The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors:
The QUARTERLY is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Manuscripts should be sent to Editor: Karsten Boye Rasmussen.

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers, Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines", Journal of the American Society for Information Science 30(2):77-82, March 1979. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Editor
Karsten Boye Rasmussen,
Department of Organization and Management,
University of Southern Denmark, SDU,
Campusvej 55, DK-5230 Odense M, Denmark
Phone: +45 6550 2115
Email: kbr@sam.sdu.dk

Production
William Block,
Minnesota Population Center,
University of Minnesota,
537 Heller Hall
271 19th Avenue South.
Minneapolis, MN 55455.
Phone: 612-624-7091
Email: block@socsci.umn.edu

Walter Plovesan
Maps/Data/GIS Library,
Simon Fraser University,
Burnaby, B.C.
Canada V5A 1S6.
Phone: (604) 291-5869.
Email: walter@sfu.ca

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 © 1997 by IASSIST. All rights reserved.
Welcome to the first issue of the IASSIST Quarterly vol. 29.

This issue contains the winner of a paper contest set up by IASSIST. Lisa Neidert from the University of Michigan wrote about the winning paper:

**Announcement: Strategic Plan Publication Award**

The review committee for the IASSIST Strategic Plan Publication Award is happy to announce a winner for this year’s competition.

The winner is Margaret Law of the University of Alberta for her paper “Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data,” which explores issues in the long-term access to data.

She discusses the costs and benefits in the secondary use of data. Her paper is based on the Canadian case. Each country in our international community of data professionals has different legal issues related to confidentiality, privacy, and ethics of data use.

The careful description of the Canadian case allows IASSIST to develop specific and more general points of advocacy on data access.

The committee was very pleased with the submissions to this competition.

The papers focused on access, documentation, and advocacy. All of the papers were excellent and we look forward to publishing them in future issues of IQ.

Lisa Neidert,
Chair, IASSIST Strategic Plan Publication Award committee

Our lead article is the paper by Margaret Law that you might have heard about earlier. Other good papers being input to the competition have been appearing in earlier issues and will be appearing in coming issues of the IQ. And “thank you” to the Publication Award committee for their work.

A trio of researchers John Adams, Hasan Al-Madffai, and Ray Thomas from Napier University, University of Glamorgan, and The Open University all in UK are here presenting a paper on “The Production and Presentation of Statistics of Unemployment: Comparability Issues”. Even though The United Nations publishes unemployment statistics for 123 countries and most of these statistics are based on International Labour Office criteria, many countries also produce unemployment statistics based on insurance records and on the basis of registered unemployment as well as sample social surveys. “This paper aims to compare the main features of the different methods. The dimensions compared include the conceptual basis for the definition of unemployment, boundaries of employment and inactivity, entry statistics and duration of unemployment, use of denominators for production of unemployment rates, and the cultural influence of the statistics. The paper identifies conflicts between achieving international comparability and national needs”. The article summarizes a full report to be published. The paper was presented at the IASSIST conference 2005 in the session “Cross-national Socio-economic Data: Boundaries of Evidence”.

Chuck Humphrey is writing on “The Preservation of Research Data in a Postmodern Culture”. The postmodernism is the compilation of an individual packaged identity rather than receiving that from an institution like in organized religion and political parties. This view transfers to data archives as dissemination turns to be a more direct on-demand access to global data resources using the Internet. The other side of this coin is that some users are “self-archiving” which is an approach unable to address the life cycle of research data. Compared to the institutional approach the individual approach will be short termed. How about you readers, do you find the institutional data archives to be under pressure from “self-archiving”?

The IASSIST website is constantly evolving so remember to pay a virtual visit to iassistdata.org and the IASSIST weblog (blog) - IASSIST Communiqué – at http://iassistblog.org.

So please visit the IASSIST website at www.IASSISTdata.org. Among other things, you can find information on previous and coming conferences. The 2006 conference will take place at Ann Arbor, Michigan (23-26 May). Furthermore, on the IASSIST website you will find access to the articles of the IASSIST Quarterly as PDF-files.

Papers for the IASSIST Quarterly are most welcome. Papers can be from IASSIST conferences, from other conferences, from local presentation, discussion input, etc. Contact the editor via e-mail: kbr@sam.sdu.dk.

Karsten Boye Rasmussen, December 2005
Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data

Introduction

“Reduce, reuse, recycle”, a phrase familiar to us from the environmental movement, can also be used to reflect on the secondary use of research data. Secondary research refers to the use of research data to study a problem that was not the focus of the original data collection. This may be data collected for administrative, health or educational purposes, census data, or data collected as part of a previous study. This secondary analysis may involve the combination of one data set with another, address new questions or use new analytical methods for evaluation (Szabo & Strang, 1997).

Both benefits and dangers have been attributed to the secondary use of research data. Distinctions are often made between large scale data collections, particularly sample survey data collected at public expense, and smaller bodies of data collected at personal expense. There is general agreement that the first should be shared and made generally available in a “timely” fashion, but little agreement about the second. There is also no sense of agreement about what would constitute timely in this situation (Clubb, Austin, Geda, & Traugott, 1985).

While the questions surrounding the secondary use of research data have always existed, they have become more pressing with the use of new technologies. New capabilities include easier data sharing, faster and more complex analysis, and the development of large scale data banks. Previously, the ability of researchers to communicate was limited by time and distance; now data can be shared globally at the click of a mouse. As we adapt to the electronic environment there are new concerns about confidentiality and the threat of security lapses. The potential for finer data resolution becomes possible with better data collection tools and technological innovations. While the fundamental ethical issues have not changed, the possibilities created by new technologies have brought them to the forefront.

Just as governments have taken a strong leadership role in developing and supporting good environmental habits, they must be encouraged to develop and support good habits concerning the storage and use of research data.

This paper summarizes ethical concerns about the secondary use of data and the arguments for encouraging or facilitating it. It includes some potential solutions and discusses the implications of the increased use of new technology. While it focuses on the Canadian regulatory environment, similar issues arise in other countries.

Concerns about data sharing and data confidentiality affect researchers and data librarians across the world. While each country may have a different regulatory environment and a different research culture, the need to find an appropriate balance between the optimal use of data and the protection of individuals is worldwide. With increasing globalization, and an increase in international research, the development and articulation of appropriate guidelines becomes paramount.

Data sharing is a fundamental value for IASSIST, and individual or random decisions about data sharing stand in the way of providing the best support for researchers. By looking at the Canadian situation, data librarians may develop and share common messages as part of an overall advocacy plan to support data sharing. There must be limits, of course, to protect respondents, but these must be delineated and managed in a coherent way that not only recognizes their rights, but also those of researchers, and of the taxpayers who frequently fund the research. This advocacy effort must be aimed at regulatory bodies, funding agencies and the researchers themselves in order to change the cultural values around secondary use of research data.

In Canada, much research involving humans is governed by the three major granting councils, who have developed a shared policy statement to govern all research involving human participants done in Canada, and by Canadians outside of Canada. Section C3 of the Canadian Tri-Council Policy Statement (Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 1998) lays out guidelines for Research Ethics Board (REB) approval of research that proposes the secondary use of data. It is clear that REB approval is required if identifying information will be involved, but leaves it to the researcher and the REB to determine exactly what constitutes identifying data.
There is sufficient legislation in most provinces to provide a framework for researchers wishing to use data that has been collected for purposes other than research. For example, the Alberta Health Information Act (Health Information Act Chapter H-5, 2000) provides direction for researchers’ use of health care data. Division 3, ‘Disclosure for Research Purposes’, defines the role of the ethics committee, including consideration of whether the researcher should be required to obtain consent from subjects, implying that there is some choice in this matter. The ethics committee must also decide whether the research is of sufficient importance to outweigh privacy concerns, whether the researcher is qualified and whether there are sufficient safeguards.

The picture is not so clear, however, when the data was originally gathered for research purposes. This data is not governed by legislation, and guidelines are open to interpretation. There is considerable discussion about who owns the data and decisions about whether to share data are often made by the original researcher and may depend on a number of personal factors. A requirement for all researchers to consider the potential for secondary use of their data, either by themselves of by others, would contribute to a more orderly use of data with resulting benefits for researchers, subjects and the community.

Concerns about the secondary use of data
Concerns about secondary use of data generally focus on the potential for harm to the individual subjects of the research and the lack of informed consent. Many writers are passionate about the primacy of informed consent for any type of research involving human subjects. Consent applies not only to a particular researcher, but also for an identified purpose. To quote Kalman (1994), ‘the requirements to seek an individual’s consent to participate and to provide data for a specific purpose must take precedence.’ Since researchers generally are not able to predict potential requests for secondary use of data that they are collecting, they are unable to fully inform subjects of the primary research about potential future uses of data. As this full disclosure of information is one of the requirements of informed consent, it follows that it is not possible to get informed consent for unanticipated uses of data.

Others argue that if the second researcher were to contact subjects to ask consent to re-use data, the original researcher must first identify the individuals thereby breaching their privacy. This situation could be managed by having the original researcher contact the subjects on behalf of the second researcher. Privacy is generally defined as a personal issue, defined by the subject. The subject may have felt comfortable disclosing information to the first researcher because of their relationship or rapport, but secondary research could leave him open to actions of researchers with whom he feels less comfortable (Homan, 1992).

Technology-driven data analysis techniques also create the potential for triangulation of data: the combining of variables that allows identification of specific individuals and organizations even though identifying information was removed from the original data sets. For example, there has been concern that the combination of census data and geographic information can allow the identification of small or unique groups. The use of GIS allows for closer identification of geographic data through the availability of differing degrees of granularity (Trainor & Dougherty, 2000).

There is additional concern for vulnerable populations that could be at particular risk if their confidentiality were breached. Current North American legislation and the media have raised awareness about profiling issues, and certain populations such as those involved in criminal activities or who are HIV-positive have a high risk of harm if they are identified.

In addition, particular forms of data such as oral histories, photographs or diaries cannot be made anonymous because the identification of the respondent is a large part of the value of the data. Researchers in these situations often feel that they have given unqualified pledges of confidentiality to participants, leading them to bar access to the material unless participants can be contacted for permission (Hedrick, 1985). Considerable commitment from the first researcher would be needed to contact the subjects for get permission for them to be approached by a second researcher.

Ethical practice requires a balancing of benefits and harms when conducting research. While this may be assumed to refer to the benefits and harms that may be experienced by the subjects of the research, it could also be interpreted as requiring a consideration of the potential harm to the original researcher. The ‘design and execution of data collection effort is a creative activity that sometimes involves innovative techniques.’ It seems reasonable to question why a secondary analyst should benefit from someone else’s work, particularly if the second research is a potential scholarly competitor (Clubb et al., 1985).

High quality data are expensive to collect, organize and store in an accessible form. If the data is to be used by someone else at a later date, additional work and documentation may be required. If this is carried out at the original researcher’s expense, it would seem to create the potential for harm with no counterbalancing benefit. This is particularly true if the original researcher is not cited, as it could have a negative effect on tenure and future funding opportunities (Sieber, 1991).

Some writers have also proposed a negative effect on “good science”, brought about as a result of too much data-sharing
IASSIST Quarterly Spring

this is a situation where the whole is greater than the sum of knowledge that none could contribute alone. In essence, the ability to link data files and to create families of data of making such a causal relationship (Dale et al., 1988). For example, some of our understanding about the causes the original data collection (Dale, Arbor, & Proctor, 1988). Secondary analysis creates an opportunity to establish relationships that were entirely unpredictable at the time of gathering intrusions into their lives. One might argue that there is, in fact, the potential for greater benefit if research with already collected data provides more opportunities to support these groups. Taking steps to ensure that interpretations of data are valid through encouraging multiple methodologies demonstrates respect for research subjects through accurate interpretations of their behaviour.

Secondary analysis creates an opportunity to establish relationships that were entirely unpredictable at the time of the original data collection (Dale, Arbor, & Proctor, 1988). For example, some of our understanding about the causes of disease has occurred through the secondary analysis of medical records that were not collected with the intention of making such a causal relationship (Dale et al., 1988). The ability to link data files and to create families of data creates possibilities that together they can contribute knowledge that none could contribute alone. In essence, this is a situation where the whole is greater than the sum of the parts (Johnson & Sabourin, 2001).

If a second research repeats original calculations to assure accuracy, it is not considered to be a secondary analysis. If, however, it analyzes the data from a different perspective or within a different theoretical framework it allows the findings to be challenged and debated, and creates an opportunity both for further discovery and for a deeper understanding of the interpretations of the data (Dale et al., 1988). Developing and implementing protocols for data sharing creates potential for testing the generality of research findings, and comparing analyses on different data sets across time or across locations allows us to ‘generalize findings about social phenomena’ (Fienberg et al., 1985).

Good science requires that data be available for scrutiny and reanalysis as part of scientific enquiry (Fienberg et al., 1985). The practice of a second researcher reanalyzing data is widespread, although this would seem to pose the same concerns about breach of confidentiality as any other access to data by a second researcher. “It seems reasonable to argue that if one is prepared to publish assertions about the nature of reality based on collected data, then one should equally be prepared to allows others to examine data to check the validity of the assertions”(Reidpath & Allotey, 2001).

Concerns about privacy and confidentiality are the most frequently raised objections to secondary use of research data. Some writers believe that it is “likely that this obstacle is cited much more frequently than is warranted” (Hedrick, 1985 p.142). Attempts to quantify the risk of identification, particularly from anonymized records (Marsh et al., 1991) support to some extent the notion that the risk is over-stated. Other writers assert that achieving informed consent for secondary research is never truly voluntary as there are pressures on subjects to agree simply because they have already agreed once before, but this appears to not have been adequately investigated.

The issue is further confused by a discussion of what is meant when the original researcher states on the consent form that personal data will not be shared. Some researchers would interpret this in the most conservative way to mean that none of the data will be shared, ever. A more sensible interpretation might be that data can be shared as long as it is properly anonymized and all identifying characteristics are removed (Johnson & Sabourin, 2001). While the real question is how the subject interprets it, not the researcher, the subject is likely influenced by the researcher’s position. This is an ethical position that must be resolved before it can be managed through improved methodology.

The Canadian Tri-Council Policy Statement allows for a breach of confidentiality in section 3.3 (c) if the individuals to whom the data refer have not objected to secondary
use. The Research Ethics Board is charged with the responsibility of evaluating the sensitivity of information, seeking consent to used the stored data, and allowing the researcher to propose an appropriate strategy. The REB is directed to pay particular attention to the possibility of “harm or stigma” that might be attached to identification. While this obviously does not preclude the secondary analysis of research data, it clearly does not take a strong position in favor of it.

The sharing of research data must also be considered under the ethical principle of balancing harms and benefits. In many situations, the original data collection was paid for through research grants, funded by the taxpayer. This ‘harm’ to the community of taxpayers should be balanced by an appropriate benefit; the most logical way of maximizing that benefit is to ensure that the optimal use is made of all data collected. The allocation of harm and benefit in this case needs to be extended to include all of the participants in the research process, not just the immediate subjects of each piece of research. To quote Davey Smith (1994), “data paid for by public money are public property.” The additional analysis of data also provides the benefit of increased confidence in the outcomes of research to the larger community.

“Publishing the findings of research in peer reviewed journals implies a high level of confidence by the authors in the veracity of their interpretation. Therefore it stands to reason that researchers should be prepared to share their raw data with other researchers, so that others may enjoy the same level of confidence in the findings” (Reidpath & Allotey, 2001).

The ethical principle of reducing harm can also be viewed as support for the secondary use of data. Subsequent use of data already collected reduces the impact on the larger population by involving a smaller number of research subjects and subjecting them to a smaller number of tests. On a more practical level, the use of previous studies can help formulate a good research question and refine the analysis carried out in subsequent studies (Davey Smith, 1994). Good methodology is one of the primary mechanisms for reducing harm.

The Tri-Council Policy Statement articulates the maximization of benefit as a guiding principle. This strongly supports the secondary use of data as a cost-effective and convenient mechanism for the advancement of knowledge. As research money becomes more restricted, increased secondary analysis will allow for ongoing research in situations where new data collection is hampered by lack of resources (Szabo & Strang, 1997). In situations where research will have a significant impact, for example in influencing public policy, it is essential that data be considered from many directions to reduce the possibility of flawed or weak conclusions. Hedrick (1985) states that secondary analysis allows for the “reinforcement of open scientific inquiry” (p.127) by providing for evaluation of research and the opportunity to replicate or reanalyze it using the same or different methods. A critical process will increase public confidence in the value of research and reduce the incidence of faked and inaccurate results. Increased public confidence may also benefit the research community by providing support for research funding.

The sharing of research data is a logical process that maximizes the benefits of research while reducing much of the potential for harm. Many of the anticipated risks and harms can be managed through improved methodologies. Once this position is understood and widely shared, those solutions will become part of the research ethos.

Solutions for anticipated risks

A number of writers have proposed solutions for the anticipated risks stated by individuals who are not in favor of secondary use of research data. While the list below is not complete, it demonstrates the breadth and ingenuity of researchers who are committed to good science and maximum benefit to the community.

A number of the solutions focus on the requirements for confidentiality from the secondary researcher. For example, Clubb, Austin et al. (1985) recommend “a form of licensing or swearing in as a condition for access to data with the possibility of legal sanctions and penalties for breaches of confidentiality” (p. 62). The British Sociological Association, cited in Heaton (1998) recommends that researchers consider obtaining consent that at least “covers the possibility of secondary analysis.”

A number of approaches to the original consent form have been proposed. In some cases the original consent form includes provision for secondary research with the requirement that the secondary study receives approval from an ethics review committee. At the very least, this raises the question of potential secondary use in the minds of both the researcher and the subject, and allows respondents the opportunity to object should they wish. While this may not strictly meet the requirement for informed consent, it demonstrates an effort to resolve the situation early in the research process. It assumes that the secondary analyst is “bound by the same confidentiality and privacy restrictions as the primary analysts”(Szabo & Strang, 1997 p.7).

Better anonymization can be built in by the original researcher as a required part of research ethics approval for gathering data concerning humans. Proposed methods include a uniform practice of removing names and substituting numeric codes, removing occasional data values that reflect rare attributes and could allow for identification of specific individuals and organizations,
aggregating data in such a way that the performance of
identifiable individuals or organizations is not obtainable,
and various forms of encryption (Clubb et al., 1985;
Johnson & Sabourin, 2001). This requires a better
understanding of which identifying items data need to be
maintained to keep the data useful while protecting the
respondents. While it can be argued that some forms of
data such as photographs or diaries have little value without
identifying information, these should be regarded as
exceptions rather than the norm and general policy should
not be based on them.

A mathematical solution has been proposed that adds
enough uncertainty to statistical analysis to prevent
the identification of individuals while not significantly
affecting the outcome of the analysis. The process, known
as “jittering”, is defined by Johnson and Sabourin (2001)
as “adding a small, normally distributed random value with
a mean of zero to all fields that might be used to identify
an individual by matching against publicly accessible
records.”

The real problem may not be a lack of potential solutions,
but a reluctance to implement them. This could be
encouraged through a number of means outlined by Sieber
(1991) in support of secondary research:

· In appealing to enlightened self interest grant bodies
could require willingness to share data as funding crite-
ron. This is already required by some funding bodies
(Davey Smith, 1994). If not a requirement, funding
priority could be given to those who create and share im-
portant data files and to research which builds on upon
existing data files. Note that this requirement only works
if it is monitored and there is a mechanism for sharing.

· To minimize the potential harm to researchers through
not having their work adequately cited, the research
community could require the implementation of clear
and enforced standards for citation of data files. Stan-
dards for authorship should include identification of the
source of data.

· In order to reduce researcher fears about secondary
use of data, research education should be enhanced to
include improved understanding about the advantages,
process, and barriers in data sharing.

Funding agency policy statements could be a stronger
advocate for secondary use of research data by including
further instructions for the original researchers. It should
work from the assumption that data sharing is standard
practice unless there are specific reasons for prohibiting
secondary analysis. It would then include, for example,
the requirement for the secondary analyst to properly
recognize the original researcher and a clear stipulation of
the conditions under which data sharing is prohibited. This
would facilitate good science by removing the potential
conflict of interest that occurs when a researcher must
decide whether or not to share data.

Research ethics approval could require a process for
coding, storing and providing access to the data in a
uniform fashion. This could be accomplished by adding an
information professional, such as a librarian or an archivist,
to the research team (Humphrey et al, 2000). A uniform
requirement would mean that the burden of this additional
work was evenly spread among researchers and would be
considered as part of the original research design.

This has primarily been a discussion of the ethical issues
surrounding data sharing. There is also the practical
consideration of whether researchers are prepared
to voluntarily share their data, or whether they have
maintained it in a form that allows it to be used by other
people. (Corti, Foster, & Thompson, 1995; Reidpath &
Allotey, 2001) At this time, to share or not is still largely an
ad hoc decision made by individual researchers. A clearly
articulated policy that evaluated the situation on scientific
merit and an analysis of harms and benefits would ensure
that the ethical principles were the basis for decision-
making.

Conclusion
The secondary analysis of existing research data provides
many exciting opportunities for the development of new
knowledge. It can be aligned with the ethical principles of
research in many countries by minimizing the respondent
burden and maximizing the potential benefits from the
data. To make a change in the research culture requires
strong advocacy on the part of data librarians, to change
the thinking of funding bodies, regulatory agencies and
researcher.

Traditionally we have not required that the potential for
data sharing be a part of every research proposal. Now
technology has provided us the opportunity to ‘build the
corpus of knowledge, not through the frenzied winnowing
that has characterized our evaluations in the past but
through an orderly interlocking of the puzzle pieces
contributed by the disparate sub-fields. We have the means,
for the first time in our history, to begin putting together
the full picture of human behaviour’ (Johnson & Sabourin,
2001). It is important that we champion the changes
needed to accept this challenge, and to advocate for the
creation of an ethical basis that requires the development
and implementation of strategies to overcome potential
barriers to data sharing.
“Reduce, reuse, recycle”. This phrase has shaped a generation of behaviors about environmental concerns. Governments and funding agencies have promoted changed behaviour through investment in infrastructure, and in policy directions. The same thinking can be used to shape our understanding about ways of reducing the costs and burdens of data collection, increasing the value of research, and maximizing the benefits for everyone involved in the research process.

* Margaret Law, University of Alberta, margaret.law@ualberta.ca.

With this article Margaret Law won the IASSIST Strategic Plan Publication Award in 2005.

The author wishes to acknowledge the advice received from Dr. G.Griener, Department of Philosophy, University of Alberta.

References


Freedom of Information and Protection of Privacy Act, RSA 2000, c. F-25, sec. 42


Contributions To The Research Team. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1(3).


Abstract
The United Nations publishes unemployment statistics for 123 countries. Most of these statistics are based on International Labour Office (ILO) criteria for the definition of unemployment. Many countries also produce unemployment statistics based on insurance records and on the basis of registered unemployment.

This paper aims to compare the main features of the different methods. The dimensions compared include the conceptual basis for the definition of unemployment, boundaries of employment and inactivity, entry statistics and duration of unemployment, use of denominators for production of unemployment rates, and the cultural influence of the statistics.

The paper identifies conflicts between achieving international comparability and national needs. Survey statistics that underpin international comparisons do not support geographically detailed analysis within countries. The value of unemployment statistics based on ILO criteria is limited by a failure to recognise the concept of entry to unemployment and the difficulties of integration with administrative unemployment statistics. The standard labour force survey (LFS) questionnaire should be modified to support the production of statistics for entrants to unemployment. The sampling frame should be modified to ensure consistency with nationally produced unemployment statistics derived from administrative records.

Introduction – Three Conceptual Bases of Unemployment Statistics
Three types of systems – insured unemployment, registered unemployment, and unemployment measured by sample social surveys – provide the basis for most unemployment statistics. The first statistical series started in 1886 when the Board of Trade asked the trade unions to provide monthly statistics of the number of their members who were not in employment (Garside, 1980). This series led to the idea of insured unemployment. The UK and many other countries have offices that help people find work or pay benefits to those without work. Such systems provided the basis for statistics of registered unemployment. In the United States, concern about mass unemployment in the 1930s led the government to develop household surveys in order to measure the extent of unemployment (see Anderson, 1988).

The International Labour Office (ILO) gives details of registered unemployment systems for 75 countries (see http://laborsta.ilo.org/). Nearly all the countries of Eastern and Western Europe have insurance and/or registered unemployment systems. The US uses insurance based statistics to help make unemployment estimates at sub-national levels (see Section 5 below). But the system that has increasingly dominated in recent decades is the sample survey.

Since 1948 the monthly Current Population Survey (CPS) has been the dominant method of measuring unemployment in the US. The main focus of the CPS is employment and unemployment, and nowadays the CPS would be described as a labour force survey. The CPS defines unemployment in terms of the numbers seeking work. In the 1980s, when the time came for an international standard, the CPS provided a model. The 13th International Conference of Labour Statisticians in 1982 adopted the seeking-work criterion of the CPS presumably because it could be applied in any country independently of any existing national systems for dealing with unemployment. ILO criteria for conduct of labour force surveys and the definition of unemployment (Hussmans et al., 1990) count the numbers seeking work in almost exactly the same way as the CPS. The ILO provides details of labour force surveys conducted in 109 countries.

The standard labour force survey (LFS) does not use the word unemployment. The crucial question in the UK LFS, for example, is “Thinking of the 4 weeks ending on Sunday. Were you looking for any kind of paid work at any time in those four weeks?” By avoiding the term unemployment, the ILO criteria aim to produce statistics that are independent of national insurance and other systems that give benefits to the unemployed and therefore use the term ‘unemployment’ in a variety of different contexts. But this pursuit of independence makes comparison with other datasets difficult or impossible. ILO unemployment statistics for the UK, for example, are not comparable to UK claimant unemployment statistics.
The next section of the paper discusses the categorisation of unemployment – the boundaries between employment and unemployment, and between unemployment and inactivity. Section 3 focuses on entry to unemployment – important because entry to unemployment is not recognised by labour force measures of unemployment, but is demonstrably important in developing policies that go beyond seeing exits from unemployment as the exclusive solution to unemployment problems.

The choice of denominators is a key theme of sections 4, 5, and 6. Sections 4 and 6 refer to problems with the conventional economically-active-population denominators for measurement of unemployment rates at the national and local levels, and the main population alternatives. Section 5 identifies a flaw in the use of current level of unemployment as a denominator in the usual measure of long-term unemployment and introduces the idea of population at risk denominators as a superior alternative. Section 6 discusses problems associated with measuring unemployment on a local scale and Section 7 points to the increasing need for such measures.

Section 8 points out that the seeking-work criterion of the ILO definition of unemployment conditions users of unemployment statistics to view unemployment as a matter that belongs exclusively to the unemployed, although this runs against many cultural traditions. The section suggests modification of the ILO criteria for the definition of unemployment and modification of labour force survey sampling methods in ways that would add value to unemployment statistics at local and national levels without reducing international comparability.

**Fixing the Boundaries**

The definition of what is considered as employment is generous both in the CPS and the standard LFS. The crucial question in the CPS questionnaire is ‘LAST WEEK, did you do ANY work for pay or profit?’ (http://www.bls.gov/cps/cps_htgm.htm). The LFS questionnaire in the UK asks first ‘Did you do any paid work in the seven days ending Sunday as an employee or as self-employed?’, and later ‘(In) the seven days ending Sunday, how many hours did you actually work ...?’ These questions support the production of statistics of employment as defined by the ILO as paid work of one hour or more per week. The motivation for the generosity of this definition is the wish to link production to total labour input (Hussmanns et al., 1990, p 71), or in other words, to produce statistics of labour productivity in terms of output per person-hour rather than just output per person. It is unlikely that this differentiation is understood or accepted by most survey respondents.

The LFS in the UK, like the CPS, establishes employment with questions that elicit the amount of paid work of more than one hour. But in the 2001 UK census of population respondents were asked if they were in employment. The resulting census statistics gave an employment level of 640 thousand, or 2.5%, below that of the corresponding LFS estimate, and an unemployment level of 204 thousand, or 14%, above the LFS figure (Heap, 2005).

It seems unlikely that the line between employment and unemployment is drawn at this one hour boundary in most systems of insured or registered unemployment statistics. In the UK the rules specify that claimants for unemployment benefits cannot work for more than 16 hours per week. It can be expected that each insurance or registered unemployment system will have individual regulations on the matter.

A similar variety of regulations can be expected to apply at the other boundary. ILO criteria specify that respondents must be seeking work within the reference period. The search period supported by the Organisation for Economic Co-operation and Development (OECD) is four weeks. In the UK both LFS and the census ask about two weeks. Surveys in Japan and Taiwan ask only about one week. Those expecting to take up a specific job can also be classified as unemployed. The CPS in the US includes as unemployed those laid off from employment who are expecting to resume their former work.

ILO criteria specify that respondents must be available to take up work within two weeks in order to be classified as unemployed. But the conditions for eligibility for Job Seekers Allowance (JSA), the name given to claimant unemployment in the UK, is tougher. A number of welfare groups describe JSA as being “about hassling people off the dole into low paid work by making it tougher to sign on” (for example, Urban 75, undated). Being available for work to qualify for JSA means being ready to start permanent or temporary work immediately.

The standard LFS questionnaire supports a major conceptual extension of unemployment by including a question ‘... Even though you were not looking for work, ... would you like to have a regular paid job at the moment ...?’ This question in the UK LFS is addressed to respondents after they have been identified as economically inactive by questions that have established that they are not classified as in employment nor as unemployed. The question is also part of the standard LFS in Europe. Statistics for the number wanting work but not classified as unemployed are published by Eurostat.

There is a follow-up question on the main reason for not seeking work. In the UK LFS the pre-coded answers are given in this order: student; waiting to take up a job; looking after family; sick; believes no jobs available. The interpretation of statistics resulting from this question may be influenced by the ordering of these pre-codes and the precise instructions given to interviewers.
That last pre-code reminds us that the ILO definition is tight in that it excludes respondents who are not looking for work because they believe that no suitable jobs are available. This clash between the logic of the individual and that of the ILO criteria is honoured by describing such respondents as discouraged workers. (Hussmans et al., 1990, p 107-8). Statistics for discouraged workers are published by the OECD (see http://www1.oecd.org/scripts/cde/queryScreen.asp?), but are rarely subjected to detailed appraisal. According to the OECD there were more than two million discouraged workers in Japan in 2000.

**Entry and Short-duration Unemployment**

Statistics relating to the duration of unemployment commonly include figures for unemployment of up to four weeks. But such figures usually relate to uncompleted spells of unemployment. It does not include completed spells of unemployment of less than four weeks. The distribution by duration is subject to left-hand censoring or truncation.

This limitation is avoidable. It is not, however, avoided by the ILO criteria for labour force surveys, nor by the CPS questionnaire. Questions on unemployment are addressed only to those who are unemployed at the time the survey is conducted. CPS and labour force survey questionnaires establish the number of those who became unemployed during the previous four weeks, but do not include those who became unemployed during the previous four weeks but had exited from unemployment by the date of the survey. The sample is biased against being representative of the whole population of working age. Newly unemployed who have re-entered employment or become economically inactive within four weeks are not included.

Kiefer (1998) described the resultant statistics for duration of unemployment as subject to ‘length-biased sampling’.

Systems that record entry to unemployment, such as the UK system of claimant unemployment, can be used to produce the numbers exiting before four weeks. The Office for National Statistics (ONS) makes available such statistics extending back to 1983 through the Nomis database. Chart 1 (see pg 14) indicates that monthly exits before four weeks are small relative to the total stock of unemployment but represent a substantial proportion of monthly entrants.

Monthly exits before four weeks show strong seasonal variation, but 12-month moving averages range between 20-30% of entrants. That range is small relative to the variation in levels of unemployment in this period. Cross-section analysis also shows a relatively small variation in exits before four weeks. The coefficient of variation (CoV) among 659 parliamentary constituency areas (PCAs) in 2004 was only 15% compared with the CoV for unemployment rates of 54% (calculations by the authors).

The omission of unemployment of less than four weeks could be considered a matter of poor survey design. We could make an analogy with a hypothetical survey of incidence of the common cold. It is to be expected that a survey of the common cold would ask people when they last suffered from a cold in order to get information relevant to catching a cold. It is unlikely that the survey would be limited to those who had colds on the day the survey was conducted. It is unlikely that respondents would first be asked ‘Do you feel healthy?’, and if they answered ‘Yes’, discarded from the sample! But the standard labour force survey creates an analogous situation by addressing questions on unemployment only to those unemployed at the time of the survey.

The omission of unemployment of less than four weeks does not affect statistics for unemployment of more than a month’s duration. But the OECD regularly publishes statistics for member countries for the percentage of unemployment of less than a month. The statistics are footnoted with the misleading comment ‘These percentages only take into account those persons for whom the duration of unemployment is known’. In fact, LFS duration statistics are based on uncompleted spells of unemployment – the difference between the date previously worked or the date started seeking and the date the survey was conducted. The only statistics of duration collected by the LFS relate to periods longer than the specified period. Exit statistics are necessary to measure known duration and exit statistics by duration are not obtainable from LFS surveys.

Countries with systems of insured and registered unemployment can be expected to have records that support the production of statistics of entry to, and exit from, unemployment. The Bureau of Labor Statistics (BLS) in the US publishes weekly statistics for initial claims and continuing claims for insured unemployment. But the BLS web site does not include any breakdown of continuing claims by duration.

The BLS also produces monthly statistics for mass layoffs. The mass layoff numbers come from establishments which have at least 50 initial claims during a 5-week period. Extended mass layoff statistics, issued quarterly, relate to a subset of such establishments where employers indicate that 50 or more workers were separated from their jobs for at least 31 days.

European countries seem to make little use of administrative data on entry to unemployment. Some implications of the failure to recognise the concept of entry to unemployment can be illustrated with statistics for claimant unemployment in the UK. Chart 2 (see pg 15) illustrates that statistics for the 659 UK parliamentary constituency areas (PCAs) in 2004 show a 90% correlation between entry to unemployment and the unemployment rate. It could be said that the chart only demonstrates
Chart 1  Exits from unemployment before four weeks compared with entrants and stock - Great Britain

Monthly figures, and 12-monthly averages

% of monthly entrants

% of stock

Chart 2  Unemployment rates and entry rates in 2004 parliamentary constituency areas of the UK
Chart 3 Measures of long-term unemployment rate (LH scale) – and numbers unemployed less than a year (RH scale)
the obvious – that the main cause of unemployment is becoming unemployed – but the scale of geographical variation is remarkable and notable.

Statistics of entry to unemployment have not been widely used in the UK. Over the past decade government labour market policy in the UK has been focused almost exclusively on exits from unemployment. Labour market policy in the UK has been dominated by programmes and slogans such as ‘new deal’ and ‘welfare to work’ as if unemployment were solely a matter of the unemployed making themselves employable.

As a result, authorities know little about the causes of unemployment or the factors that are leading to growing inequality in the geographical distribution of unemployment. In light of the large variation between areas, it is not surprising that the emphasis on exits has been associated with an increase in inequality in the geographical distribution of unemployment (Adams and Thomas, 2005).

Characteristics of the Unemployed and Long-term Unemployment

Labour force surveys can be expected to provide profile information on the unemployed on the same basis as that for the employed so that it is possible to make comparisons between the unemployed and employed population. The CPS questionnaire also includes questions on the previous job, including: ‘Did you lose or quit that job, or was it a temporary job that ended?’ These questions support the production of statistics on six alternative reasons for unemployment: temporary layoff; permanent job losers; completed temporary jobs; job leavers; re-entrants; and new entrants. But guidelines for the standard LFS do not include questions that would elicit this information.

The information available from administrative systems can be expected to vary according to the nature of the system. The BLS does not publish details on insured unemployment except for those on Federal programmes – presumably because of variations between different state schemes. Claimant unemployment in the UK included occupation when unemployment is falling the percentage of long-term unemployment will increase if most of the slack is taken by those recently out of work. (Working Party, 1995, p 387-379)

Use of a population at risk denominator provides a straightforward solution to this problem. In the case of year-or-more unemployment the population at risk (PAR) is the number unemployed a year earlier. The number unemployed a year earlier are all at risk of being unemployed a year later, and no-one not unemployed a year earlier is at risk of being unemployed a year later. But recognition of the problems with LAPU has not prevented a generation of economists from seizing on LAPU statistics to assert that the long-term unemployed have become insulated from the labour market. Stephen Nickell, a member of the Bank of England Monetary Policy Committee, writes:

“long-term unemployed still form a substantial and important group … this has a significant macroeconomic impact because the long-term unemployed tend to lose skills and motivation as well as being discriminated against by employers. This weakens their attachment to the labour market... They become ineffective in holding down wage inflation and this leads to the impact of adverse shocks to the economy … (Nickel, 1999, page 23).

Use of a population at risk (PAR) denominator reveals that year-or-more unemployment is actually more sensitive to changes in the state of the labour market than PAR rates for less than year unemployment groups (Adams and Thomas, 2004 and 2005). But Stephen Nickell was misled by LAPU statistics.

Webster (forthcoming) found that LAPU lags total unemployment by six quarters. Chart 3 shows that LAPU statistics lag the PAR rate by up to two years. Chart 3 (see pg 16) demonstrates that the PAR rate for year-or-more unemployment moves parallel to the trend in unemployment for less than a year. The parallelism suggests that levels of less-than-year and year-or-more
unemployment are influenced by the same set of factors.

**Denominators for Unemployment Rates**

In 1886, when the Board of Trade in the UK asked trade unions for the number of their members who were unemployed they also asked for the total number of members. The number of members provided an obvious denominator to support unemployment rates that could be used to make comparisons over time and between industries. When the UK introduced compulsory unemployment insurance in 1911, the insured population provided an obvious denominator. Nowadays the standard denominator for unemployment rates is the number in employment plus the number unemployed – usually described as the economically active population.

Employment, unemployment, and inactivity are usually thought of as three alternative labour market states. But the use of the economically active population as a denominator for unemployment is not consistent with the way employment, activity, and inactivity rates are usually measured. Employment rates and activity rates are usually expressed as a percentage of the working age population or, of the population in the specific age group under consideration. Use of a common population denominator would support direct comparison of unemployment rates with employment, activity, and inactivity rates.

Use of the number of trade union members as a denominator in the 19th century could be justified on the grounds that it could be assumed that the major flows over time between employment and unemployment were accounted for by trade union members. A recession could be expected to reduce employment and increase unemployment. The ratio of unemployment to employment could be expected to be an appropriately sensitive monitor of changes in labour market conditions over time. But it cannot be so easily assumed in the 21st century that the dominant flows between employment and unemployment are limited to the economically active population. Flows between employment and inactivity, and between unemployment and inactivity, detract from the value of the unemployment rate as an economic indicator.

Not taking economic inactivity into account also limits the value of the conventional unemployment rate for comparisons between different areas, age groups, or social groups. It can be expected, for example, that the scale of unemployment is often correlated with the scale of economic inactivity (for example, Beatty et al., 1997 and 2000). Where this happens comparisons based on the conventional rate could be expected to systematically underestimate the differences in economic or social conditions in different regions.

A systematic relationship does not preclude a lot of individual variation, and in making comparisons between two particular areas or two particular groups, the conventional unemployment rate can be misleading. There is, for example, wide variation in activity rates for women between different countries, especially in the older age groups. In many cases it would give a false picture to make comparisons of conventional unemployment rates without taking into account the differences in activity rates.

The ILO acknowledges this problem in regard to youth unemployment where there is great variation in the scale of economic inactivity due to variation in the proportion classified as inactive because of training or full-time education and to cultural matters such as social expectations about women working outside the home. The ILO response has been to produce statistics entitled “Youth unemployment, share of youth unemployed to youth population”, or in other words, unemployment rates with population denominators – in this case the population aged 15–24 years.

Unemployment rates with population denominators reveal substantial differences between countries. Youth unemployment rates using the conventional economically active denominator are particularly high for a number of East European countries – Bulgaria, Poland, and Slovakia. But the impression given by these statistics is mitigated by expressing unemployment using a population denominator. It might be assumed that a substantial proportion of the youth in these countries are economically inactive because they are investing in human capital by remaining in the educational system.

One of the problems with measuring unemployment rates at a local scale is that statistics for the economically active population are not available; statistics for employment are usually produced only by place of employment. Statistics for residents in employment in local areas are available only from censuses or surveys. The following section discusses the solution adopted for claimant unemployment in the UK since 2003 – to use population of working age (PWA) denominators. There do not appear to have been any disadvantages with unemployment rates measured in this way except for lack of comparability with the conventional rates used at national and regional levels.

**Local Unemployment Statistics**

There is no contest at the local level between the quality of survey statistics on one side and insurance or registered unemployment statistics on the other. Sample size limits the accuracy of survey statistics for local areas. But administrative statistics from insurance and registered unemployment systems can be produced on a 100% basis. The United States combines CPS statistics with unemployment insurance statistics to produce unemployment rates for more local areas. The CPS sample is 60,000. According to Local Area Unemployment Statistics (LAUS) as displayed on the BLS website there are 31,792 series to query for. These include series
relating to states, counties, parts of cities divided by county boundaries, and minor civil divisions.

The coverage and rules of insurance schemes vary between states. Not all of those who become unemployed are eligible for insurance benefits. Benefits do not usually extend beyond 26 weeks and the average duration is about 16 weeks. The insured unemployment rate is typically about one third of the CPS rate. The Bureau of Labor Statistics, with state authorities, makes estimates of labour force, employment, and unemployment on the basis of the CPS, the Current Employment Survey (CES) and the unemployed insurance statistics.

The scale and sophistication of the estimation processes needed to produce estimates for local areas is formidable and impressive. The CES ‘place of work’ estimates are adjusted on the basis of commuting data to ‘place of residence’. Separate estimates are implied for those who have come to the end of their period of insured unemployment, and for entrants and re-entrants to unemployment who are not covered by the insurance system. There are integral seasonal adjustment programs. The statistics are controlled to state totals.

In the UK, statistics for claimant unemployment for local areas are publicly available in considerable detail through the Nomis database at the University of Durham (http://www.nomisweb.co.uk/). But these statistics are not reconcilable with those for ILO unemployment from the Labour Force Survey (LFS). The LFS includes questions on claimant unemployment. But the grossed up statistics from the LFS are typically about 20% below the level of the administrative count of claimants (Jenkins and Laux, 1999).

The Local Labour Force Survey in the UK has an enhanced sample in low population areas to increase geographical coverage. This supports the production of unemployment statistics based, for example, on parliamentary constituency areas (PCAs) – that have on average a population of working age of 43,000 within a fairly narrow range. But little reliability can be given to most of the unemployment statistics. In 2003 the level of unemployment in 40 PCAs was too low to support any estimate of the annual average, and it was not possible to give any confidence level to the estimate of the unemployment rate for 2003 for more than half of the remaining 600 PCAs.

Statistics for claimant unemployment are produced on a 100% basis from administrative statistics, are available monthly, and are more up-to-date. The LFS, at the time of writing, can only give patchy annual unemployment statistics for PCAs for 2003. The claimant system provides monthly statistics that, at the time of writing, support analysis of the pattern of seasonal variation for individual PCAs in 2004. The PCA showing the greatest seasonal variation in 2004 was Dorset South, on the south coast, with a coefficient of variation (CoV) of 22%. Unsurprisingly seaside areas show the greatest seasonal variation. Seven PCAs have CoVs of more than 20%. But at the other extreme seven PCAs, all in major cities, have CoVs of less than 2% (calculations by the authors).

Statistics of unemployment in the UK have been available on a place of residence basis since 1983. Their development depended upon the system of postcoding that was completed in 1974 and upon computerisation in the early 1980s of the unemployment statistics based on employment office areas. An account of this development, including explanation of the abandonment of statistics of registered unemployment, is given in Brimmer (1981).

The incompatibility noted between claimant statistics and the LFS does not provide a sound basis for the production of unemployment rates with the conventional economically active population denominator. The statistics of unemployment rates for local areas first published in 2003 and available back to 1996 have, as noted in Section 5, used population of working age (PWA) denominators.

**Geography and Full Employment**

The concept of full employment as well as the concept of unemployment was more of less invented in Britain. William Beveridge’s Full Employment in a Free Society published in 1944 remains the most comprehensive single study of unemployment problems in industrial societies. Beveridge distinguished frictional, structural, and demand deficiency unemployment. Frictional unemployment was conceived as unavoidable unemployment between ending one job and starting another. Frictional unemployment can be assumed to be mostly short-term.

Beveridge would not have been surprised at the small variation in the proportion of exits before four weeks that is revealed by statistics of exits from claimant unemployment. Such short-term unemployment would have been classifiable as frictional unemployment which can be expected to exist independently of the state of the labour market. Frictional unemployment for Beveridge would constitute the minimum level of unemployment achievable. It would set the level of unemployment compatible with full employment.

Beveridge identified demand deficiency and structural unemployment, and noted the difficulty of making the distinction between them. Structural unemployment could be regarded as a form of demand deficiency unemployment. Making the distinction and identifying appropriate remedies depends upon the availability of regional and local statistics.

Several generations of economists have elaborated on the idea of full employment in a theoretical way with concepts such as Non-Accelerating Inflation Rate of Unemployment (NAIRU). The central point is that an optimum level of
full employment is achieved when labour market pressure for higher wages and salaries is not sufficient to lead to runaway inflation. The concept of NAIRU demonstrates that full employment is inseparable from the geographical distribution of unemployment. It cannot be assumed that labour market pressures that lead to wage inflation, or labour market vacuums that lead to unemployment, are likely to occur equally in all labour markets in all parts of a country.

If unemployment is unequally distributed geographically, inflationary labour market pressures will be reached first in areas of low unemployment. Areas of low unemployment will have achieved full employment or over-full employment while other areas continue to suffer from high unemployment. The proper functioning of the labour market as well as the management of the labour market by the government requires statistical information on areas within a country.

ILO/LFS statistics could be said to be adequate at a regional level, but not at local level. Over the past few decades in the UK, for example, there has been persistent growth of inner city unemployment. The main unemployment problem has become intra-regional rather than inter-regional. Survey based statistics, such as those of the LFS, are inadequate for measurement and investigation of the relatively finely-grained variation in unemployment levels now evident in every sizable urban area.

Chart 4 shows the distribution of unemployment in England among PCA areas. The map divides PCAs into quartiles according to the claimant unemployment rate. The lightly dotted PCAs are in the lowest quartile with the lowest unemployment rates. The PCAs coloured black are those in the top quartile with the highest unemployment rates. In between light grey shading denotes PCAs in the second quartile with below average levels of unemployment, and the dark grey denoted the third quartile with above average levels of unemployment. The map shows that high levels of unemployment are concentrated in urban areas. Every city and major town contains major concentrations of unemployment. With a small number of exceptions there are no major concentrations of unemployment that are not urban areas.

ILO measures of unemployment are inadequate for investigation of such a fine grained geographical distribution. Chart 4 demonstrates the need to combine ‘whole population’ information from the LFS with administrative data on unemployment, as was used to construct this chart.

The Cultural Influence

International Labour Office criteria define unemployment in terms of seeking employment. In other words unemployment is a condition found among the population. At first sight that seems unobjectionable. How can anybody be unemployed if they are not looking for employment?

One feature of this definition is that it puts the onus of being unemployed upon the individual. If individuals are unemployed, it is implied, it is their own fault. But the idea that individuals should have the right to work is a component of a number of belief systems. Islam, for example, recognises a right to work. The Cairo Declaration on Human Rights in Islam (1990) states that “Work is a right guaranteed by the State and the Society for each person with capability to work (http://www.humanrights.harvard.edu/documents/regionaldocs/cairo_dec.htm).

The Catholic church teaches that “The obligation to earn one’s bread presumes the right to do so. A society that denies this right cannot be justified, nor can it attain social peace.” (Centesimus Annus, 1991, para 43). The former Soviet Union managed to achieve full employment by insisting that everyone should work. The UN-HABITAT Human Settlements Programme has a Charter of Human Rights that specifies that male and female citizens have the right to work through worthy employment with sufficient resources to guarantee the quality of their lives.

The practical consequence of the ILO exclusive emphasis on seeking work is a lack of acknowledgement of factors that contribute to unemployment. Defining unemployment as a condition does not require investigation of cause. The LFS does not, like the CPS, include questions on reasons for unemployment, and does not allow for the production of statistics for entry to unemployment that give indications of cause (see Thomas, 2005, for elaboration).

The easiest reform would be to modify ILO guidelines for the conduct of labour force surveys. Modification would require the inclusion of a question on unemployment addressed to all respondents – not just to those unemployed on the date of the survey. For example, ‘Have you been unemployed at any time during the past three months?’ or, to more fully comply with other ILO criteria, ‘have you been without paid employment and seeking work at any time during the past three months?’

Such questions would recognise the concept of entry to unemployment and would provide statistics on the number of entrants. A follow-up question on dates of unemployment would support the production of statistics for unemployment in the four weeks prior to the survey date. Statistics for the number of entrants in the previous four weeks would provide support for the production of accurate statistics on duration of unemployment, and so deal with ‘length-biased sampling’.

Questions identifying entry could well elicit reasons for unemployment along the lines of the CPS. Data on reasons would allow for better international comparisons and would
Chart 4  Unemployment rates for PCAs in England in 2004
support more comprehensive analysis of time trends in unemployment than is possible with the statistics produced in accordance with current ILO criteria.

ILO/LFS statistics are also of limited value in investigating the geographical distribution of unemployment. The inescapable problem with ILO criteria is that they are based on survey statistics that cannot be expected to provide adequate information on local unemployment. They are difficult to integrate with national statistical systems that have the geographical detail and data on entry.

ILO guidelines for the conduct of labour force surveys followed the pattern set by the US Current Population Survey more than thirty years earlier. It is ironic that nowadays the CPS provides less information on unemployment in the US labour market than do insured unemployed statistics. Weekly statistics on entry to insured unemployment provide information on current trends. Monthly mass layoff statistics provide information on an important cause of unemployment. The Local Area Unemployment Statistics system demonstrates the value of combining administrative systems with survey statistics with the production of statistics that combine information of a few thousand CPS respondents identified as unemployed with statistics for around three million continuing claims for unemployment insurance. But the value of such combining is not expressed in the ILO criteria for the conduct of labour force surveys.

In the case of the UK the nearest equivalent insurance statistics – the claimant statistics – account for a much larger proportion of unemployment (as defined by ILO criteria) than US insured unemployment statistics. But it is known that the UK LFS data does not provide accurate information on claimant unemployment. The production of estimates of ILO unemployment by means of statistical estimates on the lines of the BLS would not be the best solution.

The general solution would be, not for labour force surveys to ignore administrative unemployment systems, but for the standard LFS to embrace administrative unemployment systems. The administrative records of insurance or registrant based unemployment systems could be used as part of the sampling frame for labour force surveys. Weighted sample figures could be grossed up to national totals in accordance with standard statistical practice and there would be no loss of representativeness. The focus on unemployment could be achieved without reducing comparability between LFS statistics for the employed, unemployed, and inactive populations at the national level, and without jeopardising international comparability of statistics relating to any of these categories. The ILO guidelines for the conduct of the standard LFS could be extended to give detailed guidance on the methods that might be followed.

Such a development could be expected to contribute to the quality of unemployment statistics as defined both by ILO criteria and by national administrative systems. Comparison of the survey results for the administrative sample with that of the general population could be expected to yield information that would support the production of estimates of ILO unemployment for local areas that would be of more ascertainable quality than the LAUS estimates in the US. The addition of a range of ILO personal profile variables to administratively defined unemployment statistics could be expected to add significant value to these statistics for national policy and decision making.

Addendum
Many of the points made in this article are supported by statistical evidence that is included here only in highly summarised form. For a more detailed report see John Adams and Ray Thomas ‘Patterns and Trends in Unemployment in Scotland 1985 to 2004’ to be published by Scotecon at the Universities of Stirling and Strathclyde. Acknowledgement is made to the Royal Statistical Society for the award of a Campion Fellowship to Ray Thomas that has supported the research for this article.

Acknowledgement is made to Scotecon for a grant to John Adams for the ‘Patterns and Trends in Unemployment in Scotland 1985 to 2004’ study that has also supported the research underlying this article.

References


Beatty, Christina, Stephen Fothergill, Tony Gore, and Alison Hetherington (1997) The Real Level of Unemployment, Centre for Regional Economic and Social Research, Sheffield Hallam University, March.


Brimmer, M.J. (1981) Review of the Statistical Services of the Department of Employment and Manpower Services Commission, Department of Employment, April (Part of
Rayner Review).


Endnotes
1 This paper was presented at the IASSIST 2005 conference in Edinburgh by John Adams (J.Adams@napier.ac.uk), Hasan Al-Madfai (hmadfai@glam.ac.uk), and Ray Thomas (r.thomas@open.ac.uk), Faculty of Social Sciences, Open University, 35 Passmore, Tinkers Bridge, Milton Keynes MK6 3DY, England. Email: r.thomas@open.ac.uk
The Preservation of Research Data in a Postmodern Culture

John Curtice, one of the plenary speakers at the 2005 IASSIST conference in Edinburgh, presented time-series evidence showing the emergence of postmodern values. Derived from major attitudinal and value surveys since the 1960’s, his research reveals a shift in the locus of personal identity formation from institutions, such as organized religion and political parties, to the “marketplace” of individually established identities. In the postmodern world, everyone supposedly shops for her or his own identity using today’s educational systems to compile a uniquely packaged identity through self-actualization. According to this theory, institutions no longer serve as touchstones in determining individual identities.

In his address, Professor Curtice also showed research findings that temper the postmodernist interpretation of today’s world. Specifically, many of today’s values represent an interaction between individuals and institutions. Nevertheless, institutions are increasingly under attack by postmodern values. These assaults even pose a threat to national data archives and how these archives function in today’s societies.

Our national data archives are not immune to fundamental changes within our cultures. After all, data archives preserve one aspect of our cultural heritages, namely, digital evidence of research value. Therefore, one would expect major shifts in culture to result correspondingly in changes in the ways in which the record of our cultures are preserved. What aspects of postmodernism present a threat to our national data archives?

The very nature of the Internet reinforces the image of individualism in today’s culture. Everyone can have her or his own domain name and an identity on the Internet for a modest fee. Web technology, including recent Web log (blog) software, has the potential of making everyone a publisher. Napster enabled peer-to-peer music distribution, much to the chagrin of the entertainment industry. In Canada, a temporarily publication ban was imposed by Justice Gomery on testimony before a Commission investigating possible misuses of public funds. Shortly after the ban, this information appeared on the Internet from a site in the United States. As a consequence, the

Chief Commissioner partially lifted his earlier ban because the evidence had been widely disseminated on the Internet. These examples illustrate why the Internet is perceived as a great leveler enabling individuals to compete against institutional powers within and outside today’s legal boundaries.

From the perspective of data services, the Internet enables us to provide access to larger numbers of users than we have been able to support in the past. Dissemination is more direct and responsive to on-demand access to data resources using the Internet. Our profession sees these as admirable qualities of this technology. Individuals are empowered to retrieve data directly and quickly to their desktops providing researchers a sense of autonomy.

An erroneous corollary of this sense of access-autonomy is the concept that everyone on the Internet is her or his own archivist. We see this arising within the discourse around digital repositories and the idea of “self-archiving.” This is an unfortunate choice of words to describe the act of contributing individual works to a database of research publications. A motivating force behind digital repositories has been to increase access to scientific findings more quickly and equitably, sometimes even circumventing traditional channels of peer-reviewed print publications. Some commentators have generalized this self-archiving concept to incorporate research data among the digital objects researchers should contribute to digital repositories.

Self-archiving to me is an oxymoron. The act of archiving involves an institutional commitment to preserve knowledge and culture beyond political and technological changes. In the case of research, data archives represent institutions dedicated to the long-term preservation of data. Ideally, data archiving is a process throughout the life cycle of research and involves the full range of contributors to a research project. While sole investigators still contribute to the overall output of research, increasingly research projects are organized around teams, especially research that is inter-disciplinary, comparative, multi-national and large in scale. Consequently, the idea of an individual being her or his own data archivist runs counter to the way major research is being performed nationally and
Internationally today. Furthermore, the process of such research engages many stakeholders, including government granting agencies, universities, researchers, data producers, publishers, libraries and data archives. All of these contributors play a role in the life cycle of research data. One function of a data archive is to identify the custodial relationships among these stakeholders throughout the various stages of the data life cycle. The concept of self-archiving is meaningless in the context of a life-cycle model.

Diminishing the value of institutions in a postmodern world and accentuating the individualistic attributes of the Internet threatens data archives as institutions. Research in Canada conducted in conjunction with the National Data Archive Consultation demonstrated the need for institutional support of data to ensure its long-term access. A study of 100 funded projects by the Social Sciences and Humanities Research Council in Canada between 1977 and 1980 found data for only three of these projects in 2001.2 The data for all three projects had been deposited with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.3 This Canadian study demonstrates the level of risk that research data face without an institution responsible for their long-term care.

Just as Professor Curtice found an interaction between institutional and individual values in his research, the challenges facing data archives consist of similar competing values. Consider the example represented by digital repositories. These systems are being built with good intentions to create better access. Unfortunately, the preservation commitment, established practices and protocols of digital repositories are less developed and tend to be based on an end-state model of preservation rather than life cycle. They are being constructed in disciplines without roots in archival science and their proponents have initiated discussions using terms that confound issues of preservation.

IASSIST members need to enter this dialogue to clarify key institutional principles about preservation. We need to educate the developers of digital repositories and the stakeholders in the research community that:

- preserving research data requires the involvement and enduring commitment of all major stakeholders in the research community.

These principles need to be apparent in the data preservation, documentation and citation standards that our profession develops as well as the services that we provide in our local institutions. Furthermore, all disciplines now struggling with new requirements to provide access to research data created through public funds need to be aware of and to embrace these preservation principles. 4

The recent OECD ministerial declaration on this issue is an opportunity for IASSIST members to educate disciplines both within and outside the social sciences about the principles of preserving research data.

What options exist to combat postmodern values threatening our data archives? We must be advocates for data archives individually and collectively through our affiliation with IASSIST and other professional organizations. Internet technology supporting individualism in today’s culture should be used to promote institutional solutions for preserving research data. Champions for data archives among the stakeholders in the research community need to be identified and supported. We need to be creative in transforming our national data archives so they remain relevant to the knowledge sector in our societies while ensuring that the functions they fulfill in preserving research data are not sacrificed or diminished. Finally, we need to monitor attitudes toward data preservation constantly and to combat indifference toward institutions with mandates to preserve data.

* Chuck Humphrey is head of data library at University of Alberta. He was president of IASSIST 1991-1995. He can be emailed on humphrey@datalib.library.ualberta.ca

Footnotes


3 The ICPSR, which was created in the 1960’s, is an institution dedicated to the preservation of social science research data.

The International Association for Social Science Information Service and Technology (IASSIST) invites your participation in its 32nd annual conference entitled Data in a World of Networked Knowledge on May 22-26, 2006 in Ann Arbor, Michigan. The conference will be preceded by workshops and followed by optional weekend activities in the Ann Arbor area. Details about the conference and the association may be found on the IASSIST website at: <http://www.iassistdata.org>

Proposals for papers, sessions and poster/demonstrations should be submitted by **16th January 2006**.

The 2006 conference theme, Data in a World of Networked Knowledge, highlights the role of empirical data in a society that wishes not only to know itself, but also to build an enduring, interconnected storehouse of knowledge for learning and research. Once again IASSIST offers a time and place to explore, enlighten, and energize the participation of data professionals in the networked information world. We seek submissions of papers, poster/demonstration sessions, and panel sessions on topics that address the full range of digital data life cycle issues, including those that focus on access, documentation, dissemination, preservation, data use and current empirical research activity.

Additional topics might also include information and statistical literacy, data confidentiality and statistical disclosure, Geographic Information Systems (GIS) and spatial data, as well as publication, annotation, curation and authentication of networked knowledge assets. For other key topics see previous IASSIST Conferences at <http://www.iassistdata.org/conferences/index.html>.

**About IASSIST**

IASSIST is an international organization of professionals working in and with information technology and data services to support research and teaching in the social sciences. The organization also explores issues of access, stewardship and the interconnections among social science, behavioral, biological, and health data. Typical workplaces include quantitative and qualitative data archives/libraries, statistical agencies, research centers, libraries, academic departments, government departments, and non-profit organizations. See the IASSIST website at <http://www.iassistdata.org> for further information.

IASSIST conferences bring together data professionals, data producers, and data analysts from around the world for presentations and workshops covering new and persistent issues relating to access to data, its documentation, and digital preservation, with special emphasis on the social sciences. The social sciences have a long history of data sharing activity which will make the conference of interest to colleagues in disciplines where improving data access practices is on the policy agenda, and where there are clear overlaps with digital curation, data publishing, e-science/cyberinfrastructure initiatives, and new interdisciplinary collaborations.

The IASSIST Quarterly (IQ), available online from the IASSIST website and in print, is another important means of communication for the data community. Each year, IQ features the papers associated with conference presentations. Of special note is the IASSIST Publication Award, involving a cash prize for the winning paper. For further details see: <http://www.iassistdata.org/publications/pubaward.html>.

The IASSIST Outreach Committee accepts applications from data professionals in countries with emerging economies for funding to attend IASSIST 2006. More information about the Outreach Committee’s work, including funding criteria and online application form, can be found at <http://www.iassist.ucdavis.edu/>

**Procedure**
The deadline for paper, session, and poster/demonstration proposals is **16th January 2006**. The Conference Program Committee will send notification of the acceptance of proposals on or before **10th February 2006**.

Individual presentation proposals and session proposals are welcome. Proposals for complete sessions, typically a panel of three to four presentations within a 90-minute session, should provide information on the focus of the session, the organizer or moderator, and possible participants. The session organizer or moderator will be responsible for securing session participants, some of whom may submit paper proposals independently.

All proposals, including proposed title and an abstract (recommended length 150 words), should be submitted using the link on the following website:

<http://www.icpsr.umich.edu/iassist/call.html>

Alternatively, proposals may be sent via email to <iassist06@gmail.com>. In this case, please use a subject heading of “Paper proposal - Your Name” or “Session proposal - Your Name” replace “Your Name” with the name of the session organizer.

Further information on travel and accommodations will be available at links from the IASSIST ‘06 Conference website: <http://www.icpsr.umich.edu/iassist/>. Online registration is scheduled to open on 1st February 2006.

**Make plans to come to Ann Arbor for the IASSIST 2006 Conference May 22-26, 2006**
Call for Papers: 2006 IASSIST STRATEGIC PLAN PUBLICATION AWARD
(Competition is not limited to current IASSIST members)

Submissions due January 10, 2006
Announcement of winner March 1, 2006
Award: $250 US and one year membership in IASSIST
http://www.iassistdata.org

Content focus: Education.

As an organization, IASSIST has a history of working to educate its members about matters of common professional interest to the social science data community. Traditionally, this education has taken the form of professional development opportunities available in member-initiated and member-taught workshops at the annual IASSIST conference. Yet as the world of social science data grows increasingly complex, staying abreast of new developments in the profession is likely to present an ever increasing challenge for IASSIST members. As a result, the Education Committee of IASSIST is receptive to recommendations for employing new instructional methods and technologies as we strive to meet IASSIST’s educational mission.

At its annual conference in May, 2004, the IASSIST membership approved a 5-year Strategic Plan that focuses upon three strategic directions: Education, Outreach, and Advocacy. This paper competition has been established as a means of exploring, articulating, and documenting topical issues related to Education. For further information about the Strategic Plan, see: http://www.iassistdata.org/membership/plan_june2004.pdf

IASSIST seeks papers that address one or more of the issues, principles, and strategies for engagement in the following subject areas:

1. IASSIST-related Educational Initiatives

2. Professional Development and Educational Opportunities of interest to IASSIST members

3. Educational outreach to research communities, including and beyond the social sciences, to promote data preservation and access.

Papers should include recommendations for action or suggestions of specific projects that can be undertaken (or are underway) to further the education goals of IASSIST.

Prospective submitters may wish to review the discussion on the IASSIST blog about the educational issues arising under the topic of the Accidental Data Librarian (available at http://iassistblog.org/?cat=5).

Criteria for evaluation:
2006 IASSIST STRATEGIC PLAN PUBLICATION AWARD

--relevance of the paper to one or all of the themes in the IASSIST Strategic Plan (specifically Strategic Direction I: Improve and Expand the Educational Component of IASSIST both Internally and Externally.

--inclusion of specific suggestions for action or specific projects that will further the education of IASSIST members or otherwise encourage progress in support of the IASSIST Strategic Plan

--potential in building a base for future IASSIST activity

--quality of writing

--bibliographic content including references to related materials

--clarity in presenting issues and viewpoints as outlined above

Competition details:

All papers are to be submitted in English.

The winning paper will be announced on the IASSIST list-serve on or around March 1, 2006. In addition to being designated as the winning paper in the IASSIST Quarterly, the author of the winning paper will receive the monetary award and a one-year membership in IASSIST and be recognized at the IASSIST conference in Ann Arbor. All other submissions meeting the criteria for evaluation will be published in the IASSIST Quarterly (IQ) (online and print).

Papers that are submitted to this Strategic Plan Publication Award competition may also be submitted for inclusion in the IASSIST conference in Ann Arbor in May 2006.

Papers must be a minimum of 5 pages in length, including bibliography and graphics as appropriate. We strongly prefer that for publication purposes all documents be submitted in Word format and each graphic be submitted as a separate file in one of the following formats: .gif .jpg .tif .bmp .png

Papers must not have copyright limitations; IASSIST Quarterly (IQ) rights will apply upon publication. All papers must be submitted by January 10, 2006 to the competition web host: David Sheaves

<sheaves@vance.irss.unc.edu>

Questions (not papers, please) may be sent to:

<iassist-reviews@mailman.srv.ualberta.ca>

Competition is not limited to current IASSIST members.

The review committee for the IASSIST Strategic Plan Publication Award will be announced on the IASSIST website www.iassistdata.org and on the IASSIST list serve.
The **International Association for Social Science Information Service and Technology (IASSIST)** is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST**.

**Membership fees are:**
- Regular Membership: $50.00 per calendar year.
- Student Membership: $25.00 per calendar year.
- Institutional Subscription: $75.00 per calendar year (includes one volume of the Quarterly)

Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details:

http://datalib.library.ualberta.ca/membership/membership.html

- $50 (US) Regular Member
- $25 Student Member
- $75 Subscription (payment must be made in US$)
- List me in the membership directory
- Add me to the IASSIST listserv

Please make checks payable, in US funds, to **IASSIST** and Mail to:

**IASSIST,**
Assistant Treasurer US
Shelagh Mackay
University of California
Davis, CA, USA
95616-8617

---

I would like to become a member of **IASSIST**. Please see my choice below:

Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details:

http://datalib.library.ualberta.ca/membership/membership.html

- $50 (US) Regular Member
- $25 Student Member
- $75 Subscription (payment must be made in US$)
- List me in the membership directory
- Add me to the IASSIST listserv

Name:

Job Title:

Organization:

Address:

City: State/Province:

Postal Code: Country:

Phone: FAX:

E-mail: URL: