

# Project Description

The project description is presented in six parts: Results of Prior Support, International Analysis, Contribution of a Remote Access Analysis System, Methods, Plan for Collaboration, and Summary of Intellectual Merit and Broader Impacts of Proposal.

## Results of Prior Support

Increasingly, addressing questions about the relationships between climate change, resource development, the establishment of new political institutions, and living conditions in the Arctic require international data sets (SEARCH 2005). With that in mind, in 1997 a group of social scientists and indigenous representatives from Greenland, Canada, Russia, the United States, Denmark, Norway, Sweden, and Finland decided to collaborate on a project entitled, **Survey of Living Conditions in the Arctic: Inuit, Saami, and Indigenous Peoples of Chukotka (SLICA)**. Below we review the results of support of this effort by the National Science Foundation (Award Number 120174) as well other national and international funding sources. The National Science Foundation grant has supported the Alaska component of the project, technical assistance by the US team to other countries, participation in international team workshops, preparation of the international data set, management of the project website, coordination of the international analysis, and ultimately (through a supplement to the grant) the Chukotka component of the project. The NSF grant will end in 2006. Other grants, such as the Canadian SSHRC grant, will end in 2008. This is a renewal proposal to extend the scope of the original project to ensure that the data are available to Arctic researchers and policy analysts while ensuring the confidentiality of respondents.

## Geographic Scope of the Study

The original intent of the study was to represent the living conditions of Inuit adults living in Inuit settlement regions of Alaska (Iñupiat settlement regions), Canada, Greenland (also including non-indigenous residents), and Chukotka (also including other indigenous peoples in Chukotka) and to represent Saami adults living in the Saami settlement regions of Norway, Sweden, Finland, and the Kola Peninsula of Russia. Funding opportunities for such research are largely national (e.g. the National Science Foundation in the US, the Social Science and Humanities Research Council in Canada, and the Greenland Home Rule Government) or regional (e.g. the Nordic Council of Ministers). All countries were successful in obtaining grant funds sufficient to support the complex task of developing an international questionnaire. Differences in funding opportunities have largely determined the timing of data collection. Funding difficulties in Scandinavian countries and Russia forced the team to proceed within the current grant cycle with an Inuit focus (and other indigenous peoples of Chukotka). Data collection is now complete in Canada, Alaska, Chukotka, and virtually complete in Greenland. A proposal has been submitted for completion of the Saami component of the international project.

## The Data Set

Table 1 describes the existing international data set constructed from the responses of over 7,000 randomly selected adults to the common set of questions contained in the international core questionnaire. Probability sampling techniques were used in all countries to ensure that the

Table 1: SLICA Sample Description

	Native Population (rounded)	Sample (rounded)	Response Rate
Canada (Inuit)	22,000	4,700	83%
Greenland (Greenlanders)	38,000	1,300	80%
Chukotka (Inuit and other indigenous peoples)	18,000	600	85%
Alaska (Iñupiat)	11,000	700	84%
Total	89,000	7,300	83%

interviews statistically represent all Inuit adults living in Inuit settlement regions as well as the indigenous peoples of Chukotka.

The dataset consists of 950 variables constructed directly from the core questionnaire plus some 250 analytic variables constructed to date. The intent of the project has been to describe living conditions in a manner relevant to Arctic peoples. Domains of living conditions represented in the dataset are described in Table 2. Concept measures include self-reports of activities, circumstances and other “objective” aspects of the lives of respondents as well as subjective assessments of these conditions.

Table 2: SLICA Questionnaire Topic Areas

employment	age, gender, cultural background	technology use
hunting, fishing, herding	identity	perceived community problems
other domestic production	social supports	political awareness, interest, knowledge
household member		
production	physical health	Native values
traditional education	mental health	religion and spirituality
western education	alcohol and drug use	public safety
language abilities	housing	resource management
residence history	income	environmental problems
moving expectations	recreation & civic activities	subjective well-being

### Questionnaire Development

An important product of prior support of the Survey of Living Conditions in the Arctic is the international core questionnaire itself. The design of the questionnaire was informed by a rich literature in the study of living conditions and the quality of life (Bauer, 1966; Sheldon and Moore, 1968; U.S. Department of Health, Education, and Welfare, 1969; Campbell and Converse, 1972; Erik Allardt 1975; Campbell, Converse, and Rogers, 1976; Andrews and Withey, 1976; Michalos, 1991; Ringen 1995; Kenneth Land 1996; Diener and Suh 1997). The international team recognized that, to be relevant to Arctic peoples, their approach had to reflect social and environmental circumstances shared by Arctic peoples that differ from the residents of most other regions. Of particular interest is the importance of forms of domestic production closely tied with natural resources (i.e. hunting, herding, fishing, and gathering). To develop an approach to measuring domestic production, the team drew upon earlier northern research (Braund, Kruse, and Andrews 1985; Usher 1992; Usher and Weinstein 1991; Kruse 1991; Duhaime 1998, 1991 and Dahl 2000). Work on the measurement of domestic production by the international team resulted in the publication of the paper, *The Household as an Economic Unit in Arctic Aboriginal Communities* (Usher, Duhaime, and Searles, 2002). Other areas of particular importance to the measurement of living conditions in the Arctic included Native values (McNabb, 1991) sustainable development and health (Clugston, 1995; Larsen, 1996), social adjustment and social support (Larsen, 1993) and ethnic identity (Spratt, 1994; Durie et al., 1995).

As important as the contribution of past theoretical studies to the development of the international core questionnaire was the collaboration of indigenous people and researchers in the project itself. Between 1997 and development of the final questionnaire in 2001 collaboration between indigenous people and researchers occurred at both national and international levels. At the national level, a Native Management Board or Steering Committee overseeing each phase of the study had the responsibility for approving questionnaire development. At the international level, representatives of the Native Management Boards and Steering Committees worked with international team researchers to ensure that a comprehensive set of living condition domains was included and that the best possible measures of these domains were identified. The project website contains an historical archive of all the workshops, teleconferences, questionnaire drafts and other materials that document the process of questionnaire development.

### Sample Frame Development and Common Methods of Measurement

The science of survey research rests on methods of obtaining representative samples with known sampling errors and minimal measurement errors. The methods of obtaining representative samples differ by country. In Greenland, for example, a national register of the population can be directly used to sample the population. In Alaska, by contrast, researchers have to develop their own sample frame through what

is known as area probability sampling (multi-stage samples of communities, blocks, households, and individuals within households). Since SLICA is the first multinational Arctic social survey of indigenous people, a major product of SLICA is the set of sample frames which collectively yield a representative sample of Inuit adults.

Language represented a major challenge to avoiding measurement errors. The team chose to develop the questionnaire in English, to translate versions into the required languages (Greenlandic, Inupiaq, Inuktitut, Saami, Russian, Danish, Norwegian, Swedish, and Finnish) and to use back translation and pretesting to identify and correct translation problems. Members of national teams worked in other countries to build a common experience and shared field procedures. The complementary sample frames and shared field methods are both research products that should prove valuable to future researchers.

## **Dissemination of Research Products**

The Survey of Living Conditions in the Arctic is directly linked to the indigenous and international policy communities through the Inuit Circumpolar Conference and the Arctic Council, both of which have endorsed the study. The international team has made presentations at seven Arctic Council meetings and two ICC meetings concerning the project. Presentations of project results are scheduled for the Inuit Circumpolar Conference in February 2006 and the Arctic Council in October 2006.

A major vehicle for informing stakeholders and researchers about progress of the Survey of Living Conditions in the Arctic is the project website, [www.arcticlivingconditions.org](http://www.arcticlivingconditions.org). The website includes a comprehensive description of the path from research design through international reporting and sharing of data. Topics include: Introduction to the study, Factors Affecting the Timing of Release of Project Reports and Data, Study concepts, Questionnaire development, Field methods, Data processing, Data review, Analysis, Reports, and Data Files.

In May 2004, the international team organized a session at the International Congress on Arctic Social Sciences to inform the Arctic social science community about progress on SLICA. Kruse also expanded the project website for this purpose. In addition, Cochran, Abryutina, and Kruse participated in an international Partnering Workshop held in conjunction with the ICASS V conference. Team members have also made presentations at the 2002 Inuit Studies Conference, 2001 World Congress for Reindeer Herders, 2001 ICASS IV Conference, 2001 1st Nordic Arctic Research Programme (NARP), 2000 12th Inuit Studies Conference, 2001 11th International Congress on Circumpolar Health, 1999 International Workshop on Key Problems of the Indigenous Peoples of the Russian North, 1999 The Sixth Circumpolar Universities Co-operation Conference, and the 1998 93rd Annual Meeting of the American Sociological Association. A major presentation of results is scheduled for the International Society for Quality-of-Life Studies Conference in July 2006. Initial publications for the project include: Andersen, Kruse, and Poppel (2002), Andersen and Poppel (2002), and Usher, Duhaime, and Searles (2002).

## **Human Resources Development**

Stephanie Martin, listed as a senior investigator in this proposal, based her Ph.D. dissertation on her analysis of SLICA results. Over 30 Inuit and Saami were trained as interviewers.

## ***International Analysis***

The intent of the Survey of Living Conditions in the Arctic is to develop a new way of measuring living conditions that is relevant to people in the Arctic, to compare living conditions among Native peoples of the Arctic, and to better understand the relationships among living conditions. There are two phases to the analysis. The first phase of analysis involves the work of the project team itself. The second phase of analysis is to facilitate the work of the broader research and policy analysis communities by making the data available. The project team's own international analysis (Phase One Analysis) began in the summer of 2005 when Statistics Canada made its Arctic Analytical File (part of the SLICA international dataset) available to researchers obtaining approval of a proposed analysis by the Social Science and Humanities Research Council (SSHRC) in Canada. Thus the analysis work of the international team began with SSHRC's approval of the team's proposal in July 2005. The potential value of the data set to future

researchers (Phase Two Analysis) is illustrated by the selection of international research themes for the Phase One analysis.

The international team has organized its analysis around five themes first developed in 2003 by Charles Dorais of Makivik Corporation, Kuujuaq, Canada and Ed Ward of Maniilaq, Inc., Kotzebue, Alaska. Dorais is a member of the Canadian Steering Committee and Ward is a member of the Alaska Native Management Board. The international team discussed the themes at a meeting held in Murmansk, Russia in February 2003. The analysis themes were further refined at a meeting of the Alaska Native Management Board held in Kotzebue, Alaska in November 2003. Brian Lyall of the Labrador Inuit Association, Goose Bay, Labrador contributed as a participant in this meeting. Lyall is a member of the Canadian Steering Committee.

To understand the research themes, it is important to understand the origin of the Survey of Living Conditions in the Arctic. Material living conditions in the Arctic, as measured by European and North American standards, are relatively poor. Especially in small settlements, people tend to experience relatively low cash incomes, poorer housing conditions, and higher costs for goods and services. Yet these measures of living conditions do not explain why many people remain in small settlements and in the Arctic as a whole. The origin of the Survey of Living Conditions in the Arctic was a 1994 survey of living conditions conducted by the Greenland Home Rule Government office of Statistics Greenland. The study was based on an approach to measuring living conditions widely used in Europe and the United States. Finding that the results did not explain settlement patterns in Greenland, Statistics Greenland sought international collaboration on a new approach which would define and measure living conditions in a way relevant to Arctic indigenous peoples.

The international comparison themes are all based on the shared view of the international team and indigenous partners that outside perceptions of living conditions in the Arctic shape policies and programs inside the Arctic. To the extent that these perceptions are wrong, the policies and programs are likely to have unintended consequences. The international analysis themes are intended to test hypotheses about living conditions.

**Theme One: The importance of social relationships and the standard of living to settlement patterns.** Our hypothesis is that social relationships are an important reason why people choose to remain in small communities, despite a lower, cash-based standard of living there.

**Theme Two: The importance of a mixed cash- and harvest/ herding- based economy to living in the Arctic.** Our hypothesis is that the combination of hunting, fishing, gathering, herding, and cash sectors of the economy continue to define the prevailing lifestyle of Arctic indigenous people.

**Theme Three: Relationships between social problems and other dimensions of living conditions.** Our hypothesis is that overall well-being and social problems are both related to a combination of other dimensions of living conditions including domestic and cash production, strength of cultural values, social participation, and a sense of local control.

**Theme Four: The influence of educators and missionaries.** During the design and pre-testing phases of this study we discovered that regions across the Arctic share a remarkable history of outside influence through schools and churches. Much of this history has been directly experienced by our survey respondents. It deserves sharing as lessons on the intended and unintended consequences of concerted policies and actions intended to improve the quality of life of Arctic people.

**Theme Five: The influence of policies on living conditions.** Governments have had a huge presence in the Arctic. They have enacted and implemented policies intended to improve housing, health, education, employment, local uses of resources, and to protect wildlife resources and the environment. Native organizations are increasingly assuming the role of outside government in developing and implementing policy. Have these policies been effective? Have there been unintended consequences? As a first step toward answering these questions, we can assess the variability of living conditions among the major policy jurisdictions of the Arctic: nations, regions and communities. Observed differences will invite comparative histories of policies in the context of geographic differences in economic resources.

The above themes obviously do not exhaust the analytic potential of the Survey of Living Conditions in the Arctic Data File. Of major interest, for example, are analyses that advance our understanding of the

interaction of climate, resource development, and other forces for change with living conditions in the Arctic. The international data set includes, for example, assessments of local environmental conditions that can provide the basis for regional comparisons.

The variability in individual responses constitutes a power tool for investigating potential relationships among living conditions and examining the dynamic relationships between such forces as climate change and well-being. The following is an illustration relevant to the study of changing political institutions. We constructed an index of perceived Inupiat influence over the management of natural resources on the basis of responses to three questions concerning perceived influence of Inuit over management of: (1) natural resources like fish and caribou; (2) management of natural resources like oil, gas, and minerals; and, (3) environmental problems in their area. We hypothesized that perceived influence is associated with overall well-being. Table 3 illustrates the strong association between perceived influence over the management of natural

Satisfaction with Life as a Whole	Index of Perceived Influence Over Natural Resource Management			
	not satisfied	somewhat satisfied	very satisfied	total
not satisfied	17%	5%	3%	10%
somewhat satisfied	41%	34%	7%	33%
very satisfied	42%	61%	90%	56%
	100%	100%	100%	100%

P = .000, Gamma=.52

resources and overall well-being among Alaska Inupiat. We would like to know more; what is the relative importance of perceived influence over the management of natural resources in explaining variations in overall well-being. Preliminary regression results shown in Table 4, again using only Alaska data, suggest that perceived influence is not the most important variable, nor is it the least:

	Unique percentage of overall satisfaction explained	Significance
satisfaction with opportunities to hunt and fish, availability of fish and game, and with local environment	4.4%	0.00
satisfaction with local health, education, recreation services	1.7%	0.00
satisfaction with influence over management of natural resources	1.5%	0.00
satisfaction with own health	1.3%	0.00
satisfaction with job opportunities	0.8%	0.00
satisfaction with job	0.3%	0.00
satisfaction with availability of goods, cost of living, and transportation	0.2%	0.00
Overall variance in satisfaction with life as a whole explained:	22%	0.00

Statistical associations such as those reported above do not establish causality. They serve as a useful hypothesis testing tool, and focus attention on potential key relationships that warrant further research. The SLICA dataset constitutes a benchmark description of living conditions in the Arctic. Its results can be compared with those of future studies, and can be extended through complementary research that produces more in-depth data on a subset of relationships. For each of these reasons - further statistical

analysis, time series comparisons, and extension through in-depth studies – data from the Survey of Living Conditions in the Arctic should be made available to Arctic researchers and policy analysts.

### **International Analysis: Implications of the Special Case of Canada**

In 1999 members of the indigenous Canadian Steering Committee requested that the SLICA team combine its survey research with the 2001 Aboriginal Peoples Survey (APS) being planned by Statistics Canada. Statistics Canada subsequently agreed, and worked with the SLICA international team in the design of the APS questionnaire. Differences in research schedules and study objectives meant that the combined APS core and Arctic questionnaires share a subset of 316 of the 950 SLICA core questionnaire variables. Statistics Canada expended more than three million dollars in interviewing and data processing associated with the Inuit settlement regions of Canada. It is doubtful that sufficient funds could have been found to implement SLICA in Canada on the scale made feasible through collaboration with Statistics Canada (approximately 4,700 interviews). Fortunately, the SLICA team has also found that it has been possible to test most hypotheses using the more restricted APS-derived dataset.

The principal implication for this proposal of the participation of Statistics Canada is the fact the data produced by Statistics Canada is subject to the Statistics Act of Canada. To implement the requirements of this law, Statistics Canada offers a combination of ways to access data holdings, including public use microdata files, remote access, and research data centres (Tambay, Godmann, and Potter 2003). Statistics Canada is currently preparing a public use microdata file derived from the APS survey. Statistics Canada has also made an Arctic Master Analytic File available through its Research Data Centre program. Under provisions of the Statistics Act, Statistics Canada cannot allow the Arctic Master Analytic File to leave the secure (offnet) environment of the research data centres. The implication of this fact is that a SLICA international master data file including the Canadian Arctic Master Analytic File will only be accessible through the Canadian Research Data Centre Program. The Canadian public use file can be combined with a comparable public use file produced from a combined Greenland, Chukotka, and Alaska international microdata master file. As discussed below, the SLICA team is also proposing to make a hybrid data set (SLICA master file plus Canadian public use file) available through a remote access analysis system.

### **Contribution of a Remote Access Analysis System**

Both professional ethics and the Office of Polar Program's *Guidelines and Award Conditions for Scientific Data* (OPP 1998), require the SLICA research team to make its data available to other researchers. The team is strongly committed to meeting the intent of this guideline – that of promoting further scientific research and informed policy decisions. The proposal stated that access to the dataset and associated materials will be available through the Web and through project team offices. The proposal also noted the key concern of how researchers accessing the files can be encouraged to follow the *Principles of Conducting Research in the Arctic* (ARCUS 1997). The team proposed to put in place a two-step process for acquiring the dataset. Prospective users would be able to download a description of the dataset, including a codebook. They would receive with this information a copy of the *Principles of Conducting Research in the Arctic* and a description of how the team applied these principles in the current study, including guidance on how to request a review and comments (not approval) by indigenous organizations on draft publications. To download the dataset itself, the researcher would be asked to register and to accept a license agreement in which they promise to follow these principles.

The proposal further stated that some variables in the microdata set would have grouped response categories in order to preserve confidentiality. This is a standard statistical disclosure limitation (SDL) technique (Gomatam, Karr, Reiter, and Sanil 2005). The proposal noted that, in some cases, this grouping could make it impossible to perform a desired analysis. In these cases, researchers would need access to the Master Data file. This file does not contain individual identifiers, but does include all responses in their original, ungrouped values. Due to the involvement of Statistics Canada and the requirements of the Statistics Act, access to the international Master Data file would have to be through one of the Statistics Canada Research Data Centers. The team itself is conducting its analysis through the Ottawa and Quebec Research Data Centres.

The SLICA team has approached the task of developing a data sharing strategy through its analysis of the combined Alaska-Canada dataset. Our goal is to understand the degree to which grouping of responses is warranted to reach an acceptable tradeoff between the usefulness of the data and the risk of disclosure of respondents. Our focus has primarily been on attribute disclosure, the situation in which a respondent could be identified by one or a combination of attributes in a record. Working within the rigorous SDL environment at Statistics Canada has been extremely informative, as they are highly experienced in evaluating the risks of disclosure. In regards to the use of APS data, for example, a key guideline for release of results is a minimum unweighted cell size of 10. By applying this guideline in the initial SLICA analysis, it became apparent that responses of many key variables would have to be grouped. In the case of geographic location, for example, it has been necessary to group within each region places above 1,000 Native population and places below 1,000 Native population. This “coarse” grouping of variables is workable in the analysis only because it is possible to access the original data to make the most appropriate grouped variable. Were the data made available only in these coarse groupings, other researchers, asking different questions, would likely find the groupings unwieldy or worse, unusable.

In the case of the SLICA dataset, the cost of disclosure could be high. The dataset includes self-reports of such sensitive information as being a victim of sexual abuse, or use of illegal drugs. Since many respondents live in small villages, even data attributes that would normally not reveal the identity of respondents (e.g. reasons for being away from a community) could do so. We have developed a high degree of trust in working with communities. Continued trust depends on maintaining confidentiality.

As pointed out in the recent National Academy of Sciences publication, *Expanding Access to Research Data: Reconciling Risks and Opportunities* (NRC 2005), there is both a growing demand for access to data and a growing concern about maintaining the confidentiality of data files. The Research Data Centre program of Statistics Canada is an example of an agency response to these dual issues. With the consent of Statistics Canada, the SLICA international data set will be available through one or more of their Research Data Centres. This approach will not, however, adequately address the need for access to the data by indigenous organizations and researchers around the Arctic. The team therefore has explored approaches to data sharing that fall between the release of a public use data file containing coarse groupings of variables and the necessity of working within a secure research data centre. A desirable approach would have the following characteristics:

- The system would permanently archive the SLICA survey datasets and its documentation
- Access would be provided under a contract in which the researcher agrees to operate within the Principles of Conducting Research in the Arctic as applied by the SLICA research team and the Native Management Boards and Committees directing the original study.
- Researchers would be able to conduct their analysis from their own institution or Native organization.
- Researchers would be able to use a full array of statistical techniques in their analysis.
- Researchers would not be able to view or download individual records.
- Data queries producing unweighted table cell counts of under a specified number would be rejected through an automated query filter.

A specific recommendation made in the National Academy of Sciences report is for research “...providing secure access through remote data access mechanisms, with the aim of increasing the availability of remote access to confidential data (NRC 2005:4). Rowland (2003a:4, 2003b) describes monitored remote access analysis systems as providing, “access to restricted microdata files on the Internet using methodology to limit or suppress output for disclosure avoidance”.

### ***Methods for Establishing the Remote Access Analysis System***

The research and system engineering involved in extending remote access analysis to restricted data sets is substantial. We have, however, identified collaborators at the Inter-university Consortium for Political and Social Research (ICPSR) at the Institute for Social Research (ISR) of the University of Michigan who can make this task feasible for the Survey of Living Conditions in the Arctic (ICPSR 2005a).

ICPSR researchers are currently designing a related project for the Office of Applied Studies, Substance Abuse and Mental Health Services Administration (SAMHSA) of the United States Department of Health and Human Services (OAS 2005). ICPSR maintains the Substance Abuse and Mental Health Data Archive (SAMHDA) for SAMHSA (ICPSR 2005b). SAMHSA would like to provide secure analytic capabilities for its National Survey on Drug Use and Health (NSDUH) restricted files that are part of the archive maintained by ICPSR. However, the focus of the SAMSHA project is making necessary analytic enhancements to the online analysis system in place at ICPSR. More specifically, the SAMSHA project will enhance the current online analysis system to include the calculation of variances to adjust for complex sample designs (those with stratification and clustering). By collaborating with ICPSR, we can make both projects cost effective.

We propose to organize the research in two phases. In the first phase, the SLICA project team and ICPSR will collaborate to archive and disseminate a public-use SLICA data set. The public use data set will have all direct and indirect identifiers removed and be disseminated in electronic form. The public use data set is not the optimal data product in that: (a) it excludes from distribution many of the rich variables that pose modest disclosure risks; (b) it creates significant barriers to the engagement of new researchers with SLICA data; and, (c) it reduces the quality of new applications by effectively preventing researchers from benefiting from the rich, detailed survey data.

The second phase of research is more ambitious. This approach will develop and test a remote access data analysis system designed for restricted data. This approach is clearly preferable. A high degree of collaboration between the SLICA research team and ICPSR is essential to the success of the more ambitious approach. The effort requires close integration of science and technology. We discuss how this integration will be accomplished under our plan for collaboration.

In the first phase of its work, ICPSR will begin with the public-use data set produced by the SLICA scientific team. This will include the preservation of the public-use data and a fully enhanced set of metadata. The second aspect of the work involves an ambitious collaboration between ICPSR and the SLICA's Scientific Directors to integrate the evolving scientific advances of restricted access data sharing. This includes the development of capacity at ICPSR to actively promote use of the data using the best available technology while fully protecting human subjects. It is essential that this capacity be sustainable over the long term based on the investment we are proposing.

## **Phase 1: Archiving and Public-Use Dissemination of Data**

### **Data Processing – SLICA Public-Use Data**

Data processing steps include preparing SAS, SPSS, and Stata data definition statements for all datasets and reviewing the dataset(s) for appropriate protection of respondents' confidentiality. The SLICA team will be responsible for providing ICPSR with public-use data that is in suitable shape for archiving, including core technical documentation in electronic format. A major challenge to the SLICA team will be to identify the combination of variable transformations (i.e. groupings of response categories) that best preserves respondent confidentiality while retaining analytical power. This phase will require extensive analysis.

Initial archival processing steps (after securing of original data and documentation; see "Preservation," below) to be performed by ICPSR will consist of checking the congruence of data and related documentation; checking for wild codes and inconsistencies in the component parts of the data collection; and harmonizing the data and documentation files. Intermediate data processing tasks will include converting ancillary technical documentation to electronic format (by scanning or converting materials to ICPSR's standard archival format for this type of material, PDF images plus text). ICPSR will also perform a confidentiality review of all contents of the dataset(s) (required by our policies for all data collections that we make publicly available), and will perform a series of disclosure analyses on the files. The latter techniques are relatively new in the world of data archiving; those that will be employed for the SLICA collection will utilize principles and automated routines developed by one of ICPSR's Topical Archives for eliminating/reducing disclosure potential on highly sensitive medical and substance abuse data. This

disclosure risk analysis will go beyond the current review of confidentiality. Final data processing steps will focus on the preparation of fully-documented public use datasets, as well as more complete versions of the same datasets that contain sensitive or confidential information that cannot be released for public use. The latter are called “Restricted Use” data; researcher access to those versions will be tightly controlled (as described below).

### **Preservation and Documentation of SLICA Data**

ICPSR’s preservation procedures are at the core of its archival mission. The goal of these procedures is to assure that the digital materials in the holdings (including both data and documentation files) remain accessible, complete, uncorrupted, and usable over time. The major challenges in this area are security, deterioration, and technological obsolescence, with the latter posing the most severe problems. Rapid technological change will always threaten the viability of digital materials produced in previous years and under technological conditions no longer in place.

To guard against accidental destruction of any of the SLICA materials, ICPSR staff will create security copies of all project materials. Current standards call for three such copies to be made, with each stored in a different physical location (including one copy stored off the ICPSR premises). As archival processing produces further “versions” of core data files, copies of those are made and secured as well, and separate indexes of the location, description, and version of all copies of files are constructed. At the present time, a main version of each public-use file in the archive (well over 90, 000 discrete data files) is stored on ICPSR’s file servers for access via the Internet by interested researchers. Security copies are currently stored on high-density DLT removable devices.

The most severe challenges to preservation of digital materials occur months or years after initial intake of such research materials. Storage media (and the hardware to access them) can rapidly become outdated. File formats supported at one time by software can become inaccessible with changes in software, and the lack of backward compatibility of newer software. To prevent data in its archival holdings from becoming unusable or inaccessible, ICPSR continually reviews and “refreshes” both the media and the storage format of all collections it archives. Staff “migrate” data from one medium to another on a periodic basis (four major migrations, and numerous minor ones, have taken place since ICPSR was founded in 1962). In addition, ICPSR converts the storage format of those collection deemed to be “at risk” when those formats are no longer supported in current software packages, or do not easily lend themselves to conversion by individual researchers. All of these services will be provided for the files from the SLICA collection, thus ensuring that these data will remain viable and accessible into the future.

Equally important to the preservation of data and documentation is its availability. In its forty-year history, ICPSR has worked to ensure that the information describing each dataset is provided in a manner that promotes effective, efficient, and accurate use of those datasets. Accordingly, ICPSR proposes a two-pronged approach to creating high-quality documentation products for the SLICA.

First, ICPSR will convert the documentation and codebooks related to the SLICA data to Portable Document Format (PDF), which can then be accessed and downloaded to a local computer. The PDF format, which has become the industry standard for the electronic distribution of documents worldwide, offers several advantages, including the retention of original appearance and format, and full-text search capabilities. These PDF files are enhanced by ICPSR staff with indexes and bookmarks, which allow the reader to easily access major sections of the document. Where applicable, ICPSR staff will also create links within the PDF files, enabling the reader to click on highlighted words on a page to access another part of the document, a Web site, or an external document.

As a second complementary format, we propose to “mark up” the documentation in XML according to the Data Documentation Initiative (DDI) specification. While the PDF format captures the original layout of the documentation for presentation purposes, the DDI format captures the substance of the document by tagging the text for content. The XML tags provide a uniform structure that can serve as input for a variety of applications. Further, the DDI-compliant core file can be repurposed to generate different types of output, for example, SAS/SPSS/Stata data definition statements or a bibliographic record.

The DDI specification has been developed by an international network of social scientists, data producers, information specialists, and data archivists to enhance the content of social science documentation and to facilitate the transport, exchange, presentation, and preservation of these critical materials. Funding for the effort, which began in 1995 at ICPSR, has been provided in part by the National Science Foundation, with contributions from ICPSR and from other participating organizations.

The DDI standard, by virtue of its being an "open," vendor-neutral standard developed collaboratively by the social science community, can grow and evolve as conditions change. SLICA can begin to benefit immediately from the development and testing work on this standard and from the expertise of ICPSR staff instrumental in this developmental work.

### **Dissemination of SLICA Public Use Data**

As mentioned above, two different versions of each SLICA dataset will be produced by ICPSR. One is intended for mounting on the ICPSR Web site, for generally unrestricted access by researchers and policy analysts. This public use version will be produced in cooperation with the SLICA Principal Investigators, and will include content that has been subjected to disclosure-avoidance measures intended to prevent identification of research subjects (either directly or through deductive analyses). Some original data items will be removed or adjusted through data-masking procedures as public-use versions of the files are prepared. These adjustments frequently impose limitations on the research uses of such files.

ICPSR will provide online data analysis capabilities for the SLICA, utilizing the Survey Documentation and Analysis (SDA) software developed and maintained by the Computer-assisted Survey Methods Program (CSM) at the University of California, Berkeley. SDA is a set of programs for the documentation and Web-based analysis of survey data (SDA 2005). This set of programs is developed and maintained by the Computer-assisted Survey Methods Program (CSM) at the University of California, Berkeley. SDA also features procedures for creating and downloading customized subsets of datasets, records, and/or variables.

Current online analysis features includes:

- **Analysis**  
Users can perform interactive data analysis online through a Web browser without downloading the dataset. All analysis programs provide options for ad hoc variable recodes, user-selected weights, filter and control variables, restricting the valid range, and question text. New variables can be saved in the dataset. Syntax can be shared among collaborators. SDA's web-based analytic capabilities include: Frequencies and crosstabulations (including charts), comparisons of means (with complex std errors), correlation matrix, comparisons of correlations, regression (ordinary least squares), and logit and probit regression. Supporting analysis capabilities are recoding one or more variables and computing new variables.
- **Codebooks**  
SDA conversions programs generate codebooks in either of two formats: (1) hypertext markup language (HTML) for browsing on the Web site and (2) ASCII text for printing and/or conversion to Portable Document Format (PDF). Codebooks in both formats document each study with the study abstract, a full description of each variable, variable frequencies, and information on survey design, sampling, and weighting.
- **Subsetting**  
After examining the data and documentation online, this feature allows users to specify and download a customized subset of variables and/or records for a study.

Online analysis is currently available for over 100 data collections housed at ICPSR, and is accessible 24 hours per day through the ICPSR Web site. SDA was specifically designed for use with social science survey data on the World Wide Web, and features ease of use, speed, and adaptability. By operating in modules, SDA accesses only the variables specified by the user rather than opening the entire dataset,

which allows for the generation of analysis result tables very quickly, within seconds, even on large datasets similar in size to the SLICA.

Data archived at ICPSR are freely available to faculty, students and staff at ICPSR's 400+ member institutions. Because the SLICA data are expected to be used heavily among research and policy groups in the Arctic (a majority are not ICPSR member institutions), we will provide 5-year memberships (with ICPSR sharing the cost of these memberships) to these organizations allowing individuals at these organizations free access not only to SLICA data, but the additional 4,000+ titles in the ICPSR collection. Approved users will also have free access to the restricted data set for five years.

## **Phase II: Developing Remote Access System Analytic Capabilities**

A second full analytic version of the SLICA file will be prepared and offered under a set of controlled conditions to bona fide researchers. This second version of a research data file is typically called a "restricted-use" version. ICPSR has created restricted-use versions of data files for several dozen research data collection in its holdings. The restricted files are secured in the manner characteristic of all data files in the ICPSR collection, but no copy is stored online or on a file server connected to any network. Gaining access to this version of the SLICA files by researchers will require a Restricted Use Data Agreement. Underlying this type of Agreement is the concept of responsible stewardship of confidential data, and assumption by a researcher (and her/his employing institution) of specific responsibility for exclusive and ethical use of those data.

ICPSR staff will develop a SLICA collection-specific version of a Restricted Use Data Agreement (ICPSR 2005c). The Agreement will include elements of the *Principles of Conducting Research in the Arctic* as applied by the indigenous steering committees directing the international study. The Agreement will also include an application form to be signed by the requesting researcher and by an official at her/his institution. A petitioning researcher will provide a curriculum vitae along with a detailed data protection plan for keeping such data secure. ICPSR staff will review each application for access to the restricted-use data file, to ensure that the data will be used in a responsible and secure fashion.

The key to researcher compliance with conditions governing access to restricted-use data files is the existence at the researcher's employing institution of mechanisms to ensure scientific integrity. Typically, such access is limited to researchers employed by an organization possessing a current NIH Multiple Project Assurances (MPA) Certification Number or Federal Wide Assurances (FWA) Certification Number (typically research-oriented universities). Access by other individuals is possible only after review of an application and consideration of relevant circumstances and the institutional setting, on a case-by-case basis. Special attention will be paid to ensure that Inuit organizations collaborating in the study are included.

The restricted-use dataset approach has been successfully employed by ICPSR (and by other organizations, including Federal agencies) for a period of some years. It is an effective way to permit access to confidential and/or sensitive research information, and has proven to be acceptable to academic researchers by permitting them to conduct research in their home environment utilizing familiar hardware, software, and readily-available supporting resources and mechanisms.

The restricted-use dataset will be available online through the Survey Documentation and Analysis (SDA), providing the same capabilities for variable creation, recoding, and analysis described above for the public use data set.

### **Survey Disclosure Limitations**

To the above capabilities, ICPSR and SDA would add disclosure limitation features for the Survey of Living Conditions dataset. The system would offer several ways to restrict analytic output. In this way, the system would have maximum flexibility and utility over time:

- Suppressing tables with fewer than X number of cases in one or more of the cells;

- Suppressing sparse tables in which the average number of cases in the cells is less than X;
- Allowing users to specify no more than X variables (including user-created variables) for a single analytic procedure (like a regression);
- Monitoring usage.

Below is a brief discussion of each of these points.

Table suppression. A common method for disclosure protection is table suppression (Cox, 1980). This prevents the user from gaining access to tables with only a small number of cases in the cells. Since such tables are usually not of great analytic utility anyway, protection is gained without substantial analytic loss. The desired minimum number of cases per cell and/or the desired minimum average cell size can be specified by the SLICA team. This decision, at least in part, will depend on other disclosure procedures applied to the file. Once the capability is built into the system, any number can be chosen.

Limiting the number of variables in one query: Another option we propose building into the system is a mechanism to limit the number of variables included in a single analytic procedure (like a single regression run). Concern is often expressed to the effect that if someone could specify enough predictor variables, it would be possible to identify individual cases. The SLICA team could decide, for example, that no more than 10 variables could be used in a regression at one time. Again, once this capability is integrated into SDA, any minimum number of variables may be specified. While the SDA system has the capability of downloading subsets of data and to list individual cases, these functions will be turned off for the restricted SLICA file.

### **Monitoring Usage**

Monitoring usage: analyses run on the system will be monitored for suspect activity. ICPSR will formulate definitions for this for review by the SLICA team. It would include queries that are well outside the range of what would be expected. This could include numerous queries using a single variable, or an unusually high number of queries from a single IP address. A protocol would be developed to deal with suspect activity when it is detected, such as by removing access to the study until the activity is investigated. The web pages for access to the restricted file could include notification that analyses are monitored as a deterrent.

### **Handling Special Analysis Needs**

SDA will satisfy most users needs. It is possible, however, that a small portion of users will want to estimate a model not offered in SDA. In these cases the user will be asked to complete an application to receive a copy of the file for analysis. Users would need to demonstrate that they have federal funding from NIH or NSF, thereby adding a layer of oversight on how the data are used and cared for. Following a successful review of an application by the ICPSR staff, a copy of the requested restricted-use dataset(s) will be made on a removable medium (such as CD-ROM). It will then be sent to the requesting researcher via registered mail, to be signed for by the addressee. After an agreed-upon period of use, the dataset will be returned to ICPSR or destroyed in a certified manner. Alternatively the user could access the file through the ICPSR secure data enclave in Ann Arbor Michigan.

An extremely secure mode of access to confidential and/or sensitive digital materials has been developed in recent years – at ICPSR this is called a Secure Data Enclave. This is a physical site that must be visited in person by scholars seeking access to research materials. This option will be available to qualified SLICA researchers requiring access to statistical packages (e.g. SPSS, SAS, STATA, LIMDEP) and who do not meet above requirements for obtaining restricted SLICA materials on a CD. The ICPSR secure data enclave contains copies of restricted-use datasets, along with hardware or equipment for use by researchers in accessing confidential materials. The equipment includes non-networked personal computers furnished with read-only CD-ROM drives. Access to this locked, attended facility is by application, and researcher activity in the facility is monitored by ICPSR staff. (The ICPSR secure data enclave is located in a restricted area at our headquarters in the Perry Building, a University of Michigan facility close to the central campus in Ann Arbor, MI.). Permitted access will include examining materials and encoding them in ways that do not identify individual respondents; performing statistical procedures on restricted variables to obtain analysis results that will be screened by ICPSR staff before being

permitted to be removed by the researcher. ICPSR will provide oversight of this form of access to researchers interested in exploring the SLICA research collection. This kind of facility imposes a charge to researchers for its use.

## Fostering Use of SLICA Data

### Finding Aids

Robust Web-based finding aids that assist and guide users in locating the data they need are critical to successful resource discovery. With funding from the National Science Foundation's Infrastructure in the Social Sciences program, ICPSR augmented its finding aids with a suite of new products and utilities and thus can offer excellent discovery capabilities for the SLICA data.

*Study description:* As with all data collections housed at ICPSR, we will catalog the SLICA and prepare a full study description to be part of our on-line catalog of holdings. The catalog entry will include not only traditional bibliographic information on Title, Principal Investigator, and ICPSR study number but also more descriptive information on the intellectual purpose and goals of the study, the sampling methodology, temporal and geographic coverage, detailed physical file specifications.

*Search engine:* ICPSR uses the Inktomi search engine, which offers support for XML, automatic search term stemming, Boolean-type operators, and extensive query and access logs. This efficient search apparatus enables users to conduct field-specific searches of the ICPSR catalog of holdings. Thus, a search across the ICPSR Archive for studies with geographic coverage of Alaska is easy to perform. Searches can be further narrowed by time period, funding agency, sampling methodology, etc. As part of the catalog of holdings, the SLICA data will be easily "discoverable" through this powerful fielded search utility.

*Indexing:* The study will also be indexed according to ICPSR's Social Science Thesaurus, developed through the Infrastructure Project. This controlled vocabulary system includes over 4,000 social science subject terms, an alphabetical authority list of name forms for individuals identified in ICPSR's metadata, and preferred and/or current names of geographic and geopolitical entities. Keywords from this thesaurus will be included in the catalog entry for the SLICA collection to provide an additional enhancement in searching.

*Links to related publications:* References to the SLICA in the scholarly literature will be identified and incorporated into ICPSR's newly created Bibliography of Data-Related Literature. The Bibliography, another component of the Infrastructure Project, now contains over 34,000 citations to ICPSR data, with two-way linking: users can view related publications from the catalog record and also link back to the underlying data from publication citations in the Bibliography itself. In addition, for ICPSR member institution with active "Link Resolvers," ICPSR can now provide click-through access to the actual journal articles. We will search relevant online journals and electronic databases for publications related to the SLICA and capture the citations in the Bibliography. This will enable SLICA users, potential users, and SLICA funders to review the existing literature to determine how the data are being used.

By providing enhanced finding aids for the SLICA data, ICPSR will ensure that these important research data are accessible by the widest possible audience in an ideal environment with all the necessary resources for teaching and research within easy reach. This environment will facilitate one of the major goals of the SLICA project: to make the data and related information accessible to a wide audience, including the members of the studied communities.

### Technical Assistance

Providing users of the SLICA data with outstanding customer service will be a major goal of all ICPSR staff members. Toward this end, ICPSR will provide researchers with technical and substantive assistance through interactive means including telephone, electronic mail, and online tutorials as well as enhanced traditional documentation.

Throughout its 40-year history, ICPSR has encouraged researchers to contact staff with questions regarding the use of data obtained through its services. A variety of means have been employed to facilitate this communication, including toll-free help lines and electronic mail groups. Mail groups are monitored by multiple staff to ensure prompt resolution to inquiries. The topics and resolutions that emerge from these interactions are then used as content for 'frequently asked questions' (FAQ) pages which appear on the Web site.

All documentation needed to analyze a dataset can be downloaded quickly and efficiently in ICPSR's Web-based environment. This documentation includes codebooks, data collection instruments, and resource guides. Reserved for complex data collections, resource guides provide additional in-depth information about important aspects of the data collection in a Web-based user-friendly format. This information will be used to highlight the intricacies of the SLICA and offer guidance for its analysis.

### **Facilitation of User Collaboration**

Approved users of the online restricted data set can register as collaborators. Such collaborators (e.g. analysts from Native organizations in Alaska and Canada sharing analysis objectives) could share syntax and variables created for the analysis.

### **Training**

The SLICA scientific team and ICPSR will conduct an intensive daylong training workshop on the use of SLICA restricted and public-use datasets during the International Association of Arctic Social Scientists Association ICASS Conference in May 2008. The workshop will be designed for social scientists in academia, and analysts in nonprofit research institutions, state and local governments and indigenous organizations. This course will be designed to familiarize Phase 2 analysts with the SLICA data and promote its effective use.

The ICPSR Summer Training Program in Quantitative Methods has been offering a comprehensive, integrated program of studies in research design, statistics, data analysis, and social methodology for thirty years. Its instructional environment stresses integration of methods of quantitative analysis within a broader context of substantive social research, and instruction is coordinated with and reinforced by active, participatory data-analytic experiences utilizing high-end, networked microcomputers. The Program has become nationally and internationally recognized as a primary mechanism for basic and advanced training in the methodologies and technologies of research and instruction in the social sciences.

The curriculum includes special workshops that provide participants with opportunities to examine the impact of various methodologies on specific substantive issues. Research scholars who have made important contributions to the development of social methodology present informal lectures focusing on their most recent research interests. In recent years, workshops have been offered on other complex data collections, including the Panel Study of Income Dynamics (PSID), the 2000 Census, and the Eurobarometer surveys.

### **Outreach and Promotion**

*ICPSR Bulletin feature article:* ICPSR will commission and publish an article highlighting the SLICA research activities in the *ICPSR Bulletin*. The *ICPSR Bulletin* is a semiannual publication featuring an article of interest to social science researchers and describing newly released and updated ICPSR data collections. Previous issues have featured articles on prominent data collections including the 2000 Census, the MacArthur Foundation Study of Successful Midlife Development, and the Comparative Study of Electoral Systems. Published in both print (with mailing to approximately 2,500 individuals) and electronic formats, the *ICPSR Bulletin* reaches a wide portion of the social science research community.

## ***Plan for Collaboration***

### **SLICA - ICPSR**

The SLICA international team (Kruse, Poppel, Duhaime, Abryutina, Martin) and ICPSR will work closely together throughout the entire project. Kruse has a long-standing working relationship with ICPSR and the

Institute for Social Research at Michigan. His graduate studies were based at ISR where he was also an Assistant Study Director. Kruse later collaborated with Frank Andrews at ISR on the development of a social indicators system for Alaska (Braund, Kruse, and Andrews 1985). All SLICA international team lead investigators will participate in a system design workshop with ICPSR and SDA investigators at the onset of the project. To meet phase-1 goals the SLICA international team will make available to ICPSR all data and documentation. To meet the goals of Phase II, the two research teams (SLICA and ICPSR) will again come together in Ann Arbor to finalize the scientific and analytic requirements of the enhancements to the SDA online analysis system. In Phase III, ISER scientists and ICPSR staff will coordinate and facilitate and the training workshop at the IASSA Conference Workshop in May 2008.

Integral to the work of SLICA international team and ICPSR is the continued oversight of the Survey of Living Conditions in the Arctic by the Alaska Native Management Board (ANMB). ANMB members are representatives of Iñupiat regional and statewide organizations. The ANMB has directed the Alaska component of the study since 1997. Major decisions made by the ANMB include deciding whether such a study is in the interests of Iñupiat, approval of all proposals, approval of a study plan for implementing the *Principles of Conducting Research in the Arctic*, approval of the questionnaire, and approval of initial analysis results. The ANMB will continue to direct the project.

### ICPSR – SDA

ICPSR and SDA have a longstanding collaboration. The SDA team, led by Tom Piazza, has made a number of enhancements to SDA over the years to meet the requirements of the social science community, many of which have been made at the request of ICPSR. Tom Piazza has reviewed the requirements of the SLICA data and has contributed to the development of aspects of the proposal dealing with the remote access data sharing system for restricted data (see letter of support). A member of the SDA team will participate in the annual project meetings in Ann Arbor.

### ***Summary of Intellectual Merit and Broader Impacts of Proposal***

**Intellectual Merit.** The Survey of Living Conditions in the Arctic is the first multinational Arctic social survey of indigenous people. The SLICA dataset was designed to cover the entire range of living conditions relevant to Arctic peoples. The 950 questionnaire variables and 250 analytic variables in the microdata set are all linked, thereby supporting analyses of multivariate relationships. The SLICA international team is committed to the goal of promoting further scientific research and informed policy decisions by making the dataset accessible to researchers and policy analysts. These data provide the basis for further statistical analyses, time series comparisons based on future studies, and extension of our understanding of living conditions through in-depth studies. Creation of such a database is one of the highest priority goals of the NSF Study of Environmental Change: “Develop a pan-arctic database of key human dimensions indicators of population, employment, and subsistence” (SEARCH 2005:xiii).

**Broader Impact.** Organizations expecting to use data from the Survey of Living Conditions in the Arctic to inform policy decisions range from Native local organizations (e.g. Kotzebue IRA) to regional organizations (Kawerak, Maniilaq, North Slope Borough, NANA, Bering Straits Foundation, Northwest Arctic Borough) to national organizations (e.g. Inuit Tapiriit Kanatami in Canada) and international organizations (Inuit Circumpolar Conference, Arctic Council). The proposed approach to handling SLICA data will ensure that these organizations can easily access the data while at the same time ensuring that confidentiality is protected. The development of a remote access analysis capability for use with restricted data sets will also advance the social sciences’ ability to handle sensitive data sets.

Of considerable importance is the impact of the remote analysis data sharing system on the state of data sharing in the social science research community. Increasingly, investigators are collecting sensitive data that pose unusually high disclosure risks (e.g. detailed geographic information, administrative data linkages, study of population groups that are relatively small). To protect the information provided by respondent, yet meet the needs of the research community, the Secure SDA will be a significant innovation benefiting not only SLICA, but also the scores of other restricted studies.