze the survey data and take corresponding actions as soon as possible. As an example of such actions, teachers in Okinawa Prefecture are now sent to Akita Prefecture to learn from teachers there because the scores were very high in Akita Prefecture and very low in Okinawa Prefecture. In addition, teachers in Akita Prefecture are invited to Okinawa Prefecture to give seminars. Such actions may appear simplistic but can bring immediate benefits. Osaka Prefecture used to be top in an academic achievement survey conducted 40 years ago, but has now fallen nearly to the bottom.

So, the Governor of Osaka Prefecture is making various attempts to improve the situation. Although many of his attempts appear off-target to me, he is making a serious effort. In this way, survey data have immediate impacts.

Let me give you another example of immediate impacts from survey data, which, in this case, came from analytical findings. I am referring to findings concerning the relationship between the economic status of parents and the academic achievement of their children. A report published by MEXT admits right at the start that there is a clear relationship between the economic status of families and the academic achievement of children. However, comparing data from different schools concerning this relationship revealed a hidden aspect of this relationship. In this analysis, the economic status of the parents of children at school was assessed on a five-point scale based on the percentage of parents who are receiving schooling aid or the like. Then, it was revealed that, with the children of those parents who were ranked at the lowest end of the scale, there was a very great dispersion in achievement test scores among different schools.

Children of affluent families tend to score high at all schools. However, children in poor families score high at some schools but very low at other schools. There is thus a very great dispersion among different schools. This was the finding, and it suggests that poverty does not always prevent children from developing academic abilities. In other words, schools can do much to prevent poverty from affecting the academic abilities of children. Because of this implication, I believe the finding has encouraged school teachers in many local communities.

As the third message, they are telling us that they have consigned groups of experts to discuss how data can be analyzed and used, and that they are studying various data with the help of such experts. In addition to a detailed analysis of the relationship between the economic status of parents and the academic achievement of their children, I am very interested in analyses of the situation at effective schools. Such analyses are being conducted by several groups. Children at such schools come from families with a very low economic status, which implies that conditions at these families are unfavorable in other respects as well. These families could be suffering in various ways (single-parent families, for example). In spite of such adversities, the academic ability of children has risen in some of such schools, which are rightly called “effective schools.” There are ongoing case studies that target such schools. The government supports these studies because their findings will be very useful in many ways when planning activities at schools.

So, there are at least three messages. The first message concerns the government’s policy behind the administration of academic achievement surveys, the second concerns the use of immediate findings from these surveys, and the third concerns the benefits of reanalyzing survey data, which will take some time.

Because we have been discussing policies, I would like to conclude my speech by saying something about government activities under the new administration. I have requested MEXT to inform the new administration of the results of the national surveys on academic achievement in the last three years and to ask them to make good use of the survey data. In addition, I have expressed several opinions to the Minister and Senior Vice Minister of MEXT. For example, I told him that it is easy to talk

about solving educational disparity, but will the proposed free education at high schools really help do this? It might not be such a simple problem. Unfortunately, I only had a short talk at that time, but still I asked them to look at the problem in more detail and think more about policy measures. They replied that they would examine plans for various actions following an evidence-based approach. Therefore, even though I still cannot say that detailed survey data are already used extensively in policy making, these national surveys on academic achievement have served as a good opportunity to turn the eyes of politicians in that direction.

(Watanabe) Thank you very much. I found your speech very interesting, which may have touched on some topics that have not yet been made public. At the same time, you responded to the earlier question from Mr. Arai, Director of the Benesse Educational Research and Development Center. Now that the panelists and our President have spoken, I would like to open the floor to questions.

(Katsuno) The national surveys on academic achievement attract very strong attention from the public. If MEXT plans to deposit data from such tests in a data archive, I think there is a problem to be addressed. As I understand from the speech by President Kajita, the data from these achievement tests have not been made available to the public or deposited in a data archive. The data, without being made available to the public, are first analyzed by experts. Before reporting the analysis results, they will have to examine the results from various viewpoints as they deduce important findings while carefully avoiding any expression that could cause misunderstanding. Once these data are deposited in a data archive and made available to the public, someone may analyze the data maliciously, deliberately focusing on some chosen parts of the data. There is a risk that the analysis results might then be reported with a great impact on the public. This is a major concern that may make it difficult to keep such data in a data archive. MEXT and other authorities may not want to deposit their data in a data archive for this reason.

What can we do about this? In my opinion, the only thing we can do is to counter biased information by correct information and try to increase the impact from the latter. I would like to know how people deal with this problem in the United States or Europe. I imagine that, with such data, we may have to allow a time lag before making it public. During that time, proper analyses can be completed and the findings can be published. Is there any example of good practices that deal with this problem in the United States or Europe? If there is, I would like to know about it.

(Mochmann) I think we have a very good and prominent example. We used to rely on the Eurobarometers in informing policies and that is still the case. The Eurobarometers have been set up in 1972 to inform the European Parliament and this has been given as a donation to the scientific community for further analysis, but the primary purpose was to inform the European Parliament. Researchers, even though they were happy and still are happy to make use of the Eurobarometers, however, I am not very satisfied with the academic quality of that information source, so in response to that observation they set up the European Social Survey as a strictly controlled academic

enterprise and here quality control is very high, cognitive testing, methodological rigor, scrutiny in samplings of data and basing everything on pre-tested theories and examples from theory and knowledge which are related to what shall be collected as data in the next round.

Those are the guiding principles for the European Social Survey and this has been very well received in the academic community and as I mentioned before, as a result, the European Social Survey is gaining recognition as the reference study in Europe. It actually was granted for the first time in the social science with the famous the Descartes Prize of the European Commission.

(Granda) I would add that in the United States in particular, there is a fairly close connection between the surveys that many of our government agencies conduct and academic institutions. So very often, academic institutions that are expert in survey work will actually do the surveys for the federal government. These institutions and their researchers provide expertise and also play a big role in how the data is processed and made available to users at the end as well.

(Watanabe) Thank you very much. Time is now up, and so I would like to close the second session.

Discussion III: Data Archives and the Educational Industry

Data Archives and Educational Industry

Kenichi Arai

Corporate Senior Vice President, Benesse Corporation, Japan
Director, Benesse Educational Research and Development Center

Activities at the Benesse Educational Research and Development Center

The Benesse Educational Research and Development Center conducts surveys and studies covering all generations from infants to adults, and from the perspectives of parents, teachers and schools. The Center also predicts the mid- to long-term evolution of the educational environment, publishes information on the results of its unique surveys and studies, and conducts sophisticated research and development on curriculums, educational methods and assessment methods.

Surveys on education are an important part of the Center's activities. For 30 years since 1980 when we started out as an educational research laboratory within Benesse Corporation, we have conducted more than 400 surveys. The results of these surveys are made widely available through reports and web sites, and are used in equally diverse ways, thus contributing to education in the home and at school, and assisting studies on education. We also handle surveys on behalf of the national and local governments.

Survey Objectives

The Benesse Educational Research and Development Center has two major objectives in conducting surveys and studies. The first is to make social contributions, and the second is to support the educational service business.

The first objective, social contributions, is an important mission of any company that does business in society because it is an essential corporate value. Today, companies are expected to do much more than simply make a profit, pay satisfactory dividends and comply with laws; companies are also expected to serve as respectable members of society by maintaining proper employment and labor conditions, keeping good relationships with consumers, taking necessary environmental actions, and contributing to local communities, for example. Benesse Corporation makes profit from services for education in the home and at school. Therefore, one of our important responsibilities is to provide information which is useful for education in the home and at school. This is a basic motivation behind our activities such as providing information on our web site and distributing free information magazines to schools.

Another objective is to support the educational service business. The results of our surveys are often analyzed by educational administrators and reported in the mass media. This interest in our surveys and studies reassures people and builds trust in our company. Therefore, our studies indirectly support our business by helping to strengthen the brand of our products and services.

In some ways, our surveys and studies contribute directly to our business. Our surveys yield valuable information on the attitudes and behaviors of children, parents and teachers. This information helps us to plan our products and services, and to determine how we should make our products and services available. The results of our surveys are made available to a broad audience including our customers. Today, people are looking for education-related

information based on objective data rather than subjective impressions. The informative magazines published by our company contain useful articles on child-rearing together with survey data. Furthermore, the network of experts we have built up through our studies is a valuable asset for us; these experts help us think more deeply about our business now and in the future.

Details of Our Surveys

Motivated by the above objectives, we conduct surveys that cover children, students, parents and teachers. In recent years, we have conducted the following surveys and reported their results.

(1) Surveys on children and students

Basic survey on learning (conducted at five-year intervals since 1990; survey on children's learning attitudes and related beliefs; including comparative studies of six cities around the world based on surveys conducted in Tokyo, Seoul, Beijing, Washington D.C., London and Helsinki); basic factual survey on children's life (since 2004; survey on facts concerning children's life and on related attitudes; the second survey conducted in this fiscal year); factual survey on the educational use of information and communication technology (ICT) by children (survey on facts and opinions concerning the use of ICT; conducted in 2008); survey on the choice of junior high school (survey on facts concerning the preparation for junior high school entrance examinations and the choice of public junior high school; conducted in 2008); survey on the use of after-school hours (study on how children use the 24 hours of the day; conducted in 2008); factual survey on the learning and life of university students (survey on learning at university and their views about the future, conducted in 2008).

(2) Surveys on parents

Basic survey on child-rearing (conducted at five-year intervals since 1997; survey on parents' views on education); questionnaire survey on the daily life of infants (conducted at five-year intervals since 1995; survey on the daily life of infants; including comparative studies of five cities around the world based on surveys conducted in Tokyo, Seoul, Beijing, Shanghai and Taipei); survey on opinions about school education (jointly conducted with Asahi Shimbun in 2008); survey on extra-school educational activities (survey on facts concerning participation in sport, art and cultural learning activities and facts concerning educational expenses; conducted in 2009).

(3) Surveys on teachers

Basic survey on teaching activities (conducted at five-year intervals since 1997; survey on facts concerning teaching activities and views on education).

In addition to the above, we conduct surveys on behalf of the Ministry of Education, Culture, Sports, Science and Technology, the Ministry of Economy, Trade and Industry, the Ministry of Public Management, Home Affairs, Posts and Telecommunications, etc.

The Use of Survey Data and Associated Benefits

We make our survey data available through leaflets and reports, as well as on our web site that can be accessed by anyone. For several years now, the web site has offered a search function that allows the user to search for data by entering a keyword or by specifying the survey target, for example. Our informative magazines, like "VIEW21"

which we distribute to school teachers, have some pages dedicated to survey data.

As mentioned, one of our purposes in providing information is boost the branding of our products and services. Therefore, we must know how well our survey data are used by people outside our company. Survey data reported by Benesse can often be seen in newspapers, magazines and television, and the level of attention from the mass media serves as an important barometer. Our survey data have also often been used by the national and local governments, particularly in recent years. Our survey data are consulted by schools, which often refer to our data when setting entrance examination questions. We collect information on such external references to our data, and consider it as a barometer of our achievements. Besides counting the number of external references, we measure our achievement by estimating the cost of advertising to gain the same coverage of the survey data in the mass media. For each survey, we establish a target for benefits accruing from providing information.

Such quantitative evaluation of benefits addresses only one aspect of the value of surveys. Nevertheless, such quantification often helps us define the targets of each survey, allows us to keep track of our achievements each fiscal year, and makes it possible to ascertain the cost-effectiveness of our surveys. Moreover, such quantification is a part of our accountability to companies that pay for our surveys, to whom we must report quantified measures of benefits rather like performance indicators.

Disclosure of Survey Data and Social Contribution

Benesse Corporation earns profits from the educational service business, and so is obliged to provide beneficial information to each family and school. Accordingly, it is our policy to make all our survey data available to the public after completing the analysis. We disclose not only our analysis results but also questionnaire data, basic summarization tables and cross tables. Benefiting from the availability of such materials, some schools and local governments conduct surveys similar to ours, and compare data from their own surveys with published data from our nation-wide surveys.

Around 60 or 70% of our surveys are carried out at schools. To each of these schools that help us conduct the surveys, we provide data for comparing that particular school against nation-wide statistics or against data from earlier years. Thus, we ensure that each school that helps us conduct the surveys receives materials useful for educational activities at the school. Data from our surveys should not only contribute to educational activities at schools that help us conduct the surveys, but also should be made available to the general public. Therefore, we ensure that our survey data are made available to the public in an easy-to-use format.

As a result, data from the Benesse Educational Research and Development Center are used in various ways in the home, at schools and at administrative organizations as basic materials that provide insights into education, and thus contribute to society.

Contributing to Research and Education by Depositing Survey Data to Archives

Certain data sets from our surveys are deposited in the SSJDA, a data archive run by the Institute of Social Science, University of Tokyo. To date, we have deposited 78 collections of survey data after suitably formatting them for disclosure. In 2007, studies based on the secondary use of our survey data deposited in the SSJDA were

promoted at the meetings of the secondary analysis study group organized by the Institute of Social Science. The outputs from these studies were compiled under the title of “Empirical Studies on the Choice of Career and on Educational Strategies,” and made available on the web site of the Institute of Social Science (<http://ssjda.iss.u-tokyo.ac.jp/rps/RPS038.pdf>). In 2008, we collaborated with the University of Tokyo in conducting secondary analyses of data from a survey on parents. The results of these studies are available on the web site of the Benesse Educational Research and Development Center under the title of “Report of the 2008 Survey Concerning Parents’ Opinions on School Education” (http://benesse.jp/berd/center/open/report/hogosya_ishiki/2008/hon/).

Our survey data at the data archive contributes to research and is used by young researchers for writing academic papers. The Institute of Social Science has agreed to send us copies of such papers written using our survey data at the data archive, and we receive several such papers each year. Some of these papers have received an encouragement award from a private association (e.g. Kousuke Sudo, “Comparison of the Levels of Understanding of School Lessons among the Three School Years at Junior High School: the Second Year as a Turning Point between Successful and Unsuccessful Students”; http://www.spss.co.jp/ronbun/archives/2008/pdf/poster_04.pdf). Since it is difficult for individual researchers to conduct a major survey, we hope that our survey data at the archive will encourage secondary studies of the data.

Another important aim in depositing data in the data archive is to contribute to education. We have agreed that university students can use the data we have already deposited in the archive for writing their graduation thesis, and that university faculty can use it for teaching students in the classroom. By thus allowing our survey data to be used for educational purposes, we contribute to university education in fields such as pedagogy and sociology.

Expectations on Data Archives and Some Comments on the Challenges

Data archives are a means for our social contributions. We expect data archives to play a key role in promoting the use of survey data, contributing to further education and studies.

Data archives bring great benefits to data depositors as well. By depositing data in an archive, we are making our survey data open to verification. This is not only good in itself for data depositors, but also allows data depositors to connect with a network of researchers who submit papers that they have written using deposited data, often leading to the discovery of new analytical perspectives.

The importance of data in the field of education appears to have been growing, particularly in recent years. For example, in the planning of educational policies and systems, the national and local governments always need data to justify their ideas. Similarly, the evaluation of ongoing administrative measures in the field of education requires quantitative data. In recent years, schools are increasingly conducting surveys on children and students concerning academic achievements and progress in learning in order to obtain information as a basis for discussing school management policies or information to facilitate planning ways to improve teaching. Moreover, many schools conduct surveys to gather necessary information for school assessments, but conducting many surveys for research is burdensome. We expect that data archives will help reduce the burden on schools.

Future challenges include sharing among survey organizations the benefits from depositing survey data in data archives and creating an environment supportive of data depositing. Data archiving organizations are expected to make progress in creating such an environment.

Chapter 2 Presentation and Discussion

Discussion III: Data Archives and the Educational Industry (Conference minutes)

(Watanabe) Thank you for your presentation, Mr. Arai. You have explained the significance of data archives and surveys from the viewpoint of the business community. It was a rare opportunity to hear about the details of surveys conducted by Benesse Corporation and their objectives. Since I am interested in comparing the approach of the business community with those of the government and academics, what struck me most as something unique to Benesse Corporation was how you evaluate the benefits of surveys. For example, you evaluate the benefits from coverage by the mass media in terms of advertising costs. It must be said, however, that it is very difficult to attract the attention of the mass media.

For example, even when a university or a group of researchers reports an epoch-making finding at a meeting of an academic society, it does not always attract the mass media. Looking at the newspapers, I often find articles on the activities of corporations involved in the education business. I have the impression that press releases from such corporations have a good chance of being reported by the mass media. What do you think? Why are the mass media eager to report on the activities of Benesse Corporation?

(Arai) I am not sure because we haven't compared ourselves with others. Still, I might point out that, separate from the Benesse Educational Research and Development Center, Benesse Corporation has a public relations department as a part of its organizational structure. This department regularly contacts the mass media to report not only the results of our surveys but also on various new initiatives by the company. I think that such regular contacts help to increase media coverage of our activities.

In addition, I think that our survey data are interesting to the mass media because we conduct multiple-year and cross-national comparative surveys that address a large population. You may be surprised to know that there are few other such surveys. Besides our surveys being unique, we are regularly helped by various expert academics when we design the surveys. Our surveys are designed to shed light on various interesting topics, and this also may explain why our surveys attract attention.

So, I think there are three reasons: our organizational effort to attract attention, the general framework of our surveys, and the attractiveness of the topics.

(Watanabe) Thank you very much. Does anyone wish to comment on the presentation by Mr. Arai, Director of the Benesse Educational Research and Development Center?

(Granda) Just one. I think it is very interesting the connection that you described between the surveys that you conduct and the ability to archive them because in the United States, the education

industry is very large also. They produce lots and lots of information and surveys, but they do not really think about archiving their data with public archives. There could be nice cooperation between these different organizations because, as you discussed, the media is more inclined to work with businesses because they can present data to them in a way that is comfortable for the media themselves.

We get lots of telephone calls too and Dr. Mochmann mentioned about this pattern, when the media call the archives, they want an answer to a specific question. They do not want to hear the answer: oh, we have this wonderful data set, you can get access to it and all you need to do is some analysis and you will get your answer. So, this cooperation between business archives and the media and the educational establishment, in general, I think could be very beneficial.

(Mochmann) I want to share one observation and I think it is very valuable for me to have you as a representative of the research industry in this group. I considered it always a particular achievement which we managed to set up in Germany that the people from the statistical sector and from the market research associations and from the academic research community are having a forum where they sit together and talk about common interest, common problems, and common methodological procedures and on top of this about adopting common standards.

So my question is, whether there is a similar mechanism over here in Japan and do you, for instance, subscribe to a kind of standard demography which is consistently used in academic, commercial, and statistical surveys? Because, if you would employ such standardst then it would be easier to compare the outcomes of the surveys across different social groupings in the society.

(Ishida) We tend to have this misconception that surveys are done mostly by academic researchers. However, large quantities of data come from surveys done by research institutions in the private sector. Listening to the presentation, I was reminded that such data is a very valuable resource. Our data archive contains large quantities of survey data from private-sector research institutions, such as the research centers of Benesse Corporation and Recruit, and the Japan Institute of Life Insurance. These data have high academic value, and the availability of such data at data archives greatly helps researchers.

Listening to Mr. Arai's presentation, what struck me as a unique strength of Benesse Corporation, which distinguishes them from us, is their remarkable public relations skills. For example, copies of their magazine "VIEW21" are distributed free of charge to school teachers. They organize press conferences four times a year, attracting much media attention. Like Prof. Watanabe, I would like to know why Benesse Corporation is so successful in attracting media attention. I believe that this success is supported by their long experience in this field. Also, I agree with Mr. Arai that Benesse Corporation attracts attention with their multiple-year and cross-national comparative surveys that cover interesting topics. We also prepare press releases, but we are not so successful in attracting media attention, so I guess that Benesse Corporation's success is partly the result of their long

experience.

Similarly, I would say that if we put more energy into public relations, it will eventually make more people willing to be covered by our surveys. That is to say, by spreading the message that we properly conduct surveys of good quality, more people will be willing to answer our surveys. Listening to the presentation, I understood the great importance of public relations.

From the standpoint of depositors, Mr. Arai said that he expects data archivists to establish a mechanism that ensures that depositors are informed of how their data are used, which is very understandable. From the standpoint of data archivists, I should add that different depositors have different expectations. Some of them wish to be informed of how their data are used, but others do not. Data archivists should provide such information to those depositors who wish to know how their data are used. By collecting information from the users, we should find out how their data have been used and how many papers were written using their data, and return such information to the depositors.

(Koyasu) I am now personally involved in a joint project with Benesse Corporation to develop a generic skill test for measuring bachelor-level (or university graduate level) academic skills of university students, addressing not only the completeness of knowledge but also the ability to apply knowledge. In this project, I was impressed by the excellence of Benesse Corporation's employees and the quality of their work. I have heard that data analysis tasks for national surveys on academic achievement were initially commissioned to Benesse Corporation, which handled the data from elementary schools, and another company, which handled data from junior high schools. Apparently there was a significant difference between the two companies in the quality of their work because, while Benesse Corporation continues to receive work under consignment, the second company was replaced by another company. So, I am also convinced of the quality of surveys conducted by Benesse Corporation and the quality of data from these surveys.

At the end of his presentation, Mr. Arai spoke about the need to strengthen the activities of the secondary analysis study group. His presentation confirmed the great importance of this effort. As I told Prof. Ishida during the break, while secondary analyses are very popular in sociology and the reputation of sociological researchers often depends on their skill in secondary analyses, this is not the case in psychology. In psychology, meta-analyses are common but secondary analyses are rare. Psychologists tend to respect the freshness of data: they are interested in catching fish in the water, so to speak. On the other hand, they are not interested in stocks of data, which may appear like dried fish. This is an understandable but incorrect attitude. It is very important for psychologists to become more interested particularly in the analysis of longitudinal data from long-term surveys because such data can never be collected by individual researchers.

So, in the future, it is important to strengthen the activities of the secondary analysis study group and similar activities to make more researchers understand the significance of secondary analysis and share with them the know-how of secondary analysis.

(Watanabe) Mr. Miyake, would you like to make a comment as one who commissions the analysis of the national surveys on academic achievement data?

(Miyake) Let me first correct a possible misunderstanding. To Benesse Corporation, we commission the technical handling of achievement test data from elementary schools, but not research work. So, at present, they are not doing any research using the achievement test data. Their research findings come solely from analyzing survey data that they have collected by their own surveys. Please note that their researches are quite independent from the work they do for us.

(Koyasu) Point taken. I understand that they are asked to do the technical handling of achievement test data.

(Miyake) I have a question to Mr. Arai. As an outsider, I used to wonder what could be the benefits for your company in conducting such a variety of research projects. Your presentation, with some very realistic examples and an explanation of how you evaluate the benefits, helped me to understand the significance of such research for a corporation.

Your research produces very interesting findings as proven by the strong media attention. But how do you decide the themes of your studies? I imagine that you first draw up a variety of research themes with some idea of their benefit to your company and then design your surveys accordingly. I am interested to know how you make a corporate decision about what topics to choose and about the survey design.

(Arai) Basic research themes do not change much because they are covered by multiple-year projects. However, we add new topics if we find it necessary even though it will not be possible to compare such new topics with earlier survey data. Usually, the same members are involved in making such decisions. Examples of new topics include the use of ICT (information and communication technologies) and after-school activities. These topics were added because we recognized the growing need to monitor these subjects. We have many other ideas for new topics, but since our human resources and budget are limited, the department in charge must first narrow down the list of topics to several choices. The final choices of new topics are discussed and approved by the company's board before they are covered by our surveys.

With regard to the use of ICT, this topic was already partially covered by our basic research on learning and our research concerning how children spend their time. In the questionnaires for such basic researches, there were a few questions like "How much is your learning helped by ICT?" or "How often in your daily life do you use ICT for learning?" Since we felt the need to ask more questions, we discussed the idea of starting an independent multiple-year survey concerning ICT utilization. This idea was approved by the company, mainly at my insistence, and the project was started. Our report on this ICT utilization survey was appreciated even more than we expected. So the company approved the idea of continuing this survey as a multiple-year project. This is the

typical procedure of our decision-making.

(Watanabe) Thank you very much. Does anyone from the audience have a question about the presentation of Mr. Arai, Director of the Benesse Educational Research and Development Center?

(Kajita) I highly appreciate the work of Benesse Corporation, which, as a private corporation, conducts surveys not only for its own benefit but also for society. Their commitment to social contribution is evident by their willingness to release the data they have collected, by publishing materials from their surveys and making their survey data available in a data archive. In fact, for a private corporation, such a daring attitude could be a double-edged sword in various business activities. For example, the published materials could provoke strange criticisms. Even though researchers may ignore such criticisms, a private corporation would suffer damage to its image. Therefore, most private corporations refrain from disclosing their materials. I guess Benesse Corporation makes its survey data available to the public because it does not fear such consequences. For quite a long time in my youth, I worked for advertising agencies and similar companies, helping them to design and conduct surveys for supporting sales. In such cases, the names of client companies are often falsified so that nobody knows who commissioned the surveys. The names of client companies remain hidden. Such surveys are conducted in secret and the survey data are never revealed to outsiders. Of course, in every kind of business, one should not depend on guesswork but start with collecting data. Today, however, it is important that private corporations begin to accurately assess how much of the data they have collected may reasonably be made public or made available for use by third parties. Listening to Mr. Arai, I was impressed by the excellent policy of Benesse Corporation.

(Watanabe) Thank you very much. President Kajita's comment reminds me of my own experience. With the earlier Director of ESSRC, I once visited Benesse Corporation to ask them to deposit their survey data in our data archive. I was surprised by the way Mr. Arai, Director of the Benesse Educational Research and Development Center, accepted our request very casually. I remember that he then reminded us of the need for careful data management and synchronization. President Kajita mentioned that a biased analysis of published data might damage a company's image. Mr. Arai, what do you think about the risk?

(Arai) We disclose our survey data because we wish to assist deeper, broader analyses. There might be a risk but, if we respect the accuracy of information, it is better to disclose our survey data, have them analyzed also by others, and invite various criticisms concerning the shortfalls of our analysis. We must adhere to this policy to sufficiently fulfill our social responsibility. The disclosure of survey data may be risky but we focus on its benefits. I think this is a part of our corporate culture. At press conferences, we often receive many questions. To those questions that we cannot answer, we simply say that we don't know. To those questions that we haven't studied, we simply say that

such subjects haven't been covered by our surveys. We don't try to invent answers, and whenever we are not sure about something, we try to cover it in future surveys. Since such openness is a part of our corporate culture, we don't think so much about the risks. I don't think we have ever suffered major damage due to this openness. We have heard that some bloggers are actively exchanging opinions about our survey data, but that is not harmful to us and we are not worried.

(Watanabe) Thank you very much. Does any panelist or anyone from the audience have a question? Well, even though we have about ten minutes left, I would like to conclude the third session of the day on "Data Archives and Education Service Business." Today, we already have spent a long time together, and we did not have a break in the morning. I hope you are not too tired. Thank you for your attention for so many hours.

General overview

Yasuo Watanabe

Professor, Hyogo University of Teacher Education
Director, Educational and Social Survey Research Center

Dear panelists and the audience, thank you for your participation and for listening to the long discussions. Today, in this symposium, we heard about the experiences of various data archives in presentations from Dr. Granda from ICPSR, Dr. Mochmann from GESIS, and Prof. Ishida from the University of Tokyo that manages SSJDA. From our side, we shared our experience with JEDI. Then Prof. Koyasu and Mr. Miyake spoke about various issues concerning data archives. Finally, we enjoyed a presentation by Mr. Arai, Director of the Benesse Educational Research and Development Center. From these presentations and discussions, it seems we have reached a great conversion on the important issues concerning data archives. I think that the challenges have been identified sufficiently well, but in reality it is not easy to overcome them.

The theme of the symposium today was: “Data Archives and International Collaboration for Research, Public Policy, and Business.” We knew this theme would not attract many people, and, therefore, I sincerely appreciate the audience who took the trouble to travel a long way to attend this symposium on such a difficult theme. I learned about the need to create a culture of data sharing, according to the words of Dr. Mochmann, in Japan through various efforts such as accumulating more data at data archives, promoting secondary analysis, and publishing valuable findings from researches conducted with the help of data archives.

Thank you all for your participation and attention for many hours.

Closing comments

Eiichi Kajita

President, Hyogo University of Teacher Education

I would also like to thank everyone for your participation and attention for many hours. Today, we enjoyed excellent presentations and good discussions. It was a surprise to find that we could remain attentive for so many hours without losing track in the middle. I think this was thanks to the many interesting stories we heard. I would like to share a few of my thoughts that came to me during the symposium.

First, I would like to say that, not only in Japan but also in the United States, emotionalism and impressionism were dominant when people talked about education in the 1970s. At that time, under the slogan of “Open Education,” flexibility of curriculum and freedom for children were encouraged. These were interesting initiatives, but the academic ability of children continued to drop. Around the 1980s, “Back to Basics” became a popular slogan, and around this time people began to criticize the school education of the 1970s that tried to support the freedom and uniqueness of children because such a style of education seemed to have caused a major drop in academic ability and a huge increase in problematic behavior.

At that time, I was working at the National Institute for Educational Policy Research. Since the Institute was well funded in those days, I was sent to the United States every year to study the situation there. So, I have many memories of what I saw in the 1970s and the early 1980s. I saw interesting initiatives, listened to interesting discussions, and heard of beautiful episodes that brought tears to my eyes. However, throughout that time, the academic ability of children continued to drop, and there was a huge increase in problematic behavior, so the slogan “Back to Basics” emerged. Then, in 1983, people were alarmed by a report called “A Nation at Risk.” This was a turning point.

Around that time, people began to discuss “Outcome Based School” and “Competency Based Education.” The first slogan asserted that schools should primarily focus on outcomes. The second slogan, “Competency Based Education,” asserted that education is not worthy of its name unless it leads to the strengthening of competency, which refers to academic ability in a broader context. These assertions highlighted the apparent need to go “Back to Basics.” This trend continues today, combined with a totally evidence-based educational policy. So, while listening to the discussions today, I was reminded of the need to go back to the basic premises of education such as the importance of evidence, the importance of outcomes from education, and the need to help children develop their academic abilities. For all these things, data is very important.

Let me tell one final story. Even though I majored in psychology, I learned about education in 1971 under Prof. Benjamin Bloom of the University of Chicago. So, even though I graduated from the department of psychology in the faculty of literature, I have received a more or less

complete training in education. At that time, I was particularly interested in the taxonomy of educational objectives, which was a field with a global reputation in those days. In those studies, I would say that what we call knowledge can be measured and represented by data in many different ways. Similarly, what we call understanding can be measured and represented by data in many different ways. The government's new guidelines for teaching stress the equal importance of learning, application and exploration. We must realize that the concept of learning, or the concept of application, refers to the functioning of the mind in many different ways. The same is true with exploration. In the taxonomy of educational objectives, the functioning of the mind in the cognitive domain alone is classified into 20, 30 or even 40 categories. The functioning of the mind in each category requires a unique method for measurement and representation by data. Such details have been described in a book of 200 to 300 pages, which is the result of joint research by American psychologists for more than a decade. This work became world-famous in the 1970s.

Since I have studied this field, I believe that data are as important as many people say, but that we must also be aware that this importance implies additional challenges. With anything that we refer to by an ordinary expression or an academic term, how shall we represent it by data? I don't think people have really begun to discuss this question in Japan. For example, it is easy to talk about "the development of academic ability." It is equally easy to say that "the concept of academic ability is similar to the concept of competency." However, the concept of academic ability or competency may refer to the functioning of the mind in many different ways that can be classified into several dozen categories. The functioning of the mind in each category requires a unique method for measurement and representation by data. What I am talking about is related to the classical choice between a descriptive test and an objective test. However, even when a descriptive test is chosen, questions may address the functioning of the mind in very different categories depending on how they are formulated.

We appreciate that data archives are being prepared in Japan. However, we must carefully examine the data in archives to clarify the relationships between data and concepts. We must clarify the rules that govern the association of concepts with data, otherwise we are allowing the random interpretation of data. This is one of the things we must do in the future.

These two things came to mind while listening to the presentations and discussions. As a general impression, I think that things are heading in the right direction. If we do not pay more respect to facts and evidence in various discussions, there will be an immediate clash of opinions. This could develop into conflicts among factions, or in the case of the academic community, complete disagreement among sociologists, psychologist and pedagogists. We do not wish to see that. If we all become more respectful of evidence, there will be fewer conflicts among factions, among different disciplines in science, and even among different groups of people in society. What I am saying is a matter of fact, but with such an ideal in our minds, we must continue to encourage the sharing of a wider variety of data, in the hope that

it will lead to new analytical findings that the depositors of data could not have imagined. This symposium strengthened my sense of commitment to achieving this goal.
Dear panelists and the audience, thank you very much. I sincerely appreciate your participation and attention.



SYMPOSIAST CVs

Challenges in Archiving Education Data: The ICPSR Experience

.....

PETER GRANDA

AFFILIATION

Archivist and Assistant Archival Director
Inter-university Consortium for Political and Social Research
(ICPSR) University of Michigan



FIELD OF SPECIALIZATION

Data Harmonization; Metadata in the Social Science Data Life Cycle

BIOGRAPHY

EDUCATION

- Summer 1986 ICPSR Summer Program
(Statistical Analysis and Data Management)
- 1974 -- 1984 University of Michigan (Ph. D. History, 1984)
- 1971 -- 1974 Kansas State University (M. A. History, 1974)
- 1965 -- 1969 Fairleigh Dickinson University (B. A. Political Science, 1969)

POSITIONS AND EMPLOYMENT

- 1994 -- present Archivist (since 2000) and Archival Assistant Director (since 1994),
Inter-university Consortium for Political and Social Research (ICPSR),
Institute for Social Research, University of Michigan.
- 1986 -- 1994 Research Associate I, II, and Senior Research Associate, ICPSR.
- 1985 -- 1986 Data Archive Specialist, ICPSR.
- 1984 Visiting Assistant Professor, Department of History,
University of Iowa (History of India).
- 1984 Adjunct Lecturer, Department of History, University of Michigan
(Comparative History).
- 1969 -- 1971 Peace Corps Volunteer, village-level agriculture (India).

HONORS AND AWARDS

- 051497 Granda (PI) 11/15/07-11/14/12
Robert Wood Johnson Foundation “Health and Medical Care Archive”
This project involves the acquisition and processing of data collections produced under the sponsorship of the Robert Wood Johnson Foundation.
Role: Principal Investigator
- R01 HD053533-02 Smock (PI) 04/01/07-03/31/12
National Institutes of Health “Integrating U.S. Fertility Surveys”
This project will produce a harmonized dataset of U.S. family and fertility surveys spanning the 1955-2002 period and facilitate temporal analyses.
Role: Co-Investigator
- 200-2000-07001 Groves (PI) 12/08/99-09/30/10
Health and Human Services “National Survey of Family Growth”
ICPSR’s role in this project involves the production of public-use data and documentation files of data collection.
Role: Project Data Manager
- 050700 Granda (PI) 06/15/05-06/14/07
Robert Wood Johnson Foundation “Health and Medical Care Archive”
This project involves the acquisition and processing of data collections produced under the sponsorship of the Robert Wood Johnson Foundation.
Role: Principal Investigator
- R01 HD42564-01 Gutmann (PI) 07/01/02-06/30/07
National Institutes of Health “NIH 2000 Census Data Project”
This project involves archiving, enhancing, creating new data products, organizing training workshops, and distributing 2000 Census data.
Role: Co-Investigator
- SES 0137019 Gutmann (PI) 02/15/02 – 01/31/05
National Science Foundation “2000 U.S. Census Data Project”
This project involves archiving, enhancing, creating new data products, organizing training workshops, and distributing 2000 Census data.
Role: Co-Investigator
- Prime K# RN95127001 Granda (PI) 07/01/01 – 12/20/05
U.S. Department of Education “International Archive of Education Data”
This project involves archiving and distributing public-use files from National Center for Education Statistics. Role: Principal Investigator
- Social Science Research Council Research Grant (with Thomas R. Trautmann and Richard D. Saran), 1982-1983
- General University Fellowship, University of Michigan, 1980-1981

- Foreign Language Area Studies Fellowship (Hindi), University of Michigan, 1979-1980
- Horace H. Rackham School of Graduate Studies, University of Michigan, Pre-Doctoral Fellowship, 1978-1979
- Fulbright-Hays Doctoral Dissertation Research Abroad Fellowship, 1977-1978
- History Faculty Award, Department of History, University of Michigan, 1977-1978 (declined)
- National Defense Foreign Language Fellowship (Hindi), University of Michigan, 1976-1977
- Orla B. Taylor Scholarship, Department of History, University of Michigan, 1976-1977 (declined)
- National Defense Foreign Language Fellowship (Hindi), University of Michigan, 1975-1976
- American Institute of Indian Studies Language Fellow (Tamil), 1972-1973
- B.A. Summa Cum Laude, Fairleigh Dickinson University, 1969

PUBLICATIONS / PAPERS / PRESENTATIONS

SELECTED PEER-REVIEWED PUBLICATIONS

Erik W. Austin, Jerome M. Clubb, William H. Flanigan, Peter A. Granda, and Nancy H. Zingale. "Electoral Participation in the United States, 1968-1986." *Legislative Studies Quarterly*, Vol. 16, Number 1, February 1991, pp. 145-164.

Peter Granda, Thomas R. Trautmann, and Richard D. Saran, "The Study of South Indian Inscriptions," in R. E. Frykenberg and Pauline Kolenda, eds., *Studies of South India: An Anthology of Recent Research and Scholarship*, Madras: New Era Publications, 1985.

"The 'Gift after Purchase' in Vijayanagara Inscriptions", Vol. 6 of the *Journal of the Epigraphical Society of India (Bharatiya Purabhilekha Patrika)*, 1979, pp. 25-31.

OTHER SELECTED PUBLICATIONS / PAPERS / PRESENTATIONS

"*DDI Across the Lifecycle: One Data Model, Managing the Metadata Life Cycle: The Future of DDI at GESIS and ICPSR*" with Mary Vardigan, Joachim Wackerow, Meinhard Moschner, and Wolfgang Zenk-Moeltgen. Presentation at the 35th Annual Conference of the International Association for Social Science Information Service and Technology (IASSIST), Tampere, Finland, May 26-29, 2009.

“DDI Across the Lifecycle: One Data Model, Many Products” with Mary Vardigan, Sue Ellen Hansen, Sanda Ionescu, and Felicia LeClere. Presentation at the 35th Annual Conference of the International Association for Social Science Information Service and Technology (IASSIST), Tampere, Finland, May 26-29, 2009.

“Common Workflows: Health and Social Science Data Curation Collaborations”. Panel presented at DigCCurr 2009: Digital Curation Practice, Promise and Prospects: Chapel Hill, NC, April 1-3, 2009.

“DDI Across the Lifecycle: One Data Model, Many Products” with Mary Vardigan, Sue Ellen Hansen, Sanda Ionescu, and Felicia LeClere. Presentation at the Sixth Annual Workshop on Comparative Survey Design and Implementation, Ann Arbor, MI, March 5-7, 2009.

“ICPSR and the Fenway Institute Partnership”. Presentation at the Faculty Meeting of the Fenway Institute/Center for Population Research in LGBT Health, Boston, July 10-11, 2008.

“Harmonizing Survey Data” with Christof Wolf. Presentation at the International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts (3MC), Berlin-Brandenburg Academy of Sciences and Humanities, Berlin Germany, June 25-28, 2008.

“Harmonization of Survey Data: An Introduction”. Presentation at the 2nd Meeting of the European Survey Research Association (ESRA). Prague, June 25-29, 2007.

“Want Data?” Presentation at the Robert Wood Johnson Foundation New Connections Initiative Annual Grantee Symposium. Princeton, NJ, May 17-18, 2007.

“Best Practices in the Dissemination of Survey Data”. Presentation at Workshop on Planning for the Future of the General Social Survey. National Science Foundation. Arlington, VA, May 2-3, 2007.

“ICPSR and the Gerald R. Ford Presidential Library: Two Decades of Collaboration”. Presentation at the Annual Meeting of the Society of American Archivists. Washington, DC, July 31-August 5, 2006.

“Archival Partners: Handling “Born Digital” Materials”. Presentation at the Annual Meeting of the International Association for Social Science Information Service and

- Technology (IASSIST). Ann Arbor, MI, May 23-26, 2006.
- “Integrating Fertility Surveys Conducted in the United States since 1955”*. Presentation at the Fourth International Workshop on Comparative Survey Design and Implementation, Social and Cultural Planning Office (SCP). The Hague, March 16-18, 2006.
- “Data Archives and Data Producers: A Cooperative Partnership”*. Presentation at the 4th Annual Scientific Conference of the Independent Institute for Social Policy, Moscow, Russian Federation, December 8-9, 2005.
- RM Groves, G Benson, WD Mosher, J Rosenbaum, P Granda, W Axinn, JM Lepkowski, A Chandra. *Plan and Operation of Cycle 6 of the National Survey of Family Growth*. Vital and Health Statistics, Series 1, Number 42. Hyattsville, MD: National Center for Health Statistics. August 2005.
- “The Data Documentation Initiative (DDI) and Comparative Survey Data”*. Presentation (with Meinhard Moschner) at the Third International Workshop on Comparative Survey Design and Implementation, Ortega y Gasset Foundation, Madrid, March 10-12, 2005.
- “WEBDOC and Survey Data”*. Panel at the Second International Workshop on Comparative Survey Design and Implementation (CSDI), Institut National D’Etudes Demographiques, Paris, April 1-3, 2004.
- “Geographic Elements and Social Science Data: The Archival Perspective”*. Panel at the Annual Meeting of the Social Science History Association, St. Louis, October 24-27, 2002.
- “The Future of Data Dissemination and Research: The Continuing Evolution of Web-Based Data Access Tools”*. Panel at the Annual Meeting of the Social Science History Association, Chicago, November 15-18, 2001.
- “From Many, One: The Archival Infrastructure for U.S. Social Science Data Research”*. Panel at the Annual Meeting of the Social Science History Association, Pittsburgh, October 26-29, 2000.
- “DDI and Experiences at ICPSR”, “Data Documentation and DDI”, “DDI and the Data Produce”, and “ICPSR and Data Security”*. A series of presentations at the Expert Seminar of the Council of European Social Science Data Archives (CESSDA), Tampere, Finland, September 1-2, 2000.

“Old and New Technical Documentation: End User Access to the National Survey of Family Growth” Presentation at INFOCUS 2000, the International Health Information Conference, Vancouver, British Columbia, June 24-27, 2000.

“DDI and NESSTAR: Integrated Tools for Metadata-driven Resource Discovery”. Workshop presented (with Mary Vardigan) at the joint meeting of the International Association for Social Science Information Service and Technology (IASSIST) and the Canadian Association of Public Data Users (CAPDU), Toronto, May 17-21, 1999.

“DDI - Data Documentation Initiative”. Presentation (with Peter Joftis) at the Expert Seminar of the Council of European Social Science Data Archives (CESSDA), Zentralarchiv, Cologne, October 28 - 31, 1998.

“Archival Resurrections: Old Data Remastered”. Paper presented at the annual meeting of the Association of Public Data Users, Washington, October 19-22, 1997.

“ICPSR: The Changing Face of Archival Services”. Paper presented at the annual meeting of the Social Science History Association, New Orleans, October 10-13, 1996.

ABSTRACT:

This paper will describe the experiences of the Inter-university Consortium for Political and Social Research (ICPSR), a large social science data archive located in the United States, with the acquisition, preservation, and dissemination of education data to the research community. Focusing on data from both governmental and academic sources, the presentation will also stress the importance of confidentiality issues and international efforts to promote the availability of data as widely as possible.

From National Data infrastructures to International Data Infrastructure Networks

.....

EKKEHARD MOCHMANN



AFFILIATION

Former Director, International Data Service
GESIS-Leibniz Institute for the Social Sciences
Former Administrative Director, Central Archive for
Empirical Social Research (ZA) at University of Cologne

FIELD OF SPECIALIZATION

1. International comparative research and its data infrastructure
2. Computer aided content analysis: Democracy concepts of political elites in the language of politics
3. Social Sciences and Humanities Futures
4. e - Social Science

BIOGRAPHY

- 1944: Born in Lutherstadt Eisleben, Germany
- 1969: Diplom Kaufmann (MBA), Economics and Social Sciences Department, University of Cologne
- 1969: Assistant professor, Zentralarchiv für Empirische Sozialforschung, University of Cologne
- 2003: PhD h. c., University of Gothenburg
- 2006/2007: Visiting professor Sciences Po, Paris
- 2008: Visiting professor ISS, University of Tokio

PROFESSIONAL CAREER

Ekkehard Mochmann studied Social Sciences and Law at the Universities of Cologne and Bonn. He was executive manager (1977-2002), then administrative director of the Central Archive for Empirical Social Research at the University of Cologne (ZA) (2003 -

2008), Director of GESIS, the German Social Science Infrastructure Services (1998 – until retirement 2008) and founder and project leader of ZA-EUROLAB, the European data laboratory for comparative social research(1996 - 2008). He taught methods courses in various international summer schools.

He has been expert advisor to the European Science Foundation- Standing Committee for the Social Sciences (ESF-SCSS) for the European Data Base and was member of the Methodology group for the European Social Survey. On the national level he contributed to the German National Science Foundations Memorandum on Quality Criteria for Survey Research. Ekkehard Mochmann was President of the International Federation of Data Organisation for the Social Sciences (IFDO) as well as President of the Council of European Social Science Data Archives (CESSDA).

Currently he serves on the Conseil scientifique du Comité de concertation pour les données en sciences humaines et sociales of the French Research Department for a second term and has served on the Infrastructure Committee "Towards a future data-infrastructure for the social sciences" of the Royal Netherlands Academy of Arts and Sciences as well as on the International infrastructure panel at the Swedish Research Council and the OECD drafting group "Access to publicly funded data".

EU related project activities include the project management of the European Data Laboratory for comparative social research (ZA-EUROLAB) under the TMR /LSF activity and the IHP / Access to Research Infrastructures Programme, the MetaDater Project to create a data model and tools for best practice documentation of Comparative research data, as well as the EU Network of Economic and Social Science Infrastructure in Europe (NESSIE).

Mochmann is member of the Board of the German Association for Communication Research and was steering committee member of the European Consortium for Communication Research (ECCR).

In 2006 the International Social Science Council appointed Mochmann as senior advisor on data issues. Currently Dr. Mochmann is working with the Interdisciplinary Centre for Comparative Research in the Social Sciences (ICCR, Vienna) on the international project "Social Sciences and Humanities Futures" funded by the European Commission and is member of the High Level Expert Panel for the Lithuanian Research Infrastructure Roadmap.

HONORS AND AWARDS

- Honorary board member of the Council of European Social Science Data Archives (CESSDA)
- Doctor honoris causa in Philosophy awarded by the Political Science Department of the University Gothenburg, Sweden (2003)

SELECTED PUBLICATIONS / PAPERS / PRESENTATIONS

Mochmann, Ekkehard: Improving the evidence base for international comparative research, in: *International Social Science Journal*, Volume 59, Issue 193-194, pages 489 - 506, October 2009.

Mochmann, Ekkehard: e-Science Infrastructure for the Social Sciences, Council for Social and Economic Data Working Paper Series WP115, Berlin, July 2009

Kaase, Max; Mochmann, Ekkehard; Frank, Sybille: On the development of political culture in the Federal Republic of Germany between 1952 and 1962 (in German) in: Christian H.C.A. Henning und Christian Melbeck (Hrsg.): *Interdisziplinäre Sozialforschung. Theorie und empirische Anwendungen*, Frankfurt a.M. 2003.

Mochmann, Ekkehard: On the Institutionalisation of International Comparative Electoral Research (in German) In: Dieter Fuchs, Edeltraud Roller, Bernhard Wessels (Hrsg.): *Bürger und Demokratie in Ost und West. Studien zur Politischen Kultur und zum Politischen Prozess*. Westdeutscher Verlag 2002, S. 227-241

Mochmann, Ekkehard: International Social Science Data Service: Scope and Accessibility. Report for the International Social Science Council (ISSC), Cologne 2002

Mochmann, Ekkehard.

Infrastructure for Comparative Social Research in Europe (in German) In: Uwe Hasebrink, Christiane Matzen (Hrsg.): *Forschungsgegenstand Öffentliche Kommunikation. Funktionen, Aufgaben und Strukturen der Medienforschung*. Symposium des Hans-Bredow-Instituts Nr. 20. Nomos Verlagsgesellschaft: Baden-Baden/Hamburg 2001, S. 161-173

Mochmann, Ekkehard: The Infrastructure of academically organised Social Research: Development and Problems (in German) in Heinz Sahner (ed.)

Wissenschaftliche Jahrestagung 2001 der ASI, in Kooperation mit ADM und BVM
Fünfzig Jahre nach Weinheim: Empirische Markt und Sozialforschung gestern, heute,
morgen. Weinheim 26.10.2001

Klein, Markus; Jagodzinski, Wolfgang; Mochmann, Ekkehard; Ohr, Dieter (eds.):
50 Years Electoral research in Germany (in German) (incl. CD with data base),
Wiesbaden: Westdeutscher Verlag 2000

Klingemann, Hans-Dieter; Mochmann, Ekkehard; Newton, Kenneth (eds.): Elections in
Central and Eastern Europe. The First Wave. Wissenschaftszentrum Berlin für
Sozialforschung, Abteilung: Institutionen und Sozialer Wandel. Edition Sigma 2000

Mochmann, Ekkehard; Oedegaard, Ingvill; Mauer, Reiner (in cooperation with the
ICORE and the CESSDA: Inventory of National Election Studies in Europe
1945-1995: Belgium, Denmark, France, Germany, Great Britain, Hungary, The
Netherlands, Norway and Sweden. Bergisch Gladbach: Edwin Ferger Verlag 1998

Mochmann, Ekkehard; Gerhardt, Uta (eds.). Violence in Germany: Social findings and Lines
of interpretation. Herausgegeben im Auftrag der Arbeitsgemeinschaft Sozialwissen-
schaftlicher Institute e.V. (ASI). München: Oldenbourg 1995

Gerhardt, Uta; Mochmann, Ekkehard (eds.). Social Transformation 1945 –1990.
Re-Democratisation and Living Conditions, München: Oldenbourg 1992

Best, Heinrich; Mochmann, Ekkehard; Thaller, Manfred (eds.).
Computers in the Humanities and the Social Sciences. Achievements of the 1980s -
Prospects for the 1990s. München, London, New York, Paris: Saur 1991

de Guchteneire, Paul; Mochmann, Ekkehard (eds.). Data Protection and Data Access.
Reports from Ten Countries on Data Protection and Data Access in Social Research, with an
Annotated International Bibliography. Amsterdam: SWIDOC 1990

Mochmann, Ekkehard (ed.): Computer Strategies for Communications Analysis, Frankfurt
(in German): Campus 1980

ABSTRACT

Starting from observations of the impact data infrastructures had on their scientific
communities, this papers outlines the functions of a social science data archive using the

GESIS data archive at Cologne in Germany as an example. Founded in 1960 as first social science data archive in Europe it contributed much to the development of archival tools and methods of secondary analysis. Working in close interaction with researchers and research networks, the data archive emphasized ex ante and ex post harmonisation of national data sets and their integration into international comparative data bases allowing cross cultural and longitudinal analyses.

The early years of the data archive were characterized by a severe data shortage. So systematic efforts were implemented to locate and acquire relevant data via the research project documentation on empirical social research in its scientific community. It was obvious that data literacy had to be unfolded. To this end annual spring seminars were started to train scholars in advanced methods of data analysis and enabling them to transfer the acquired skills together with analysis packages and related data to their home universities. The exponential growth of data holdings required information retrieval tools that allow direct access to data on variable and file level, as well as interactive statistical analysis of the related individual micro data. So metadata standards and the development of data bank systems came into focus.

Since international comparative research was cumbersome given the shortage of resources for trans-national research and difficult transatlantic communication in these decades, the pioneers of the data infrastructure worked towards a network of internationally co-operating data archive in the International Federation of Data Organisations for the Social Sciences (IFDO) and the Council of Social Science Data archives (CESSDA). The archives became central nodes supporting comparative research networks and increasingly included different types of empirical evidence, such as statistical micro data and aggregate data to support multi level analyses and likewise materials for qualitative and historical social research such as party manifestoes or life biographies. EUROLAB was created as access point for scholars to make use of the comparative data resources, training in advanced data analyses and data management with support from the access to research infrastructures programme of the European Commission.

Privacy legislation called to balance the right of informational self determination and the research access to relevant data through instruments and concepts such as informed consent, factual anonymity and research data centres serving as safe centres for access to sensitive data.

Then this paper informs about a documentation of longitudinal studies held by the archive and will proceed to two recently created research data centres for education in

Germany. These are the “Research Data Centre at the Institute for Educational Progress in Berlin” and the “Research Data Centre of the Federal Institute for Vocational Education and Training” (BIBB-FDZ) in Bonn, both operating under principles established by the German Council for Social and Economic Data (RatSWD).

Finally this paper addresses declarations such as the OECD Declaration and the Berlin Declaration for open access to publicly financed data that are targeted to foster a culture of open data sharing.

SSJDA (Social Science Japan Data Archive) and Activities at the Center for Social Research and Data Archives

.....

HIROSHI ISHIDA

AFFILIATION

Professor Institute of Social Sciences, University of Tokyo
Vice President Japanese Association for Mathematical Sociology



FIELDS OF SPECIALIZATION

Comparative social stratification and mobility, Japanese studies,
education and the labor market

BIOGRAPHY

Professional and Academic Positions

Professor	Institute of Social Sciences, University of Tokyo, 1999-present
Visiting Professor	Department of Sociology, University of Michigan, 2001-02, 2004, 2005
Visiting Research Scientist	Institute for Social Research, University of Michigan, 2001-02, 2004
Associate Professor	Institute of Social Sciences, University of Tokyo, 1995-99
Associate Professor	Department of Sociology and the East Asian Institute Columbia University, 1991-95
Assistant Professor	Department of Sociology and the East Asian Institute Columbia University, 1989-91
Senior Associate Member	St. Antony's College, University of Oxford, 1985-89
Associate Member	Nuffield College, University of Oxford, 1985-89 (John Swire Centenary Scholar 1985-87, Spencer Fellow 1987-89)

Education

Harvard University	Ph.D. in Sociology, 1986
Harvard University	M.A. in Sociology, 1983
Sophia University, Tokyo	B.A. in Sociology, 1979
University of Arizona	Sociology and Anthropology, 1976-78

HONORS AND AWARDS

- Grant-in-Aid for Scientific Research, Japanese Ministry of Education, Science and Technology, and Japan Society for Promotion of Science
1997-2000, 1998-2001, 2000-2004, 2002-2003, 2002-2005, 2004-2008, 2006-2011
- National Research Grant, Japanese Ministry of Health, Labor and Welfare, 2004-2008
- Research Grant, Center for International Business Education, Columbia University, 1995
- Abe Fellowship, Social Science Research Council, New York, U.S.A., 1993-94
- Fellowship, Council for Research in the Social Sciences, Columbia University, 1990 and 1991
- Spencer Fellowship, National Academy of Education, U.S.A., 1987-88
- John Swire Centenary Scholarship for Post-Doctoral Study at Oxford, John Swire and Sons (Japan) 1985-87
- Toyota Foundation Grant, Toyota Foundation, Tokyo, 1985-86
- Dissertation Research Grant, Edwin O. Reichauer Institute of Japanese Studies, Harvard University 1985
- Dissertation Research Award, Department of Sociology, Harvard University, 1983-84
- ITT International Scholarship for Graduate Study at Harvard, Fulbright Commission of Japan and the Institute of International Education, U.S.A., 1979-81
- Phi Beta Kappa, University of Arizona, 1978
- General International Scholarship, University of Arizona, 1977-78
- Special Scholarship for Academic Achievement, Sophia University, 1975-76

PUBLICATIONS / PAPERS / PRESENTATIONS

Books

- Hiroshi Ishida and David Slater (eds.). 2009. *Social Class in Contemporary Japan: Structures, Sorting and Strategies (Nissan Institute/Routledge Japanese Studies)*, (Routledge).
- Hiroshi Ishida (ed.). 2008. *Social Stratification and Social Mobility in Late-Industrializing Countries*. (The 2005 SSM Research Committee).
- Hiroshi Ishida (ed.). 2006. *Educational and Career Choices and Attitudes Changes among High School Students*, (in Japanese) (Institute of Social Sciences, University of Tokyo).
- Hiroki Sato, Hiroshi Ishida, and Ken'ichi Ikeda (eds.). 2000. *Secondary Data Analysis in Social Research*, (in Japanese) (University of Tokyo Press).
- Takehiko Kariya, Shinji Sugayama, and Hiroshi Ishida. 2000. *Schools, Public Employment Offices and the Labor Market in Post-war Japan*, (in Japanese) (University of Tokyo Press).
- co-winner of the 2000 Labor Studies Book of the Year Award
- Hiroshi Ishida (ed.). 1998. *Social Stratification and Mobility: Basic Analysis and Cross-national*

- Comparison*, (in Japanese) (The 1995 SSM Research Committee).
- Hiroshi Ishida. 1993 (paperback edition 1995). *Social Mobility in Contemporary Japan: Educational Credentials, Class, and the Labour Market in a Cross-national Perspective*, (Macmillan Press and Stanford University Press).

Selected Articles and Chapters

- Hiroshi Ishida. 2009. "Does Class Matter in Japan," in *Social Class in Contemporary Japan*, edited by Hiroshi Ishida and David Slater, (Routledge), pp. 31- 56.
- Hiroshi Ishida and David Slater. 2009. "Social Class in Japan," in *Social Class in Contemporary Japan*, edited by Hiroshi Ishida and David Slater, (Routledge), pp. 1-30.
- Hiroshi Ishida. 2009. "Japanese Life Course Panel Survey and Educational Attainment of Youth," *Social Science Japan* 40, pp. 3-6.
- Hiroshi Ishida and Satoshi Miwa. 2009. "Intergenerational Class Mobility and Japanese Society: Long-term Trends and Cross-national Comparisons," (in Japanese) *Japanese Sociological Review* 59(4): 648-662.
- Hiroshi Ishida, Satoshi Miwa, and Akane Murakami. 2009. "Japanese Life Course Panel Survey 2008 and Life Styles and Consciousness of Japanese People," (in Japanese) *Chuo Chosaho (Central Research Report)* 616: 1-7.
- Hiroshi Ishida. 2008. "Trends in Intergenerational Class Inheritance: Survival Analysis Approach," (in Japanese) *Riron to Hoho(Sociological Theory and Methods)* 23(2): 41-63.
- Hiroshi Ishida and Satoshi Miwa. 2008. "Trends in Intergenerational Class Mobility and Education in Japan," (with Satoshi Miwa) pp 1-48 in *Social Stratification and Social Mobility in Late-Industrializing Countries*, edited by Hiroshi Ishida. Tokyo: The 2005 SSM Research Committee.
- Hiroshi Ishida, Satoshi Miwa, and Masao Ohshima. 2008. "The Panel Survey of the Institute of Social Sciences, University of Tokyo: Results from the Japanese Life Course Panel Survey 2007," (in Japanese) *Chuo Chosaho (Central Research Report)* 604: 1-7.
- Hiroshi Ishida. 2008. "International Comparison and Trends in Social Mobility," (in Japanese) *Koza Shakaigaku* (Sociology of Social Stratification). Tokyo: University of Tokyo Press.
- Hiroshi Ishida. 2007. "Japan: Educational Expansion and Inequality in Access to Higher Education," in *Stratification in Higher Education: A Comparative Study*, edited by Yossi Shavit, Richard Arum, and Adam Gamoran (Stanford University Press), pp. 63-86.
- Hiroshi Ishida. 2007. "Changes and Prospects of Career Formation during the Stage of Transition from School to Work," (in Japanese) in *Development of Occupational Competency and Adjustment of Educational Training Foundation in Japan*, edited by the Japan Institute for Labour Policy and Training, (Japan Institute for Labour Policy and Training), pp. 38-75.
- Hiroshi Ishida. 2006. "The Persistence of Social Inequality in Postwar Japan," *Social Science Japan* 35 (October), pp. 7-10.
- Hiroshi Ishida. 2006. "Inequality in Health," (in Japanese) in *Inequality in a Changing Society*:

- Hidden Disparities Behind the Demographic Shift in Japan*, edited by Sawako Shirahase (University of Tokyo Press), pp. 137-163.
- Hiroshi Ishida. 2006. "Background of Educational Choices among High School Students and High School Student Survey," (in Japanese) in *Educational and Career Choices and Attitudes Changes among High School Students*, edited by Hiroshi Ishida (Institute of Social Sciences, University of Tokyo).
 - Hiroshi Ishida. 2006. "Transition from School to Workplace," (in Japanese) in *Career Analysis of Contemporary Japanese from Diverging Perspectives*, edited by the Japan Institute for Labour Policy and Training, (Japan Institute for Labour Policy and Training), pp.9-38.
 - Hiroshi Ishida. 2005. "Education: The Transition from School to Work," (in Japanese) in *Contemporary Japanese Enterprise, Volume 2*, edited by Akira Kudo, Takeo Kikkawa, and Glenn D. Hook (Yuhikaku Publishing), pp. 208-234.
 - Hiroshi Ishida. 2005. "Perspectives for Career Formation among the Youth: Lessons from the Panel Survey," (in Japanese) *Business Labor Trend* 11 (November), pp. 5-7.
 - Hiroshi Ishida. 2005. "Stratification and the Labor Market in the Stage of Extended Adolescence," (in Japanese) *Journal of Educational Sociology* 76 (May): 41-57.

Book Reviews

- Hiroshi Ishida. 2009. "Book Review of Uchinawareta Bawo Sagashite (Lost in Transition), by Mary C. Brinton," *Nihon Keizai Shinbun (Japan Economic Newspaper)*, January 25.
- Hiroshi Ishida. 2008. "Book Review of Kyariya Rada toha Nanika (Moving Up in the New Economy), by Joan Fitzgerald," *Nihon Keizai Shinbun (Japan Economic Newspaper)*, November 9.
- Hiroshi Ishida. 2007. "Book Review of Social Mobility in Europe edited by Richard Breen," *European Sociological Review* 23 (Number 3, July): 405-407.
- Hiroshi Ishida. 2006. "Book Review of Changes in the Labor Market of High School Graduates and the Discrepancy between Structure and Cognition in High School Guidance, by Miki Tsutsui," (in Japanese) *Japanese Journal of Labour Studies* 48 (Number 10, October): 91-94.
- Hiroshi Ishida. 2006. "Book Review of The Transformation of British Life, 1950-2000, by Andrew Rosen," (in Japanese) *Riron to Hoho (Sociological Theory and Methods)* 21 (2): 343-344.

Research Experience

- Research Project on "Japanese Life Course Panel Survey," funded by the Grant-in-Aid for Scientific Research (S), Japan Society for Promotion of Science (principle investigator), Institute of Social Sciences, University of Tokyo, 2006-present
- Research Project on "A Cross-national Study of Socio-Economic Differentials in Low Fertility Aging Societies," funded by the Grant-in-Aid for Scientific Research, Japanese Ministry of Health, Labor, and Welfare (co-investigator) Graduate School of Sociology and Humanities, University of Tokyo, 2004-2007
- Research Project on "Employment Behavior and Attitudes Among the Youth and Their Relationship

- to Low-Birth-Rate/Aging Society,” funded by the Grant-in-Aid for Scientific Research, Japanese Ministry of Health, Labor, and Welfare (co-investigator), Institute of Social Sciences, University of Tokyo, 2004-2007
- Research Project on “Comprehensive Study of the Structure and Process of Japanese Stratification System,” funded by the Grant-in-Aid for Scientific Research, Japanese Ministry of Education (co-investigator), Faculty of Humanities, Tohoku University, 2002-present
 - Research Project on “White-collar Work and Organization,” funded by the Grant-in-Aid for Scientific Research, Japanese Ministry of Education (co-investigator), Institute of Social Sciences, University of Tokyo, 2000-present
 - Research Project on “Changes in the Transition from High School to Work,” funded by the Grant-in-Aid for Scientific Research, Japanese Ministry of Education, Science and Technology (co-investigator), Institute of Social Sciences, University of Tokyo, 2000-present

Participation in Conferences, Workshops and Invited Lectures (selected since 2007)

- “Youth in Transition in Japan: Inequality and Diversity among Japanese Youth,” paper presented at the Inequality Conference, Chung-Ang University, South Korea, July 2009
- “School, Discipline, and Achievement in Japan,” (with Satoshi Miwa) paper presented at the meeting of Comparative School Discipline Project, Milan, Italy, June 2009
- “Social Inequality in Health in Japan,” paper presented at International Sociological Association, Research Committee on Social Stratification, Renmin University, Beijing, China, May 2009
- “Class Origin and Early Career Progressions among the Youth,” International Sociological Association, Research Committee on Social Stratification, Florence, Italy, May 2008
- “Social Mobility among Women in Japan” (with Satoshi Miwa) presentation at the meeting of Social Stratification and Social Mobility in Late-Industrializing Nations Project, La Pietra International Conference, New York University, Florence, Italy, May 2008
- “Comparative Social Mobility among Early and Late-Industrializing Nations” (with Satoshi Miwa) presentation at the meeting of Social Stratification and Social Mobility in Late-Industrializing Nations Project, La Pietra International Conference, New York University, Florence, Italy, May 2008
- “Trends in Class Mobility and Education,” (with Satoshi Miwa) paper presented at the workshop on Social Mobility in Late-industrializing Nations, Masaryk University, Brno, Czech Republic, May 2007
- “Trends in Intergenerational Class Mobility in Japan in the late 20th Century,” (with Satoshi Miwa) paper presented at the meeting of the International Sociological Association, Research Committee on Social Stratification, Masaryk University, Brno, Czech Republic, May 2007

Professional Activities

Social Science Japan Journal, associate editor, 1997-2005, editor-in-chief, 2005- present
Journal of Political Science and Sociology, editorial board member, 2005- present
British Journal of Sociology, international advisory board member, 2001- present

International Sociology, editorial board member, 2001-2004
European Sociological Review, editorial board member, 2000- present
International Journal of Japanese Sociology, editorial board member, 1998- present
Riron to Hoho (Sociological Theory and Methods), editorial board member, 1997-2001
American Journal of Sociology, consulting editor, 1996-98
Member of the American Sociological Association, 1981- present
Member of the Japanese Sociological Society, 1981- present
Member of the International Sociological Association (Research Committee on Social Stratification) 1985- present
Member of the British Sociological Association, 1985- present
Member of the Japanese Association for Mathematical Sociology, 1988 - present
Member of the Japanese Society of Educational Sociology, 1997- present
Member of the Japanese Society of Health and Medical Sociology, 2002- present

ABSTRACT

1. Roles and Significance of Data Archives
2. Outline of SSJDA (Social Science Japan Data Archive)
3. Activities concerning SSJDA
4. Major Collections of Survey Data Available on SSJDA and Major Depositors
5. Data Sets Available on SSJDA and Number of Data Utilization Requests
6. Center for Social Research and Data Archives
7. Further Challenges

Launching of Educational Data Archive “JEDI” and the Networking of Academic Data Archives

YASUO WATANABE

AFFLICTION

Professor, Graduate School of Education, and
Director, Educational and Social Survey Research Center,
Hyogo University of Teacher Education



FIELD OF SPECIALIZATION

- Organization theory, network theory and analysis, economic sociology, social system theory and sociological theory
- Development of social organization theory and empirical study focusing on the correlations between four types of structural reasoning and network patterns in organization

BIOGRAPHY

- 1955: Born in Nagoya
- 1979: Graduated from Waseda University, School of Law
- 1993: Joined Ph.D. program of the Department of Sociology at the Graduate School of Arts and Sciences, Columbia University
- 1998: Completed the program
- 2004: Earned a Ph.D. degree in sociology from the Graduate School of Arts and Sciences, Columbia University

PROFESSIONAL CARRER

- 1979: Joined the Tokai Bank, Ltd. Worked at the Bakurocho branch, Yurakucho branch and International Business Department (Tokyo) of International Headquarters, and worked at the New York branch after international finance training in London and Hong Kong.
- 1987: Transferred to the Tokai Trust Company of New York, and served as Vice President, Manager, and Assistant Treasurer.
- 1993: Retired from the Tokai Bank, Ltd. after serving as Researcher in International Planning Department of International Headquarters.
- 2007: Served as Professor and Assistant Director, Educational and Social Survey Research Center, Hyogo University of Teacher Education.
- 2009: Promoted to Director at the Educational and Social Survey Research Center.

PUBLICATION

2004: “Why Do Organizations Form Groups? Complex Structure and Behaviors of Japanese Business Groups from 1977 to 1998,” pp. 396, Ph.D. Dissertation, Columbia University.

1988: Joint author of “A Quick Introduction to Foreign Currency Deposits and Impact Loans (New Edition)” (in Japanese), Banking Education Co., Ltd.

ABSTRACT

1. Introduction: Objective and Basic Policies

- Objective of the Educational and Social Survey Research Center (ESSRC): *“Setting up and running a worldwide data archiving organization to play a pivotal role in the field of research on education in the age of globalization and information technology”*
- Basic policies of JEDI:
 - ① To handle the storage and disclosure of valuable data on behalf of data depositors.
 - ② To make data as freely accessible as possible.
→ Maximizing the ease of data depositing and the accessibility of deposited data

2. “JEDI system” for the Educational Data-archive

- The JEDI system is based on DSpace, which was jointly developed by the Massachusetts Institute of Technology and Hewlett-Packard Development Company, L.P.
- DSpace is freely available as open-source software.
- The JEDI system has been introduced by many universities, libraries and research institutions in Japan.
- DSpace is the global standard software used by many institutional data repositories.
- The software is easy to use and supports standardized control and management practices.
- DSpace archiving systems are expected to remain in use for a long time.

3. Free Distribution of the JEDI Systems and the Networking of Archives

- Hyogo University of Teacher Education (HUTE) is ready to distribute the JEDI system free of charge, complete with related documentation such as agreements, protocols and procedure descriptions.
- HUTE proposes the establishment of the Japan Academic Data-archive Network (tentative name).

4. Features of the JEDI System: Four Types of Access Controls

- Level 1: Open data
 - Level 2: Data requiring user-verification before access
 - Level 3: Data requiring approval for access
 - Level 4: Data under surveillance by archive staff in a strictly controlled environment
 - Undisclosed: unavailable up to a specified date or for an indefinite period
-
- The diagram illustrates the relationship between access levels and their characteristics. On the left, a list of five access levels is shown. Brackets on the right group these levels into two categories: 'Online' and 'Offline'. The 'Online' category includes Level 1 and Level 2. The 'Offline' category includes Level 3, Level 4, and Undisclosed. Further brackets on the right group these into two copyability categories: 'Can be copied' (which includes Online and Level 3) and 'Cannot be copied' (which includes Level 4 and Undisclosed).

5. Other Features of the JEDI System

- Simple Google-like portal window
- Search operation facilitated by tagging and a metadata scheme specifically designed for cataloging educational data
- Support of R, freely distributed statistical analysis software (this year)
- Preparation of help screens and support for major languages in addition to English (in the future)
- Addition of “shopping cart” and “my favorites” features (in the future)
- Periodical system vulnerability testing by experts

6. Three Categories of Collected Data and Their Secondary Use

- Three categories of collected data:
 - ① Researchers’ data: Data from surveys conducted by individual researchers, reported at academic conferences and published in academic journals. Data in this category are handicapped by the limited scale and scope of surveys, but excel in the robustness of theoretical and analytical frameworks.
 - ② Flagship data: Data systematically collected by institutional efforts (e.g. census data). Ideal for secondary use because they come from major continuous survey programs.
 - ③ Teachers’ data: Data collected by teachers for practical purposes (i.e., identifying problems, achievements in the classroom). Data in this category are handicapped by the limited scale and scope of surveys, but are valuable because they contain detailed information on each pupil.
- Significance and possibility of secondary use and the expected roles of data archives:
 - ① Flagship data: Data archives should support supplemental researches and provide data merging services.
 - ② Researchers’ data: Data archives should support follow-up surveys and help solve problems of meta-analysis and multi-population analysis.
 - ③ Teachers’ data: Data archives shall make available data that can be used as training materials for students to learn about survey techniques and statistical analysis, or as reference materials for students to learn about survey design when planning a survey as part of their thesis.
 - ④ Data archives lower the risks of data loss by saving duplicates in geographically separated locations.

7. Conclusion: the Future of the Educational Data Archive

- Request for data depositing:
 - ① Flagship data: Administrative agencies and education support industries are invited to deposit their survey data in the data archive:
 - They may serve as the core to encourage others to deposit various types of data, and also may strongly motivate secondary use.
 - For data that must remain undisclosed for the time being, owners are invited to use JEDI like a safety deposit box at a bank.
 - ② Researchers’ data: Individual researchers and academic societies are invited to deposit their survey data in the data archive or encourage such depositing as a policy:
 - Researchers are invited to deposit their survey data to publicize the

results of surveys conducted using public funds, aided by a scientific research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology, for example.

- Academic societies are encouraged to make the depositing and disclosure of survey data as a condition for publishing papers in their journal.
- ③ Teachers' data: Teachers, as well as graduates and undergraduates, are invited to deposit their survey data to JEDI:
 - To stimulate the “evidence-based” arguments on the subject of education referring the results of empirical studies
 - To help children, teachers and researchers in the future

The significance of data preservation and dissemination in the area of longitudinal cohort studies

.....

MASUO KOYASU

AFFILIATION

Professor, Graduate School of Education, Kyoto University
President, Japan Society of Developmental Psychology



FIELDS OF SPECIALIZATION

Developmental psychology

BIOGRAPHY

Masuo Koyasu was born in Kyoto, in 1950. He graduated from the Faculty of Education of Kyoto University (majoring in educational psychology) in 1973 and studied at Kyoto University's Graduate School of Education until leaving the school partway through the doctorate course in 1977. After serving as assistant professor and associate professor at Aichi University of Education, he has been working at Faculty of Education, Kyoto University, as associate professor since 1988 and as professor since 1997. Masuo Koyasu has a PhD in psychology. He currently specializes in "cognitive psychology in education", conducting researches on the development of cognitive ability in young childhood and adulthood. He is known for his recent publications and papers on the "theory of mind."

PUBLICATIONS / PAPERS / PRESENTATIONS

Lewis, C., Koyasu, M., Oh, S., Ogawa, A., Short, B. & Huang, Z. (2009). Culture, executive function and social understanding. In Lewis, C. & Carpendale, J.I.M. (Eds.), *Social Interaction and the Development of Executive Function* (pp.69-85). Monograph in the series New Directions in Child and Adolescent Development, Issue 123. San Francisco: Jossey Bass.

Ando, H., & Koyasu, M. (2008). Differences between acting as if one is experiencing pain and acting as if one is pretending to have pain among actors at three expertise levels. Itakura, S., & Fujita, K. (Eds.), *Origins of social mind: Evolutionary and developmental views* (pp.123-140). Tokyo: Springer.

Mizokawa, A., & Koyasu, M. (2007). Young children's understanding of another's apparent crying and

its relationship to theory of mind. *Psychologia*, **50**, 291-307.

Manalo, E., Koyasu, M., Hashimoto, K., & Miyauchi, T. (2006). Factors that impact on the academic motivation of the Japanese University students in Japan and in New Zealand. *Psychologia*, **49**, 114-131.

Cox, M.V., Koyasu, M., Hiranuma, H., & Perara, J. (2001). Children's human figure drawings in the UK and Japan: The effects of age, sex, and culture. *British Journal of Developmental Psychology*, **19**, 275-292.

Furnham, A., Ota, H., Hosoe, T., & Koyasu, M. (2000). Beliefs about overcoming psychological problems among British and Japanese students. *Journal of Social Psychology*, **140**, 63-74.

ABSTRACT

1. Intent of the Topic Presentation

- Emphasis on the following message concerning the purpose of the project: *“In pursuing the educational data archive project, our aim is to gather and save the valuable survey data concerning education that would otherwise be scattered, lost, or obsolete, to make these data available on the Internet and elsewhere in properly compiled or edited formats, and to encourage their secondary use for empirical research.”*
- Summary of request: The topic presenter wishes to speak for 10 minutes on topics related to survey data concerning education, data archives, academic researches and academic societies.
- Suggestion from the standpoint of developmental psychology: emphasis on issues that concern longitudinal cohort studies.

2. Archiving of Data from Longitudinal Cohort Studies

- The term “longitudinal cohort study” refers to a research study that involves long-term tracking of the target population. Such a project, for example, may involve the implementation of a certain educational, training or treatment program as a part of the project, followed by monthly or annual surveys on how the benefits from the program may change after the program ends. More broadly, the same term may refer to study projects that involve the long-term tracking of developmental changes manifested by the target population with or without the implementation of a certain program as a part of the project. If the target population shares the same birth period or experience of a particular social event, the researches fall into the category of cohort studies.

- *Examples of cohort studies in the world (1) - Canada:*

In Canada, two longitudinal study projects are in progress. The National Longitudinal Study of Children and Youth, which started in 1994, targets 22,831 children (born 1983-1994, divided into 10 cohort groups), and involves a survey once every two years. The Québec Longitudinal Study of Child Development, which started in 1998, targets about 2,000 children.

- *Examples of cohort studies in the world (2) - Great Britain:*

In Great Britain, the UK Millennium Cohort Study was started in 2000 by the Longitudinal Study Center of the Institute of Education, University of London. The project targets 18,818 children in 398 localities of Great Britain, and involves the tracking of these children at nine months from birth and at the ages of three, five and seven.

- *Examples of cohort studies in the world (3) - Australia:*

- Data collection activities for the Longitudinal Study of Australian Children started in 2003. The project targets about 5,000 infants of up to 1 year old and about 5,000 young children at the age of four or five, and involves a survey of these children once every two years.

- *Examples of cohort studies in the world (4) - “Study of Broadcasting Programs Good for Children” by the NHK Broadcasting Culture Research Institute*

The NHK (Japan Broadcasting Corporation) Broadcasting Culture Research Institute started in 2002 a 12-year project for studying broadcasting programs good for children. The target population was selected as follows: out of about 6,000 infants born in Kawasaki City, Kanagawa Prefecture, in the six months from February to July, 2002, about 1,600 infants were chosen at random, out of which the infants in about 1,368 households that agreed with the project were finally selected as targets. The project will run for 12 years, and involves annual questionnaire surveys (by mail), usually in January. The topic presenter is involved in the project.

3. Need for the Disclosure of Data Archives

The importance of disclosing data archives, containing data from completed study projects, is as follows:

- Data verifiability: Data should be made available for reanalysis by researchers other than the main researchers.
- Data accessibility: The main researchers alone will not be able to study the enormous amount of data in full detail.
- Historical importance of data: The archives contain data that would have remained unavailable if they had not been collected at the time they were.
- Making research results available to research sponsors: Researches are conducted using public funds (e.g. taxes). Research results, therefore, should be disclosed to the research sponsors (e.g. the public).

Using Data to Formulate Education Policies

–With reference to nationwide surveys of academic ability and learning context–
.....

TAKANORI MIYAKE



AFFILIATION

Senior Specialist,
Office for Assessment of Academic Ability,
School Management Support Division,
Elementary and Secondary Education Bureau,
Ministry of Education,
Culture, Sports, Science and Technology

RESPONSIBILITIES

Responsible for all aspects of nationwide surveys of academic ability and learning context, including planning, implementation, analysis, and utilization of the surveys.

BIOGRAPHY

February 1979	Born in Kanagawa
March 2001	Graduated from the School of Science, Tokyo Institute of Technology
March 2003	Completed master's course at the Graduate School of Information Science and Technology, The University of Tokyo

PROFESSIONAL CAREER

April 2003	Joined the Ministry of Education, Culture, Sports, Science and Technology
April 2006	Unit Chief, Office for Coordination and Planning, Research and Coordination Division, Science and Technology Policy Bureau, Ministry of Education, Culture, Sports, Science and Technology – Japan
May 2007	Research Officer, The Economic and Social Research Institute, Cabinet Office, Government of Japan

July 2009

Senior Specialist,
Office for Assessment of Academic Ability,
School Management Support Division,
Elementary and Secondary Education Bureau,
Ministry of Education, Culture, Sports, Science and Technology

ABSTRACT

The current circumstances of acquiring and using data to formulate education policies and using knowledge in the policy deliberation process will be presented, then the problems and improvement measures will be discussed with reference to nationwide surveys of academic ability and learning context.

Data Archives and Educational Industry

.....

KENICHI ARAI



AFFILIATION

Corporate Senior Vice President, Benesse Corporation
Director, Benesse Educational Research and Development Center
Chair, Center for Research on Educational Testing, a nonprofit organization (NPO)

WORK RESPONSIBILITIES

- Research and Development
- Management

BIOGRAPHY

1956 Born in Saitama
Graduated from Tokyo Gakugei University

PROFESSIONAL CAREER

- 2003 Joined Benesse Corporation
- 2004 Corporate Senior Vice President of Benesse Corporation, and Director of Education Research & Development Headquarters
- 2005 Also appointed Director of Benesse Educational Research and Development Center when it was established
- 2007 Established the Center for Research on Educational Testing (NPO), and serves as Chair

ABSTRACT

Expectations and issues for data archives from the viewpoint of a private educational research institute

情報交換会

