###### Full Proposal

1. **Project title:** School Attendance Boundary Information System
2. **Researcher status and title:** Associate Professor of Sociology at the College of William and Mary
3. **Statement of purpose, research problem, and rationale for conducting research**

This data infrastructure proposal seeks to assemble and distribute a new, nationally comprehensive spatial

database consisting of public school attendance boundaries in an effort to improve the quality of

geography and demographic data available to researchers. Currently the Census Bureau’s administrative

units are “coin of the realm” for researchers interested in understanding a core social science question: the

impact of the social context on life chances. Thus researchers who study how neighborhood contexts

affect educational outcomes, crime, disease, and related social processes base their analyses on units such

as census tracts or block groups. While useful, these areas have important limitations. They do not “map

on to” socially meaningful boundaries. Nor do the boundaries have an impact on access to important local

services—particularly access to educational facilities. It is a common lament that census tracts are a poor

operational definition of neighborhoods (Sampson, Morenoff and Gannon-Rowley 2002) and tracts are

deeply flawed proxies for school attendance boundaries themselves.

Even as the quantity and accessibility of statistical information has grown exponentially with the advent

of the Internet, the geographic infrastructure of the country has not maintained pace by assembling and

distributing nationally comprehensive and consistent small area database that delineate boundaries that

social scientists and members of the public view as meaningful. While thousands of socially-relevant

small-area geographic boundaries such as school attendance boundaries and property maps are available

from local entities, these data are piecemeal, inconsistent and incomplete. Social scientists who demand

geo-referenced demographic data for small areas within an entire region are limited to using census tracts

or zip codes as proxies for school attendance boundaries, neighborhoods and other socially relevant areas

that impact the lives of people who live, work and play within them. Researchers who study social

processes that are impacted by local contexts have by necessity confined their analyses to a local area.

Assembling relevant geography for broader studies is simply too expensive and time-consuming.

For example, it is hypothesized that families make housing decisions based on information about local

schools. Economists have sought to understand the relationship between home values and the

demographic composition and educational quality of schools that serve specific school catchment areas

with a larger school district. While economists have explored these issues within a handful of school

districts (Black 1999), school attendance boundaries have been simply unavailable at the state or national

level. With so few cases, it has not been possible to draw definitive conclusions regarding the role of

school quality and demographic composition in shaping housing values. Indeed, some economic studies

have settled for using census geography as proxies for school attendance boundaries in models that

explore the relationship between school quality and neighborhood housing prices (Brunner, Imazeki and

Ross forthcoming; Brunner, Murdoch and Thayer 2002; Brunner, Sonstelie and Thayer 2001; Downes

and Zabel 2002; Ioannides 2004; Weimer and Wolkoff 2001) while other scholars have relied on

boundaries within a small district (Bogart and Cromwell 2000).

The novel data we propose to collect will be called the School Attendance Boundary Information System

(SABINS) and will enable researchers to develop new areas of scientific inquiry by allowing them to

explore social conditions and processes within areas that have meaning to the people who chose to live,

work and play within them. The first SABINS data will consist of elementary, middle and high school

attendance boundaries for the 2009-2010 school year. Each school attendance boundary will be integrated

with data from the 2010 decennial census, the American Community Survey (ACS), and the Common

Core of Data (CCD). Linking population characteristics with school attendance boundaries will provide

researchers with an abundance of unique information describing detailed population characteristics of

children and households living in school attendance boundaries. SABINS data will be distributed to

scholars and the general public through the National Historic Geographic Information System’s web

interface, allowing simple, free and fast access to digital maps, statistical data, and metadata.

1. **Research questions and hypotheses : Not Applicable**
2. **Benefit/cost to Denver Public Schools, the profession of education, and/ or educational setting**

The availability of nationally comprehensive geography delineating school attendance boundaries will inand-of-itself spawn many new areas of research related to the configuration of school zones. At present,

social scientists know little about school attendance boundary characteristics such as variations in their

compactness, how many schools share coincident boundaries or, by contrast, how many of them have

islands that are “split” from the main boundary. Researchers do not know the average travel time and

distance from home-to-school for the typical child. Similarly, researchers do not know how many school

districts have open enrollment school attendance boundaries or how many school districts completely lack

formal school attendance boundaries. Further, little is known about the social and educational amenities

or hazards that school attendance boundaries contain. Scholars do not know what types of school

boundaries contain, private, magnet or charter school or where libraries, parks, swimming pools, or other

landscape and land-use features are located. Researchers have yet to conduct large-scale studies that

document incidents of communicable childhood diseases within school attendance boundaries and other

hazards such as crime rates. While a few national studies of racial and economic segregation across

schools have been conducted (Frankenburg, Lee, and Orfield 2003; Logan and Oakley 2004; Reardon and

Yun 2003) comparable studies of inequality across school attendance boundaries have yet to be

completed. Thus scholars and policy-makers have little information describing the variety of ways that

school attendance boundaries could be redrawn to optimize the delivery of educational services or

minimize social inequality. These are fertile areas of research that cry out for exploration.

School Attendance Boundaries will have even broader application when combined with other data

sources. School attendance boundaries collected during this project will be linked with a wide array of

statistical information from the 2010 Census, the ACS and the CCD. The census provides basic

denominators for an array of studies across the social sciences, including such diverse fields as criminal

justice, demography, economics, education, epidemiology, geography, political science, public policy and

regional science. Social scientists have become increasingly aware that individuals’ life chances, choices

and attitudes are shaped not only by their own characteristics, but also by the characteristics of their

neighbors and communities. SABINS encourages and simplifies the use of techniques such as multi-level

analysis that draw upon such insights. The aggregate-level SABINS data dovetail with and complement

widely used micro-level datasets. Researchers can link SABINS data with individual level data from

sources such as the Panel Study of Income Dynamics (PSID), the ACS Micro Data, and many other

specialized surveys of children, adolescents, and households. These data sets and many others identify the

location of respondents to these surveys (typically with a block identifier). Researchers who are granted

access to the geographic location of survey respondents would be able to generate new research questions

by linking the specialized data sets with SABIN geography and data.

Delineating school attendance boundaries and integrating them with existing statistical data will provide

new and critical data for academic researchers who study issues of segregation. For nearly a century,

scholars have used census tracts as proxies for neighborhoods, yet census tracts are statistical and ***not***

socially constructed entities. This is problematic for several reasons. Several decades ago, census tract

boundaries were delineated by local groups of census data users in accordance with Census Bureau’s

*Geographic Reference Manual*. At that time, the manual required that census tracts “to be homogeneous

with respect to population characteristics, economic status, and living conditions.” Therefore, the areas

covered by the Census Bureau’s small area geography did not necessarily conform to the conceptual

model that people who live in an area have of their neighborhood (Grannis, 1998; 2005). Yet, hundreds of

studies have used census tracts as proxies for neighborhoods. In particular, numerous studies of racial and

economic segregation rely on the census tract (Massey and Denton 1993). This is problematic given that

census tracts were designed to be racially homogonous. The conclusion that “neighborhoods” are racially

segregated may simply be an artifact of the way they were drawn. To further complicate matters, the

Census Bureau has now changed the criteria for delineating census tracts so 2010 tracts are census tract

definitions: a.) place a priority on setting a standard population or housing unit threshold and b.)

maintaining comparability of tract boundaries over time to facilitate longitudinal data analyses. Though

these new criteria will keep tracts stable over time—a goal that is clearly important to social scientists—

tracts will still not reflect entities that are socially meaningful. Thus, it would be desirable for researchers

to conduct studies of racial and economic segregation across school attendance boundaries since these

boundaries and the populations they contain are the result of a variety of sociological and economic

processes. These processes likely include the residential preferences and choices of people from different

racial and economic groups, the decision-making processes of school boards who delineate school

attendance boundaries and discrimination in the housing market. While it would seem that school

attendance boundaries are delineated “arbitrarily,” the degree to which school boundaries differ across

school districts is a fundamental research question that has yet to be explored. Indeed, it is likely that a

series of novel studies will use SABINS school attendance boundary geography to determine how much

school catchment areas deviate from minimal and maximal levels of racial or economic segregation. This

research question can be explored with automated zone design methods and software developed by

geographers (Openshaw and Alvanides 1999). These methods entail aggregating a given number of small

areas (e.g., census blocks) into larger zones (e.g., school attendance boundaries) so as to optimize a

criteria (such as compactness, population size and racial or economic integration) defined by the user.

Although it is a common lament among social scientist that census tracts are less than ideal

representations of neighborhoods, many studies nevertheless rely upon them (Clapp and Wang 2006;

Jargowsky 1997). For example, Sampson, Morenoff and Gannon-Rowley (2002) reviewed 40 studies of

the review “neighborhood effects” on social processes related to deviant behavior, health-related

outcomes, and found that virtually all of them used census areas as “An imperfect operational definition

of neighborhoods” (p. 445). Many of the findings regarding neighborhood effects are modest and not as

powerful as some researchers had anticipated; different measures of neighborhoods may results different

results (although, the operational definition of a neighborhood is one of several methodological issues that

impact such results). Similarly, researchers who study inter-neighborhood migration patterns recognize

that “Even though census tracts are imperfect operationalizations of neighborhoods, they likely come the

closest of any commonly available spatial entity in approximating the usual conception of a

neighborhood, and their use in this capacity is widespread in sociological and demographic research”

(South, Crowder and Chavez 2006, p 77). Even epidemiologists who use census tracts as proxies for

school boundaries or neighborhoods in studies of child-well being find them inadequate (Brooks-Gunn,

Duncan, Klebanov and Sealand 1993; Crane 1991; Krieger, Chen, Waterman, Soobader and Carson 2003;

Cunradi, Clark and Schafer 2000).

Finally, educational researchers often use census tracts as proxies for school attendance boundaries even

though the two are not aligned closely. For example, educational researchers who require measures of

school context often “Assume that each school serves students in the census tract within which the school

is located” (Ladd and Ludwig 1997, p 274). Yet this assumption is flawed and highly problematic. Still,

prominent scholars have simply made do with census geography as proxies for school attendance

boundaries in studies school on segregation and school effectiveness (Card and Krueger 1992; Entwisle,

Doris, Karl Alexander and Lisa Olson 1997; Frankenburg, Lee, and Orfield 2003; Logan and Oakley

2004; Reardon and Yun 2001; Reardon, Yun, and Eitle 2000). Since the passage of No Child Left Behind,

schools are mandated to publish important statistics including school-level drop-out rates and

standardized test scores (from the National Longitudinal School-Level State Assessment Score Database).

This recent advance is critical. But the potential benefit of these new data has been limited. To be useful,

the test score data needs to linked with the detailed characteristics of student populations that live in a

school attendance boundary (or who attend the school itself). SABINS will link school attendance

boundaries with ACS data, allowing researchers to explore the relationships between detailed

characteristics of students in school catchment areas with school test scores and student outcomes.

Some students that live within a school attendance boundary do not attend the public school. Tabulating

census data from the ACS would apparently lead to invalid measure of the characteristics of children who

are actually enrolled in the school serving the area. This is a legitimate concern, but one that can be

addressed in two ways. First, researchers can gain access to restricted micro census data available at the

Census Bureau and create their own create custom tabulations for public school children. For example,

researchers can determine how many children living in a school attendance boundary and enrolled in

public school live in single-parent households. Second, as we describe in more detail below, the Census

Bureau will create custom tabulations of public school students living in school attendance boundaries.

While it is clear that some students living in a school attendance boundary are enrolled in magnet or

charter schools (information unavailable in census data) these numbers would likely be small in the vast

majority of school boundaries. The Census Bureau has allowed publication of custom tabulations for

school attendance boundaries and other small-area geography very similar to those we propose here

(Geverdt 2005). Space does not permit an overview of the technical issues involved in the release of such

data. We will provide reviewers with details of this process upon request.

The lack of a national database describing student and household populations living in school attendance

boundaries makes it difficult to accurately describe the educational landscape. Consequently, rigorous

empirical testing of hypotheses on the linkages between student outcomes and educational contexts has

proved difficult because analysts lack systematic, reliable, and detailed data on the children and families

living in appropriate geographic boundaries. This lack of reliable data hinders the practice of

professionals working on childhood development issues, whether in local, state, and national

governments, in academic circles, or in policy institutes. The proposed SABINS project seeks to remedy

many of these data problems and help researchers to fill gaps in the literature.

1. **Methods**

We have subdivided the project into six closely interrelated work components. The *data collection*

component will entail the collection of school boundary maps from school districts. The *mapping*

component will require integrating all school attendance boundaries into a common topological

framework through digitizing, projecting, edge-matching, and other GIS techniques that will ultimately

ensure full alignment with the most recent census geography. The *CCD data integration* component will

link each school boundary with school information from the Common Core of Data so that each school

boundary will have information on students, teachers and free- and reduced-priced lunch counts for the

school serving each attendance boundary. The *census tabulation* component will entail using areal

weighting interpolation to reallocate block-level population characteristics to school attendance

boundaries, thereby allowing users to easy access population characteristics of educational geography.

The *ACS tabulation* component will entail using target density weighted interpolation to reallocate

detailed block-group level population characteristics using the American Community Survey to school

attendance boundaries. The *user support, outreach and dissemination* component will ensure that the

research community has the support, information, and training needed to take full advantage of the

scientific opportunities made possible by SABINS. The following sections detail the methods and

procedures of each component.

*Phase 1: School Attendance Boundary Data Collection Process*

Perhaps the largest challenge to assembling data is merely acquiring school attendance boundaries for

every school district in the country. It is a fair question to ask whether collecting all school boundaries in

all districts is possible. Three projects serve as evidence that collecting boundaries for an entire state (or

for major metropolitan areas) can be accomplished in a reasonable time period with a 20 undergraduate

students working full-time during the summer 10 to 15 hours per week during the school year. The first

example is the Minnesota State Department of Education which has collected and distributed school

attendance boundaries annually for every school and district in the state since the 1998-99 school year.

Mr. Scott Freburg, who is the GIS Administrator for the Minnesota Department of Education, collected

all school boundaries for all school districts. These maps were then assembled into a single GIS files. This

took one person eight months. At the timed these data were assembled, roughly half of the districts in

Minnesota maintained a digital file of their attendance boundaries. The remaining districts provided paper

maps and a staff member assembled and integrated the digital and analogue school boundary maps into a

single GIS file in a six month span. For the 2008-2009 school year, there were 339 school districts in

Minnesota which represent roughly 1/50th of the school districts in the United States. Minnesota has 730

elementary, 408 middle and 385 high school attendance boundaries.

Based on his extensive experience, Mr. Freburg has agreed to serve as a key advisor to the project and has

already provided valuable advice on ways to collect school boundary maps for those districts that do not

currently maintain a map. For those school districts without a school attendance boundary map, the

proposed SABINS project will post a web-based digital map of streets, waterways, railroads and other

line features on an Internet site that can be accessed by local school districts. With support from SABINS

personnel, school districts without school attendance boundary maps will delineate their school

attendance boundaries online. These newly created web-maps can be converted to GIS digital files,

processed and returned to local districts. Although it is faster and less expensive to create a web interface

that school districts can use to digitize their maps online, school districts without sufficient Internet access

or adequate computers may need to hand-draw their boundaries on a paper map. In these cases we will

georeference the data, and using existing US Census Bureau TIGER line files as a reference, construct the

school attendence boundaries within a GIS. This process will ensure consistency of the data produced

with all US Census Bureau and NHGIS produced products. On rare occasions a school district will treat

maps of their school attendance boundaries as proprietary information (although most are strikingly

cooperative in providing information). For example, Saporito spent several hours over the course of two

weeks negotiating with a school district for its GIS shapefile of school attendance boundaries. The school

district’s personnel—after gently informed of their state’s public record laws—were eventually

cooperative and emailed the maps after they were paid a small customary service fee.

Despite some of the challenges involved in obtaining maps, school districts and state educational offices

have a strong incentive to provide their information to the project. In return for supplying maps to the

project, each district would receive a digital copy with all statistical information linked to each school in

their district. Many school districts contract companies to perform similar services and our project will

save local school districts this expense. Working with our project to produce digital maps would likely

cost school districts less money than continuing to contract a private company to digitize maps and create

demographic profiles student populations living in school attendance boundaries. Indeed, as one statelevel

administrator in California indicated in an email to the Principal Investigator, “I think I have had a

dream of collecting school attendance boundaries for California. If attendance zones in California were in

shapefiles we could stretch our analysis into another dimension--a more interesting one at that.”

The second example of successful school attendance boundary collection and data integration efforts are

based on the recent experiences of the Principal Investigator of this proposal. In preparation for this

proposal, the PI worked with an undergraduate to conduct searches of school district and county web

pages in Virginia. The search took place in November and December of 2008. The goal was to determine

how many school districts in Virginia had school attendance boundaries posted on the Internet as either a

digital GIS file or as a paper map. (Importantly, paper maps were typically posted as Adobe Acrobat PDF

files but many of these maps were themselves produced with GIS software.) Including Washington DC,

there are 135 school districts in Virginia. Of the 26 school districts with 20 or more schools, 25 had a

digital or analogue school attendance boundary map available on the Internet. Maps for these larger

districts were located and download to a PC in a three week span. Among the remaining 109 school

districts, 70 had digital or paper maps posted on a school district or county web site. In all, 95 of the 135

of school districts in Virginia**—or 70 percent—**had school attendance boundary maps visible on a web

page. It is likely that phone calls to district administrators would produce maps for many of the remaining

districts that do not post their maps.

The final and most powerful example of the proposed SABINS project is also exemplified by the

Principal Investigator’s recent publications. This research collected and used school attendance boundary

maps from the largest school districts in the US for the 1999-2000 school year and integrating those maps

with the Common Core of Data and census variables tabulated at the block-group level. This data

collection and integration process serves as a “proof of concept” for the proposed SABINS project.

Based on the extensive experience of the Principal Investigator and Minnesota Department of Education

staff in collecting and analyzing school attendance boundary data sets, we have devised a data collection

plan that will unfold as follows. The US will be divided into six regions, each of which will contain

roughly 17 percent of the nation’s school districts. In the first phase of the data collection process, six

SABINS staff members will lead a team of five undergraduate students in collecting school attendance

boundaries for the largest school districts within each region. Within the first year of the project, we will

collect 2009-2010 school year school attendance boundary maps for the largest 500 US school districts.

These districts contain roughly 44 percent of US public school children and just over 32 percent of all

public schools. Further, the goal of collecting maps for the largest 500 districts is achievable as

experience demonstrates that larger school districts typically maintain a digital map. Thus, within one

year, a great deal of information will have been collected for largest school districts. (It is certain that

many more school attendance boundaries for hundreds of other school districts will have been collected in

the process of obtaining data for the 500 largest districts.)

The remaining three years of the project will be devoted to collecting those school attendance boundaries

for smaller suburban and rural districts that were not collected within the first year of the project. Each of

the four data collection teams will proceed to assemble school attendance boundaries for an entire state to

avoid creating a patchwork of school attendance boundaries for the US To this end, Saporito will

undertake an extensive outreach program to state-level educational GIS personnel and state-level

administrative authorities to coordinate data collection efforts for the mid-sized and smaller school

districts that may not have maps available. Much of this effort will be supported by Robert Warren and

Paul Manna of William and Mary’s Department of Government. In particular, Manna is a member of the

SABINS advisory board and has extensive ties with local, state and national educational entities such as

the National Association of State Boards of Education, the Council of Chief State School Officers, the US

Department of Education and related agencies. Virtually all states have agencies that require local school

districts to report school-level test scores and other educational data to the state. This administrative

infrastructure can be exploited to collect school attendance boundaries and Manna’s knowledge of this

terrain will be a valuable asset in supporting Saporito’s efforts in developing an outreach program to

state-level departments of education that have a vested interest in contributing to SABINS project.

By then end of the project we will have built a database of school district contacts that include the names,

web addresses, email addresses and telephone number of school district and county GIS personnel who

are responsible for creating or maintaining school attendance boundary files for the local school district.

While a large upfront investment is necessary to collect the 2009-2010 school attendance boundaries,

subsequent data collection efforts will be far less costly. Our contact list will allow us to locate the

necessary school districts offices and personnel efficiently. Moreover, we will not have to digitize paper

maps from scratch but merely make adjustments to school boundaries as districts make yearly

adjustments to their boundaries. This is a sustainable project that becomes more cost-effective over time.

*Phase 2: Mapping Procedures*

Much time and effort will be devoted to creating a single, topologically consistent map from school

attendance boundaries collected from numerous school districts. Most school districts will provide

SABINS with maps as either a digital GIS file or as a paper map. Digital GIS files will likely be provided

to us in a variety of map projections (which are ways the surface of the earth can be displayed on a flat

surface). Modern GIS software can readily convert maps in different projections into a common

framework. Paper maps will require much more processing. Students will digitize school boundaries. This

process (which Saporito’s undergraduates have performed for his published research) requires students to

download free digital GIS maps of streets, waterways, and railroads from the Internet and use GIS

software to trace school attendance boundaries along these line features. Digitizing will be completed in

W&M’s Center for Geospatial Analyses, where up to 20 students can work with the support of Saporito

and Hamilton. Students will work in pairs to digitize school attendance boundaries supplied on paper

maps. A pair of students can accurately digitize 10 school catchment areas in an hour. We anticipate that

up to 40,000 school boundaries will need to be digitized.

In addition to digitizing, school attendance boundaries from neighboring school districts will need to be

“edged-matched.” This means, for example, that adjacent elementary school attendance boundaries from

two neighboring school districts cannot have overlapping boundaries (as every square foot of surface

must be served by one and only district). The opposite is also the case: there cannot be gaps between

borders. This means, for example, that adjacent elementary school attendance boundaries from two

neighboring school districts must meet and share a single border. If gaps existed between school

attendance boundaries then many children living in blocks and block-groups could not be assigned to a

school catchment area. Edge matching will be the most painstaking and time-consuming process of the

project. We estimate that it will take 14,000 student hours to edge match all of the districts collected over

the duration of the project.

*Phase 3: Linking Educational Data from the US Department of Education*

This project will integrate the Common Core of Data (from the National Center for Educational Statistics)

with school attendance boundaries. The CCD contains the number of students by grade, race, gender and

poverty status of virtually every public school in the US Every school in the CCD also has a school name

and a unique identification number. Importantly, all schools in the most recent versions of CCD are georeferenced. In other words, each school has a latitude and longitude making it possible to pin-point its

location on a map. Because each school is geo-referenced, Saporito’s experience with Minnesota maps

demonstrates that GIS software can be used to link school attendance boundaries with the identification

information of each school as it is appears in the CCD. Of course, there may be schools in the CCD that

are not geo-referenced accurately. One benefit of this project is assuring the accuracy of the georeferenced data in the CCD by determining whether schools lie within or in close proximity to their

attendance boundaries. Thus, a by-product of linking school attendance boundaries with the CCD is that

SABINS staff will be able to further refine the CCD. Even though GIS streamlines the entry of

identification numbers, students will spend time manually entering data and verifying its accuracy. On

average, students can enter 60 CCD codes per hour and 79,000 codes need to be entered. Thereafter,

students can use statistical software to link all of the information for each school that is contained in the

CCD with each of the 79,000 polygons in the school attendance boundary shapefiles that we collect.

Integrating school attendance boundaries with the CCD will allow SABINS users to construct custom

school attendance boundaries with statistical information that correspond with those data. Users will have

two data creation and extraction options. The first option is to construct school boundaries by school

level. SABINS will classify schools into one of six main school types: 1.) elementary schools that have a

low grade of Pre-K and a high grade of 5; 2.) middle schools that have a low grade of six and high grade

of eight; 3.) high schools that have a low grade of nine and high grade of 12; 4.) elementary/middle

schools that have a low grade of Pre-K and high grade of eight; 5.) middle/high schools that have a low

grade of six and high grade of 12 and; 6.) elementary through high schools that have a low grade of Pre-K

and a high grade of 12. Users will be able to extract maps and the corresponding statistical information

for all elementary schools, or all middle schools or all high schools based on their research needs. These

six classifications are based on the typical grade structures identified in the CCD.

Users will also be able to generate and extract a school boundaries based on a single grade level (or userspecified combinations of grade levels). For example, users will be able to create school attendance

boundaries for all schools with grade one in them. Such a system is necessary as there are a wide variety

of grade configurations across schools (there are 103 unique grade combinations across US schools).

Thus, if a research question demands school boundaries and data for schools with a first grade, users can

generate the necessary boundaries (for a school district, county, state or the country).

*Phase 4: Linking School Attendance Boundaries with Population Data from the 2010 Census*

We will tabulate 2010 decennial census complete count data (i.e., the “short form”) for school attendance

boundaries. As in the past, the 2010 Census will collect basic population characteristics describing age,

race, gender and household composition and distribute these tabulations at the census block level. As

noted, virtually all census block nest entirely within school attendance boundaries, allowing for accurate

counts to be made for school attendance boundaries. Still, some blocks are split by school attendance

boundaries and, to address this challenge, we will use areal weighting methods to interpolate census data

from blocks to school attendance boundaries. The tabulations we produce will be the same as that made

available to the public for population characteristics summarized for other census geography. For

example, researchers will have access to familiar tabulations such as “age by gender” for school

attendance boundaries. Typical “age by gender” tabulations would allow researchers to count the number

of girls and boys in the following age categories: less than 5; 5 to 9; 10 and 14; and 15 to 17.

An added benefit of the Census Bureau will likely complete a custom tabulation that will describe the

characteristics of ***public school*** populations living in school attendance boundaries. These tabulations for

school attendance boundaries will correspond with the age/grade structure of school serving the boundary.

For example, if a school teaches children in grades 1 to 5, the Census Bureau will create a tabulation of

population totals for children whose ages correspond with that school’s grade structure. Similarly, a

school with children in grades 1 to 4 will have data tabulated for children whose ages are appropriate for

this grade structure. These custom tabulations would focus on urban areas (which have much higher

population densities) and enable the Census Bureau to avoid disclosure issues that would arise with

custom tabulations in rural areas. Because the next decennial census will be collected in April of 2010

time is of the essence. If the utility of school attendance boundaries is to be put to optimal use, the vintage

of the boundaries must be for the 2009-2010 school year. The sooner we can begin the project, the more

likely it is that we can ensure that maps correspond with April 1, 2010 census day. School districts do not

necessarily archive their old school attendance boundaries and by the summer of 2010, we are at risk of

losing boundaries for the 2009-2010 school year.

*Phase 5: Integrating Ongoing American Community Survey Data*

As in the past, statistical information from the 2010 decennial census will consist of complete count

population data from the census short form. However, the American Community Survey has supplanted

the decennial census’ long form which was the instrument for collecting detailed household and

population characteristics from a sample of one in six persons. The ACS is taken continuously and

questionnaires are distributed throughout the year. The reference period for ACS questions varies from

respondent to respondent (and area to area) since it depends on the month in which the survey was

mailed. The ACS will provide detailed information for block groups and census tracts beginning in 2010.

To provide estimates for these smaller areas, the ACS aggregate five years of ACS data to obtain

statistically meaningful estimates. The first of these smaller-area estimates will be made available in

2010, and will cover the period 2005-2009. The Census Bureau will update the statistics of smaller areas

annually. For example, the 2011 ACS will represent cover the five year period from 2006-2010. Thus,

integrating the American Community Survey data into school boundaries will be more challenging than it

will be for the decennial census.

Still, the five-year moving average of small-area ACS detailed population characteristics presents a

unique opportunity for the proposed SABINS. Importantly, school attendance boundaries are updated

annually by school districts. We anticipate that by the completion of the project that school attendance

boundaries for the largest school n the US will be assembled; thereafter, SABINS will refine school

attendance zones on an annual basis as school districts alter the boundaries of the schools within them.

The data will be distributed through the NHGIS web site. Thus, the proposed project presents the unique

opportunity to collect novel geographic information that moves hand-in-hand with the annual release of

the ACS by the US Census Bureau.

Another critical way in which school attendance boundaries can improve upon delivering tabulations by

census geography is that school boundaries: 1) encompass multiple census tracts even while they maintain

their relevance to the people who live within them and; 2) vary in size as schools move up the grade

ladder. These two geographic features will allow the US Census Bureau to create special tabulations from

the ACS for school attendance boundaries (on an annual basis) that produce smaller confidence intervals

and more reliable data than will exist for census tracts and block groups. In particular, the larger areas

covered by the typical high school will produce more accurate population estimates than will most census

tracts. The same is true for middle schools.

A pilot study based on the educational and administrative geography for the state of Minnesota

demonstrates the greater statistical accuracy (i.e., smaller confidence intervals) that can be achieved

compared to existing small area administrative geography distributed currently by the US Census Bureau.

In Minnesota there are 1.77 census tracts for every elementary school attendance boundary. In the major

metropolitan areas of Minnesota (i.e, Ramsay and Hennepin Counties that contain Minneapolis and St.

Paul) there are 2.30 census tracts for every elementary school. In the remaining rural areas of the state

there are 1.59 census tracts for every elementary school boundary. Minnesota is likely representative of

other rural areas in the country. As discussed above, in the US there fewer elementary school attendance

boundaries than there are census tracts. Therefore, statistical information summarized from the ACS to

the typical elementary school boundaries will substantially improve upon the statistical estimates of the

typical census tract. These same principles apply even more so for middle and high school attendance

boundaries. In Minnesota, there are 3.17 tracts for every middle school attendance boundary and this ratio

is 3.35 for high school attendance boundaries. These ratios vary between urban and rural areas: in rural

areas outside of the Ramsey and Hennepin Counties there are 2.63 tracts for every middle school and 2.47

tracts for each high school; within Ramsay and Hennepin Counties there are 5.35 tracts per middle school

and 7.23 tracts per high school.

In short, the larger regions covered by the average middle or high school attendance boundary will have

smaller statistical errors than will population estimates for elementary school attendance boundaries.

Since school elementary, middle and high school boundaries vary in size, researchers can balance the

accuracy of population estimates with their research questions. In the end, the proposed SABINS

infrastructure project will develop geographic units that many researchers would prefer over census tracts,

zip codes and other administrative geography because: 1.) of their social significance; 2.) school

boundaries change yearly and thus hand-in-hand with the ongoing ACS and; 3.) of the greater precision

with school attendance boundaries estimate population statistics relative to census tracts.

To supplement the interpolated data that we will generate, there are strong indications that the Census

Bureau will immediately exploit the school attendance boundaries to create custom tabulations. While the

success of the SABINS project does not hinge on their continued cooperation, the Census Bureau will

likely create custom tabulations voluntarily and for free (see Doug Geverdt’s support letter in the

appendix). Custom tabulations would result in population characteristics for school-aged children enrolled

in ***public schools***. The tabulation for each school would also correspond with the grade structure of the

school. For example, if a school’s grades range from K to 4 the corresponding tabulations of 2010 census

and ACS data will be based on children who are aged 5 to 9 enrolled in public school. These custom

tabulations would result in an additional refinement of the tabulations that SABINS staff will produce

with ACS data using interpolation techniques.

1. **Human subject protection: Not applicable**
2. **If applicable, briefly describe how the project is funded.**

The project is funded through a two year grant from the National Science Foundation.

1. **Describe plans for the dissemination of research findings.**

SABINS data will be disseminated through the existing National Historic Geographic Information System

web site, eliminating the need to duplicate the infrastructure for data delivery at another university. The

NHGIS user interface makes it easy for users to navigate the mass of documentation, locate and select

tables and variables, and explore geographic boundaries. It will help users to extract downloadable

datasets and boundary files in a variety of formats. Users will create customized subsets of both data and

documentation tailored to their particular research questions.

All geographic data will contain complete FGDC compliant metadata attached to the geographic files.

FGDC metadata documents: a.) a dataset’s source, resolution, spatial reference information, and b.) its

attribute information, lineage and distribution information. Users of the SABINS system will be able to

request their spatial date in numerous differing formats including CSV format for use in database or

spreadsheet software, ESRI shapefiles for use in off-the-shelf GIS software, and open source GML files

for use in advanced spatial data systems. Users will receive not only the geographies, but also the 2010

Decennial Census, ACS, and CCD data requested as attributes corresponding to these geographies.

The Minnesota Population Center will provide user support for SABINS by means of email and

telephone. MPC currently manages a steady demand for user support for the NHGIS and IPUMS projects

and MPC is equipped to receive significantly more questions when the first series of SABINS data are

released. Extensive experience with the IPUMS and NHGIS databases demonstrates that high-quality user

support not only improves the quantity and quality of research based on the data, but it also helps us to

improve data quality. Responding to user queries often helps us to identify errors in data or documentation. We will also publicize SABINS. Large investments in social science infrastructure can be

justified only if the data are used intensively. The MPC routinely rents exhibit space at major conferences,

and SABINS can be represented at these exhibits at minimal cost.